MODELING AND APPLICATIONS OF OPTOELECTRONIC DEVICES FOR ACCESS NETWORKS

EDITED BY: Chongfu Zhang, Nico de Rooij, Zichuan Yi, Guofu Zhou and Raffaele Gravina
PUBLISHED IN: Frontiers in Physics

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-099-7
DOI 10.3389/978-2-88971-099-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

MODELING AND APPLICATIONS OF OPTOELECTRONIC DEVICES FOR ACCESS NETWORKS

Topic Editors:
Chongfu Zhang, University of Electronic Science and Technology of China, China
Nico de Rooij, École Polytechnique Fédérale de Lausanne, Switzerland
Zichuan Yi, University of Electronic Science and Technology of China, China
Guofu Zhou, South China Normal University, China
Raffaele Gravina, University of Calabria, Italy

Topic Editor Raffaele Gravina is a founder and a co-owner of company SenSysCal S.R.L. Topic Editor Guofu Zhou is a founder and a director of Electronic Paper Display Institute of South China Normal University and science advisor of Eindhoven University of Technology. All other Topic Editors declare no competing interests with regards to the Research Topic subject.

Citation: Zhang, C., de Rooij, N., Yi, Z., Zhou, G., Gravina, R., eds. (2021). Modeling and Applications of Optoelectronic Devices for Access Networks. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-099-7

Table of Contents

frontiers in Physics

EDITORIAL
published: 07 May 2021
doi: 10.3389/fphy.2021.678269

Editorial: Modeling and Applications of Optoelectronic Devices for Access Networks

Chongfu Zhang[1]*, *Zichuan Yi*[2], *Nico de Rooij*[3], *Guofu Zhou*[4] and *Raffaele Gravina*[5]

[1] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China, [2] College of Electron and Information, Zhongshan Institute, University of Electronic Science and Technology of China, Zhongshan, China, [3] Institute of Microengineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [4] South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China, [5] Department of Informatics, Modeling, Electronics, Informatics and Systems, University of Calabria, Cosenza, Italy

Keywords: optoelectronic devices, access networks, optical processing devices, information detection, modeling

Editorial on the Research Topic

Modeling and Applications of Optoelectronic Devices for Access Networks

With the development of optical devices and advanced materials, optoelectronic devices can be newly reconsidered and improved since these are indispensable for access networks. Specifically, the modeling and applications of optoelectronic devices play an important role in promoting the development of access networks. Therefore, this Research Topic reports on the latest developments in the modeling and applications of optoelectronic devices.

MODELING OF OPTOELECTRONIC DEVICES FOR ACCESS NETWORKS

It is beneficial to improve the accuracy and performance of optoelectronic detections in access networks by establishing models of optoelectronic devices [1, 2]. A differential entropy feature signal recognition model based on a gated recurrent unit network was proposed to improve the recognition accuracy of optoelectronic detections (Zhu and Zhong). In this model, differential entropy and power spectral density feature signals were used to reduce electrode channels, which provided a basic theory for the optoelectronic detections. Similarly, an improved WC-KNN model based on frequency modulation (FM) signals was proposed to improve the accuracy of photoelectric devices (Duan et al.). Specifically, the complexity of these devices in indoor scenarios was reduced effectively and the positioning accuracy was improved, which could expand their plethora of practical indoor applications for access networks.

The efficiency of access networks is related to the performance of optoelectronic devices. Moreover, their performance can be analyzed and optimized by establishing specific models, which can improve the efficiency of access networks [3, 4]. A model based on the density functional theory was proposed to describe the electrochemical performance of cathode materials (Gao et al.). The conductivity and the potential energy of cathode material were also calculated, and theoretical proposal for the electrochemical performance of doping could be provided by this model. In addition, this model was applied to the study of optoelectronic devices, which could improve the performance analysis of optoelectronic devices. Similarly, a model was applied to analyze the device performance (Zhuang et al.), in which a severe degradation was found at short wave lengths. Then, by using the

OPEN ACCESS

Edited and reviewed by:
Lorenzo Pavesi,
University of Trento, Italy

***Correspondence:**
Chongfu Zhang
cfzhang@uestc.edu.cn

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *09 March 2021*
Accepted: *01 April 2021*
Published: *07 May 2021*

Citation:
Zhang C, Yi Z, de Rooij N, Zhou G and Gravina R (2021) Editorial: Modeling and Applications of Optoelectronic Devices for Access Networks. Front. Phys. 9:678269. doi: 10.3389/fphy.2021.678269

model, suggestions to improve the optoelectronic efficiency were provided.

Non-invasive detection of Majorana bound states (MBSs) is one of the core issues in the modeling of optoelectronic devices [5]. Conductance properties in double quantum dots (QDs) were studied by the Green's function method (Chi et al.). The information of MBSs can be inferred by comparing properties of linear conductance. Then, resonances at zero and positive energy regimes would display the typical Fano line-shape when MBSs are overlapped, which was useful for quantum information processing and design of new quantum devices. Similarly, spin-dependent current and tunneling magnetoresistance (TMR) in a QD were investigated by the non-equilibrium Green's technique (Tang and Mao). The TMR was sandwiched between two ferromagnetic leads, and the MBSs were formed at the opposite ends of a superconductor nanowire. Furthermore, it was found that the intensity of the TMR depended on the spin polarization of electrodes in a non-monotonic way, and it was positive for a large spin-polarization regardless of the existence of MBSs. In addition, MBSs are promising in the context of high-efficiency and energy-saving quantum devices [6]. The photon-assisted transport through a QD side-coupled to a topological nanowire hosting MBSs at its two ends was studied (Chi et al.). It was found that the photon-induced peaks in the transmission function were split by the existence of MBSs, and the value of central peaks was suppressed to zero. Meanwhile, positions of additional peaks induced by the MBS-MBS overlapping in the presence of the photon field were quite different from the case of the zero-photon field, which could propose a new application for quantum devices.

APPLICATIONS OF OPTOELECTRONIC DEVICES FOR ACCESS NETWORKS

The fabrication of optoelectronic devices is needed for the development of access networks [7]. Large-scale growth of zinc oxide (ZnO) nanorod arrays on graphene sheets was realized by hydrothermal techniques, and the Fowler-Nordheim theory was used to build a model to describe the properties of the arrays' field emissions (Yang et al.). The morphological characteristics of ZnO nanorods can also be tuned by varying reaction time and concentrations of the solution. These specific nanorod arrays with enhancement of emission properties would be useful for modulator units of access networks. Ultrasmall precious metal clusters have attracted extensive attention for promoting electron transfer in optoelectronic devices [8]. Ultrasmall Au clusters based on defective TiO_2 nanosheets (Au/D-TiO_2) were fabricated (Zhang et al.). Different defects of TiO_2 nanosheets (D-TiO_2) were induced by using a heating process. This method could be effective for enhancing the stability of electron transfer in optoelectronic devices.

Applications of optoelectronic devices can provide good prospects for the development of access networks [9, 10]. A photonic-assisted channelized receiver, which is based on the spectrum analysis, was developed for multi-band microwave signals (Huang et al.). Instantaneous spectral analysis was introduced to determine the frequency and bandwidth of the dynamic wideband signal. Then, the dynamic wideband signal was received by a multi-band coherent channelizer. This new receiver provided a low-complexity and low-cost implementation, which showed a tremendous advantage and potential in access networks. An application based on data access of a thermal camera was proposed to distinguish vehicle headlight (Li S. et al.). In this application, details of thermal images were enhanced by adjusting the temperature display dynamically. Then, features of a vehicle headlight were extracted by YOLOv3, high beams and low beams were further distinguished by the filter. This application provided an effective method for distinguishing headlight, which can be applied to transportation-based access networks. The performance of optoelectronic devices can be improved by new applications [11]. A new driving waveform with a direct current (DC) and an alternating current (AC) was proposed to improve the performance of Electrowetting displays (EWDs) (Liu et al.). In this driving waveform, the DC was used to reduce the response time of EWDs, and the AC was used to suppress the oil backflow. Similarly, a multi adaptive driving waveform was proposed to reduce the influence of hysteresis effect in EWDs (Li W. et al.). Then, different driving waveforms were designed and implemented according to different driving stages of EWDs. This could effectively reduce the hysteresis effect of EWDs.

Optical information detection is of great significance to promote application development of optoelectronic devices [12]. A microcomputer was used for the real-time intelligent reading of a pointer instrument, it could accurately locate the panel area and read corresponding values (Lin et al.). At the same time, it was convenient to collect and analyze historical data for system optimizations, which had a high feasibility and a practical value for access networks. In addition, an adaptive specular highlight detection was proposed for endoscopic images (Yu et al.). Specifically, a criterion for specular highlight detection based on the ratio of red channel was designed for each pixel. This detection application provided accuracy for an endoscopic imaging system. Open electrowetting on dielectric (EWOD) devices have been widely used in optical information detection [13]. An EWOD digital microfluidic operating device, which was convenient in manufacture based on a printed circuit board, was realized. A Paraflim M and silicone oil were used as the dielectric hydrophobic layer (Yi et al.). This EWOD optoelectronic device not only had simple preparation steps, but also reduced production cost, with clear practical impact for access networks.

CONCLUSION

The development of society puts forward higher requirements for access networks, and modeling and applications of optoelectronic devices have an important impact on the performance of access networks. The contributions of this Research Topic describe

the new progress of optoelectronic devices, which improve the efficiency and accuracy of an access network system. The rapid development of optoelectronic devices is expected to bring other theoretical and practical breakthroughs for access networks in the next few years.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

This Research Topic was supported by National Key R&D Program of China under Grant 2018YFB1801302； Project for Innovation Team of Guangdong University under Grant 2018KCXTD033； Project for Zhongshan Social Public Welfare Science and Technology (2019B2007). We express our gratitude to the Frontiers team and reviewers for supporting our organization works. We also sincerely thank all the authors for their contributions.

REFERENCES

1. Li Y, Zheng W, Wang L, Zong Y, Cui Z. From regional to global brain: a novel hierarchical spatial-temporal neural network model for EEG emotion recognition. *IEEE Trans Affect Comput*. (2019) 99:1. doi: 10.1109/taffc.2019.2922912
2. Yu Z, Guo G. Improvement of positioning technology based on RSSI in ZigBee networks. *Wireless Pers Commun*. (2017). 95:1943–62. doi: 10.1007/s11277-016-3860-1
3. Lei QQ, Aierken A, Sailai M, Heini M, Shen XB, Zhao XF, et al. 150 KeV proton irradiation effects on photoluminescence of GaInAsN bulk and quantum well structures. *Opt Mater*. (2019) 97:109375. doi: 10.1016/j.optmat.2019.109375
4. Shen DF, Guo JJ, Lin GS, Lin JY. Content-aware specular reflection suppression based on adaptive image inpainting and neural network for endoscopic images. *Comput Methods Prog Biomed*. (2020) 192:105414. doi: 10.1016/j.cmpb.2020.105414
5. Karzig T, Knapp C, Lutchyn RM, Bonderson P, Hastings MB, Nayak C, et al. Scalable designs for quasiparticle-poisoning-protected topological quantum computation with Majorana zero modes. *Phys Rev B*. (2017) 95:235305. doi: 10.1103/physrevb.95.235305
6. Ricco LS, Souza MD, Figueira MS, Shelykh IA, Seridonio AC. Spin-dependent zero-bias peak in a hybrid nanowire-quantum dot system: distinguishing isolated Majorana fermions from Andreev bound states. *Phys Rev B*. (2019) 99:1–9. doi: 10.1103/PhysRevB.99.155159
7. Li L, Yu H, Xu J, Zhao S, Liu Z, Li Y. Rare earth element, Sm, modified graphite phase carbon nitride heterostructure for photocatalytic hydrogen production. *New J Chem*. (2019) 43:1716–24. doi: 10.1039/C8NJ05619F
8. Kumaravel V, Mathew S, Bartlett J, Pillai SC. Photocatalytic hydrogen production using metal doped TiO_2: a review of recent advances. *Appl Catal B Environ*. (2019) 244:1021–64. doi: 10.1016/j.apcatb.2018.11.080
9. Li C, Peng Z, Huang TY, Fan T, Wang FK, Horng TS, et al. A review on recent progress of portable short-range noncontact microwave radar systems. *IEEE Trans Microw Theor Tech*. (2017) 65:1692–706. doi: 10.1109/TMTT.2017.2650911
10. Yi ZC, Chen ZB, Peng B, Li SX, Bai PF, Shui LL, et al. Vehicle lighting recognition system based on erosion algorithm and effective area separation in 5g vehicular communication networks. *IEEE Access*. (2019) 7:111074–83. doi: 10.1109/access.2019.2927731
11. Yi Z, Huang Z, Lai S, He W, Wang L, Chi F, et al. Driving waveform design of electrowetting displays based on an exponential function for a stable grayscale and a short driving time. *Micromachines*. (2020) 11:313. doi: 10.3390/mi11030313
12. Liu Y, Liu J, Ke Y. A detection and recognition system of pointer meters in substations based on computer vision. *Measurement*. (2020) 152:122–6. doi: 10.1016/j.measurement.2019.107333
13. Samiei E, Tabrizian M, Hoorfar M. A review of digital microfluidics as portable platforms for lab-on a-chip applications. *Lab Chip*. (2016) 16:2376–96. doi: 10.1039/C6LC00387G

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 05 May 2020
doi: 10.3389/fphy.2020.00147

Detection of Majorana Bound States by Sign Change of the Tunnel Magnetoresistance in a Quantum Dot Coupled to Ferromagnetic Electrodes

Li-Wen Tang and Wei-Guo Mao*

Department of Materials Engineering, Xiangtan University, Xiangtan, China

We study quantum transport and tunneling magnetoresistance (TMR) through an interacting quantum dot (QD) attached to ferromagnetic electrodes in the Coulomb blockade regime. The QD is also side-coupled to a superconductor nanowire hosting Majarana bound states (MBSs). It is found that when the electrodes' magnetic moments are arranged in antiparallel, the current's intensity will be enhanced to be larger than that of the parallel configuration by the hybridization between the QD and the MBSs. This change in the current induces anomalous negative TMR unique to the existence of MBSs, providing an efficient detection way of the MBSs. The negative TMR is weakened by the overlap between the two bound states but obviously enhanced by the left-right asymmetry between the QD and the electrodes. We also find that the TMR value changes non-monotonously with the spin polarization of the electrodes. Our results may find real use in energy saving spintronic devices and quantum information processing.

Keywords: quantum dot, tunneling magnetoresistance, ferromagnetic electrodes, Majorana bound states, superconductor nanowire

OPEN ACCESS

Edited by:
Zichuan Yi,
University of Electronic Science and Technology of China, China

Reviewed by:
Feng Chi,
University of Electronic Science and Technology of China, China
Qiang Xu,
Nanyang Technological University, Singapore
Zhen-Guo Fu,
Institute of Applied Physics and Computational Mathematics (IAPCM), China

****Correspondence:***
Li-Wen Tang
qdtang2020@163.com

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *21 March 2020*
Accepted: *14 April 2020*
Published: *05 May 2020*

Citation:
Tang L-W and Mao W-G (2020) Detection of Majorana Bound States by Sign Change of the Tunnel Magnetoresistance in a Quantum Dot Coupled to Ferromagnetic Electrodes. Front. Phys. 8:147. doi: 10.3389/fphy.2020.00147

1. INTRODUCTION

Tunneling of spin-polarized electrons in nano- and micro-scale devices are at the core of quantum information data processings [1, 2] and future spintronics [3]. Spin-dependent current is conveniently generated in a system with ferromagnetic electrodes, and its intensity changes when the magnetic configuration of the device changes from parallel to antiparallel alignment [4]. Generally, the electronic current is larger in parallel configuration when transport occurs between the majority-majority and minority-minority spin bands, than in the antiparallel configuration, where electrons tunnel between majority and minority spin bands. The change in the currents is measured by the quantity of tunnel magnetoresistance (TMR) defined as $\mathrm{TMR} = (J_P - J_{AP})/(J_P + J_{AP})$, where $J_{P/AP}$ is the electronic current in parallel (antiparallel) configuration. Usually, the TMR value is positive as $J_P > J_{AP}$ [4]. The TMR effect has been extensively studied in various systems, including the relatively simple planar junctions [5], and in more complex mesoscopic double-barrier junctions [6], junctions with nano-scale granular systems [7, 8], as well as single-electron transistors [9], etc. When the central region between the electrodes is sufficiently small that the charging energy is larger than the thermal energy k_BT, some interesting effects due to quantization of energy levels becomes visible, such as the Coulomb blockade of electric current characterized by the Coulomb steps in the current-voltage curves [10–13]. Especially, behaviors of the TMR have been investigated both theoretically and experimentally in systems composing of ferromagnetic

electrodes and quantum dots (QDs) [14–19], which is a nano-scale zero-dimensional structure [13]. This is because such a kind of device is ideal candidate to study the fundamental interactions between spins and charges [3, 13], and is being considered for applications in energy saving devices as well as for quantum computing [1, 2].

Recently, there is much interest in preparation and detection of Majorana fermions, which is a kind of elementary particle being of its own antiparticle, due to its promising applications in quantum computation free from decoherence [20–22]. It has been be prepared in the edges of topological superconductors as a pair of zero-energy Majorana bound states (MBSs) that obeys non-Abelian statistics enabling topologically protected quantum computation and high-efficiency spintronic devices [20, 22]. The MBSs has also be prepared in some other systems, including the non-centrosymmetric superconductors [23], three- or two-dimensional topological insulators coupled to superconductors [24], electrostatic defects in topological superconductors [25], p-wave superconductors [26], the semiconducting [27], or ferromagnetic [28] nanowires with native strong spin-orbit interaction proximitization to a conventional s-wave superconductors, and Josephson junctions [29], etc. The detection of MBSs is quite challenging because the unique chargeless and zero-energy properties, and then most of conventional detection techniques for elementary particles failed. For example, the MBSs formed in the nanowire induces a zero-bias anomaly in the differential conductance when a bias voltage is applied [30], which has been demonstrated in experiments. However, it can also arise from some other mechanisms, such as the quantum interference and Kondo effects [31]. Anomalous change of the thermoelectric effect in systems composing of a QD side-coupled to topological superconductor hosting a Majorana edge state has also been found and proposed as an alternative detection means in recent years [32–34]. The hybridization between QD and MBSs breaks the particle-hole symmetry of the system and leads to large enhancement and sign reversion of the thermopower. These, unfortunately, can also not been solely attributed to the existence of the MBSs.

In recent experiments, semiconductor spacers of InAs QD has been inserted in between nickel or cobalt leads [15, 35]. The dot's size and discretized energy levels can be fully adjusted by the length and thickness of the spacers or the gate voltages. Moreover, the spin polarization of the current injected from the ferromagnetic electrodes and TMR can also be electrically adjusted by a gate nearby the QD. Some new characteristics, such as the anomalies of the TMR caused by the intradot Coulomb repulsion energy in the QD, were explained in the subsequent theoretical work [36]. In the present manuscript, we study the properties of current-voltage and TMR in a QD sandwiched between the left and right ferromagnetic electrodes and side-coupled to a semiconducting nanowire hosting a pair of MBSs at the opposite ends of the wire (see the schematic plot of the system in **Figure 1**). We assume that the MBSs are coupled only to spin-up electrons in the QD under a strong magnetic field in experiments. Our calculation results show that due to the presence of the MBSs, the current of the antiparallel configuration can be larger than that of the parallel one, inducing a negative TMR. Such an effect is a good evidence of the existence of the Majorana fermions. The negative TMR will be weakened by the coupling between the two MBSs but obviously enhanced by the left-right asymmetry between the QD and the electrodes. Moreover, we find that the TMR strength depends non-monotonously on the spin polarization of the electrodes.

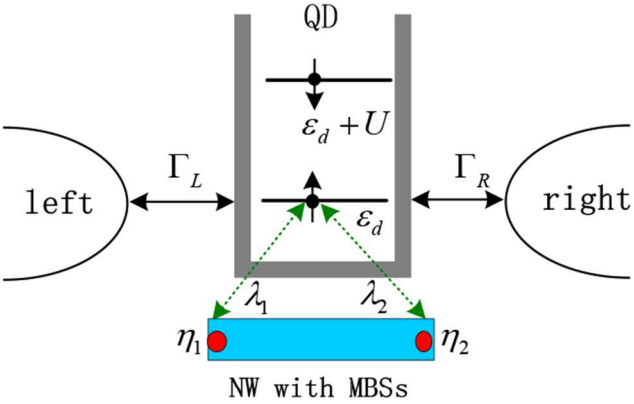

FIGURE 1 | Schematic diagram for a QD coupled to MBSs and the left and right ferromagnetic leads with coupling strength $\Gamma_{L/R}$. Due to the chiral nature of the Majorana fermions, only spin-up is assumed to couple to the MBSs, formed at opposite ends of a nanowire. The two MBSs are denoted by η_1 and η_2 and coupled to the QD with strength of λ_1 and λ_2, respectively. If the dot level ε_d is occupied by an spin-up electron, then the spin-down electron will by pushed up to the level of $\varepsilon_d + U$ by the intradot Coulomb repulsion U.

2. MODEL AND METHODS

The Hamiltonian of the QD coupled to MBSs and ferromagnetic electrodes takes the following form [18, 32, 37, 38]

$$H = \sum_{k\beta\sigma} \varepsilon_{k\beta} c^{\dagger}_{k\beta\sigma} c_{k\beta\sigma} + \sum_{\sigma} \varepsilon_d d^{\dagger}_{\sigma} d_{\sigma} + U d^{\dagger}_{\uparrow} d_{\uparrow} d^{\dagger}_{\downarrow} d_{\downarrow} + \sum_{k\beta\sigma} (T_{k\beta\sigma} c^{\dagger}_{k\beta\sigma} d_{\sigma} + H.c) + H_{MBSs}, \quad (1)$$

where $c^{\dagger}_{k\beta\sigma}$ ($c_{k\beta\sigma}$) creates (annihilates) an electron of momentum k, energy $\varepsilon_{k\beta\sigma}$ and spin $\sigma = \uparrow, \downarrow$ in the ferromagnetic electrode $\beta = L, R$. For the QD, $d^{\dagger}_{\sigma}$ (d_{σ}) is the creation (annihilation) operator of an electron having energy level ε_d, spin-σ and intradot Coulomb interaction U, which has usually been neglected in previous work [32]. In experiments, ε_d and U are tunable by external gate voltages. The coupling strength between the QD and the electrodes is described by $T_{k\beta\sigma}$, which is spin-dependent due to the ferromagnetism on the electrodes. The last term H_{MBSs} in Equation (1) stands for the zero-energy MBSs located on the opposite ends of the semiconducting nanowire and their coupling to the QD [39, 40]:

$$H_{MBSs} = i\delta_M \eta_1 \eta_2 + \lambda_1 (d_{\uparrow} - d^{\dagger}_{\uparrow}) \eta_1 + i\lambda_2 (d_{\uparrow} + d^{\dagger}_{\uparrow}) \eta_2, \quad (2)$$

in which δ_M is the overlap strength between the two MBSs with operator satisfying both $\eta_j = \eta^{\dagger}_j (j = 1, 2)$ and $\{\eta_i, \eta_j\} = \delta_{i,j}$. The hopping amplitude between MBSs and spin-$\uparrow$ electrons in

the QD is accounted by λ_j. Following previous work [39, 40], we write η_j in terms of the regular fermionic operators f as $\eta_1 = (f^\dagger + f)/\sqrt{2}$ and $\eta_2 = i(f^\dagger - f)/\sqrt{2}$, and then H_{MBSs} becomes

$$H_{MBSs} = \delta_M(f^\dagger f - \frac{1}{2}) + \frac{\lambda_1}{\sqrt{2}}(d_\uparrow - d_\uparrow^\dagger)(f^\dagger + f) - \frac{\lambda_2}{\sqrt{2}}(d_\uparrow + d_\uparrow^\dagger)(f^\dagger - f). \quad (3)$$

Within the standard Keldysh Green's function technique, the spin-dependent electric current is obtained as [37, 38]

$$J_\sigma = \frac{e}{\hbar}\int T_\sigma(\varepsilon)[f_L(\varepsilon) - f_R(\varepsilon)]\frac{d\varepsilon}{2\pi} \quad (4)$$

where e is the electron charge, $\hbar$ the reduced Planck's constant, $f_{L/R}(\varepsilon) = \{1+e^{(\varepsilon-\mu_{L/R})/k_BT}\}^{-1}$ the Fermi distribution function of the left/right electrode with chemical potential $\mu_{L/R}$, temperature T and Boltzmann constant k_B. The transmission coefficient $T_\sigma(\varepsilon)$ can be expressed with the help of the retarded Green's function $G_\sigma^r(\varepsilon)$ as [37, 38]

$$T_\sigma(\varepsilon) = \frac{\Gamma_\sigma^L\Gamma_\sigma^R}{\Gamma_\sigma^L + \Gamma_\sigma^R}[-2\mathrm{Im}G_\sigma^r(\varepsilon)], \quad (5)$$

where $\Gamma_\sigma^\beta = 2\pi\sum_k |T_{k\beta\sigma}|^2\delta[\varepsilon - \varepsilon_{k\beta\sigma}]$ is the spin-dependent line-width function. Ferromagnetism on the electrodes can be expressed by defining a spin-polarization parameter P_β as $P_\beta = (\Gamma_\uparrow^\beta - \Gamma_\downarrow^\beta)/(\Gamma_\uparrow^\beta + \Gamma_\downarrow^\beta)$. In the present paper, we consider the parallel ($P_L = P_R$) and antiparallel ($P_L = -P_R$) configurations of the two electrodes. In both of the two cases, we have $\Gamma_{\uparrow(\downarrow)}^\beta = (1 \pm P_\beta)\Gamma_\beta$.

By applying the equation of motion method, the retarded Green's function in Equation (5) is obtained as (we have truncated the higher-order Green's functions by following reference [39, 40], in which the simultaneous tunneling of the electron of opposite spin has been neglected):

$$G_\sigma^r(\varepsilon) = \frac{\tilde{\varepsilon}_{-,\sigma} - \Sigma_\sigma^1 - U\{1 - < n_{\tilde{\sigma}} > [1 - (\lambda_1^2 - \lambda_2^2)^2\tilde{M}_\sigma\tilde{M}_{U,\sigma}]\}}{(\tilde{\varepsilon}_{-,\sigma} - \Sigma_\sigma^0)(\tilde{\varepsilon}_{-,\sigma} - U - \Sigma_\sigma^1)}, \quad (6)$$

where the self-energies considering the MBSs are given by $\Sigma_\sigma^0 = M_1 + (\lambda_1^2 - \lambda_2^2)^2 B\tilde{M}_\sigma$ and $\Sigma_\sigma^1 = M_1 + (\lambda_1^2 - \lambda_2^2)^2 B\tilde{M}_{U,\sigma}$ with $B = \varepsilon/(\varepsilon^2 - \delta_M^2)$ and $\tilde{M}_\sigma = B/(\varepsilon_{+,\sigma} + M_2)$, $\tilde{M}_{U,\sigma} = B/(\varepsilon_{+,\sigma} + U - M_2)$ in which

$$M_{1(2)} = \frac{1}{2}[\frac{\lambda_1^2 - \lambda_2^2}{\varepsilon - (+)\delta_M} - \frac{\lambda_1^2 + \lambda_2^2}{\varepsilon + (-)\delta_M}], \quad (7)$$

and $\tilde{\varepsilon}_\pm = \varepsilon \pm \varepsilon_d + i(\Gamma_\sigma^L + \Gamma_\sigma^R)/2$. The occupation number in Equation (6) is calculated self-consistently from

$$< n_\sigma > = \int \frac{d\varepsilon}{2\pi}\frac{\Gamma_\sigma^L f_L(\varepsilon) + \Gamma_\sigma^R f_R(\varepsilon)}{\Gamma_\sigma^L + \Gamma_\sigma^R}[-2\mathrm{Im}G_\sigma^r(\varepsilon)]. \quad (8)$$

To our knowledge, there are three kinds of schemes that are mainly used for studying the transport phenomena [41–43]. The first kind is the relatively rough Hartree-Fock-like scheme used in reference [41, 42]; the second kind is the Hubbard-I scheme adopted in the present work as well as other reference [39, 40] that can deal with relatively higher-order transport processes; the third kind of scheme is the one that can account for multiple transitions from contacts to the central site, such as the subtle Kondo effect. Detailed introduction of these schemes can be found in of reference [43]. In the Coulomb blockade regime, these three schemes give essentially the same results, see for example reference [43]. From theoretical point of view, here we adopt the Hubbard-I scheme for the sake of sophistication, and the procedure to evaluate the QD Green's function will be relatively simplified.

3. RESULTS AND DISCUSSION

In this section, we present our numerical results for the spin-dependent current, TMR and differential conductance varying as functions of the bias voltage. The intradot Coulomb interaction $U = 1$ is chosen as the energy unit with symmetric bias voltage V as $\mu_L = -\mu_R = eV/2$. **Figure 2** shows the spin-polarized current, differential conductance and the current's spin-polarization $p = (J_\uparrow - J_\downarrow)/(J_\uparrow + J_\downarrow)$ for parallel and antiparallel configurations, respectively. The currents exhibit typical Coulomb blockade effect as shown in **Figures 2A,B,D,E** [3, 10, 13]. The reason is that by applying a bias voltage between the electrodes, the QD can be charged by single electrons when the energy levels ε_d are within the transport window. The current then is increased. With an increase from n to $n + 1$ electrons on the QD, an additional potential of e/C is induced with C denoting the capacitance of the device [10, 13]. The transfer of an electron through the QD is inhibited when the potential of the charged dot is larger than the applied bias voltage. Now the electrons are blockaded on the QD and the current will not increase with increasing bias voltage. With further increased bias voltage, electrons may tunnel through the dot level of $\varepsilon_d + U$ again when the potential is overcome. In this way, the electrons are added on the dot one by one and the current shows Coulomb staircases. The Coulomb blockade effect can be clearly seen from the differential conductance, which develops two sharp peaks around ε_d and $\varepsilon_d + U$ as given in the figure.

In the parallel configuration, the current intensity of $J_\uparrow$ is obviously larger than that of $J_\downarrow$ in the absence of QD-MBSs. The reason is that the spin-up electrons will enter and leave the QD faster than the spin-down ones because of the ferromagnetism on the electrons, i.e., $\Gamma_\uparrow^L = \Gamma_\uparrow^R > \Gamma_\downarrow^L = \Gamma_\downarrow^R$. In the antiparallel configuration, however, the current intensities of both the two spin components are the same because the transmission coefficient is spin-independent in the absence of hybridization between the QD and MBSs ($\lambda_1 = \lambda_2 = 0$), which is not shown in the figure. Turning on the coupling between the QD and MBSs, the intensity of spin-up currents in both parallel and antiparallel configuration are enhanced as shown in **Figures 2A,D**, whereas that of the spin-down electrons are unchanged as the MBSs is coupled only to the spin-up electrons. Moreover, the enhancement of the intensity of $J_\uparrow$ in the

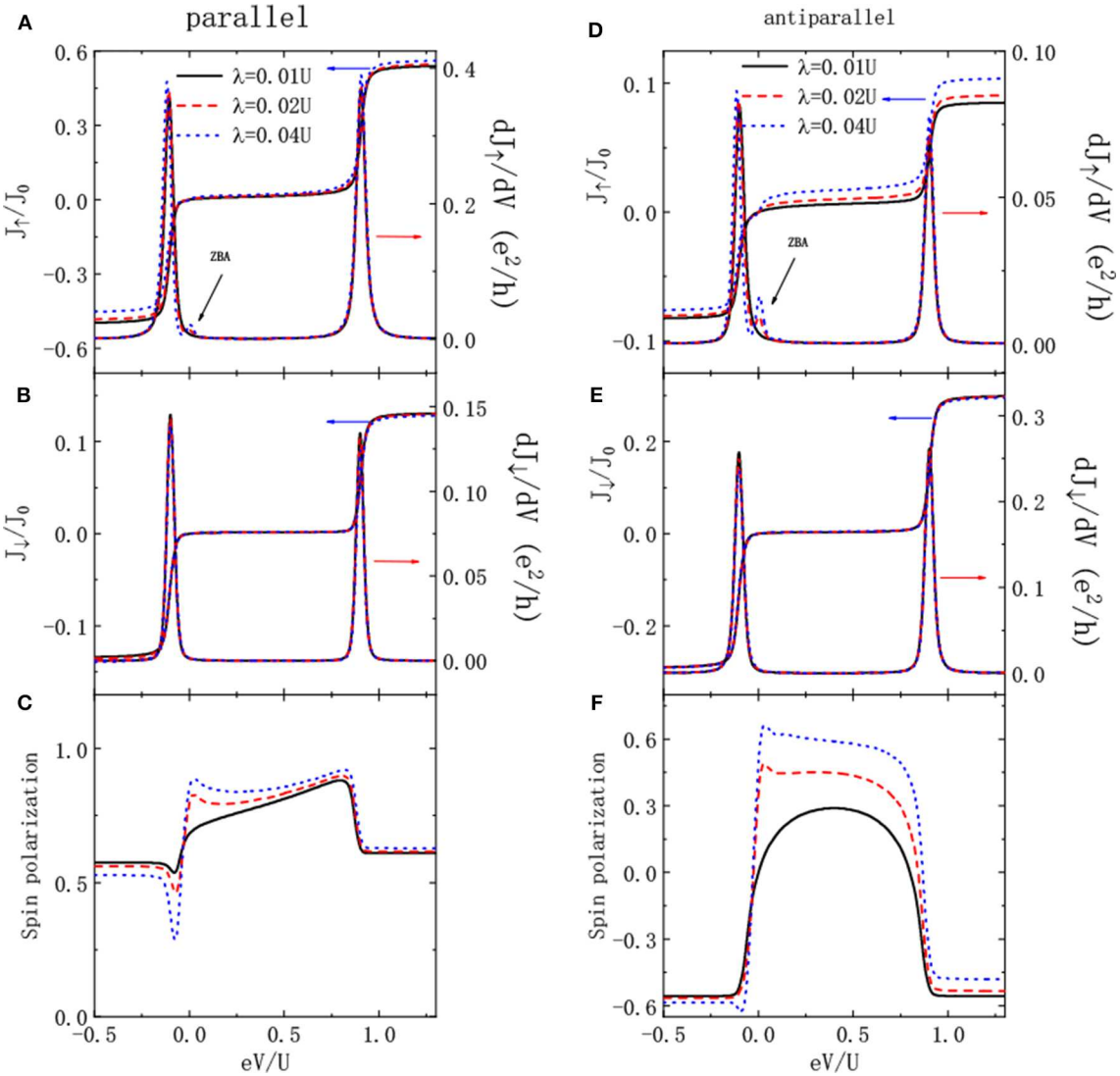

FIGURE 2 | Spin-resolved current, non-linear differential conductance, and current' spin polarization as functions of the bias voltage when the magnetic moments of the leads are arranged in parallel **(A–C)**, and antiparallel **(D–F)** configurations for indicated QD-MBS coupling strength $\lambda_1 = \lambda_2 = \lambda$. Other parameters are: $\delta_M = 0$, $\varepsilon_d = -0.1$, $\Gamma_L = \Gamma_R = k_B T = 0.01U$, $P_L = \pm P_R = 0.6$ (± for parallel and antiparallel cases, respectively).

antiparallel is more obvious than that of the parallel one, which results in negative TMR as will be shown in the following. For the differential conductance, an obvious zero bias anomaly (ZBA) emerges in addition to the increase in the intensity [39, 40]. The ZBA, which has been theoretically predicted and experimentally observed, originates from the QD-MBSs coupling and serves as a signal of the existence of the Majornana fermions [39, 40]. In the absence of bias voltage and Coulomb interaction, the conductance is 0.5 e^2/h by the coupling between the QD and the MBSs [44–46]. **Figures 2C,F** shows that the current's spin polarization is also enhanced in the presence of the MBSs in both parallel and antiparallel configurations. This is because the intensity of the spin-up current is increased by the MBSs whereas that of the spin-down one remains unchanged. Moreover, the spin polarization of the current changes more drastically near the dot level of ε_d as compared to $\varepsilon_d + U$ due to the spin blockade effect.

As demonstrated by previous work, the influence of the MBSs on the current is rather weak since a pair of Majarana fermions are charge neutral [20–22]. This can be seen from the total current $J = J_\uparrow + J_\downarrow$ in **Figures 3A,B**. The ZBA in the differential conductance, which is measured by the electrical tunnel spectroscopy, then is experimentally interesting (indicated by the arrows in the figure). To enlarge the signal of the MBSs in the current, we present the TMR in **Figure 3C**. For weak QD-MBSs hybridization strength $\lambda = 0.01U$, the TMR is positive as indicated by the solid line. It has two plateaus in the bias regimes of $eV < \varepsilon_d$ and $eV > \varepsilon_d + U$, in which the current remains almost unchanged [3]. In the Coulomb blockade regime ($\varepsilon_d < eV < \varepsilon_d + U$), the intensity of TMR increases with

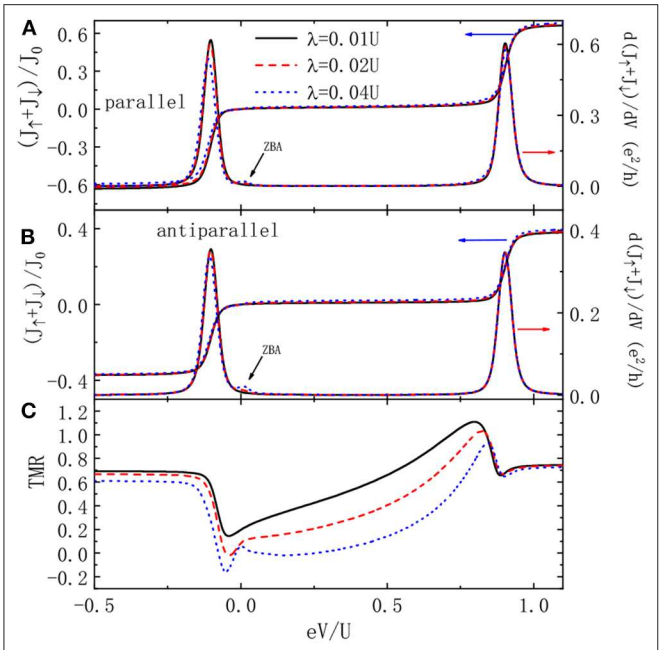

FIGURE 3 | Total current, non-linear differential conductance for the magnetic moments' configurations of parallel **(A)** and antiparallel **(B)**, and the corresponding TMR **(C)**. Other parameters are as in **Figure 2**.

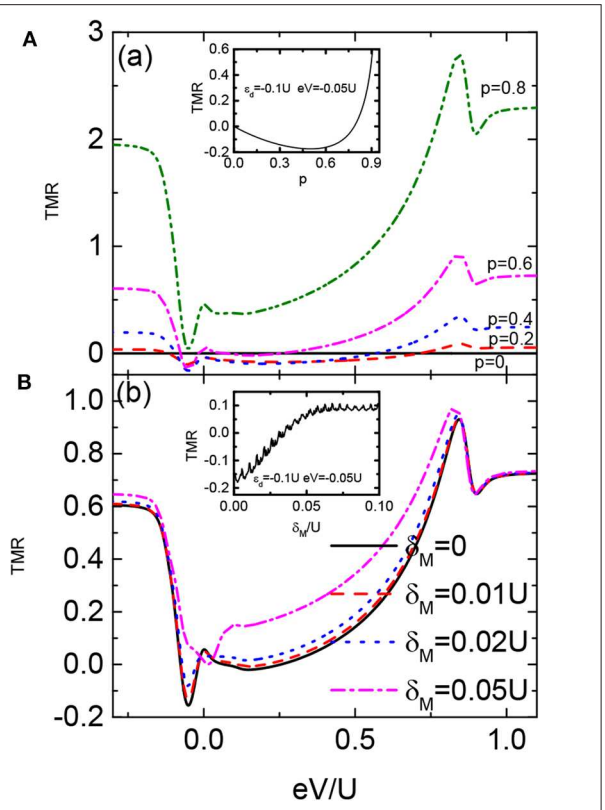

FIGURE 4 | TMR as a function of the bias voltage for different values of spin polarization of the leads p in **(A)**, and direct overlap between the MZMs δ_M in **(B)**. The insets in **(A,B)** are the TMR varying with the p and δ_M, respectively. Other parameters are as in **Figure 2**.

increasing bias voltage, which is quite different from the behavior of the current. This is because the current from the parallel configuration originates from transport between the majority-majority and minority-minority spin bands, and that in the antiparallel configuration between majority and minority spin bands. In the absence of the QD-MBSs coupling, the former transport process is rather easier than the latter case. When the electrons on the QD is coupled to the MBSs, however, there is more transport channels in the antiparallel configuration and then the current intensity is increased accordingly. With increasing bias voltage, the leakage current in the parallel configuration then will be larger than that of the antiparallel one, resulting in increased TMR. Around the two transport channel ε_d and $\varepsilon_d + U$, the TMR develops a dip and peak, respectively. The reason is that the current of the antiparallel configuration originating from transport through ε_d ($\varepsilon_d + U$) is enhanced (suppressed). With increased QD-MBSs coupling strength, the strength of the TMR is weakened as shown by the dashed and dotted lines in **Figure 3C**. For sufficiently large λ, the TMR around transport channel ε_d is negative (see the dotted line), and then serves as a detection method for the existence of the Majorana fermions in the superconductor nanowire side-coupled to the QD. We emphasize that this phenomenon originates from the fact that the current intensity of the antiparallel configuration is more sensitive to the MBSs as compared to the parallel one.

In **Figure 4** we present the TMR varying with respective to the bias voltage for different values of the spin-polarization of the electrodes ($p = P_L = P_R$) and overlap amplitude between the two MBSs δ_M. For $p = 0$, the TMR is zero as indicated by the solid line in **Figure 4A**. With increasing spin polarization, the line-shape of the TMR essentially resembles those in **Figure 3C**. The strength of TMR changes with respective to the spin polarization in a non-monotonic way. **Figure 4A** shows that the TMR is negative around dot level ε_d for $p <= 0.6$, and is positive for $P = 0.8$. The TMR in the inset of **Figure 4A** has a minimal value in the regime of $0.5 < p < 0.6$ for the indicated dot level and bias voltage. This is because that with increasing ferromagnetism on the electrodes p, the transport processes will be more and more dominated by only one spin-component electrons, and then the difference between the two kinds of the current becomes less prominent. In the extreme case of $p \sim 1$ (half-metallic), electron transport in the antiparallel is totally blockaded and the TMR equals to 1. Under this condition, the signal of the MBSs disappears. **Figure 4B** presents the function of δ_M on the TMR for fixed spin polarization of the electrodes $p = 0.6$. When the two states are decoupled from each other ($\delta_M = 0$), an obvious negative TMR emerges around the dot level ε_d. With increasing δ_M, we find that the strength of the TMR is enhanced, especially in the Coulomb blockade regime. For strong enough overlap between the two states $\delta_M = 0.05U$, the TMR is positive in the whole bias voltage regime. The reason is that when the two MBS are coupled, their functions on the TMR are canceled out and the abnormal increase of the current in the antiparallel configuration induced by the MBSs disappears.

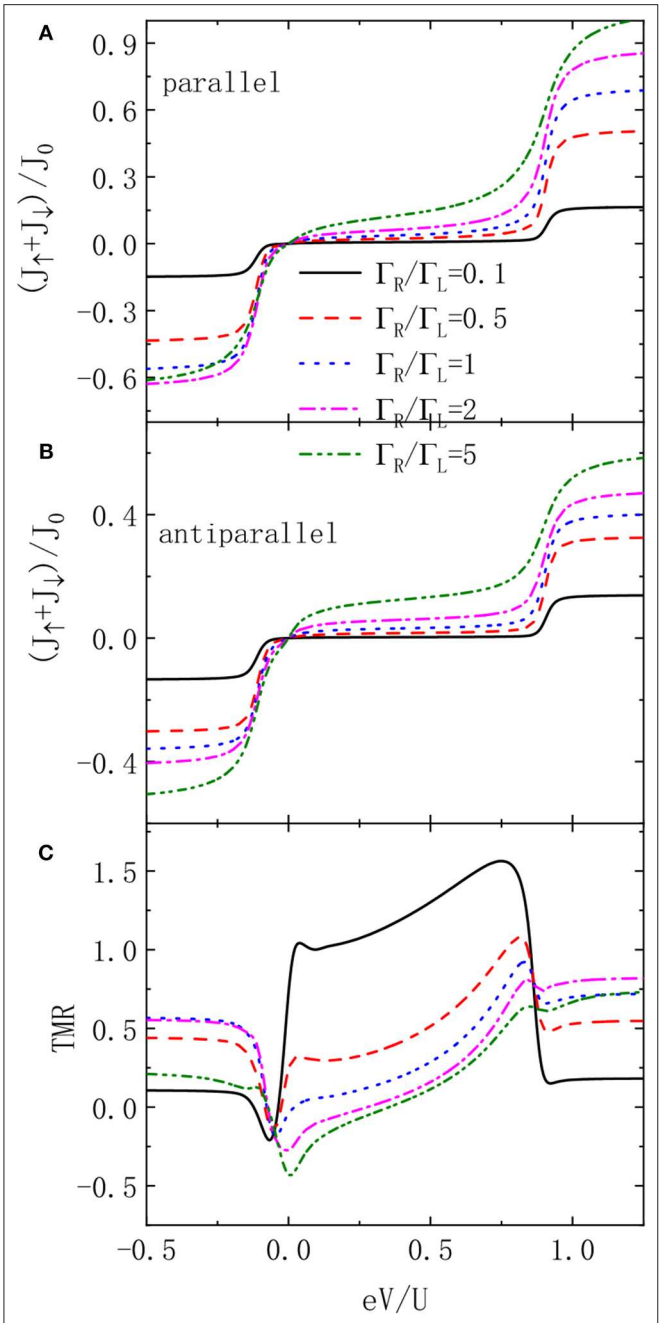

FIGURE 5 | Total currents for parallel **(A)** and antiparallel **(B)** configurations, and TMR **(C)** as functions of the bias voltage for different values of QD-lead coupling asymmetry q with $\Gamma_0 = 0.02U$. Other parameters are as in **Figure 2**.

In **Figure 5**, we present the total currents in both of the two configurations and the TMR varying with the bias voltage for different values of left-right asymmetry Γ_R/Γ_L. First of all, the currents' strength is monotonously enhanced by increasing Γ_R/Γ_L. This is because that for fixed ingoing tunneling rate Γ_L, the electrons will tunnel through the dot faster with increased outgoing tunneling rate Γ_R. Obviously, this holds true for both of the two spin component electrons regardless of the magnetic configuration. The TMR in **Figure 5C**, however, changes in a non-monotonic way with varying Γ_R/Γ_L. In the regimes of $eV < \varepsilon_d$ and $eV > \varepsilon_d + U$, the TMR is positive and its strength increases with increasing Γ_R/Γ_L. This is the usual case in the absence of the QD-MBSs. In the Coulomb blockade regime $\varepsilon_d < eV < \varepsilon_d + U$, however, the value of TMR decreases with increasing left-right asymmetry. As a consequence of it, the negative TMR becomes more prominent and becomes a good means to detect the existence of the MBSs. As mentioned above, this is induced by that fact that the QD-MBSs coupling strengthes the current intensity of the antiparallel configuration as compared to the parallel one, which is an unique function of the MBSs on the electron transport. Finally, we briefly discuss the experimental realization of the present devices. The nanowire hosting the MBSs can be fabricated with InAs grown by molecular beam epitaxy with several nanometers of epitaxial Al layer. It has been experimentally proven that an hard superconducting gap can be induced on such a kind of nanowires by applying a critical magnetic field exceeding 2T along the wire axis. A QD is formed in the bare InAs segment at the end of the wire due to density of state gradients at the edges of the Al shell.

4. SUMMARY

In summary, spin-dependent current and TMR in a QD sandwiched between two ferromagnetic leads and side-coupled to a pair of MBSs formed at the opposite ends of a superconductor nanowire is investigated within the non-equilibrium Green's technique. An unique negative TMR induced by the hybridization between the QD and the MBSs is found, which serves as a detection means of the Majorana fermions. This negative TMR is more likely to emerge in longer nanowire in which the two MBSs are well-separated from each other and the overlap between them is weak. By increasing the left-right asymmetry of the coupling strength between the QD and electrodes, the negative TMR becomes more obvious. It is also found that the intensity of the TMR depends on the spin-polarization of the electrodes in a non-monotonic way and is positive for large spin-polarization regardless of the existence of the MBSs.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

W-GM and L-WT contributed the ideas equally and performed the numerical calculations. L-WT derived the formulae in the paper and wrote the original manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos. 11772287 and 11572277).

REFERENCES

1. Loss D, DiVincenzo DP. Quantum computation with quantum dots. *Phys Rev A*. (1998) **57**:120–6. doi: 10.1103/PhysRevA.57.120
2. Fujisawa T, Austing DG, Tokura Y. Allowed and forbidden transitions in artificial hydrogen and helium atoms. *Nature*. (2002) **419**:278–81. doi: 10.1038/nature00976
3. Hanson R, Kouwenhoven LP, Petta JR. Spins in few-electron quantum dots. *Rev Mod Phys*. (2007) **79**:1217–65. doi: 10.1103/RevModPhys.79.1217
4. Julliere M. Tunneling between ferromagnetic films. *Phys Lett A*. (1975) **54**:225–6. doi: 10.1016/0375-9601(75)90174-7
5. Rudziński W, Barnaś J. Tunnel magnetoresistance in ferromagnetic junctions: tunneling through a single discrete level. *Phys Rev B*. (2001) **64**:085318:1–10. doi: 10.1103/PhysRevB.64.085318
6. Barnaś J, Fert A. Magnetoresistance oscillations due to charging effects in double ferromagnetic tunnel junctions. *Phys Rev Lett*. (1998) **80**:1058–61. doi: 10.1103/PhysRevLett.80.1058
7. Schelp LF, Fert A, Fettar F. Spin-dependent tunneling with Coulomb blockade. *Phys Rev B*. (1997) **56**:R5747–50. doi: 10.1103/PhysRevB.56.R5747
8. Mitani S, Fujimori H, Takanashi K. Tunnel-MR and spin electronics in metal-nonmetal granular systems. *J Magn Magn Mater*. (1999) **198**:179–84. doi: 10.1016/S0304-8853(98)01041-5
9. Zhang X, Li BZ. Spin-polarized tunneling and magnetoresistance in ferromagnet/insulator(semiconductor) single and double tunnel junctions subjected to an electric field. *Phys Rev B*. (1997) **56**:5484–8. doi: 10.1103/PhysRevB.56.5484
10. Beenakker CWJ. Theory of Coulomb-blockade oscillations in the conductance of a quantum dot. *Phys Rev B*. (1991) **44**:1646–56. doi: 10.1103/PhysRevB.44.1646
11. Averin DV, Korotkov AN, Likharev KK. Theory of single-electron charging of quantum wells and dots. *Phys Rev B*. (1991) **44**:6199–211. doi: 10.1103/PhysRevB.44.6199
12. Li SS, Chang K, Xia JB, Hirose K. Spin-dependent transport through Cd1-xMnxTe diluted magnetic semiconductor quantum dots. *Phys Rev B*. (2003) **68**:245306:1–5. doi: 10.1103/PhysRevB.68.245306
13. žutić I, Fabian J, Sarma SD. Spintronics: fundamentals and applications. *Rev Mod Phys*. (2004) **76**:323–410. doi: 10.1103/RevModPhys.76.323
14. Barnaś J, Martinek J, Michałek G. Spin effects in ferromagnetic single-electron transistors. *Phys Rev B*. (2000) **62**:12363–73. doi: 10.1103/PhysRevB.62.12363
15. Hamaya K, Masubuchi S, Kawamura M. Spin transport through a single self-assembled InAs quantum dot with ferromagnetic leads. *Appl Phys Lett*. (2007) **90**:053108:1–3. doi: 10.1063/1.2435957
16. Weymann I, Barns J. Transport through two-level quantum dots weakly coupled to ferromagnetic leads. *J Phys Condens Matter*. (2007) **19**:096208:1–21. doi: 10.1088/0953-8984/19/9/096208
17. Ma MJ, Jalil MBA, Tan SG. Effect of interface spin-flip scattering on the spin polarized transport through a quantum dot: Master equation approach. *J Appl Phys*. (2009) **105**:07E907:1–3. doi: 10.1063/1.3075982
18. Chi F, Zeng H, Yuan X. Flux-dependent tunnel magnetoresistance in parallel-coupled double quantum dots. *Superlattice Microst*. (2009) **46**:523–32. doi: 10.1016/j.spmi.2009.04.002
19. Michalek G, Bulka BR. Tunnel magnetoresistance in quantum dots in the presence of singlet and triplet states. *J Phys Condens Matter*. (2011) **23**:175305–15. doi: 10.1088/0953-8984/23/17/175305
20. Nayak C, Simon SH, Stern A. Non-Abelian anyons and topological quantum computation. *Rev Mod Phys*. (2008) **80**:1083–159. doi: 10.1103/RevModPhys.80.1083
21. Alicea J, Oreg Y, Refael G. Non-Abelian statistics and topological quantum information processing in 1D wire networks. *Nat Phys*. (2011) **7**:412–7. doi: 10.1038/nphys1915
22. Torsten Karzig RML Christina Knapp. Scalable designs for quasiparticle-poisoning-protected topological quantum computation with Majorana zero modes. *Phys Rev B*. (2017) **95**:235305:1–32. doi: 10.1103/PhysRevB.95.235305
23. Sato M, Fujimoto S. Topological phases of noncentrosymmetric superconductors: Edge states, Majorana fermions, and non-Abelian statistics. *Phys Rev B*. (2009) **79**:094504:1–16. doi: 10.1103/PhysRevB.79.094504
24. Qi XL, Zhang SC. Topological insulators and superconductors. *Rev Mod Phys*. (2011) **83**:1057–110. doi: 10.1103/RevModPhys.83.1057
25. Wimmer M, Akhmerov AR, Medvedyeva MV. Majorana bound states without vortices in topological superconductors with electrostatic defects. *Phys Rev Lett*. (2010) **105**:046803:1–4. doi: 10.1103/PhysRevLett.105.046803
26. Sau JD, Lutchyn RM, Tewari S. Generic new platform for topological quantum computation using semiconductor heterostructures. *Phys Rev Lett*. (2010) **104**:040502:1–24. doi: 10.1103/PhysRevLett.104.040502
27. Lutchyn RM, Sau JD, Sarma SD. Majorana fermions and a topological phase transition in semiconductor-superconductor heterostructures. *Phys Rev Lett*. (2010) **105**:077001:1–4. doi: 10.1103/PhysRevLett.105.077001
28. Choy TP, Edge JM, Akhmerov AR. Majorana fermions emerging from magnetic nanoparticles on a superconductor without spin-orbit coupling. *Phys Rev B*. (2011) **84**:195442:1–6. doi: 10.1103/PhysRevB.84.195442
29. San-Jose P, Prada E, Aguado R. AC Josephson effect in finite-length nanowire junctions with Majorana modes. *Phys Rev Lett*. (2012) **108**:257001:1–5. doi: 10.1103/PhysRevLett.108.257001
30. Mourik V, Zuo K, Frolov SM. Signatures of Majorana fermions in hybrid superconductor-semiconductor nanowire devices. *Science*. (2012) **336**:1003–7. doi: 10.1126/science.1222360
31. Wang R, Su W, Zhu JX. Kondo signatures of a quantum magnetic impurity in topological superconductors. *Phys Rev Lett*. (2019) **122**:087001:1–6. doi: 10.1103/PhysRevLett.122.087001
32. Lo´pez R, Lee M, Serra L. Thermoelectrical detection of Majorana states. *Phys Rev B*. (2014) **89**:205418:1–7. doi: 10.1103/PhysRevB.89.205418
33. Ricco LS, Dessotti FA, Shelykh IA. Tuning of heat and charge transport by Majorana fermions. *Sci Rep*. (2018) **8**:2790–7. doi: 10.1038/s41598-018-21180-9
34. Smirnov S. Universal Majorana thermoelectric noise. *Phys Rev B*. (2018) **97**:165434:1–14. doi: 10.1103/PhysRevB.97.165434
35. Hamaya K, Kitabatake M, Shibata K. Oscillatory changes in the tunneling magnetoresistance effect in semiconductor quantum-dot spin valves. *Phys Rev B*. (2008) **77**:081302(R):1–4. doi: 10.1103/PhysRevB.77.081302
36. Stefański P. Tunneling magnetoresistance anomalies in a Coulomb blockaded quantum dot. *Phys Rev B*. (2009) **79**:085312:1–9. doi: 10.1103/PhysRevB.79.085312
37. Sun QF, Xie XC. Spontaneous spin-polarized current in a nonuniform Rashba interaction system. *Phys Rev B*. (2005) **71**:155321:1–6. doi: 10.1103/PhysRevB.71.155321
38. Sun QF, Wang J, Guo H. Quantum transport theory for nanostructures with Rashba spin-orbital interaction. *Phys Rev B*. (2005) **71**:165310:1–11. doi: 10.1103/PhysRevB.71.165310
39. Ricco LS, de Souza M, Figueira MS. Spin-dependent zero-bias peak in a hybrid nanowire-quantum dot system: distinguishing isolated Majorana fermions from Andreev bound states. *Phys Rev B*. (2019) **99**:155159:1–9. doi: 10.1103/PhysRevB.99.155159
40. Piotr S. Properties of the Majorana-state tunneling Josephson junction mediated by an interacting quantum dot. *J Phys Condens Matter*. (2019) **31**:185301:1–18. doi: 10.1088/1361-648X/ab052a
41. Prada E, Aguado R, San-Jose P. Measuring Majorana nonlocality and spin structure with a quantum dot. *Phys Rev B*. (2017) **96**:085418:1–10. doi: 10.1103/PhysRevB.96.085418
42. Deng MT, Vaitiekenas S, Prada E, San-Jose P, Nygard J, Krogstrup P, et al. Nonlocality of Majorana modes in hybrid nanowires. *Phys Rev B*. (2018) **98**:085125:1–10. doi: 10.1103/PhysRevB.98.085125
43. Pals P, MacKinnon A. Coherent tunnelling through two quantum dots with Coulomb interaction. *J Phys Condens Matter*. (1996) **8**:5401–14. doi: 10.1088/0953-8984/8/29/015

44. Liu DE, Baranger HU. Detecting a Majorana-fermion zero mode using a quantum dot. *Phys Rev B*. (2011) **84**:201308R:1–4. doi: 10.1103/PhysRevB.84.201308
45. Vernek E, Penteado PH, Seridonio AC, Egues JC. Subtle leakage of a Majorana mode into a quantum dot. *Phys Rev B*. (2014) **85**:165304:1–5. doi: 10.1103/PhysRevB.89.165314
46. Ruiz-Tijerina AD, Vernek E, da Silva LGGVD, Egues JC. Interaction effects on a Majorana zero mode leaking into a quantum dot. *Phys Rev B*. (2015) **91**:115435:1–5. doi: 10.1103/PhysRevB.91.115435

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 03 June 2020
doi: 10.3389/fphy.2020.00193

Design of an Open Electrowetting on Dielectric Device Based on Printed Circuit Board by Using a Parafilm M

Zichuan Yi[1], *Haoqiang Feng*[2], *Xiaofeng Zhou*[2] *and Lingling Shui*[2]*

[1] College of Electron and Information, University of Electronic Science and Technology of China, Zhongshan Institute, Zhongshan, China, [2] South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China

In recent years, the open electrowetting on dielectric (EWOD) device has been widely used in biomedical detection, chemical synthesis analysis, and so on. However, the cost of using ITO glass as surface material is difficult to meet the requirement for large-scale array chip production. So, a low-cost, easy-to-manufacture open EWOD platform is designed in this paper. In hardware platform, an operation platform is prepared by using a printed circuit board (PCB) as a substrate. The electrode shape is designed as zigzag, and its surface is optimized by organic solderability preservatives (OSP). In addition, Parafilm M and silicone oil are used as a dielectric hydrophobic layer to prepare the open platform. In software, the system program is designed by C programming language, including initialization program, serial port communication program, high-voltage output port program, and interrupt program, which can be used to drive droplets. The system can achieve an effective driving voltage of 180–240 V. The moving speed of droplets can reach 15 mm/s when the droplet volume is 1850 μL and the electrode voltage output frequency is 10 Hz.

Keywords: electrowetting on dielectric, printed circuit board, Parafilm M, dielectric hydrophobic layer, open platform

OPEN ACCESS

Edited by:
Mohamed Swillam,
American University in Cairo, Egypt

Reviewed by:
Feng Chi,
University of Electronic Science and Technology of China, China
Yong Deng,
Dalian University of Technology, China
Ji-Pei Chen,
Guangzhou University, China

***Correspondence:**
Lingling Shui
shuill@m.scnu.edu.cn

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *29 March 2020*
Accepted: *30 April 2020*
Published: *03 June 2020*

Citation:
Yi Z, Feng H, Zhou X and Shui L (2020) Design of an Open Electrowetting on Dielectric Device Based on Printed Circuit Board by Using a Parafilm M. Front. Phys. 8:193. doi: 10.3389/fphy.2020.00193

INTRODUCTION

The open EWOD system has the advantages of saving samples, easy integration, multi-function, and precise operation [1–3], and so on. It has been used in many fields and has great development potential. Most of digital microfluidic systems are made on the basis of ITO glass currently [4], and it can realize the smooth movement of droplets, which is favored by the majority of scientific researchers and the market. However, the cost of ITO glass surface material is difficult to meet the needs of large-scale array chip production. It has been reported that a simple ITO glass EWOD device (4.5 × 1.5 cm^2) currently costs several hundred dollars [5]. In addition, a gap of several micrometers in width can be etched on the surface of the ITO glass by using a mask, but this means that all traces and electrodes can only be laid out on one side of the glass [6]. So, it has serious limitations on the compatibility and expansibility of devices. Therefore, in order to overcome those shortcomings, PCBs were used to prepare EWOD devices.

With the rapid development of electronics manufacturing industry, the PCB manufacturing industry has a unified and standardized manufacturing facility and process [7]. Moreover, it has the characteristics of low cost and high integration, which is suitable for making EWOD devices [8–10]. In current PCB manufacturing process, a PCB with a thickness of several millimeters can accommodate up to 30 wiring layers. It can achieve electrical performance connection by the via

process, which can improve wiring efficiency [11]. So, the PCB has been used as the base of a two-dimensional electrode array to fabricate EWOD devices as early as 2005 [12]. Once this scheme was proposed, many researchers did a lot of design research on PCB-based EWOD devices. However, there are still many shortcomings in the PCB-EWOD devices. For example, a closed PCB-based EWOD device manufacturing method was proposed to drive 1 μL droplet, and the manufacture of a dielectric layer (SU-8) and a hydrophobic layer (Teflon AF1600) on the PCB is completed by chemical vapor deposition (CVD) technology [13]. However, the droplet size is too small, and the closed structure brings inconvenience for droplets replacement. In addition, an open PCB-based EWOD device manufacturing method was proposed, and spin coating technology was used to deposit a layer of polydimethylsiloxane (PDMS) on the electrode, a 6 μL liquid drop can be driven at the voltage of 380 V after adding silicone oil, but the movement speed is only 2.71 mm/s [14]. So, the driving efficiency is low. In order to improve the performance of EWOD devices, a lot of research of electrode shape and material selection for the dielectric layer and the hydrophobic layer have been done [15, 16]. Among them, different electrode shapes (crescent, interdigitated, rectangular, notched, and zigzag edges) have been designed on the EWOD device. The results showed that the zigzag edge can drive droplets more effectively than other types [17]. But the driving voltage must be higher than 300 V for obtaining a droplet speed of 10 mm/s, which can cause the breakdown for general dielectric materials. In addition, a PCB-EWOD device which was produced by a 2 × 4 electrode array has been proposed, polyethylene films, and edible oil were used as dielectric materials. However, the device can only drive a 2.5 μL droplet with a high driving voltage and the number of electrodes was too small to be used in the large-scale manipulation [18].

In this paper, an open EWOD device is proposed based on a PCB, and a kind of zigzag edge is used for the design of electrode shape, Parafilm M film and silicone oil are used as the dielectric hydrophobic layer. The device can realize the driving of large droplets, and reduce costs at the same time. In addition, once the surface of the dielectric material is damaged, it can be torn off and replaced for further use. So, the proposed EWOD device can improve the integration of the system and has a certain cost advantage.

PRINCIPLE

Electrowetting Principle of Droplets

By adjusting the electric potential applied between the liquid and the electrode, the surface tension is changed, thus the contact angle is changed, which is called as electrowetting phenomenon. As shown in **Figure 1A**, it is the shape of the droplet when no voltage is applied, and a droplet is tiled as **Figure 1B** when voltage is applied. Its contact angle can be determined by Young's Equation [19], as shown in Equation (1).

$$\cos\theta_0 = \left(\gamma_{sg} - \gamma_{sl}\right)/\gamma_{\mathrm{lg}} \tag{1}$$

Where, γ_{sg}, γ_{sl}, γ_{lg} are surface tensions of solid-gas, solid-liquid, and liquid-gas, respectively, and θ_0 is the initial contact angle of the droplet. For hydrophobic surfaces, the contact angle of discrete droplets is always >90°. The hydrophobic solid-liquid interface tension becomes smaller when the driving voltage U is applied between the electrode and the droplet, and the relationship is described by the Lippmann Equation [20], as shown in Equation (2).

$$\gamma_{slU} = \gamma_{sl} - \varepsilon_0\varepsilon_r U^2/2d \tag{2}$$

Where, γ_{slU} is the surface tension of the solid-liquid interface when the voltage is applied, γ_{sl} is the surface tension when the voltage is canceled. ε_0 and ε_r are the dielectric constant in vacuum and the effective dielectric constant of the dielectric layer, respectively, d is the thickness of the dielectric layer. As the interfacial tension of the solid-liquid interface becomes smaller, the contact angle of the droplet becomes smaller, as shown in **Figure 1B**, the contact angle θ_U can be derived from (1) and (2), as shown in Equation (3) [21].

$$\cos\theta_U = \cos\theta_0 + \varepsilon_0\varepsilon_r U^2/2d \cdot \gamma_{\mathrm{lg}} \tag{3}$$

In Equation (3), the original contact angle θ_0 is a constant, the greater of the applied voltage, the smaller of the contact angle. In addition, the change of the contact angle is related to the dielectric constant and the thickness of the dielectric layer.

Droplet Movement Principle

The contact angle of a droplet on the electrode could decrease gradually when the voltage applied on the electrode increases gradually. The contact angle between two sides of the droplet will vary greatly when the voltage reaches a certain value. So, a large unbalance force ΔF can be formed in the droplet. When the force is greater than the friction between the droplet and the electrode, the droplet can be driven to one side gradually, which toward the direction of voltage [22]. As shown in **Figure 2**, and ΔF can be expressed as Equation (4).

$$\Delta F = \gamma\left(\theta_1 - \theta_2\right) = \frac{1}{2}CU^2 = \frac{1}{2}\frac{\varepsilon_r\varepsilon_0 A}{d}U^2 \tag{4}$$

Where, γ is the surface tension of the droplet; θ_1 is the initial contact angle of the droplet and θ_2 is the contact angle when the voltage is applied; C is a constant; ε_0 and ε_r are the dielectric constant of the dielectric layer and the dielectric constant in vacuum, respectively. d is the thickness of the dielectric layer, A is the area of the droplet; U is the driving voltage.

Bottom Area Calculation of a Droplet on an EWOD

In order to obtain the relationship between the size of an electrode and the minimum droplet size which can be driven, the volume of a droplet is quantified as the bottom radius of the droplet. The bottom radius of the droplet can be calculated by using the spherical void formula [23], and the corresponding minimum droplet volume can be estimated by comparing with the size of the electrode, as shown in Equation (5).

$$V = \pi H^2\left(R - \frac{H}{3}\right) \tag{5}$$

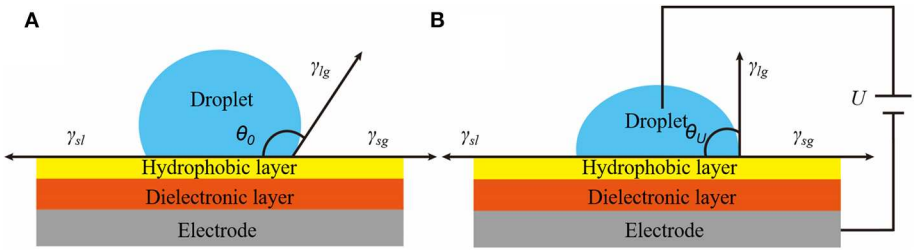

FIGURE 1 | Schematic diagram of the droplet electrowetting. **(A)** Schematic diagram of a droplet contact angle without voltage application. **(B)** Schematic diagram of a droplet contact angle when the voltage is applied.

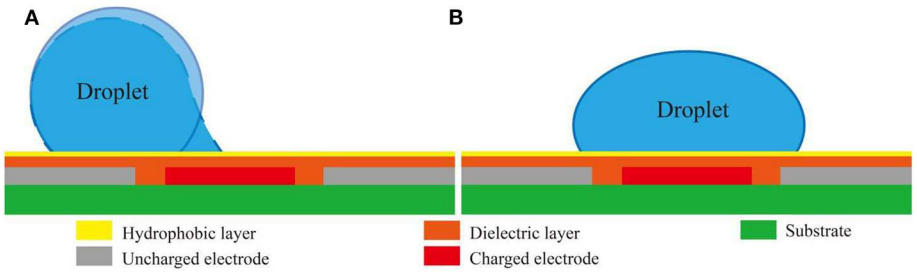

FIGURE 2 | Schematic diagram of droplets moving on two adjacent electrodes. **(A)** The schematic diagram of a droplet moving toward a charged electrode from a stationary initial state. **(B)** The schematic diagram of a droplet when it reaches the charged electrode.

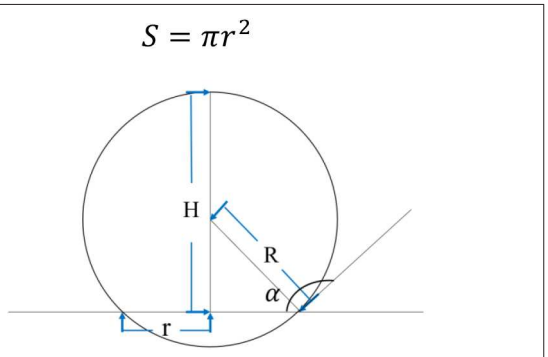

FIGURE 3 | Schematic diagram of the shape of a droplet on the EWOD.

Where, V is the volume of a droplet, H is the height of the droplet on the EWOD, R is the radius of the droplet. The shape diagram of a droplet on the EWOD is shown in **Figure 3**. The volume of a droplet can be obtained by using a pipette gun to fill the electrode in each experiment. So, in the static state, the relationship among the bottom radius of the droplet, the radius of the droplet and the contact angle can be calculated, as shown in Equation (6).

$$r = R\sin\left(180^\circ - \alpha\right) \tag{6}$$

Where, r is the radius of a droplet on the bottom of an electrode and α is the contact angle between the droplet and the electrode, and α can be obtained by a contact angle meter. Then, the bottom area S can be calculated as Equation (7).

$$S = \pi r^2 \tag{7}$$

DEVICE FABRICATION

Preparation of a PCB-Based Electrode Array

The PCB supports flexible or rigid substrate, and electrode arrays can be fabricated on rigid or flexible substrate. Rigid FR4 has good mechanical and electrical properties, and it is more cost-effective and attractive. In order to develop low-cost and high-compatibility EWOD devices, the rigid FR4 material is used to produce the PCB. Comparison of different materials which can be used to produce the PCB are shown in **Supplementary Table 1**. The table of detailed information is shown in **Supplementary Material**.

The edge zigzag is applied to the electrode, as shown in **Figure 4A**. The electrode surface treatment process is also very important. The electrode surface treatment process has a key impact on the movement of droplets. Electrode treatment processes include: lead tin spraying, lead-free tin spraying, gold deposition, gold plating, OSP and tin deposition, etc. The materials involved include tin, gold and antioxidants. All surface treatment processes are completed on the base copper of the PCB, as shown in **Figure 4B**. The electrode with different surface treatment has no effect on the electrical performance of PCBs, but its roughness which has a bad effect on droplets movement cannot be ignored. The comparison of electrode

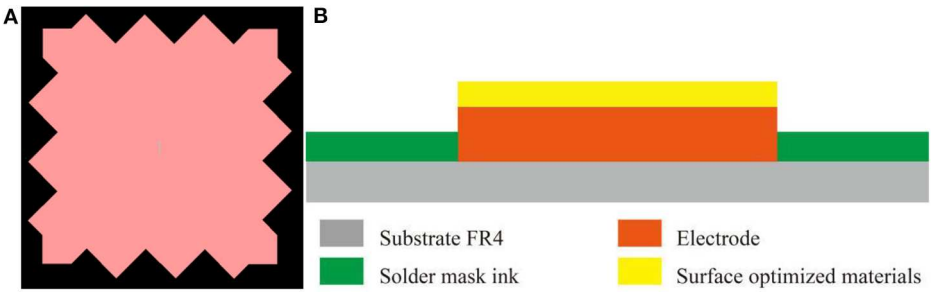

FIGURE 4 | (A) The zigzag electrode schematic diagram, every "tooth" is 500 μm and the depth of "tooth" is 350 μm. **(B)** Single PCB electrode schematic diagram, and the surface treatment material is covered on electrode.

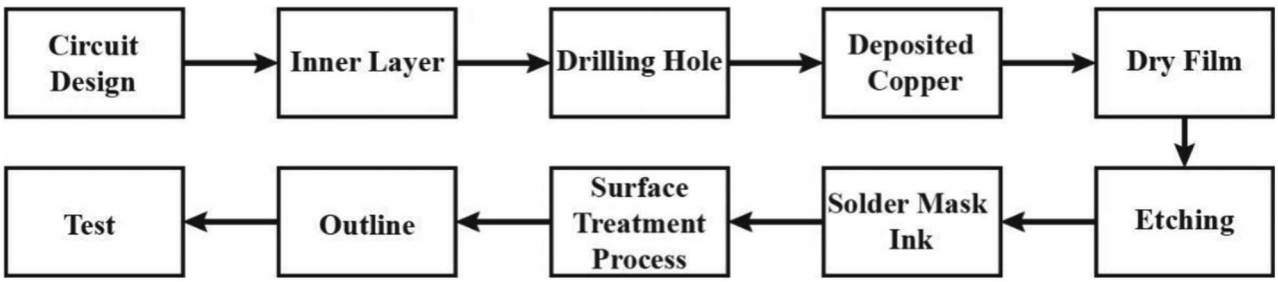

FIGURE 5 | The steps for manufacturing PCBs. There are main ten steps for manufacturing PCBs. Firstly, a correct circuit diagram is generated a Gerber film. According to the diagram, the inner layers are bonded together, the electrical connection is conducted by drilling, the copper is deposited in the hole as the base of electroplated copper, a layer of dry film is pasted on the layout, and the desired pattern can be obtained by etching, and then, the electrode is surface-treated for outline and testing.

surface roughness and price of three common surface treatment processes is shown in **Supplementary Table 2**. The table of detailed information is shown in **Supplementary Material**.

It can be seen that the surface roughness of the electrode obtained by the lead-free tin spray process is the largest, while the roughness of the immersion gold process and the OSP process is relatively small, so the immersion gold and OSP process are preferred. But OSP has the lowest price, and it is selected for surface treatment process in this paper. The steps for manufacturing PCBs are shown in **Figure 5**.

According to the design requirements, a zigzag electrode with an electrode size of 2×2 mm^2 is designed. The number of electrodes is 8×8, and an electrode array with a distance between the two electrodes of 140 μm is shown in **Figure 6.**

Dielectric Layer

In EWOD devices, the dielectric layer is an important part, which can determine the device performance. At present, the reported dielectric materials used in EWODs are mainly divided into polymer dielectric material and inorganic dielectric material [24]. Because the ITO glass surface is flat. The Teflon can be used for EWODs based on ITO glass by a spin coating process [25].

However, in a standard PCB process, the height difference between the electrodes and the substrate is in the range of 18–70 μm, and the thickness is measured by using an optical microscope. The gap between the two electrodes is 140 μm and the surface roughness of an electrode is about 3.1 μm, which is measured by using a step profiler. The diagram of the two adjacent electrodes is shown in **Figure 7**. Surface defects can reduce the fluidity of droplets, but the dielectric layer made by ordinary spin coating process cannot fill this huge height difference, and then, the movement of the droplet can be blocked, which could make the EWOD device unusable. So, the Parafilm M is used to realize a "bridge" between two adjacent electrodes, so as to eliminate the influence of the "huge" gap between electrodes in the droplet movement process. The electrowetting performance of a Parafilm M is shown in **Figure 8**. It can be seen that the contact angle change of a Parafilm M is 43° in the voltage range of 0–340V, and it can be increased with the increase of driving voltage. So, the Parafilm M can produce enough contact angle changes to drive the droplets. Its parameters are shown in **Supplementary Table 3**, The table of detailed information is shown in **Supplementary Material**. The initial contact angle of the Parafilm M is more than 90° before and after stretching, which shows hydrophobicity. And the Parafilm M after stretching is used as the dielectric layer by us.

RESULTS AND DISCUSSION

Hardware Design of the Driving System

According to the principle of the electrowetting, the driving system is designed to achieve the manipulation of micro-droplets,

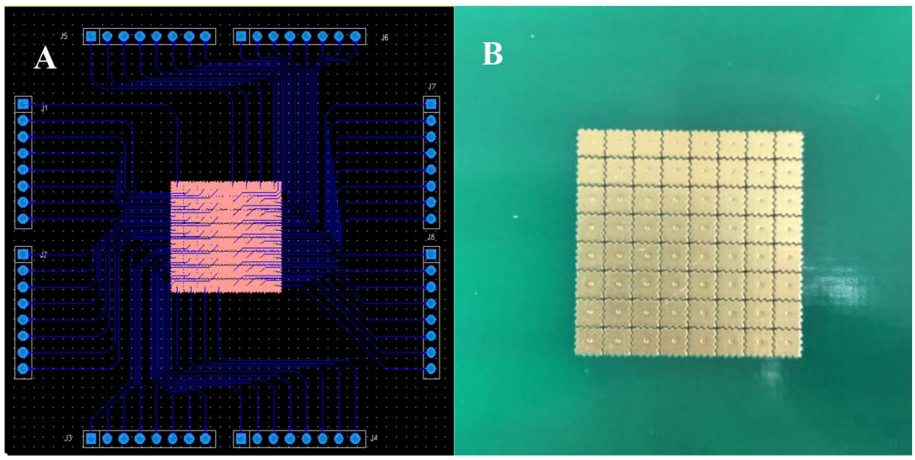

FIGURE 6 | PCB-based EWOD operating platform. **(A)** The design of a PCB electrode array. **(B)** The physical picture of the PCB electrode array.

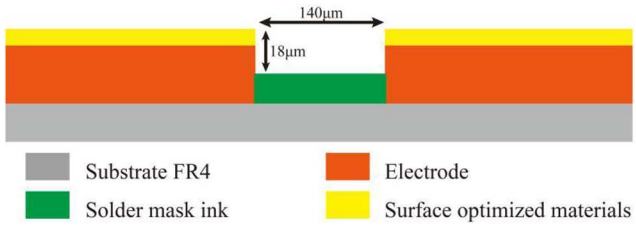

FIGURE 7 | Schematic diagram of two adjacent electrodes on a PCB-based EWOD.

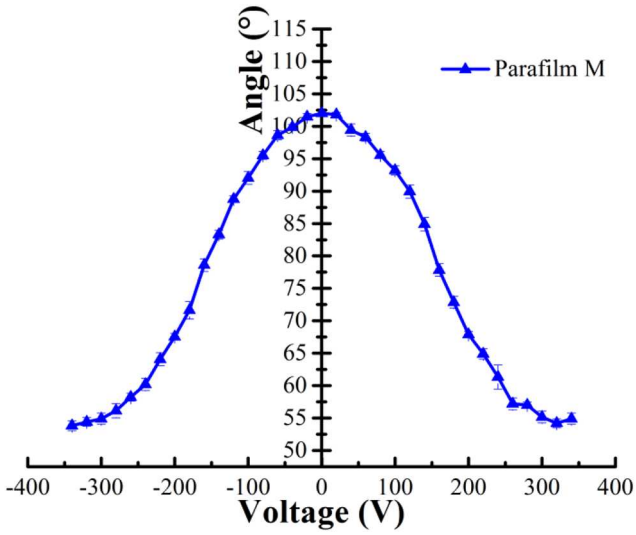

FIGURE 8 | Electrowetting characteristic curve of the Parafilm M.

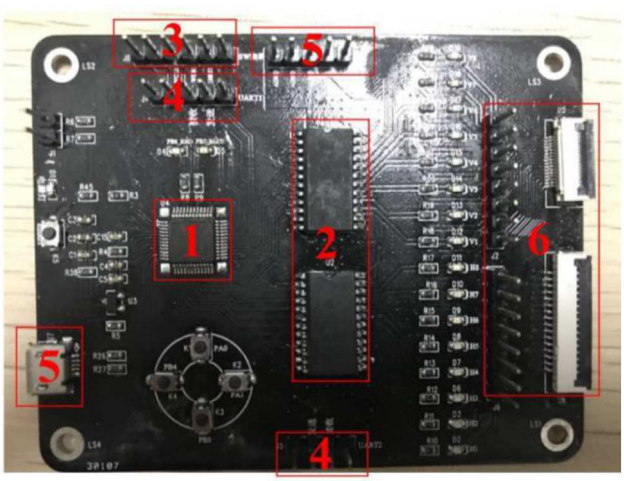

FIGURE 9 | Driving system hardware physical picture.

the driver board is shown in **Figure 9**. The driver board can accept an input voltage of 5–250 V. It has 16 editable high-voltage outputs without setting output circuits. In addition, a built-in protection circuit and a short-circuit detection section are designed in the driver. Compared with a driving circuit using a large number of relays and logic components [26], the proposed driving system is simple. As shown in **Figure 9**, ① is the control IC which is the core part of the driving system and a STM32 single-chip is used as the main chip. Its core model is Cortex-M3 and it can run up to 72 MHz. ② is a driving module, which mainly consists of a HV513 chip developed by Supertex company and some external circuits. Its function is to provide high output voltage for driving droplets. The HV513 chip can output the external power supply voltage in parallel according to the program, and it can output high resistance state at all ports. At the same time, it can control the output voltage within the range of 5–250 V, and has the function of short circuit automatic monitoring. ③ is a burning module, which uses ST-LINK V2 simulation programmer to burn the program into STM32 chip; ④ is a function serial port module, which can use DuPont line

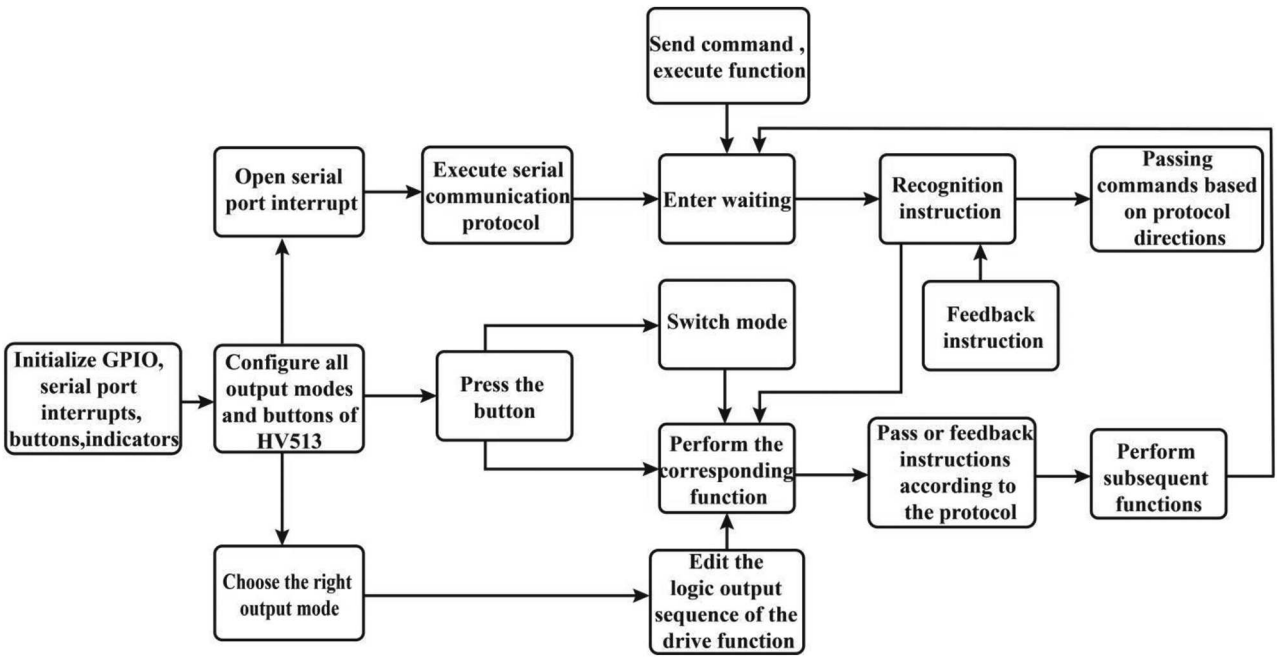

FIGURE 10 | System software flow chart. Each function is in the original standby state, and then, the system starts to configure the output mode of the HV513, which can be divided into three ways to execute the function. The first is to directly execute the original output sequence to complete the corresponding functions. The second is to switch to other output sequences by switching mode keys to perform corresponding functions. The third is to execute the serial program to perform the corresponding functions.

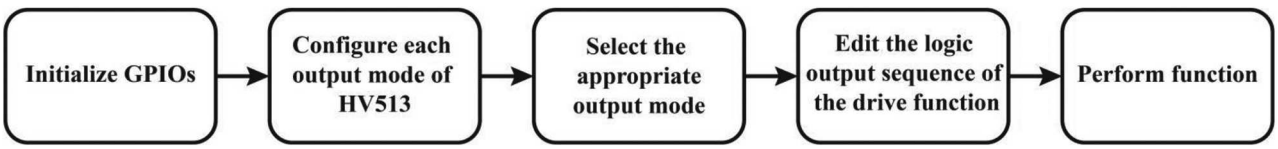

FIGURE 11 | High-voltage parallel output function sequence.

to connect two or more driver boards for achieving instruction transfer; ⑤ is a power module, a PSW800-1.44 DC power supply is used to drive the system. ⑥ is a high-level output port module, which is connected to the EWOD device to provide a high-level voltage. The driving system has two sets of communication serial ports, which are used for the connection between the systems. The maximum communication rate of the system can reach 4.5 Mbps.

Design of Software for EWODs

The completion of the hardware circuit provides a basis for the software system. In order to accurately control the movement of droplets, the software system flowchart is shown in **Figure 10**.

In system initialization, the port parameters and function initialization settings of each module must be configured in the system, including the initial configuration of GPIOs, communication serial ports, interrupts, and so on.

In order to control the movement of droplets, the output logic sequence of the high-voltage port of the system must be set. So, we initialize the function serial port, then configure the output mode, edit the logic output sequence, and output the driving voltage finally. The specific procedure steps are shown in **Figure 11**.

Performance Testing

In order to achieve a smooth movement of droplets, a layer of silicone oil is added on the Parafilm M, whose organic groups are methyl. Methyl silicone oil has good chemical stability, insulation, and good hydrophobic properties. Then, a pipette gun is used to suck up 3 μL of silicone oil and drip it right in the middle of the electrode array, then, the whole device is left for 5 min in a dust-free box to ensure that the silicone oil is spread evenly. The entire EWOD system connection is shown in **Figure 12**. This system consists of four driving boards and an EWOD platform which are connected by wires. The four driving boards provide the driving voltage for the operation platform by the communication interconnection, and the output sequence of the driving voltage can be adjusted by programming in the

STM32. Any one driver board can send completion signal to another driver board when its own instructions are completed. Then, the next driver board can continue to send instructions in order. The system has a total of 64 output voltage terminals which can drive all electrodes in the EWOD. In the process of the operation, STM32 output an instruction signal to a HV513 chip according to the program code, and then, the HV513 can output the high voltage to control the droplet movement on the EWOD device by executing the received command signal.

In the testing process, the control variable method is used for obtaining the relationship between the driving voltage and the droplet moving speed with the same droplet volume, and the relationship between the droplet volume and the moving velocity with the same driving voltage.

Firstly, we only changed the driving voltage and tested the relationship between driving voltage and droplet moving speed.

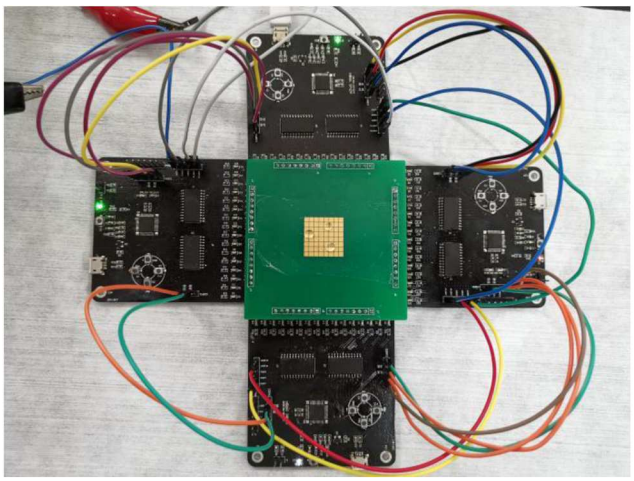

FIGURE 12 | Connection diagram of the EWOD system.

As shown in **Figure 13A**, the output frequency of the electrode driving voltage which refers to the inverse time of two adjacent electrodes is 5 Hz, a 25 μL droplet can be driven to 6.32 mm/s when the driving voltage is >150 V. Then, the droplet velocity can be increased when the voltage is gradually increased. And the droplet moving speed can reach 15.03 mm/s when the driving voltage is 240 V. In addition, the movement stability of the droplets is high during the experiment, and there is no phenomena such as droplet stop, sideslip loss and so on. So, the effective driving voltage of the EWOD device is within 150–240 V, and the speed of a droplet can be increased with the increase of the driving voltage. And then, a fixed driving voltage of 240 V is applied, the size of different droplet volumes and different electrode output frequencies are adjusted, and the droplet volume-velocity relationship curve of the Parafilm M is shown in **Figure 13B**.

The curve in **Figure 13B** is measured at three different frequencies. It can be seen that the higher the frequency of the output voltage is, the greater the average moving speed of the droplets is when the droplet volume is the same as each other. Because of a low frequency, a droplet moved to the second electrode, and the third electrode has not yet been applied driving voltage, so the droplet can stay on the second electrode, resulting in "stuck" phenomenon, as shown in **Figure 14**. However, the droplet has no time to contact the third electrode if the frequency is too high, and the voltage has been switched to the fourth electrode at this time, then, the droplet is "lost." As shown in **Figure 15**, a droplet with a volume of 15 μL can be driven normally at a frequency of 2 Hz, but it cannot be driven continuously at a frequency of 5 Hz.

The continuous driving of the droplet cannot be realized when the droplet volume is <10 μL (electrode size is 2 × 2 mm^2), no matter how much the driving output frequency is. This is because the bottom area of the droplet is less than the electrode area, and the droplet cannot always contact the next electrode during the moving process, so the droplet cannot move continuously. When

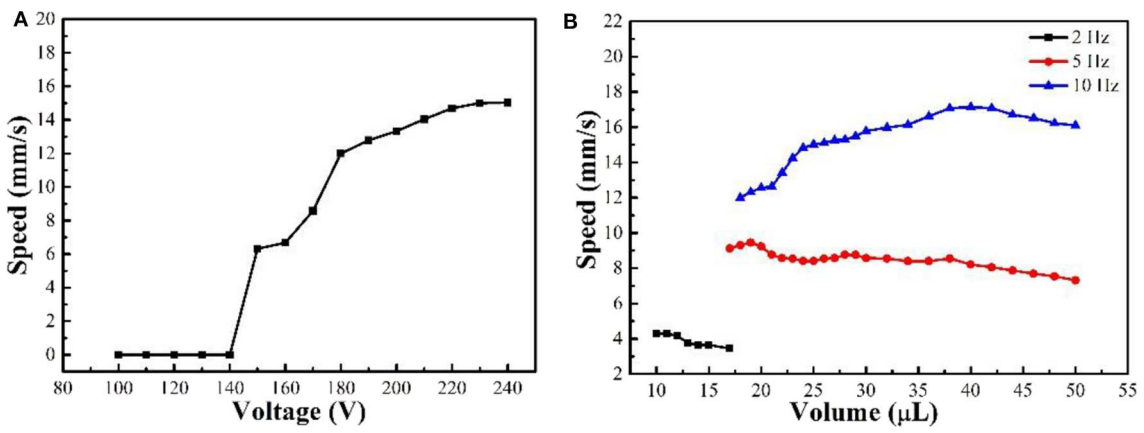

FIGURE 13 | Device performance testing. **(A)** when the droplet volume is 25 μL, and the dielectric hydrophobic layer is Parafilm M & silicone oil, the relationship between the driving voltage and the droplet moving speed. **(B)** When the driving voltage is 240 V, the relationship between the droplet volume and the droplet moving speed.

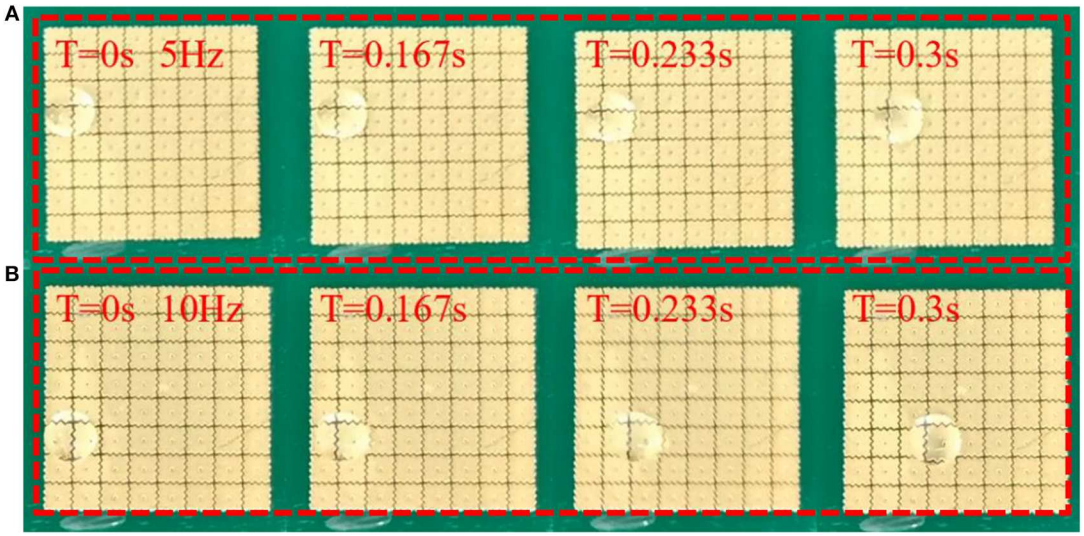

FIGURE 14 | Position comparison of droplets which have the same volume during the driving process when switching frequencies between two adjacent electrodes are 5 and 10 Hz. **(A)** When the electrode switching frequency is 5 Hz, the droplet is driven from the initial position to the third electrode after 0.3 s. **(B)** When the electrode switching time is 10 Hz, the droplet is driven from the initial position to the fourth electrode after 0.3 s.

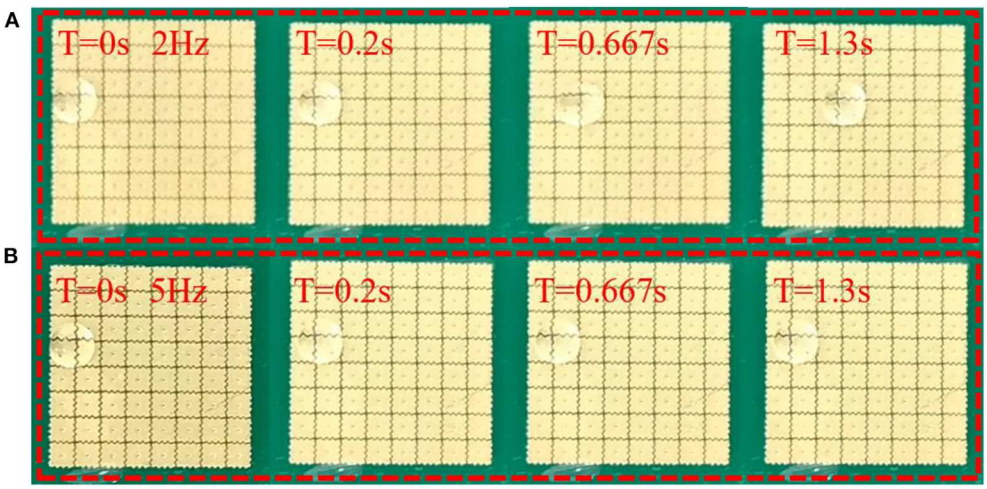

FIGURE 15 | When the droplet volume is 15 μL, the comparison chart of droplets when the electrode switching frequency are 2 and 5 Hz. **(A)** The droplet moved to the fourth electrode from the initial position after 1.3 s when the electrode switching frequency is 2 Hz. **(B)** The droplet has no movement from the initial position after 1.3 s when the electrode switching frequency is 5 Hz.

the output frequency is 10 Hz and the droplet volume is >18 μL, the droplet velocity can be increased with the volume increase of the droplet, and the velocity can be decreased until the droplet volume is >40 μL. The main reason is that the 240 V voltage can provide enough driving force to the 18–40 μL droplets. As the volume is increased, the effective contact area between the droplet and the electrode also can be increased, so the droplet can obtain a greater driving force. However, the driving force provided by the 240 V is not enough to drive the droplet when the volume is increased to a certain extent, and then, the velocity of the droplet could be decreased. In order to improve the driving speed, the driving voltage and output frequency must be increased at the same time. In summary, the EWOD prepared in this paper can drive the droplets of 40 μL at the driving voltage of 240 V with an output frequency of 10 Hz, and the moving speed of the droplet can reach 16 mm/s.

CONCLUSION

An EWOD digital microfluidic operating platform based on the PCB which is convenient in manufacture is proposed in this

paper, the Paraflim M and silicone oil are used as the dielectric hydrophobic layer for low cost production, and a driver board composed of STM32 and HV513 chips are designed as the control system. Then, the EWOD device can drive large droplets, and the using of removable media material can eliminate cross contamination and expand chip compatibility. So, the proposed EWOD system not only has simple preparation steps, but also reduces production cost, which has a certain practical value.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

ZY and LS designed this project. HF and XZ carried out most of the experiments and data analysis.

FUNDING

This research was funded by the Guangdong Basic and Applied Basic Research Foundation (no. 2020A1515010420), the Key Research Platforms and Research Projects in Universities and Colleges of Guangdong Provincial Department of Education (no. 2018KQNCX334), the Zhongshan Innovative Research Team Program (no. 180809162197886), the Zhongshan Institute high-level talent scientific research startup fund project (no. 416YKQ04), the Project for Innovation Team of Guangdong University (no. 2018KCXTD033), and the National Key R&D Program of China (no. 2018YFB0407100-02).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.00193/full#supplementary-material

REFERENCES

1. Samiei E, Tabrizian M, Hoorfar M. A review of digital microfluidics as portable platforms for lab-on a-chip applications. *Lab Chip*. (2016) **16**:2376–96. doi: 10.1039/C6LC00387G
2. Decrop D, Ruiz E, Kumar PT, Tripodi L, Kokalj T, Lammertyn J. Digital microfluidics assisted sealing of individual magnetic particles in femtoliter-sized reaction wells for single-molecule detection. *Methods Mol Biol*. (2017) **1547**:85–101. doi: 10.1007/978-1-4939-6734-6_7
3. Torabinia M, Asgari P, Dakarapu US, Jeon J, Moon H. On-chip organic synthesis enabled by engine-and-cargo in an electrowetting-on-dielectric digital microfluidic device. *Lab Chip*. (2019) **19**:3054–64. doi: 10.1039/C9LC00428A
4. Shen HH, Fan SK, Kim CJ, Yao DJ. EWOD microfluidic systems for biomedical applications. *Microfluidics Nanofluidics*. (2014) **16**:965–87. doi: 10.1007/s10404-014-1386-y
5. Moschou D, Tserepi A. The lab-on-PCB approach: tackling the μTAS commercial upscaling bottleneck. *Lab Chip*. (2017) **17**:1388–450. doi: 10.1039/C7LC00121E
6. Abdelgawad M, Wheeler AR. Low-cost, rapid-prototyping of digital microfluidics devices. *Microfluidics Nanofluidics*. (2008) **4**:349–55. doi: 10.1007/s10404-007-0190-3
7. Meimandi A, Seyedsadrkhani N, Jahanshaki A. Development of an electrowetting digital microfluidics platform using low-cost materials. In: *2019 27th Iranian Conference on Electrical Engineering (ICEE)*. Yazd: IEEE (2019). doi: 10.1109/IranianCEE.2019.8786395
8. Zou FX, Ruan QY, Lin XY, Zhang MX, Song YL, Zhou LJ, et al. Rapid, real-time chemiluminescent detection of dna mutation based on digital microfluidics and pyrosequencing. *Biosensors Bioelectron*. (2018) **126**:551–7. doi: 10.1016/j.bios.2018.09.092
9. Jain V, Raj TP, Deshmukh R, Patrikar R. Design, fabrication and characterization of low cost printed circuit board based ewod device for digital microfluidics applications. *Microsyst Technol*. (2015) **23**:389–97. doi: 10.1007/s00542-015-2680-7
10. Jain V, Vasavi D, Rajendra P. Study of two-dimensional open EWOD system using printed circuit board technology. *Global J Res Eng*. (2017) **17**:59–66.
11. Alistar M, Pop P. Towards droplet size-aware biochemical application compilation for AM-EWOD biochips. In: *Symposium on Design IEEE*. (2015). doi: 10.1109/DTIP.2015.7161024
12. Gong M, Kim CJ. Two-dimensional digital microfluidic system by multilayer printed circuit board. Micro electro mechanical systems. In: *2005 18th IEEE International Conference on Micro Electro Mechanical Systems*. Miami Beach, FL: IEEE (2005).
13. Macro N, Nicola L, Pablo RL, Domenico C, Giampiero DC, Augusto N. 2-D digital microfluidic system for droplet handling using printed circuit board technology. In: *2015 XVIII AISEM Annual Conference*. Trento: IEEE (2015).
14. Jain V, Hole A, Deshmukh R, Patrikar R. Dynamic capacitive sensing of droplet parameters in a low-cost open ewod system. *Sens Actuators A: Phys*. (2017) **263**:224–33. doi: 10.1016/j.sna.2017.06.014
15. Jang LS, Hsu CY, Chen CH. Effect of electrode geometry on performance of ewod device driven by battery-based system. *Biomed Microdevices*. (2009) **11**:1029–36. doi: 10.1007/s10544-009-9320-x
16. Xu XW, Sun LN, Chen LG, Zhou ZZ, Xiao JJ, Zhang YL. Electrowetting on dielectric device with crescent electrodes for reliable and low-voltage droplet manipulation. *Biomicrofluidics*. (2014) **8**:064107. doi: 10.1063/1.4902554
17. Jain V, Hole A, Deshmukh R, Patrikar R. Effect of electrode geometry on droplet velocity in open EWOD based device for digital microflfluidics applications. *J Electrostatics*. (2017) **87**:11–8. doi: 10.1016/j.elstat.2017.02.006
18. Zulkepli SNIS, Hamid NH, Shukla V. A low cost open droplet-based microfluidic decvices on printed circuit board. In: *2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*. Penang: IEEE (2017). doi: 10.1109/ICCSCE.2017.8284405
19. Yamaguchi Y, Kusudo H, Surblys D, Omori T, Kikugawa G. Interpretation of Young's equation for a liquid droplet on a flat and smooth solid surface: mechanical and thermodynamic routes with a simple Lennard-Jones liquid. *J Chem Phys*. (2019) **150**:044701. doi: 10.1063/1.5053881
20. Wang J, Zhang J, Liu S. The accurate control of the La2O3 hierarchical structure of the flower-like microspheres accompanied with vertical-standing nanopetals and the electrowetting responses with low voltage actuation. *Ceramics Int*. (2020) **46**:607. doi: 10.1016/j.ceramint.2020.01.147
21. Dreyer W, Guhlke C, Landstorfer M, Muller R. New insights on the interfacial tension of electrochemical interfaces and the lippmann equation. *Eur J Appl Math*. (2018) **29**:708–53. doi: 10.1017/S0956792517000341
22. Gong J, Kim CJ. Direct-referencing two-dimensional-array digital microfluidics using multilayer printed circuit board. *J Microelectromechanical Syst*. (2008) **17**:257–64. doi: 10.1109/JMEMS.2007.912698

23. Abdelgawad M, Park P, Wheeler AR. Optimization of device geometry in single-plate digital microfluidics. *J Appl Phys*. (2009) **105**:094506. doi: 10.1063/1.3117216
24. Liu H, Dharmatilleke S, Maurya DK, Tay AAO. Dielectric materials for electrowetting-on-dielectric actuation. *Microsyst Technol*. (2009) **16**:449–60. doi: 10.1007/s00542-009-0933-z
25. Chao JP. An Q. Liu Z P. Electrowetting on liquid-infused membrane for flexible and reliable digital droplet manipulation and application. *Sens Actuators B: Chem*. (2019) **291**:470–7. doi: 10.1016/j.snb.2019.04.102
26. Li CQ, Zhang KD, Wang XB, Zhang J, Liu H, Zhou J. Feedback control system for large scale 2D digital microfluidic platforms. *Sens Actuators B:Chem*. (2017) **255**:3616–22. doi: 10.1016/j.snb.2017.09.071

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer FC declared a shared affiliation with one of the authors, ZY, to the handling editor at time of review.

frontiers in Physics

ORIGINAL RESEARCH
published: 22 July 2020
doi: 10.3389/fphy.2020.00254

Photon-Assisted Transport Through a Quantum Dot Side-Coupled to Majorana Bound States

Feng Chi[1], Tian-Yu He[1,2], Jing Wang[1,2], Zhen-Guo Fu[3]*, Li-Ming Liu[1], Ping Liu[1] and Ping Zhang[3]*

[1] School of Electronic and Information Engineering, University of Electronic Science and Technology (UEST), Zhongshan Institute, Zhongshan, China, [2] South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China, [3] Institute of Applied Physics and Computational Mathematics, Beijing, China

Transmission function of a system composing of a quantum dot (QD) subjected to a photon field and side-coupled to a topological superconductor nanowire hosting a pair of Majorana bound states (MBSs) is calculated by using the non-equilibrium Green's function technique. We find that a series of photon-induced peaks emerge and are split by the coupling between the QD and the MBSs. Moreover, the peaks' height are suppressed to zero because the MBSs absorb (emit) the photon energy. Under this condition, the MBSs may be shifted to the non-zero energy mode, and thus provide another detection scheme for its existence which is quite different from the currently adopted ones depending on the zero-energy mode of the MBSs. In the presence of MBS-MBS overlaping, the central photon-assisted peaks in the transmission function reappear due to the fact that the photon absorbed (emitted) by one mode of the MBSs are subsequently emitted (absorbed) by another MBSs' mode. We also find that the positions of the additional peaks induced by the MBS-MBS overlaping in the presence of the photon field are quite different from the case of zero photon field.

Keywords: quantum dot, photon, Majorana bound states, topological superconductor, transmission function

OPEN ACCESS

Edited by:
Raffaele Gravina, University of Calabria, Italy

Reviewed by:
Qiang Xu, Nanyang Technological University, Singapore
Congcong Ma, Nanyang Institute of Technology, China
Qimeng Li, Shenzhen Institutes of Advanced Technology (CAS), China

****Correspondence:***
Feng Chi
chifeng@semi.ac.cn
Zhen-Guo Fu
zhenguo@iapcm.ac.cn

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *19 April 2020*
Accepted: *09 June 2020*
Published: *22 July 2020*

Citation:
Chi F, He T-Y, Wang J, Fu Z-G, Liu L-M, Liu P and Zhang P (2020) Photon-Assisted Transport Through a Quantum Dot Side-Coupled to Majorana Bound States. Front. Phys. 8:254. doi: 10.3389/fphy.2020.00254

1. INTRODUCTION

In recent years, there is much interest in the newly emerged issue of how to prepare and detect Majorana bound states (MBSs) in solid platforms. In condensed matter physics, the MBSs refer to a kind of quasi-particle of Majorana fermions being of their own antiparticles with zero energy [1]. Due to these unique characters, the MBSs obey non-Abelian statistics instead of the usual Fermi-Dirac one. The quantum bits designed by MBSs then have some exotic properties besides all the merits of traditional bits. It has been demonstrated that the MBSs enable topologically protected quantum information with potential applications in quantum computation free from decoherence [2–6]. Besides, the MBSs are also promising in the research field of high-efficiency and energy-saving electronic devices [7]. The MBSs have been successfully prepared in some kinds of systems, such as the mainstream p-wave superconductors [1, 8], and non-centrosymmetric superconductors, topological insulators coupled to superconductors defects in topological superconductors, the semiconducting or ferromagnetic nanowires with intrinsic strong spin-orbit interaction proximitization to a conventional s-wave superconductors, Josephson junctions [5, 6, 8–13], and so on.

Due to the massless properties of the MBSs having no charge, the detection of it is also quite challenging. A vast of experimental works have been devoted to the detection of MBSs, relying on the signatures probably induced by the MBSs, including the 4π periodic Josephson current-phase in junctions between topological superconductors [13], half-integer conductance plateau at the coercive field in a hybrid structure composing of topological superconductors and topological quantum anomalous Hall insulator [6], tunneling spectroscopy using the Rashba nanowires coupled to the bulk s-wave superconductors [10], zero-bias of the differential conductance at the edges of the wires [14, 15], the non-locality [16–19] and Kondo effect [20] influenced by MBSs. Since the above signatures can also be brought about by other mechanisms, some alternative schemes then have been continually put forward, such as the optical detection ones [21–23]. In [21], the authors found that the MBSs will absorb(emit) photons and result in photon-assisted tunneling side band peaks. This process can split the MBSs and then induces non-zero MBSs mode, providing another detection scheme for the existence of MBSs which is quite different the currently adopted ones concerning zero-energy MBSs mode. They also found that the height of the photon-assisted tunneling side band peaks is related to the intensity of the microwave field, and the time-varying conductance induced by the MBSs shows negative values for a certain period of time. Three all-optical detection schemes for the MBSs were proposed in [22], including a single QD, a hybrid QD-nanomechanical resonators system, and a carbon nanotube resonator implanted in a single electron spin system with optical pump-probe, respectively. They investigated the signatures of the MBSs in terms of the probe absorption spectrum and non-linear optical Kerr effect, the coupling strength between MBSs and the QD or the single electron spin. In the hybrid QD-resonantors system, they found that the vibration of the nanomechanical resonators will enhance the non-linear optical effect, which makes the MBSs more sensitive for detection. In the carbon nanotube resonator with a single electron, the single electron spin can be considered as a sensitive probe, and the nanotube resonator behaved as a phonon cavity is robust for detecting of MBSs. The MBSs signatures in these optical schemes are quite different from electrical ones and are promising for the detection MBSs, as well as for new applications in manipulation of MBSs or quantum information processing based on MBSs.

In the present manuscript, we study electronic transport through a QD shelled by a photon field as shown in **Figure 1**. Quite different from the above optical detection schemes for the MBSs, we propose to insert a QD, which is coupled to a topological nanowire hosting MBSs at its two ends, between two normal metal leads. One of the advantages of our system, as compared to the ones in [21, 22], is that the transport processes occur between the leads and the QD. The electrons will not enter into the MBSs mode existed in the topological superconductor nanowire, and thus our system provides a signature of the MBSs without direct electron transport through it. Our numerical results show that the transmission function develops photon-induced peaks, which are split by the coupling between the QD and MBSs. The central peaks' height are suppressed to zero because of the photon absorbtion (emission) by MBSs. As a result of it, the MBSs are shifted to the non-zero energy mode, and thus provides another detection scheme based on non-zero MBSs mode. If the MBSs are overlapped, the central photon-assisted peaks in the transmission function reappear because the photon absorbed (emitted) by one mode of the MBSs are subsequently emitted (absorbed) by another MBSs' mode. Under this condition, the positions of the peaks induced by the MBS-MBS overlaping in the presence of the photon field are quite different from those without photon field.

2. MODEL AND METHODS

The Hamiltonian of the QD coupled to MBSs and is subjected to a photon field can be written as the following form [15, 24–26]

$$H = \sum_{k\alpha} \varepsilon_{k\alpha} c^{\dagger}_{k\alpha} c_{k\alpha} + \varepsilon_d(t) d^{\dagger} d + \sum_{k,\alpha=L,R} (V_{k\alpha} c^{\dagger}_{k\alpha} d + H.c) + i\varepsilon_M \eta_1 \eta_2 + \lambda(d - d^{\dagger})\eta_1. \quad (1)$$

where $c^{\dagger}_{k\alpha}$ ($c_{k\alpha}$) creates (annihilates) an electron of momentum k, energy $\varepsilon_{k\alpha}$ in the lead $\alpha = L, R$. For the QD, $d^{\dagger}$ (d) is the creation (annihilation) operator of an electron having energy level $\varepsilon_d(t) = \varepsilon_d + \Delta_d \cos(\omega_p t)$ [21]. Here, we have assumed that the photon field with strength Δ_d and frequency ω_p is applied only on the QD and the leads are free from its irradiation, and then only the dot's energy level is time-varying. In experiments, the dot level in the absence of photon field ε_d is tunable by external gate voltages. The coupling strength between the QD and the leads is described by $V_{k\alpha}$. The last two terms in Equation (1) stand for the zero-energy MBSs with operators η_1 and η_2, as well as overlap strength between them ε_M. The MBSs locate on the opposite ends of the topological superconductor nanowire and are coupled to the QD with strength of λ [1, 6, 15]. In the present manuscript we assume that the dot is coupled only to the mode of the MBSs nearby to it with strength λ. The Majorana operators satisfy the relations of $\{\eta_i, \eta_j\} = 2\delta_{ij}$ and $\eta_i = \eta_i^{\dagger}$. We now follow previous work to switch from the Majorana fermion representation to the completely equivalent regular fermion one by defining [15] $\eta_1 = (1/\sqrt{2})(f + f^{\dagger})$, and $\eta_2 = (-i/\sqrt{2})(f - f^{\dagger})$, the Hamiltonian concerning about the MBSs in Equation (1) is rewritten as

$$H_{MBSs} = \varepsilon_M (f^{\dagger} f - \frac{1}{2}) + \frac{\lambda}{\sqrt{2}}(d - d^{\dagger})(f + f^{\dagger}). \quad (2)$$

The time-averaged electric current through the system is calculated by following the standard Keldysh Green's function technique as [21, 25, 26]

$$< J(t) >= \frac{e}{\hbar} \int [f_L(\varepsilon) - f_R(\varepsilon)] T(\varepsilon) \frac{d\varepsilon}{2\pi} \quad (3)$$

where e is the electron charge, $\hbar$ the reduced Planck's constant, $f_{L/R}(\varepsilon) = \{1 + e^{(\varepsilon - \mu_{L/R})/k_B T}\}^{-1}$ the Fermi distribution function of the left/right electrode with chemical potential $\mu_{L/R}$, temperature T and Boltzmann constant k_B. The transmission coefficient $T(\varepsilon)$

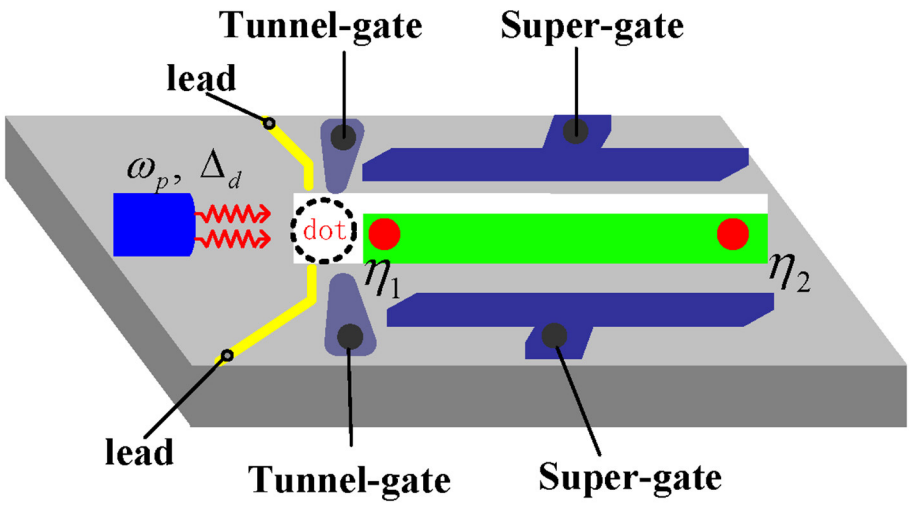

FIGURE 1 | Schematic plot of a QD (white circle) coupled to MBSs (red circles) formed at the two ends of a topological superconductor nanowire (green). The dot is connected to two normal metal leads (golden), and is shelled by a microwave field of strength Δ_d and photon frequency ω_p. The dot and the superconductor nanowire are individually defined by tunnel- and super-gates.

can be expressed with the help of the retarded Green's function $G^r(\varepsilon)$ as [25, 26]

$$T(\varepsilon) = \frac{\Gamma^L\Gamma^R}{\Gamma^L+\Gamma^R}[-2\mathrm{Im}\sum_k \tilde{G}^r_{dd;k}(\varepsilon)J_k^2(\frac{\Delta_d}{\omega_p})], \tag{4}$$

where $\Gamma^\alpha = 2\pi\sum_k |V_{k\alpha}|^2\delta[\varepsilon-\varepsilon_{k\alpha}]$ is the line-width function for coupling strength between the dot and the leads. $J_k(x)$ is the k-th ($k=-\infty,\infty$) order Bessel function of argument x, and $\tilde{G}^r_{dd;k}(\varepsilon)$ the retarded Green's function in the presence of photon field and QD-MBSs coupling. By applying the equation of motion method in [25] and adopting the truncation scheme introduced in [26], $\tilde{G}^r_{dd;k}(\varepsilon)$ can be obtained as (detailed calculation processes are neglected here for the sake of conciseness):

$$\tilde{G}^r_{dd;k}(\varepsilon)=\frac{1}{g^{r-1}_{dd;k}(\varepsilon)-\sum_m J^2_{k+m}(\frac{\Delta_d}{\omega_p})\Sigma^r_{mm}(\varepsilon)-\sum_{k'}\frac{\tilde{\Sigma}^{r2}_{kk'}(\varepsilon)}{\tilde{g}^{r-1}_{dd;k'}(\varepsilon)-\tilde{\Sigma}^r_{k'}(\varepsilon)}}, \tag{5}$$

where the free electron and hole Green's functions are respectively given by,

$$g^r_{dd;k}(\varepsilon) = \frac{1}{\varepsilon-\varepsilon_d-k\omega_p+i(\Gamma^L+\Gamma^R)/2}, \tag{6}$$

and

$$\tilde{g}^r_{dd;k}(\varepsilon) = \frac{1}{\varepsilon+\varepsilon_d+k\omega_p+i(\Gamma^L+\Gamma^R)/2}. \tag{7}$$

The self-energies in Equation (5) are given by $\Sigma^r_{mm}(\varepsilon)=\lambda^2(\varepsilon+m\omega_p)/[(\varepsilon+m\omega_p)^2+\varepsilon_M^2]$,$\tilde{\Sigma}^r_k(\varepsilon) = \sum_k J_{k-m}(\frac{\Delta_d}{\omega_p})\Sigma^r_{mm}$, and $\tilde{\Sigma}^r_{kk'}(\varepsilon)=\sum_{k'}J_{k+m}(\frac{\Delta_d}{\omega_p})J_{k'-m}(\frac{\Delta_d}{\omega_p})\Sigma^r_{mm}(\varepsilon)$. In the absence of the photon field ($\Delta_d=0$), because the Bessel function is $J_k(0)=\delta_{k,0}$, then the self-energies becomes $\Sigma^r_{00}(\varepsilon)=\delta_{m,0}\Sigma^r_{mm}(\varepsilon)=\lambda^2\varepsilon/(\varepsilon^2+\varepsilon_M^2)$, and $\tilde{\Sigma}^r_k(\varepsilon)=\tilde{\Sigma}^r_{kk'}(\varepsilon)=\Sigma^r_{00}(\varepsilon)$, now the above Green's function reduces to

$$\tilde{G}^r_{dd;0}(\varepsilon) = \frac{1}{g^{r-1}_{dd;0}(\varepsilon)-\Sigma^r_{00}(\varepsilon)-\frac{\Sigma^{r2}_{00}(\varepsilon)}{\tilde{g}^{r-1}_{dd;0}(\varepsilon)-\Sigma^r_{00}(\varepsilon)}}, \tag{8}$$

which is just the retarded Green's function derived in [15] with coupling between the QD and topological superconductor nanowire hosting MBSs at its ends.

3. RESULTS AND DISCUSSION

In the following numerical calculations, we choose the photon frequency $\omega_p=1$ as the energy unit ($\hbar=1$), and fix the values of $\Gamma=\Gamma^L=\Gamma^R=0.1$ unless noted. We do not consider the case of finite bias voltage and then the leads' chemical potentials are se to be $\mu_L=\mu_R=0$ as the zero-point of the energy. We first study in **Figure 2** the case of zero photon field ($\Delta_d=0$). Under this condition, the retarded Green's is given in Equation (8). If there is no coupling between the QD and MBSs existed at the ends of the topological nanowire $\lambda=0$, the transmission function T in **Figure 2A** shows the typical resonant tunneling feature, i.e., it develops a Lorentzian peak with height of 1 at zero energy state [27], $T(\varepsilon\longrightarrow 0)=1$. This can be seen from the black solid line in **Figure 2A**, which denotes that the electrons can transport from one lead through the QD to the other lead if the electron energy in the leads (Fermi level μ_α) equals to the dot level ε_d. Turning on the coupling between the QD and MBSs at the ends of topological superconductor nanowire $\lambda\neq 0$, we find

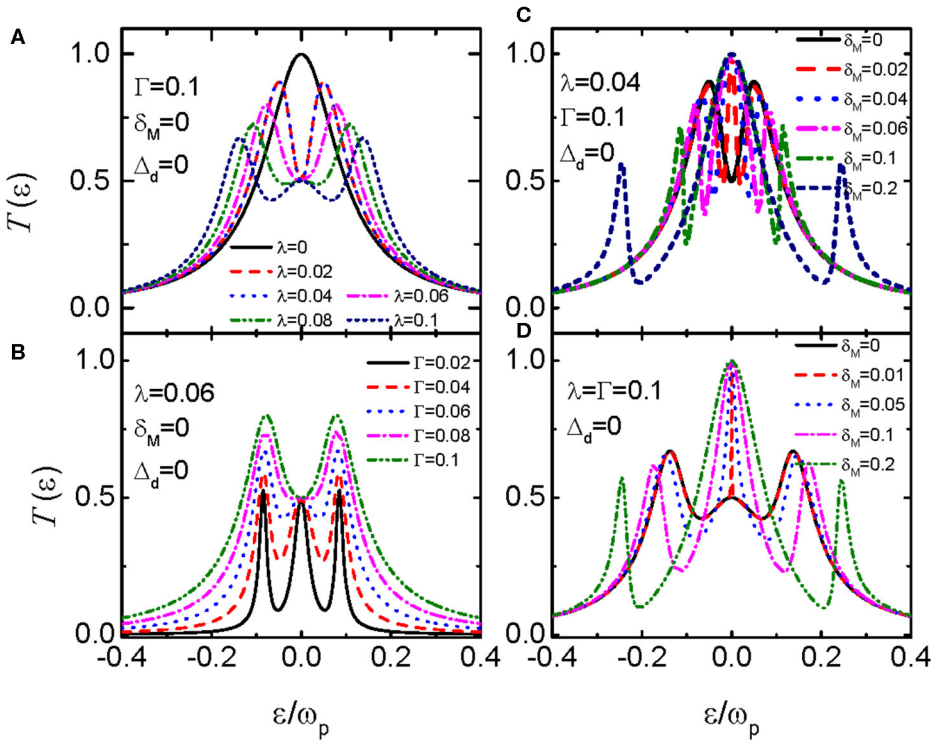

FIGURE 2 | Transmission function T as a function of the electron energy ε in the absence of the photon field $\Delta_d = 0$. Panels **(A–D)** correspond to different parameters, which are given in the figure.

that value of the zero-energy transmission function is reduced to half of its quantum value 1, i.e., $T(0) = 1/2$, showing the half-fermionic character of the MBSs. This is because now the retarded Green's function is $\tilde{G}_{dd;0}(\varepsilon \longrightarrow 0) = 1/2(\varepsilon + i\Gamma)$, and then the value of the zero-energy transmission function is reduced by a factor of 1/2, accordingly. This change of the value oftransmission function is believed to be the signature of existence of MBSs [15], and is also responsible for the zero-bias anomaly of the conductance peak, a kind of main detection means for the MBSs in tunneling spectroscopy. It is worth noting that as long as the coupling strength between the QD and MBSs $\lambda \neq 0$, the result of $T(0) = 1/2$ remains unchanged regardless of the variation of it. For weak QD-MBSs coupling $\lambda < \Gamma$, the transmission function develops two peaks centered around $\varepsilon \sim \pm\lambda$, induced by the splitting of the dot energy-level in the presence of coupling between the QD and the MBSs [15]. With increasing λ, the positions of the two peaks are shifted away from the zero-energy state, and the double-peak configuration in the transmission function evolves to a triple-peak one for $\lambda \geq \Gamma$ as shown in **Figure 2A**. The evolution of the peaks' configuration is a clear signature of the existence of MBSs. To show the evolution of the peak configuration in the transmission function by the relative strength between λ and Γ, we show the behavior of $T(\varepsilon)$ in **Figure 2B** for fixed $\lambda = 0.06$ and different values of Γ. Under the condition of $\lambda \geq \Gamma$, the transmission function shows the triple-peak configuration as indicated by the black solid, red dashed and blue dotted lines in **Figure 2B**. Whereas for $\Gamma > \lambda$, the transmission function shows the double-peak configuration.

The two modes of the MBSs at the two ends of the topological superconductor nanowire will overlap with each other, and the overlap strength between them δ_M depends on the length of the nanowire. In **Figures 2C,D**, we examine the influences of δ_M on the peak configurations of the transmission function $T(\varepsilon)$. For the double-peak configuration in **Figure 2C** in which $\lambda < \Gamma$, the zero-energy transmission function shows peak of height $T(0) = 1$ not 1/2 even for very weak MBS-MBS hybridization $\delta_M = 0.02$. The width of this zero-energy peak is proportional to the value of δ_M. Meanwhile, two additional peaks in $T(\omega)$ emerge around $\varepsilon \sim \pm(\lambda + \delta_M)$ corresponding to the energy of the effective Dirac fermionic state f. For large coupling strength $\delta_M \geq 0.1$, the zero-energy transmission function reduces to the resonant level result. For sufficiently long nanowire in which the overlap strength between the two MBSs is weak enough as compared to the value of the QD-MBSs coupling λ and thermal energy k_BT, the signature of the MBSs $T(0) = 1/2$ will emerge. The behavior of the triple-peak configuration in the transmission function $T(\varepsilon)$ in **Figure 2D** resembles that in **Figure 2C** because the peaks around $\varepsilon \sim \pm\lambda$ are merged into $\varepsilon \sim \pm(\lambda + \delta_M)$ if the two modes of the MBSs are overlapped [15].

If the QD is shelled by a photon field of strength $\Delta_d = \omega_p$, a series of photon-assisted additional channels are opened [21–23, 25, 26, 28, 29]. Correspondingly, the transmission function $T(\varepsilon)$ develops peaks at $\varepsilon = \varepsilon_d \pm n\omega_p$, in which $n = 0, \pm1, \pm2, \cdots\cdots$.

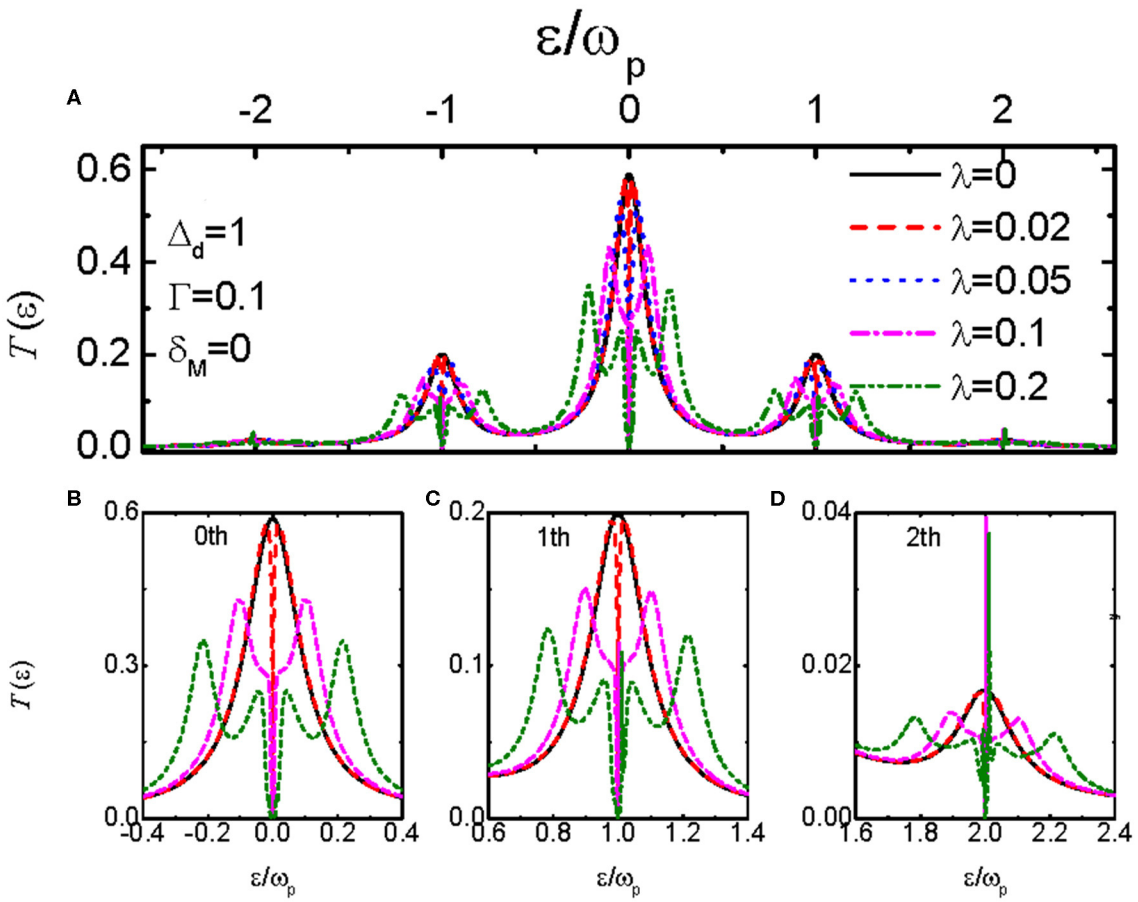

FIGURE 3 | Transmission function T and its blowups as a function of the electron energy ε under fixed strength of the photon field $\Delta_d = \omega_p$, and different values of the coupling between the dot and the MBSs λ. Panel **(A)** correspond to the transmission coefficient in a large regime of the electron energy, and **(B–D)** denote the peaks in **(A)** for n=0,1,2, respectively.

Now electrons can tunnel from one lead to the other through the dot whenever a channel enters into the conduction window. As a result of photon-induced additional transport channels (dot levels), the electron transport probability through each channel is weakened and the peak height of the transmission function in **Figure 3A** is lowered, accordingly. In the presence of coupling between the QD and MBSs ($\lambda \neq 0$), we find that the peaks' value originally at the states of $\varepsilon = \varepsilon_d \pm n\omega_p$ are suppressed to zero accompanied by the emergence of two additional peaks roughly at $\varepsilon = \varepsilon_d \pm n(\pm\lambda + \omega_p)$ [21–23, 25, 26, 28, 29]. It indicates that the two MBSs have absorbed the photon energy and then the character of its zero-energy mode is destroyed. The reason of the change of the transmission function can be explained as follows: Under the irradiation of the microwave field, the MBSs will absorb or emit n photons from the microwave field and jump to the states of $\pm n\omega_p$. Accordingly, the energy levels of electrons on the QD is shifted to $\varepsilon_d \pm n\omega_p$. If the chemical potential of the leads is aligned to these states, the electrons at them will transport through the QD, leading to peaks at $\varepsilon = \varepsilon_d \pm n(\pm\lambda + \omega_p)$. The blowups of the peaks in the transmission function are given in **Figures 3B–D**, in which the disappearance and splitting of the peaks are clearly presented. **Figures 3B–D** also indicate that the application of the photon field may excite the zero-energy MBSs to non-zero-energy mode. This enables the probability of detection of the MBSs in not only the usual zero-energy mode but also in the non-zero one. Here we find that even the photon-induced peaks are destroyed by the existence of the MBSs, the splitting of these peaks still can serve as the signature of the MBSs, and provides another detection scheme. The peaks' height of the transmission function shown in **Figures 3B–D** are obviously lowered with increasing n because the electron transport probability through those channels are suppressed. For $n = \pm 2$, the transmission function in **Figure 3D** develops a sharp peak at $\varepsilon = \varepsilon_d \pm 2\omega_p$ [25, 26, 28, 29]. The reason may be attributed to the fact that the MBSs have less probability of absorbing or emitting more than one photons and then the photon-induced peaks of $n > 1$ are relatively less influenced. The splitting of the peaks by the QD-MBSs coupling, however, is still prominent in **Figure 3D**. We have examined the cases of varying photon field strength Δ_d, and found that it will change the height of the peaks but not their splitting and suppression. Variation of the photon frequency induces the change of the peaks' positions, and will not change the essential influences of the photon field on the transmission function.

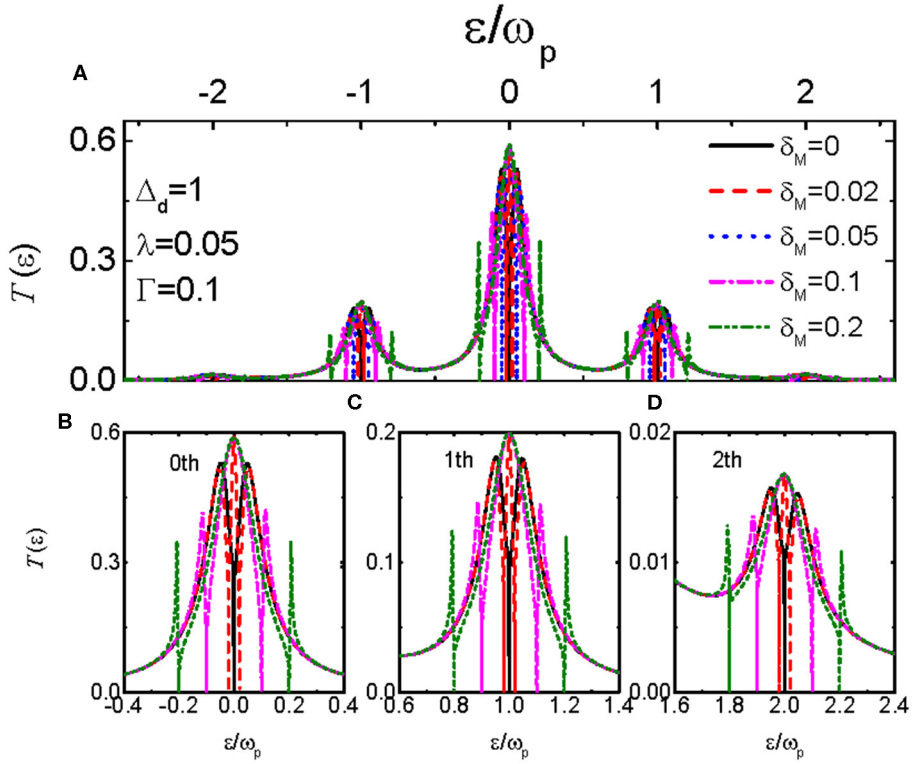

FIGURE 4 | Transmission function and its blowups as a function of ε for $\Delta_d = \omega_p$, $\lambda = 0.05$ and different values of overlap strength between the MBSs δ_M. Panel **(A)** correspond to the transmission coefficient in a large regime of the electron energy, and **(B–D)** denote the peaks in Figure 4A for n=0,1,2, respectively.

We now study the impacts of the MBS-MBS direct overlap δ_M on the properties of the transmission function in **Figure 4**. The strength of the photon field is also set to be $\Delta_d = 1$ and the coupling strength between the QD and the MBSs is weak than that between the QD and the leads, i.e., $\lambda < \Gamma$. In the absence of overlap between the MBSs $\delta_M = 0$, the photon-assisted peaks in the transmission function are split by finite value of QD-MBSs coupling λ, which can be seen from the solid black lines in **Figures 4A–D**. Moreover, the value of the central peaks at $\varepsilon = \varepsilon_d \pm n\omega_p$ are suppressed to zero due to the photon absorbtion (emission) processes by the MBSs. For non-zero δ_M, we find that the peaks of the transmission function at $\varepsilon = \varepsilon_d \pm n\omega_p$ reappear, which resembles the result in **Figure 2C** in the absence of photon field. In other words, the impacts of the MBSs on the disappearance of the photon-assisted central peaks in the transmission function are eliminated by the overlap between the MBSs. We attribute this phenomenon to the fact that the photon absorbed (emitted) by one mode of the MBSs are immediately emitted (absorbed) by the other mode of the MBSs. Just by these two opposite processes, the photon energy are unchanged and then the MBSs remain the same. As compared to the case in **Figure 2C**, we find that the additional peaks induced by δ_M are positioned exactly at $\varepsilon = \varepsilon_d \pm n(\pm\delta_M + \omega_p)$, but not the expected positions of $\varepsilon = \varepsilon_d \pm n[\pm(\lambda + \delta_M) + \omega_p]$. It may originates from the fact that the MBS-MBS overlaping dominates the transport processes and then the peaks in the transmission function are then determined by δ_M, accordingly.

4. SUMMARY

In summary, we have studied photon-assisted transport through a QD side-coupled to a topological nanowire hosting MBSs at its two ends. It is found that the photon-induced peaks in the transmission function are split by the existence of the MBSs, and the value of the central peaks are suppressed to be zero. Such an abnormal change of the photon-assisted peaks may serve as the detection means for MBSs. If the two MBSs modes at the opposite ends of the nanowire are overlapped, the photon-assisted peaks are further split with the restoration of the central peaks, which may be induced by the opposite photon absorbtion (emission) processes of the two modes of the MBSs.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

FC, Z-GF, and PZ contributed the ideas equally. FC, Z-GF, T-YH, JW, L-ML, and PL derived the formulae in the paper. FC, T-YH, JW, L-ML, and PL performed the numerical calculations. FC and

Z-GF wrote the original manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

FC was supported by the Initial Project of UEST of China, Zhongshan Institute (No. 415YKQ02), and Science and Technology Bureau of Zhongshan (Grant Nos. 2017B1116, 2017B1016, and 180809162197886). Z-GF was also supported by the Innovation Development Fund of the China Academy of Engineering Physics (CAEP; Grant No. ZYCX1921-02) and the Presidential Foundation of the CAEP (Grant No. YZ2015014). PL was supported by the NSF of China under Grant No. 11775047.

REFERENCES

1. Fu L, Kane CL. Superconducting proximity effect and Majorana Fermions at the surface of a topological insulator. *Phys Rev Lett*. (2008) **100**:096407. doi: 10.1103/PhysRevLett.100.096407
2. Nayak C, Simon SH, Stern A. Non-Abelian anyons and topological quantum computation. *Rev Mod Phys*. (2008) **80**:1083–159. doi: 10.1103/RevModPhys.80.1083
3. Alicea J, Oreg Y, Refael G. Non-Abelian statistics and topological quantum information processing in 1D wire networks. *Nat Phys*. (2011) **7**:412–7. doi: 10.1038/nphys1915
4. Torsten Karzig RML Christina Knapp. Scalable designs for quasiparticle-poisoning-protected topological quantum computation with Majorana zero modes. *Phys Rev B*. (2017) **95**:235305. doi: 10.1103/PhysRevB.95.235305
5. Sato M, Fujimoto S. Topological phases of noncentrosymmetric superconductors: edge states, Majorana fermions, and non-Abelian statistics. *Phys Rev B*. (2009) **79**:094504. doi: 10.1103/PhysRevB.79.094504
6. Qi XL, Zhang SC. Topological insulators and superconductors. *Rev Mod Phys*. (2011) **83**:1057–110. doi: 10.1103/RevModPhys.83.1057
7. Smirnov S. Universal Majorana thermoelectric noise. *Phys Rev B*. (2018) **97**:165434. doi: 10.1103/PhysRevB.97.165434
8. Wimmer M, Akhmerov AR, Medvedyeva MV. Majorana bound states without vortices in topological superconductors with electrostatic defects. *Phys Rev Lett*. (2010) **105**:046803. doi: 10.1103/PhysRevLett.105.046803
9. Sau JD, Lutchyn RM, Tewari S. Generic new platform for topological quantum computation using semiconductor heterostructures. *Phys Rev Lett*. (2010) **104**:040502. doi: 10.1103/PhysRevLett.104.040502
10. Lutchyn RM, Sau JD, Sarma SD. Majorana Fermions and a topological phase transition in semiconductor-superconductor heterostructures. *Phys Rev Lett*. (2010) **105**:077001. doi: 10.1103/PhysRevLett.105.077001
11. Choy TP, Edge JM, Akhmerov AR. Majorana fermions emerging from magnetic nanoparticles on a superconductor without spin-orbit coupling. *Phys Rev B*. (2011) **84**:195442. doi: 10.1103/PhysRevB.84.195442
12. San-Jose P, Prada E, Aguado R. AC Josephson effect in finite-length nanowire junctions with Majorana modes. *Phys Rev Lett*. (2012) **108**:257001. doi: 10.1103/PhysRevLett.108.257001
13. Mourik V, Zuo K, Frolov SM. Signatures of Majorana Fermions in hybrid superconductor-semiconductor nanowire devices. *Science*. (2012) **336**:1003–7. doi: 10.1126/science.1222360
14. Ricco LS, de Souza M, Figueira MS. Spin-dependent zero-bias peak in a hybrid nanowire-quantum dot system: distinguishing isolated Majorana fermions from Andreev bound states. *Phys Rev B*. (2019) **99**:155159. doi: 10.1103/PhysRevB.99.155159
15. Liu DE, Baranger HU. Detecting a Majorana-fermion zero mode using a quantum dot. *Phys Rev B*. (2011) **84**:201308R. doi: 10.1103/PhysRevB.84.201308
16. Vernek E, Penteado PH, Seridonio AC, Egues JC. Subtle leakage of a Majorana mode into a quantum dot. *Phys Rev B*. (2014) **85**:165304. doi: 10.1103/PhysRevB.89.165314
17. Ruiz-Tijerina AD, Vernek E, da Silva LGGVD, Egues JC. Interaction effects on a Majorana zero mode leaking into a quantum dot. *Phys Rev B*. (2015) **91**:115435. doi: 10.1103/PhysRevB.91.115435
18. Prada E, Aguado R, San-Jose P. Measuring Majorana nonlocality and spin structure with a quantum dot. *Phys Rev B*. (2017) **96**:085418. doi: 10.1103/PhysRevB.96.085418
19. Deng MT, Vaitiekenas S, Prada E, San-Jose P, Nygard J, Krogstrup P, et al. Nonlocality of Majorana modes in hybrid nanowires. *Phys Rev B*. (2018) **98**:085125. doi: 10.1103/PhysRevB.98.085125
20. Wang R, Su W, Zhu JX. Kondo signatures of a quantum magnetic impurity in topological superconductors. *Phys Rev Lett*. (2019) **122**:087001. doi: 10.1103/PhysRevLett.122.087001
21. Tang HZ, Zhang YT, Liu JJ. Photon-assisted tunneling through a topological superconductor with Majorana bound states. *AIP Adv*. (2015) **5**:127129. doi: 10.1063/1.4939096
22. Chen HJ, Zhu KD. All-optical scheme for detecting the possible Majorana signature based on QD and nanomechanical resonator systems. *Sci China Phys Mech Astron*. (2015) **58**:050301. doi: 10.1007/s11433-014-5637-4
23. Väyrynen JI, Rastelli G, Belzig W, Glazman LI. Microwave signatures of Majorana states in a topological Josephson junction. *Phys Rev B*. (2015) **92**:134508. doi: 10.1103/PhysRevB.92.134508
24. Piotr S. Properties of the Majorana-state tunneling Josephson junction mediated by an interacting quantum dot. *J Phys Condens Matter*. (2019) **31**:185301. doi: 10.1088/1361-648X/ab052a
25. Jauho AP, Wingreen NS, Meir Y. Time-dependent transport in interacting and noninteracting resonant-tunneling systems. *Phys Rev B*. (1994) **50**:5528–44. doi: 10.1103/PhysRevB.50.5528
26. Sun QF, Wang J, Lin TH. Photon-assisted Andreev tunneling through a mesoscopic hybrid system. *Phys Rev B*. (1999) **59**:13126–38. doi: 10.1103/PhysRevB.59.13126
27. Pals P, MacKinnon A. Coherent tunnelling through two quantum dots with Coulomb interaction. *J Phys Condens Matter*. (1996) **8**:5401–14. doi: 10.1088/0953-8984/8/29/015
28. Chi F, Sun LL. Photon-assisted heat generation by electric current in a quantum dot attached to ferromagnetic leads. *Chin Phys Lett*. (2016) **33**:117201. doi: 10.1088/0256-307X/33/11/117201
29. Chi F, Liu LM, Sun LL. Photon-mediated spin-polarized current in a quantum dot under thermal bias. *Chin Phys B*. (2017) **26**:037204. doi: 10.1088/1674-1056/26/3/037304

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 10 September 2020
doi: 10.3389/fphy.2020.00354

Dynamic Adjustment and Distinguishing Method for Vehicle Headlight Based on Data Access of a Thermal Camera

Shixiao Li[1], *Pengfei Bai*[1,2]* *and Yuanfeng Qin*[1]

[1] Guangdong Provincial Key Laboratory of Optical Information Materials and Technology & Institute of Electronic Paper Displays, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China, [2] National Center for International Research on Green Optoelectronics, South China Normal University, Guangzhou, China

In recent years, more and more traffic accidents have been caused by the illegal use of high beams. Therefore, the distinguishing of the vehicle headlight is vital for night driving and traffic supervision. Then, a method for distinguishing vehicle headlight based on data access of a thermal camera was proposed in this paper. There are two steps in this method. The first step is thermal image dynamic adjustment. In thermal image dynamic adjustment, the details of thermal images were enhanced by adjusting the temperature display dynamically and fusing features of multi-sequence images. The second step is vehicle headlight dynamic distinguishing, and features of vehicle headlight were extracted by YOLOv3. Then, the high beam and low beam were further distinguished by the filter based on the position and proportion relationship between the halo and the beam size of vehicle headlights. In addition, the accessed thermal image dataset during the night was used for training purposes. The results showed that the precision of this method was 94.2%, and the recall was 78.7% at a real-time speed of 9 frames per second (FPS). Compared with YOLOv3 on the Red Green Blue (RGB) image, the precision was further improved by 11.1%, and the recall was further improved by 5.1%. Dynamic adjustment and distinguishing method were also applied in single-shot multibox detector (SSD) network which has good performance in small-object detection. Compared with the SSD network on the RGB image, the precision was improved by 8.2% and the recall was improved by 4.6% when SSD network was improved by this method.

Keywords: dynamic distinguishing, dynamic adjustment, vehicle headlight, thermal images, multi-sequence images

OPEN ACCESS

Edited by:
Chongfu Zhang,
University of Electronic Science and Technology of China, China

Reviewed by:
Lelun Jiang,
Sun Yat-sen University, China
Xiao Shu,
McMaster University, Canada

***Correspondence:**
Pengfei Bai
baipf@m.scnu.edu.cn

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *23 June 2020*
Accepted: *27 July 2020*
Published: *10 September 2020*

Citation:
Li S, Bai P and Qin Y (2020) Dynamic Adjustment and Distinguishing Method for Vehicle Headlight Based on Data Access of a Thermal Camera. Front. Phys. 8:354. doi: 10.3389/fphy.2020.00354

INTRODUCTION

In recent years, traffic accidents have become a common problem for vehicle drivers. The risk of traffic accidents on an unlit road is about 1.5–2 times higher than that during the day [1]. Due to the complexity of the road and the negligence of drivers, the high and low beam of the vehicle cannot be switched correctly in time, which can lead to a series of traffic accidents. In addition, oncoming headlight glare can reduce the visibility of objects on the road, which may have a bad effect on safety at night. For cataracts, the impact of oncoming headlight glare is more severe [2]. So, it is necessary to realize the vehicle headlights distinguishing.

Currently, vehicle detection is mostly based on visual images [3–10]. The visual image is not clear at night, and the detail of the vehicle is also unclear. In order to overcome the problem, a number of papers have been published to detect vehicles at night by identifying the shape and track of the headlights [3–10]. Many studies were both detecting vehicles via headlights pairing and trajectories matching [3, 4]. For extracting the nighttime image details, image enhancement was used for preprocessing before vehicle detection [5, 6]. Given that the headlights were usually white in color, the inputted images were usually transformed into different color spaces. The dominant color components in Red Green Blue (RGB) images were then processed by the threshold to extract blobs for the headlight [7]. However, this nighttime vehicle detection method depended on clear headlights or taillights shape [5, 8–10], and the existence of high beams glares was ignored. When the vehicle headlight was captured by a camera, it could produce a halo, which would affect the judgment and measurement of the vehicle headlight. The minuscule vehicle details can be retained in a dim environment by the thermal image. Simultaneously, the temperature of vehicles can be collected by thermal cameras. So, it could not be interference by the halo. Thermal imaging technology has been used for vehicle detection at night [11]. The temperature difference between the object and the ambient is insignificant, and it is impossible to separate the object from the environment. Moreover, the value of the temperature was conversed to a pseudo-color image, it can increase the difficulty for object detection. The adaptive histogram equalization method was used to enhance the counters of images [11]. However, when the image content was enhanced, the background information was also continuously enhanced, which can increase the difficulty of recognition. In addition, the thermal image was affected by the resolution, so that the details of distant objects cannot be captured. In object detection, machine learning and deep learning have been applied to various research fields. The unsupervised learning was applied for the classification of vehicles successfully [12, 13]. Furthermore, Convolutional Neural Networks (CNNs), YOLO [14], and other neural networks have made outstanding contributions to vehicle detection in both RGB images and thermal images [11, 15, 16]. However, more relevant optimization and adjustment are needed to obtain a more suitable training model. Recent work showed that multi-sequence images and deep neural networks can match the vehicle types [17]. The deep neural network YOLOv3 has a good detection effect on the COCO data set [18, 19]. But the detection model needs to be further improved to achieve distinguishing similar objects.

In this paper, a vehicle headlight distinguishing method based on the thermal image dynamic adjustment and dynamic distinguishing was proposed. The thermal image enhancement and multi-sequence image feature fusion were contained by the thermal image dynamic adjustment. The YOLOv3-Filter operation was applied as the thermal image dynamic distinguishing. The target can be separated effectively by the thermal image enhancement from the environment. Simultaneously, the details of the thermal image were supplemented by multi-sequence image feature fusion. Lastly, the vehicle headlight distinction model was realized by the YOLOv3-Filter operation.

PRINCIPLE

Thermal Image Dynamic Adjustment

Thermal Image Enhancement

In the case of low illumination at night, the vehicle characteristics can be disturbed by the halo of the headlights, so that the camera cannot capture the contour of a vehicle. The thermal camera cannot be disturbed by such a high light source because the thermal imaging map generates a visual image by acquiring the temperature of the object. In addition, thermal imaging technology has many disadvantages. The color difference between the color of the object and the environment is not obvious. The thermal camera can also be affected by the external environment [20], such as sky radiation, ground background radiation, radiation reflections, temperature changes, wind speed, and geographic latitude. In order to reduce these interferences of headlight distinguishing, the thermal image enhancement was used in this paper.

As shown in **Figure 1B**, the thermal histogram shows that the temperature of the vehicle and the ambient temperature can be changed within a certain interval. The data set used in this paper was captured by us at 25°C, and the relative humidity was 55%. The maximum vehicle temperature in the data set was 125°C. The object at −20 to 25°C and 125–400°C does not need to be displayed in thermal images. As shown in **Figure 2**, the color scale range is 0–255; it allows as many objects as possible to be displayed in this interval.

To extract the object information, the thermal image dynamic adjustment method has been adopted. Firstly, the ambient temperature information is obtained from the thermal camera. Secondly, the ambient temperature is subtracted from each temperature pixel value in a thermal image to obtain an object which is different from the ambient temperature. Finally, the image is multiplied with the device parameters. The pixel value of the thermal image is defined as Equation (1).

$$P_{(x,y)} = \lambda \left(\left| T_{(x_0,y_0)} - T_{environment} \right| \right) \tag{1}$$

In Equation (1), λ is the device parameters, and it can be calculated through Equation (2).

$$\lambda = \frac{\left(T_{(x_0,y_0)\,\max} - T_{(x_0,y_0)\,\min} \right) \left(T_{MAX} + |T_{MIN}| \right)}{256} \tag{2}$$

where $T_{(x0,y0)max}$ is the maximum temperature value on a thermal map. T_{MAX} is the maximum acceptable temperature value of the thermal image sensor. The object temperature is first subtracted from the ambient temperature value to obtain an object, which is different from the ambient temperature. $T_{(x0,y0)}$ is a temperature pixel value, and $T_{environment}$ is an ambient temperature in Equation (2). Then, the temperature difference can be multiplied by an appropriate ratio λ, and the characteristics of the object can be obtained obviously.

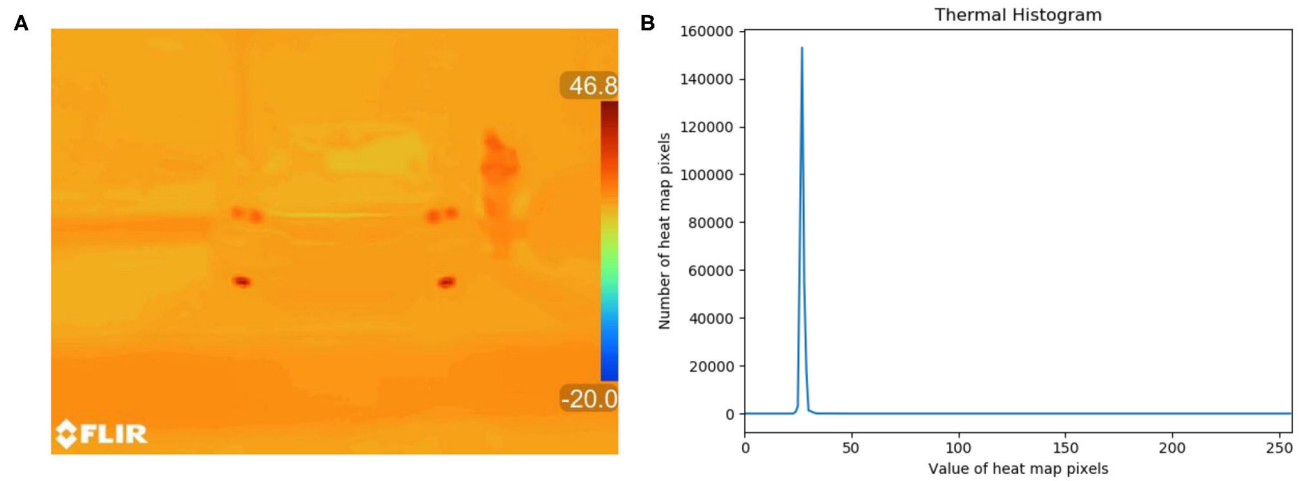

FIGURE 1 | Thermal image and thermal histogram. **(A)** An original thermal image captured by a thermal camera. **(B)** Thermal histogram of the original thermal image. The thermal histogram represents the distribution of the pixel values in the thermal diagram.

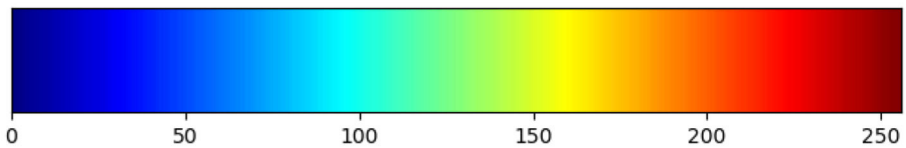

FIGURE 2 | Color scale of the thermal image. The temperature of an object was displayed with the corresponding color in the thermal image.

Multi-Sequence Image Feature Fusion

After thermal image enhancement, the next step is to fuse the thermal image with the RGB image. As shown in **Figure 3**, the RGB image extracted from the original image data is reduced to the same size as the thermal image with a resolution of 640 × 480. In this paper, the contour features of vehicle headlight can be extracted by the Sobel operator, as shown in Equation (3). Because it can obtain the edge of target which has a great gradient with background, the Sobel operator on a preprocessed image in order to retrieve an edge image is used to find and extract a rectangular area in the original image which represents the license plate [21, 22].

$$G = |G_x| - |G_y| \tag{3}$$

In a horizontal variation, the image value I is convolved with an odd-sized kernel G_x. In the vertical variation, the image value I is convolved with an odd-sized kernel G_y.

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I \tag{4}$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * I \tag{5}$$

Finally, the contour features of a vehicle and vehicle headlight, which are extracted from the RGB image, are fused with the thermal image. Then, a multi-sequence image can be obtained. The multi-sequence image not only has the information about the thermal image but also the contour information of the RGB image.

In addition, the halo areas of vehicle headlight S_{Light} in the RGB image can be obtained after threshold processing [22]. In the same way, the areas of the lamp in the image S_{Lamp} can be obtained. These parameters are used in Equation (9).

Vehicle Headlight Dynamic Distinguishing

In order to realize the distinguishing between high and low beam lights, the following steps must be done. Firstly, the YOLOv3 is used to initially identify the candidate areas of a vehicle and its headlights. Secondly, the distance between the vehicle and the camera can be determined by the size of the bounding box. Then, the halo and the contour of the headlight are extracted from the RGB image and the thermal image, respectively. Finally, the distance between the high and low headlights can be determined by calculating the relationship between the halo and the headlight profile.

Deep Network Model for Beam Detection

The deep neural network YOLOv3 is used as a preliminary screening model, as shown in **Figure 4**. The coordinate of the vehicle in the image is chosen as input data. Then, a likelihood

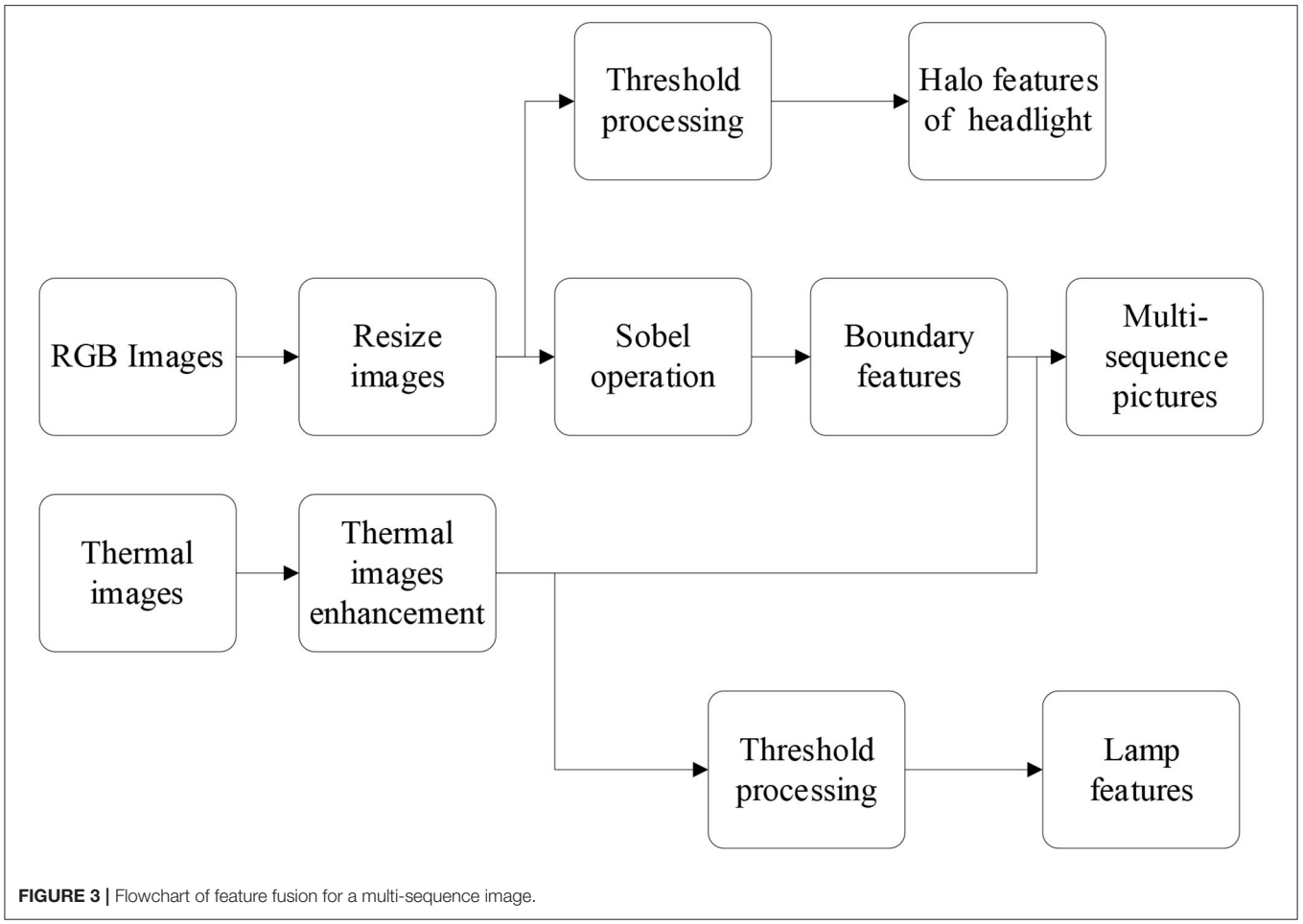

FIGURE 3 | Flowchart of feature fusion for a multi-sequence image.

score of the candidate about the high beam and low beam is outputted by the model. The network contains 23 residual blocks and three times up-sampling. The model is detected at 32x, 16x, and 8x subsampled which can be used to make multi-scale measurements. Leaky Relu, which gives all negative values, can be used as an activation function for all residual blocks. The total number of parameters of the network is about 110,536.

Low Probability Candidate Filter

The accuracy of vehicle light distinguishing can be obtained by adding discriminant conditions to the YOLOv3. The low probability candidate filter is used as a discriminant condition filter in this paper.

In order to design the low probability candidate filter, the conversion relationship between the image and the three-dimensional (3D) space is required to be found. The pinhole imaging model can be used to obtain the actual location of the object in the image. As shown in **Figure 5**, the target size is converted to the actual target size in the image. *A'B'* is the straight line of road *AB* mapped to the image at *Y*. In the same way, *C'D'* is the straight line of road *CD* mapped to the image at *Y*. The relationship between the actual distance of the road and the pixel width of the road in the image can be written as Equation (6).

$$D_{PicRoad}(Y) = D_{A'B'} + (D_{C'D'} - D_{A'B'}) \frac{Y - Y_1}{Y_2 - Y_1} \tag{6}$$

where D_{AB} and D_{CD} are the actual distances of the road. $D_{A'B'}$ and $D_{C'D'}$ are the pixel widths of the road in the image. Therefore, we can obtain Equation (8).

$$\Delta x = \Delta X \cdot \frac{D_{AB}}{D_{PicRoad}(Y)} \tag{7}$$

As shown in **Figure 6**, Y_1 and Y_2 are the vertical distances of the road mapped on the image. In Equation (6), $D_{PicRoad}(Y)$ is the length of the road displayed on the image from the origin *O* to the height *Y*. In Equation (7), ΔX is the width of the target object on the image. Δx is the width of the actual target. The actual halo size of vehicle headlights and vehicle beams can be gotten by this method.

The Zhang's calibration method was used to calibrate the camera for rebuilding a 3D space, as shown in Equation (8) [23].

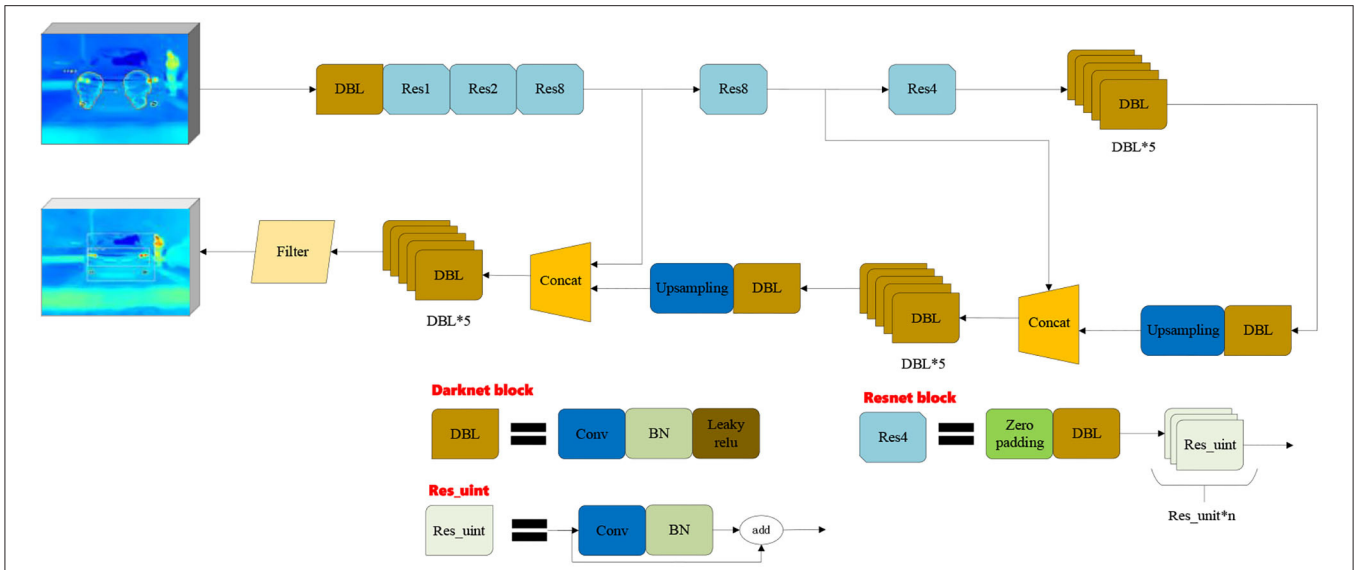

FIGURE 4 | The YOLOv3 with adding a filter (YOLOv3-Filter).

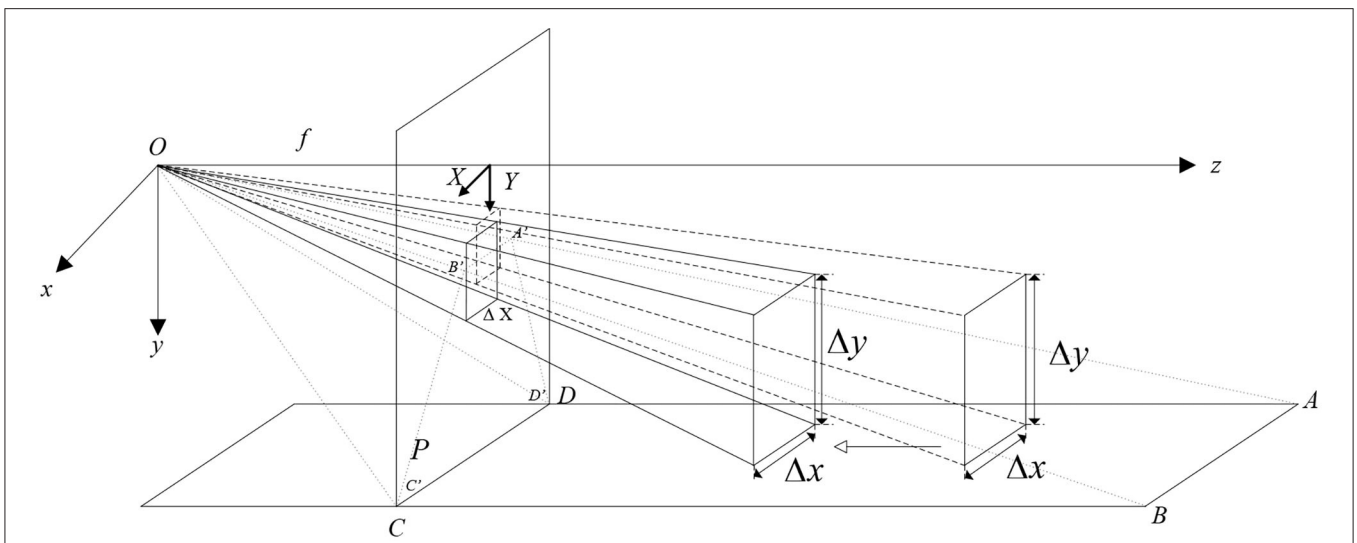

FIGURE 5 | Demonstration of the perspective projection.

$$Z_c \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \cdot \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} \frac{\Delta x}{\Delta X} & 0 & u_0 \\ 0 & \frac{\Delta y}{\Delta Y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \cdot \widetilde{M} = A \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \widetilde{M} \quad (8)$$

where u, v are the horizontal and vertical coordinate values in the image coordinate system; Z_c is the distance from the camera surface to the object along the optical axis. d_x, d_y are the horizontal and vertical dimensions of the pixel. u_0 and v_0 are the center positions of the image plane. f is the camera focal length. R is the rotation matrix of the calibration object. t is the translation matrix. X_w, Y_w, and Z_w are the positions of the feature points in the world coordinate system. According to Equation (6), the distance D can be obtained between the vehicle and the camera. ΔX is the width of the target object in the image. Δx is the width of the actual target. ΔY is the height of the target object in the image. Δy is the height of the actual target.

According to Equation (8), the distance D can be obtained between the vehicle and the camera. The high beam of the vehicle can be defined by looking for the relationship among S_{Light}, S_{Lamp}, and D. The halo area of vehicle headlight S_{Light} and the area of the lamp S_{Lamp} can be obtained by threshold processing.

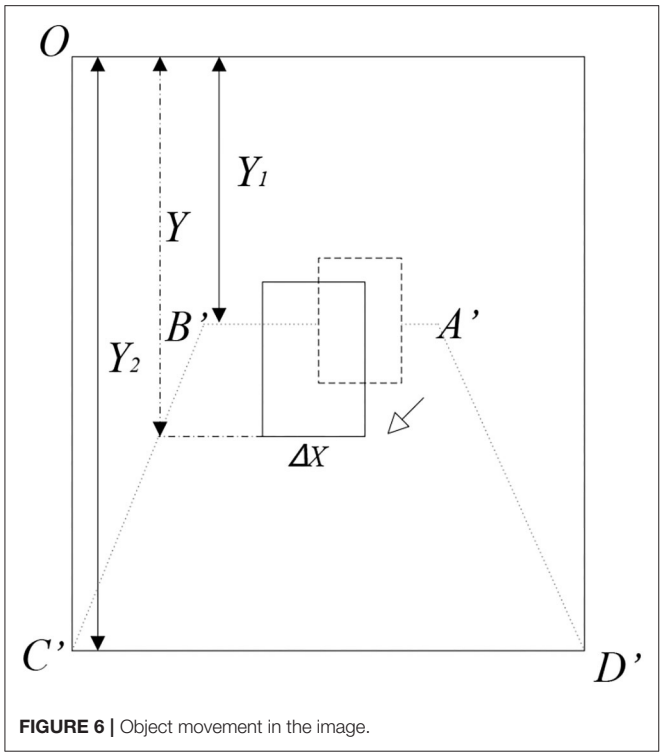

FIGURE 6 | Object movement in the image.

As shown in **Figures 7**, **8**, two halos of the headlights are just distinguished when the vehicle is in $D_{tangent}$ position. If the distance between the vehicle and the camera is less than $D_{tangent}$, the halo of the headlights is separated. If the distance between the vehicle and camera is greater than $D_{tangent}$, the halo of the vehicle headlights is in a coincident state. Therefore, two situations can be classified and then discussed. The discrimination conditions of the low and high beams satisfy the following relationship in Equation (9).

$$Result = \begin{cases} LowBeam & \frac{S_{Lamp}}{S_{Light}} > \frac{\delta \pm \Delta E_c}{\Delta E_m}, 0 \le D \le D_{tangent} \\ HighBeam & \frac{S_{Lamp}}{S_{Light}} \le \frac{\delta \pm \Delta E_c}{\Delta E_m}, 0 \le D \le D_{tangent} \\ LowBeam & \frac{S_{Lamp}}{S_{Light}} > \frac{\delta' \pm \Delta E_c}{\Delta E_m}, D_{tangent} < D \\ HighBeam & \frac{S_{Lamp}}{S_{Light}} \le \frac{\delta' \pm \Delta E_c}{\Delta E_m}, D_{tangent} < D \end{cases} \quad (9)$$

where D is the real distance between the camera and the vehicle. δ is the ratio between S_{Light} and S_{Lamp} when D is in $[0, D_{tangent}]$. δ' is the ratio between S_{Light} and S_{Lamp} when D is over $D_{tangent}$. ΔE_c is a calculation error. ΔE_m is a measurement error. *LowBeam* and *HighBeam* can be output as the *Result*. $D_{tangent}$ is the distance between the camera and the vehicle when two halos of headlights are just distinguished. According to the pinhole imaging and theorem about similar triangle, it can be calculated by Equation (10).

$$D_{tangent} = \frac{d}{2\tan\frac{\theta}{2}} \quad (10)$$

where θ is the headlight angle, and d is the actual distance of the headlight.

According to the theorem about similar triangle, Equation (11) can be obtained.

$$\frac{L_{DE}}{L_{GH}} = \frac{L_{AE}}{L_{AH}} \quad (11)$$

$$L_{DE} = R_{RealLamp}, L_{GH} = R_{RealLight} \quad (12)$$

$$L_{AE} = D_{RealLamp}, L_{AH} = D_{RealLight} \quad (13)$$

where L_{DE} and $R_{RealLamp}$ are the actual radius width lamp in **Figure 7A**. L_{GH} and $R_{RealLight}$ are the actual radius width halo. L_{AE} and $D_{RealLamp}$ are the distance between headlight focus and lamp. L_{AH} and $D_{RealLight}$ are the distance between headlight focus and halo.

Combining Equations (11)–(13), then δ can be obtained as Equation (14).

$$\delta = \frac{S_{RealLamp}}{S_{RealLight}} = \frac{\pi R_{RealLamp}^2}{\pi R_{RealLight}^2} = \frac{L_{DE}^2}{L_{GH}^2} = \frac{L_{AE}^2}{L_{AH}^2} = \frac{D_{RealLamp}^2}{D_{RealLight}^2} \quad (14)$$

Two halos of the headlights are intersected when the distance between the vehicle and the camera is greater than $D_{tangent}$. The vehicle headlight halo area $S_{RealLight}$ is expressed by Equation (15).

$$\begin{aligned} S_{RealLight} &= 2\pi R_{RealLight}^2 - S_{Intersect} \\ &= 2\pi R_{RealLight}^2 - \left(\frac{\alpha\pi R_{RealLight}^2}{360} - \frac{R_{RealLight}^2 \sin\alpha}{2} \right) \\ &= R_{RealLight}^2 (2\pi - \frac{\alpha\pi}{360} + \frac{\sin\alpha}{2}) \end{aligned} \quad (15)$$

$$\frac{\frac{d}{2}}{R_{RealLight}} = \cos\frac{\alpha}{2} \quad (16)$$

where α is $\angle JI'K$in **Figure 7B**. $S_{Intersect}$ is the intersect area of two halos.

Combining Equations (15) and (16), the vehicle headlight halo area $S_{RealLight}$ can be obtained as Equation (17).

$$\begin{aligned} S_{RealLight} = R_{RealLight}^2 (2\pi &- \frac{\arccos(\frac{d}{2R_{RealLight}})\pi}{360} \\ &+ \frac{\sin(\arccos(\frac{d}{2R_{RealLight}}))}{2}) \end{aligned} \quad (17)$$

$$\delta' = \frac{2S_{RealLamp}}{S_{RealLight}} \quad (18)$$

$$= \frac{2\pi R_{RealLamp}^2}{R_{RealLight}^2 (2\pi - \frac{\arccos(d\cdot(2R_{RealLight})^{-1})\pi}{360} + \frac{\sin(\arccos(d\cdot(2R_{RealLight})^{-1}))}{2})}$$

where δ' is the ratio between S_{Light} and S_{Lamp}, it can be gotten when the distance between the vehicle and camera is greater than $D_{tangent}$.

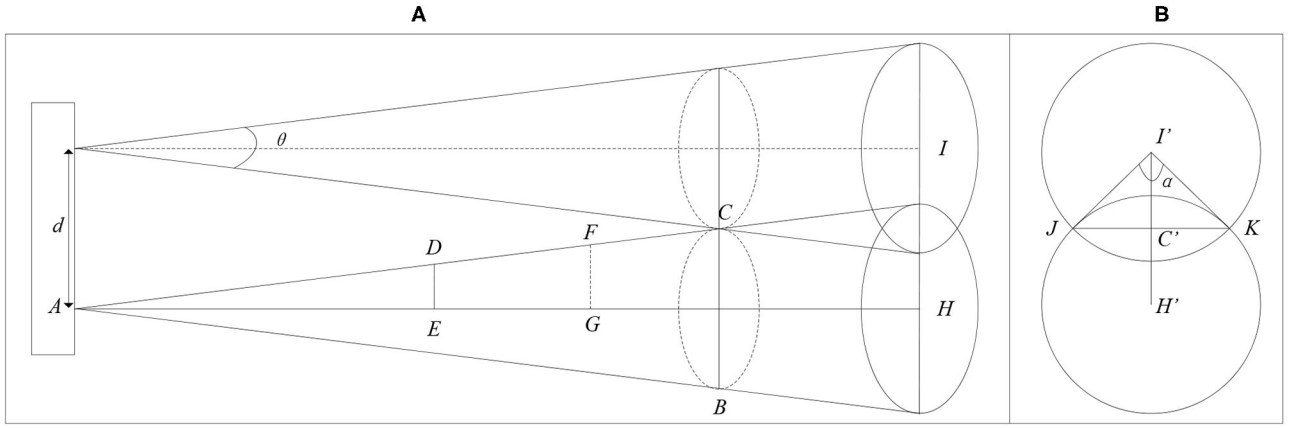

FIGURE 7 | Schematic diagram of the high beam. **(A)** A schematic diagram of cross section about the high beam. **(B)** A schematic diagram of the vertical section about the high beam when the two halos are intersected.

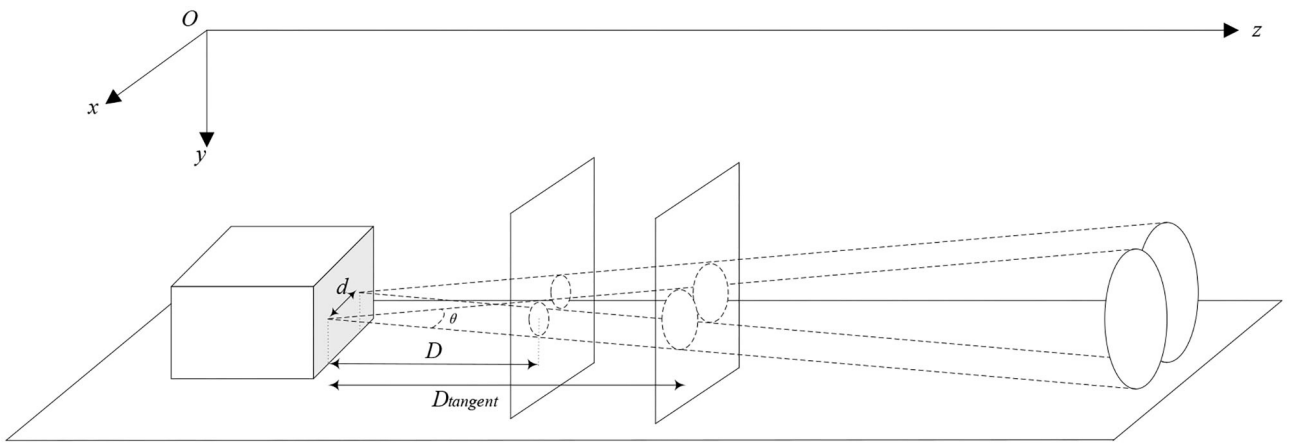

FIGURE 8 | Schematic diagram of the high beam light line space.

Testing Method

For the vehicle headlight distinguishing, an Intersection Over Union (IOU) score of more than 50% is considered a correct detection. Our evaluation method is F-Score ($\beta = 1$) which is defined as Equations (16)–(18) [24]:

$$F - Score = \left(1 + \beta^2\right) \times \frac{Precision \times Recall}{\left(\beta^2 \times Precision\right) + Recall} \quad (19)$$

$$Precision = \frac{TP}{TP + FP} \quad (20)$$

$$Recall = \frac{TP}{TP + FN} \quad (21)$$

where TP is the true-positive. FP is the false-positive. FN is the false-negative.

RESULTS AND DISCUSSION

Dataset and Experimental Platform

For training and testing purposes, the data were accessed from thermal cameras in the urban nighttime road. It is meaningful for supervise whether headlight is used legally by drivers. The thermal stream and RGB stream were both obtained from FLIR ONE PRO camera with 160 × 120 thermal resolution and 1,440 × 1,080 RGB resolution in an 8.7-Hz frame rate. The scene dynamic range is −20 to 400°C. The spectral range of the thermal sensor is about 8–14 μm, and thermal sensitivity (NETD) is 70 mK. The visual resolution, which is 640 × 480 with an iron color scale, was obtained.

In this paper, a computer platform was used for training the deep neural network model and testing. The deep neural network model training was performed by using Slim with TensorFlow v1.13 as the backend on a desktop equipped with 16 GB RAM. The computation was accelerated by using an NVIDIA RTX2080Ti GPU with 12 GB memory. The network

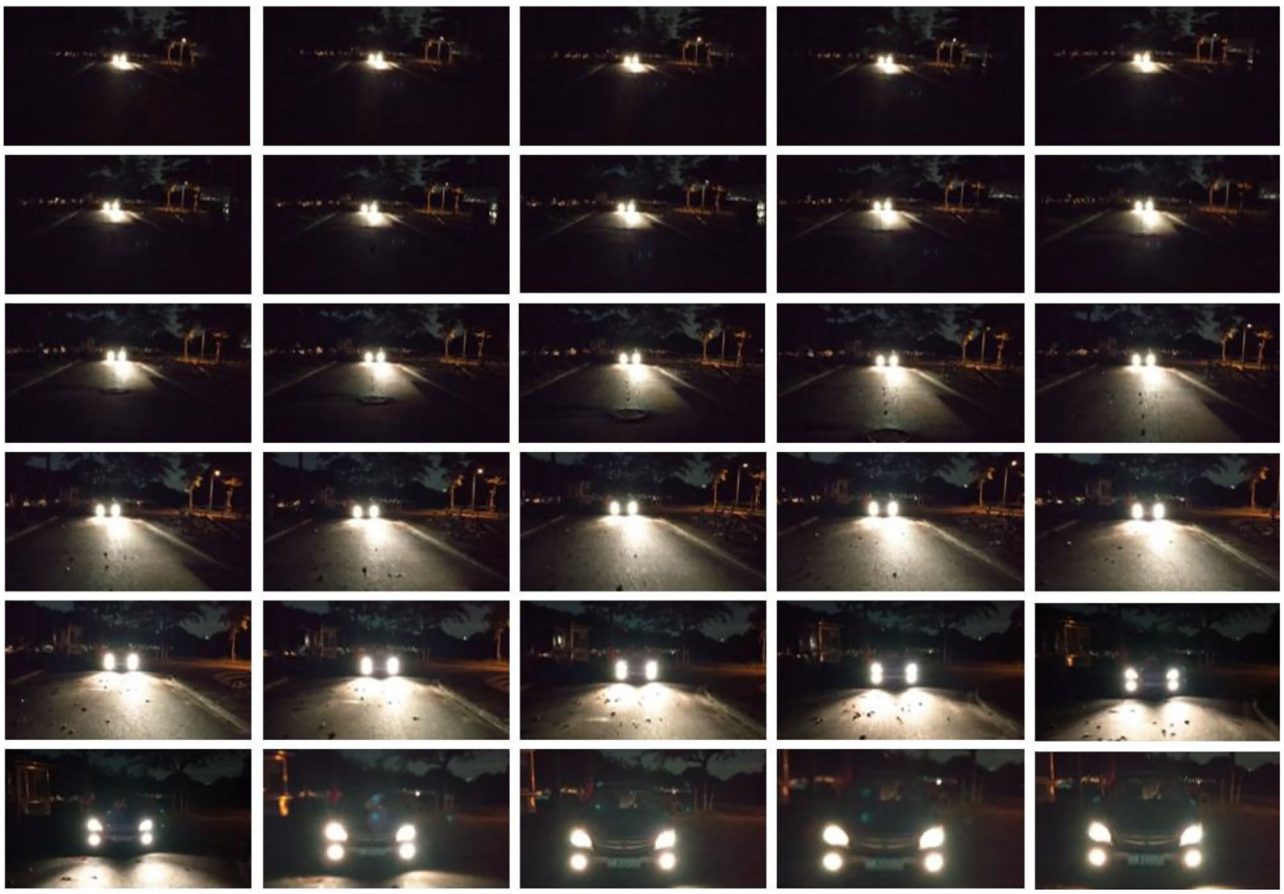

FIGURE 9 | Diagram of the dynamic change process of low beam headlights from far to near.

was trained for 150,000 iterations with a batch size of 8. The optimizer algorithm was "Adam" with a learning rate of 0.001 and a learning rate decay factor of 0.94. In order to avoid over-fitting, local data augmentation was performed through a 2D rotation, translation, and random left-to-right or up-to-down flipping. The ranges of the rotation were $[-45^{\circ}, 45^{\circ}]$ and $[-180^{\circ}, 180^{\circ}]$. After transforming and resizing, the training samples were cropped to $640 \times 480 \times 3$ and were inputted in the deep neural network model.

Performance Testing

For designing the low probability candidate filter, the relationship between the halo of vehicle headlight and vehicle lamp was analyzed. In **Figures 9**, **10**, images were intercepted from the 30-frame live shooting video and intercepted every five frames. As shown in **Figure 9**, when the vehicle headlight dynamically changes from far to near on the RGB image, the halo of the low beam remains clear. As a result, the area of the lamps can be obtained easily. Compared with low beam light, it was more difficult to distinguish the vehicle and its headlight in **Figure 10** than in **Figure 9**, because the halo of the high beam was always in a fusion state in the RGB image. Because the halo of the high beam was always in a fusion state in the RGB image. When the distance between the vehicle and the camera is close enough, the shape of the vehicle headlight can be distinguished easily. Therefore, the low probability candidate filter was designed based on the distance among the vehicle, camera, area of lamp, and headlight.

In order to realize vehicle headlight distinguishing, the dynamic adjustment and distinguishing method for vehicle headlight were designed as shown in **Figure 11.** This method had two parts: thermal image dynamic adjustment and vehicle headlight dynamic distinguishing.

The thermal image enhancement plays an important part in dynamic adjusting. The object detection can be interfered by the ambient temperature and the target object temperature which were displayed in thermal images. After thermal image enhancement, the value of the thermal image was adjusted to a suitable range on the thermal histogram, as shown in **Figure 12B**. Compared with **Figure 1A**, the lights in the thermal image after dynamic adjustment are more prominent, as shown in **Figure 12A**. By using the method of thermal image enhancement, not only the interference features on the image were eliminated but also the target features were enhanced.

After thermal image enhancement, the next step was the thermal image feature fusion. The halo contour features of

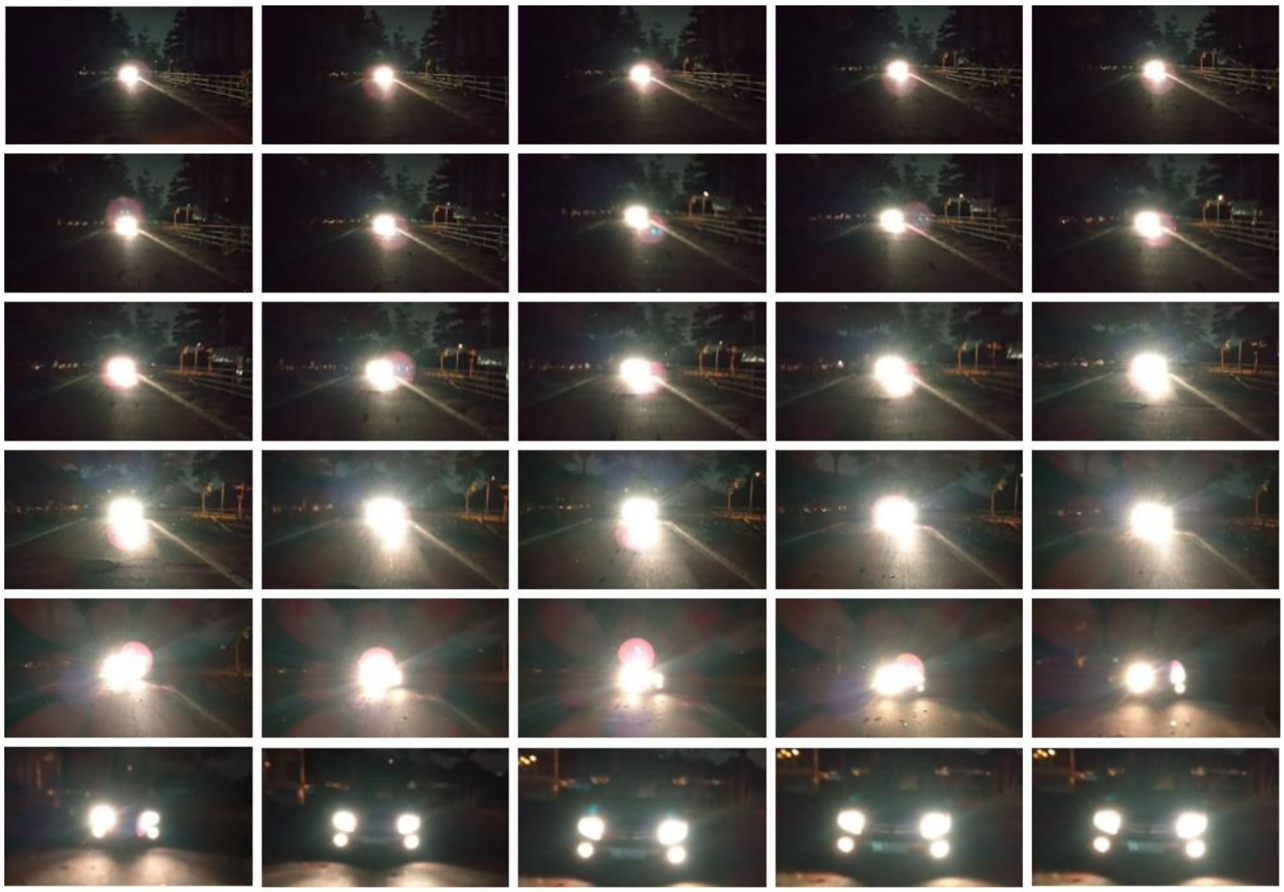

FIGURE 10 | Diagram of the dynamic change process of high beam headlights from far to near.

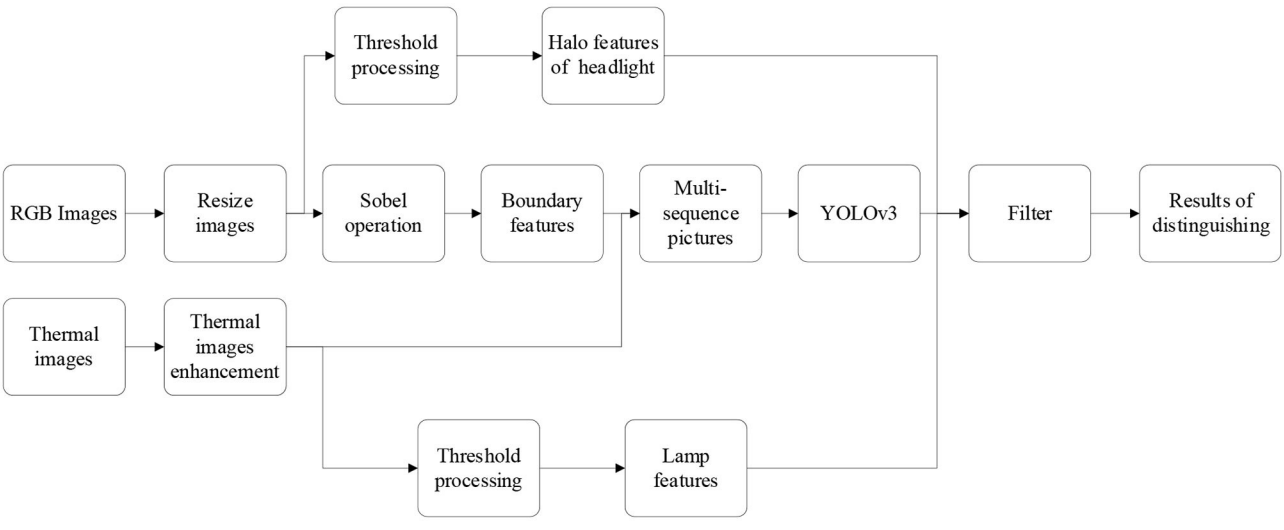

FIGURE 11 | Flowchart of dynamic adjustment and distinguishing method.

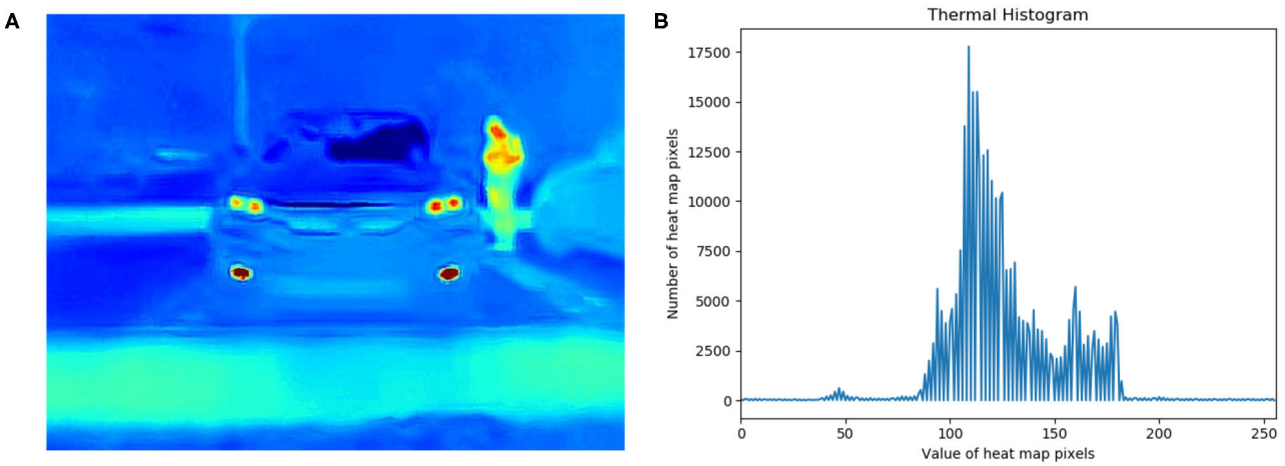

FIGURE 12 | Thermal histogram and thermal image after dynamic adjustment. **(A)** Thermal image after dynamic adjustment. **(B)** Thermal histogram after dynamic transformation.

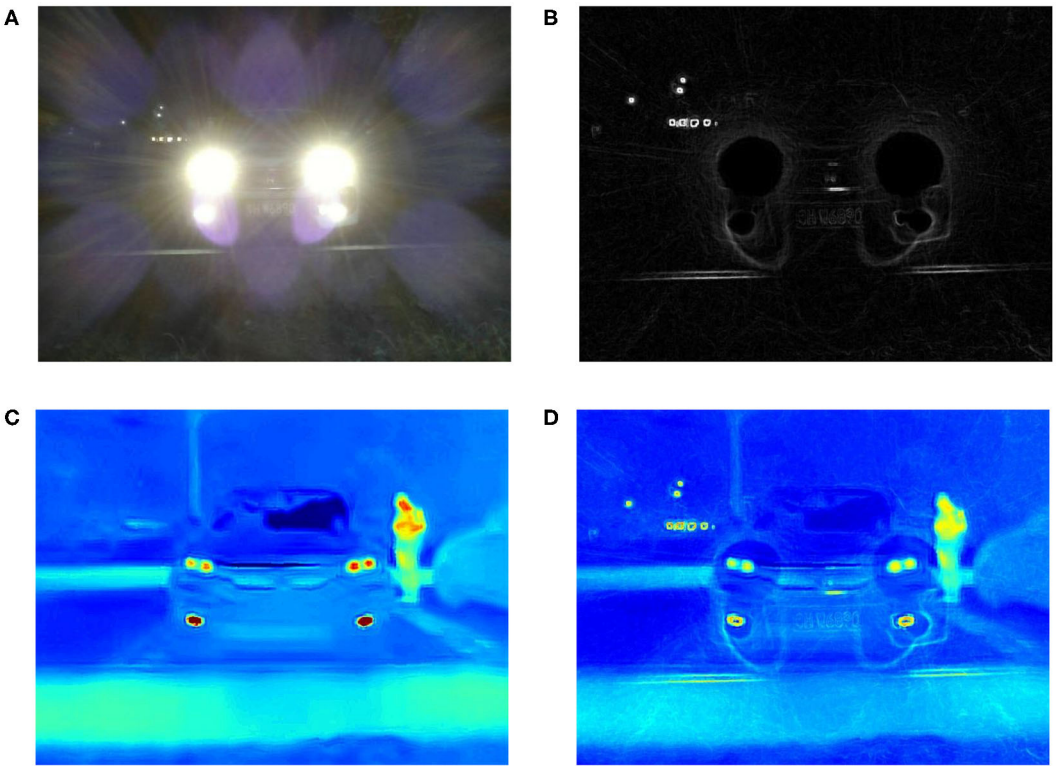

FIGURE 13 | Vehicle headlight feature extraction and fusion. **(A)** The Red Green Blue (RGB) image extracted from the original thermal image. **(B)** The RGB image after Sobel operation. **(C)** The thermal image extracted from the original thermal image. **(D)** A feature map synthesized by a thermal image and an RGB map.

vehicle headlight were extracted by the Sobel operation, as shown in **Figure 13B**. The contour features of the vehicle headlight extracted from the RGB image (**Figure 13A**) were fused with the thermal image (**Figure 13C**). As shown in **Figure 13D**, this figure not only has the information about the thermal image but also the contour information of the RGB image. Furthermore, contour information of the object was strengthened in the thermal image.

Then, the areas of headlight halo and lamp were extracted by the threshold processing for designing the low probability candidate filter. As shown in **Figure 14A**, the people next to the vehicle and other interference were filtered, and only the halo of the vehicle headlight and pixels similar to the vehicle beam can be obtained. As shown in **Figure 14B**, the position information of the vehicle and the headlight on the image is obtained by the

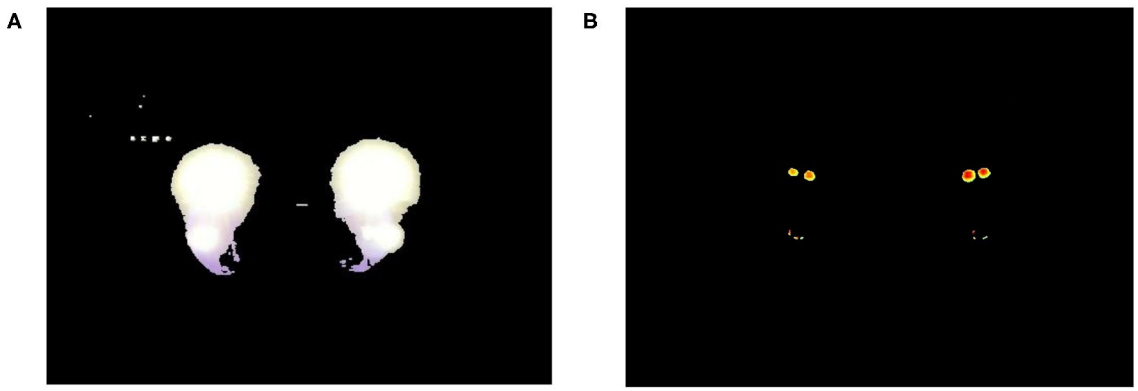

FIGURE 14 | Results of threshold processing. **(A)** A Red Green Blue (RGB) image after threshold processing. **(B)** A thermal image after threshold processing.

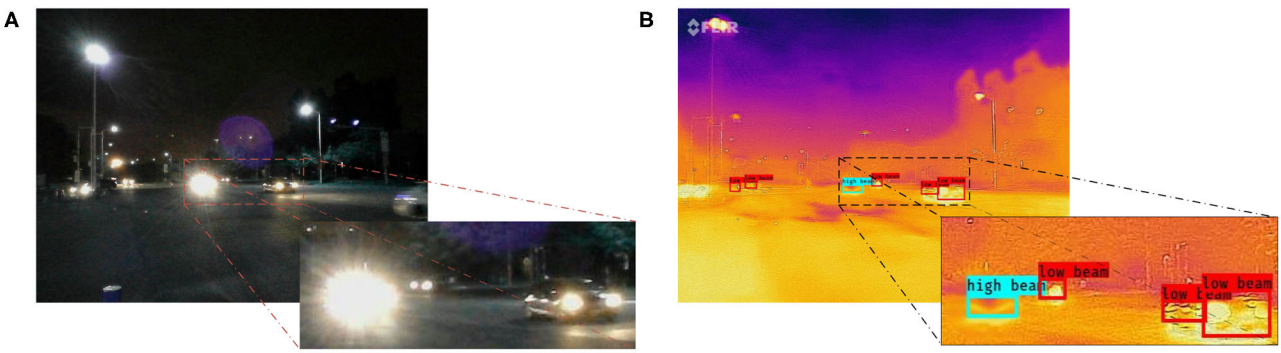

FIGURE 15 | Results of distinguishing. **(A)** A Red Green Blue (RGB) original image. **(B)** A result image with distinguishing.

TABLE 1 | Performance of methods.

Method	Precision	Recall	F-Score	FPS
SSD Net (RGB image) [25]	0.701	0.645	0.671	8
SSD Net-Filter (multi-sequence images)	0.783	0.691	0.734	4
YOLOv3 (RGB image) [18]	0.831	0.736	0.781	13
YOLOv3 (thermal image)	0.847	0.584	0.691	14
YOLOv3 (multi-sequence images)	0.894	**0.792**	0.840	11
YOLOv3-Filter (multi-sequence images)	**0.942**	0.787	**0.858**	9

FPS, frames per second; RGB, Red Green Blue; SSD, single-shot multibox detector. The mean of the bold values is the best valve in current column.

preliminary distinguishing of the deep neural network model, and then the contour of the lamp in the thermal image was extracted by the fixed threshold processing, and the contour of the headlight halo was extracted from the RGB image. The extracted contour was retained when it was within the vehicle candidate box; otherwise, it was discarded. Therefore, we can get the headlight features in **Figure 14B**.

After testing, not only had this method made an achievement of distinguishing between high beam and low beam, but it also overcame the interference caused by halo efficaciously in **Figure 15**. The precision, recall, and F-Score have been improved effectively by our methods. As shown in **Table 1**, we found that the recall rate on the RGB image was 15.2% which was higher than that on the thermal image. The reason is that the resolution of the thermal image was low, and it was impossible to separate the high beam headlight from the low beam headlight.

For the training image data, the recall and the precision of the YOLOv3 on multi-sequence images obtained by dynamic adjustment were 5.6 and 6.3% higher than the RGB image, respectively. The precision and the recall were both increased effectively by using the dynamic adjustment on the thermal image. The halo of headlight information was retained by the multi-sequence images. Moreover, the contour information of distant vehicles and the contour of the vehicle beam can be obtained from the thermal image. For the performance of training models, the precision of the model with adding the filter (YOLOv3-Filter) was improved by 4.8% effectively, and F-Score of YOLOv3-Filter was increased by 1.8%, compared with YOLOv3 on multi-sequence images. In this situation, the filter played a decisive role in the model.

Finally, the dynamic adjustment and distinguishing method have been tested. For the method performance, YOLOv3-Filter

(Multi-sequence images) is the best among these three methods. The precision and the recall were increased by 11.1 and 5.1% compared with YOLOv3 on the RGB image, respectively. Our method has been tested based on single-shot multibox detector (SSD) network which has good performance in small-object detection [25]. After improved, the precision and the recall were increased by 8.2 and 4.6% compared with SSD network on the RGB image, respectively. The data show that the method of this paper has greatly improved the ability of vehicle headlight distinguishing.

In order to confirm the feasibility of using the YOLOv3-Filter method in real-time applications, we carried out comparative experiments on different networks. The single forward inference time for the YOLOv3-Filter (Multi-sequence images) method is 111 ms, which is 34 ms longer than that of the YOLOv3 (RGB image). The main reason for the slightly decreased speed is a complex filter structure and thermal image dynamic adjustment used in the YOLOv3-Filter. Our method shows a big advantage over the SSD network in the performance of detection with a similar term of the operation speed. Overall, YOLOv3-Filter (Multi-sequence images) method does not compromise on the operation time while considerably improving the detection accuracy.

However, there were many factors that lead to low recall in a series of testing methods. Firstly, due to the low resolution of the thermal imager, the thermal image was distorted after being amplified to a certain extent, and the entire contour cannot be effectively restored. Therefore, the information obtained by the thermal imager was inaccurate. Secondly, there were various types of vehicles, and the size of the vehicle is determined by the type of vehicle. As a result, this method has certain errors because of the uncertain size of vehicles. It was necessary to establish a complete database of vehicle types and sizes to solve this problem. Finally, the calculation error ΔE_c was only effective when the camera and the vehicle were in the same straight line. When the deviation angle of the vehicle and the camera can be changed, ΔE_c would also be changed. Digital cameras are projected on a complex system of lens and sensor array susceptible to a variety of undesirable effects. The main effects can be described considering the exposure triangle: aperture, shutter speed, and sensitivity (ISO) [26]. The size of the halo can also be affected by exposure settings of the RGB camera. As the exposure time, aperture, and ISO decreases, the area of the headlight halo captured by the camera decreases. The measurement error ΔE_m can correct for this effect. In this paper, these parameters of the camera have been set before leaving the factory.

CONCLUSION

A vehicle headlight dynamic adjustment and distinguishing method based on data access of a thermal camera was proposed in this paper. Thermal image enhancement and multi-sequence image feature fusion were used as the dynamic adjusting for extracting object features clearly, and the YOLOv3 with adding a filter (YOLOv3-Filter) was used for the dynamic distinguishing. The features between high beam and low beam can be distinguished easily by the filter. So, the proposed dynamic adjustment and distinguishing method not only can enhance the thermal image but also can realize an accurate distinguishing of the high beam and low beam, which provides an effective method for distinguishing the vehicle headlight in night driving and traffic supervision.

DATA AVAILABILITY STATEMENT

The data analyzed in this study was subject to the following licenses/restrictions: the thermal stream and RGB stream were both obtained from FLIR ONE PRO. The data set used in this paper was captured by us at 25°C and the relative humidity was 55%. Requests to access these datasets should be directed to lishixiao0727@outlook.com.

AUTHOR CONTRIBUTIONS

SL: work concept and design and draft paper. YQ: data collection. PB: make important modifications to the paper and approve the final paper to be published. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the Project of Zhongshan Innovative Research Team Program (No. 180809162197886), Special Fund for Guangdong University of Science and Technology Innovation Cultivation (No. pdjh2019b0135), Science and Technology Program of Guangzhou (No. 2019050001), Program for Guangdong Innovative and Enterpreneurial Teams (No. 2019BT02C241), Program for Chang Jiang Scholars and Innovative Research Teams in Universities (No. IRT17R40), Guangdong Provincial Key Laboratory of Optical Information Materials and Technology (No. 2017B030301007), and MOE International Laboratory for Optical Information Technologies and the 111 Project.

REFERENCES

1. Zahran ESMM, Tan SJ, Yap YH, Tan EH, Pena CMF, Yee HF, et al. An investigation into the impact of alternate road lighting on road traffic accident hotspots using spatial analysis. In: *2019 4th International Conference on Intelligent Transportation Engineering (ICITE).* Singapore: IEEE (2019). doi: 10.1109/ICITE.2019.8880263
2. Hwang AD, Tuccar-Burak M, Goldstein R, Peli E. Impact of oncoming headlight glare with cataracts: a pilot study. *Front Psychol*. (2018) **9**:164. doi: 10.3389/fpsyg.2018.00164
3. Zou Q, Ling H, Pang Y, Huang Y, Tian M. Joint headlight pairing and vehicle tracking by weighted set packing in nighttime traffic videos. *IEEE Trans Intell Transport Syst*. (2018) **19**:1950–61. doi: 10.1109/TITS.2017.2745683

4. Dai X, Liu D, Yang L, Liu Y. Research on headlight technology of night vehicle intelligent detection based on hough transform. In: *2019 International Conference on Intelligent Transportation, Big Data and Smart City (ICITBS)*. Changsha: IEEE (2019). p. 49–52. doi: 10.1109/ICITBS.2019.00021
5. Kuang H, Yang KF, Chen L, Li YJ, Chan LLH, Yan H. Bayes saliency-based object proposal generator for nighttime traffic images. *IEEE Trans Intell Transport Syst*. (2018) **19**:814–25. doi: 10.1109/TITS.2017.2702665
6. Lin CT, Huang SW, Wu YY, Lai SH. GAN-based day-to-night image style transfer for nighttime vehicle detection. *IEEE Trans Intell Transport Syst*. (2020) 99: 1–13. doi: 10.1109/tits.2019.2961679
7. Yi ZC, Chen ZB, Peng B, Li SX, Bai PF, Shui LL, et al. Vehicle lighting recognition system based on erosion algorithm and effective area separation in 5g vehicular communication networks. *IEEE Access*. (2019) **7**:111074–83. doi: 10.1109/access.2019.2927731
8. Wu JT, Lee JD, Chien JC, Hsieh CH. Nighttime vehicle detection at close range using vehicle lamps information. In: *2014 International Symposium on Computer, Consumer and Control (IS3C)*. Vol. 2. Taichung: IEEE (2014). p. 1237–40. doi: 10.1109/IS3C.2014.320
9. Pradeep CS, Ramanathan R. An improved technique for night-time vehicle detection. In: *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. Bangalore: IEEE (2018). p. 508–13. doi: 10.1109/ICACCI.2018.8554712
10. Chen XZ, Liao KK, Chen YL, Yu CW, Wang C. A vision-based nighttime surrounding vehicle detection system. In: *2018 7th International Symposium on Next-Generation Electronics (ISNE)*. Taipei: IEEE (2018). p. 1–3. doi: 10.1109/ISNE.2018.8394717
11. Chang CW, Srinivasan K, Chen YY, Cheng WH, Hua KL. Vehicle detection in thermal images using deep neural network. In: *2018 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE (2018). p. 7–10. doi: 10.1109/VCIP.2018.8698741
12. Satzoda RK, Trivedi MM. Looking at vehicles in the night: detection and dynamics of rear lights. *IEEE Trans Intell Transport Syst*. (2019) **20**:4297–307. doi: 10.1109/TITS.2016.2614545
13. Shan Y, Sawhney HS, Kumar R. Unsupervised learning of discriminative edge measures for vehicle matching between nonoverlapping cameras. *IEEE Trans Pattern Anal Mach Intell*. (2008) **30**:700–11. doi: 10.1109/TPAMI.2007.70728
14. Birogul S, Temur G, Kose U. YOLO object recognition algorithm and "buy-sell decision" model over 2D candlestick charts. *IEEE Access*. (2020) **8**:91894–915. doi: 10.1109/ACCESS.2020.2994282
15. Chien SC, Chang FC, Tsai CC, Chen YY. Intelligent all-day vehicle detection based on decision-level fusion using color and thermal sensors. In: *2018 International Conference on Advanced Robotics and Intelligent Systems (ARIS)*. Taipei: IEEE (2018). doi: 10.1109/ARIS.2017.8297189
16. Cygert S, Czyzewski A. Style transfer for detecting vehicles with thermal camera. In: *2019 Signal Processing - Algorithms, Architectures, Arrangements, and Applications Conference Proceedings (SPA)*. Vol. 9. IEEE (2019). p. 218–22. doi: 10.23919/SPA.2019.8936707
17. Zheng Y, Blasch E, Cygert S, Czyzewski A, Sangnoree A, Chamnongthai K, et al. Robust method for analyzing the various speeds of multitudinous vehicles in nighttime traffic based on thermal images. *IEEE Trans Intell Transport Syst*. (2018) **9871**:7–10.
18. Wei X, Wei D, Suo D, Jia L, Li Y. Multi-target defect identification for railway track line based on image processing and improved YOLOv3 model. *IEEE Access*. (2020) **8**:61973–88. doi: 10.1109/ACCESS.2020.2984264
19. Vinyals O, Toshev A, Bengio S, Erhan D. Show and tell: lessons learned from the 2015 MSCOCO image captioning challenge. *IEEE Trans Pattern Anal Mach Intell*. (2017) **39**:652–63. doi: 10.1109/TPAMI.2016.2587640
20. Kargel C. Thermal imaging to measure local temperature rises caused by hand-held mobile phones. In: *IEEE Instrumentation and Measurement Technology Conference*. Vol. 2. Como: IEEE (2004). p. 1557–62. doi: 10.1109/imtc.2004.1351363
21. Israni S, Jain S. Edge detection of license plate using Sobel operator. In: *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. Chennai: IEEE (2016). p. 3561–3. doi: 10.1109/ICEEOT.2016.7755367
22. Todd J. Digital image processing (second edition). *Opt Lasers Eng*. (1988) **8**:1–71. doi: 10.1016/0143-8166(88)90012-7
23. Zhang Z. A flexible new technique for camera calibration. *IEEE Trans Pattern Analy Mach Intell*. (2000) **22**:1330–4. doi: 10.1109/34.888718
24. Heaton J. Deep learning. In: Goodfellow I, Bengio Y, Courville A, editors. *Genetic Programming and Evolvable Machines*. Massachusetts: MIT press (2018). p. 424–5. doi: 10.1007/s10710-017-9314-z
25. Qu J, Su C, Zhang Z, Razi A. Dilated convolution and feature fusion SSD network for small object detection in remote sensing images. *IEEE Access*. (2020) **8**:82832–43. doi: 10.1109/ACCESS.2020.2991439
26. Steffens CR, Drews-jr PLJ, Botelho SS, Grande R. Deep learning based exposure correction for image exposure correction with application in computer vision for robotics. In: *2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)*. Joao Pessoa: IEEE (2018). doi: 10.1109/LARS/SBR/WRE.2018.00043

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 01 October 2020
doi: 10.3389/fphy.2020.562456

Photonics-Assisted Multi-Band Microwave Receiver Based on Spectrum Analysis and Coherent Channelization

Huan Huang[1], Chongfu Zhang[1,2], Wei Zheng[1], Yong Chen[1], Haifeng Yang[3], Zichuan Yi[2], Feng Chi[2] and Kun Qiu[1]*

[1] School of Information and Communication Engineering, Zhongshan Institute, University of Electronic Science and Technology of China, Chengdu, China, [2] School of Electronic Information, University of Electronic Science and Technology of China, Zhongshan Institute, Zhongshan, China, [3] Southwest Institute of Electronic Technology (SIET), Chengdu, China

Due to the limitation of analog-to-digital/digital-to-analog converters, photonics-assisted channelized receivers are thought to be a promising approach to receiving wideband microwave signals. Herein, based on the spectrum analysis and the coherent channelization, we develop a photonics-assisted channelized receiver for multi-band microwave signals. In the proposed channelized receiver, the instantaneous spectral analysis is introduced to determine the frequencies and bandwidth of a dynamic wideband signal. The dynamic wideband signal is then received by a multi-band coherent channelizer. By exploiting transparency of the optoelectronic devices, we equivalently build a multi-band coherent channelizer that can work for dynamic microwave signals, where a few optoelectronic devices are required. Compared with the existing coherent channelizers, the operative bandwidth of the proposed multi-band coherent channelizer is much larger (up to 28 GHz). The proposed photonics-assisted channelized receiver doesn't need tunable optical comb generators and radio frequency sources, and it also doesn't require knowing the spectrum information in advance. Moreover, the designed example of the proposed photonics-assisted channelized receiver for a 4 GHz original signal in 2–30 GHz microwave bands is given and discussed. The spectrum information of the dynamic original signals is obtained by monitoring the optical power in each sub-channel. We verify the feasibility of the multi-band coherent channelizer by channelizing two 4 GHz linearly-chirped signals with center frequencies of 4 and 28 GHz into four 1 GHz sub-channels, respectively.

Keywords: optical signal processing, optoelectronic device, spectrum analysis, radio frequency photonics, channelization

OPEN ACCESS

Edited by:
Olivier J. F. Martin, École Polytechnique Fédérale de Lausanne, Switzerland

Reviewed by:
Nirmal Mazumder, Manipal Academy of Higher Education, India
Youbin Yu, Zhejiang Sci-Tech University, China

****Correspondence:***
Chongfu Zhang
Zhangcfzhang@uestc.edu.cn

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *15 May 2020*
Accepted: *11 September 2020*
Published: *01 October 2020*

Citation:
Huang H, Zhang C, Zheng W, Chen Y, Yang H, Yi Z, Chi F and Qiu K (2020) Photonics-Assisted Multi-Band Microwave Receiver Based on Spectrum Analysis and Coherent Channelization. Front. Phys. 8:562456. doi: 10.3389/fphy.2020.562456

INTRODUCTION

Microwave signal receivers with larger bandwidth and higher frequency are a pressing need in both commercial and defense applications such as modern radars, communication satellites, radio over fiber systems, and the 5th generation mobile network [1–5]. However, wideband signal receivers are limited to analog-to-digital/digital-to-analog converters [6]. To handle the wideband signal reception with this limitation, channelized receivers that are a specialized class of pre-processors have attracted

much attention. A channelized receiver is employed to channelize a wideband signal into N narrowband slices and then these narrowband slices are processed in parallel [7–11], where the processing band of each narrowband is reduced to $1/N$. Generally, the operative bandwidth of an electric channelized receiver is limited to a few GHz, because the "Electronic Bottleneck" resists in it [12, 13]. In a photonics-assisted channelized receiver, the operative bandwidth can reach several GHz (tens of GHz) [14–18]. Moreover, photonics-assisted channelized receivers have other advantages such as high compactness, low loss, and immunity to electromagnetic interference [19, 20].

The existing works about photonics-assisted channelized receivers can be classified into three typical categories. In the first channelization category, a radio frequency (RF) signal is first modulated on an optical carrier (OC), and a series of optical filters with consecutive passbands are then introduced to slice the modulated optical signal [10]. The optical filters such as integrated Bragg grating Fabry–Perot devices [10], diffraction gratings [21], and acousto-optic crystals [22] are available. This channelization scheme requires narrowband, flat-top, steep-edge, and precisely centered filter banks to ensure effective separation of slices, hence it suffers from limitations in spectral slice resolution and the available number of channels [19, 23]. Therefore, the second channelization category has been proposed, where multicast copies were generated by modulating the original RF signal on a series of OCs. Then an optical comb filter such as a Fabry-Perot filter (FPF) with a special free spectrum range (FSR) was employed to separate different slices from multicast copies [7–9, 17, 24]. This photonic channelizer is also limited to the comb filter, so the comb filter is desired to have narrowband, flat-top, and steep-edge transmission peaks (TPs). In practice, optical filters with ideal TPs are difficult to achieve. Therefore, application scenarios for these two channelization categories are mostly the instantaneous spectral analysis [7, 25].

Recent approaches to photonics-assisted coherent channelization without optical filters have also been demonstrated [11, 23, 26–30]. This channelization scheme, i.e., the third photonic channelization category demands no optical filters, but it requires a couple of coherent comb light sources with detuning comb line spacing [23, 30]. Specifically, one coherent comb light source provides comb lines for generating optical multicast copies, and the other is employed to generate local oscillators (LOs). It is worth mentioning that the LOs in the third photonic channelization category can be replaced by dual-polarization LOs [31, 32], where the required optical comb lines (OCLs) are reduced. Moreover, other photonics-assisted channelized receivers have also been reported, for instance, the approach relying on wavelength scanning structures [33]. The third-category photonic channelization requires no optical filters, so it can reach the maximum suppression ratio of spurs in each sub-channel. For different scenarios, microwave signals are required to have different bands such as the S-band, C band, and K band [34–37]. Therefore, a microwave signal receiver working at many microwave bands is demanded. However, in existing works about the third photonic channelization category, the operative bandwidth is generally several GHz, and it is difficult to achieve tens of GHz operative bandwidth. Moreover, most of the existing coherent channelized receivers are designed for fixed microwave bands, where the spectrum information should be known in advance. Although the work in [11] has presented a reconfigurable coherent channelized receiver, tunable RF sources are required.

In this work, a photonics-assisted channelized receiver based on the spectrum analysis and the coherent channelization is proposed for multi-band microwave signals. The implementation of this photonics-assisted channelized receiver is twofold including the instantaneous spectral analysis and the coherent channelization. First, the instantaneous spectral analysis is employed to obtain the spectrum information. In the instantaneous spectral analysis, an efficient second-category photonic channelizer is built to obtain the spectrum information because of the large operative bandwidth and low-complexity of the second-category photonic channelizer. Second, a multi-band coherent channelizer belonging to the third-category photonic channelization is designed for a virtual signal covering the entire frequency range. Based on the transparency of photonic devices, we equivalently build a multi-band coherent channelizer that can work for dynamic microwave signals by using a few devices. After determining the spectrum information of the original dynamic signal, this wideband signal with dynamic microwave bands is received by the multi-band coherent channelizer.

PRINCIPLES

Photonics-Assisted Channelized Receiver for Multi-Band Microwave Signals

Figure 1 schematically depicts the proposed photonics-assisted channelized receiver for multi-band microwave signals. The implementation of this channelized receiver is twofold including the instantaneous spectral analysis and the coherent channelization. First, the instantaneous spectral analysis is performed for an original wideband signal to determine its frequencies and bandwidth, i.e., the spectrum information, as shown in the upper part of **Figure 1**. Second, introducing two ultra-flat optical comb generators (U-FOCGs), the original signals are received by the multi-band coherent channelizer, which is illustrated in the lower part of **Figure 1**. By introducing a controller to control the selected OCLs, we equivalently build a coherent channelized receiver by a few devices for multi-band microwave signals.

In the instantaneous spectral analysis, an efficient photonic channelizer belonging to the second category is designed to analyze the spectrum information of the original dynamic signal.

Firstly, an U-FOCG is implemented to provide OCLs, where two intensity modulators (IMs) and one phase modulator (PM) are cascaded in turn. A demultiplexer called DeMux 1 is then employed to select a series of OCLs from the output of U-FOCG, where adjacent OCLs have the wavelength spacing of δ_c. The center frequencies (CFs) of the selected OCLs are termed as $f_1^c, f_2^c \ldots$ and f_n^c, where n is equal to $[(f_H - f_L)/2\Delta\delta]$, the original

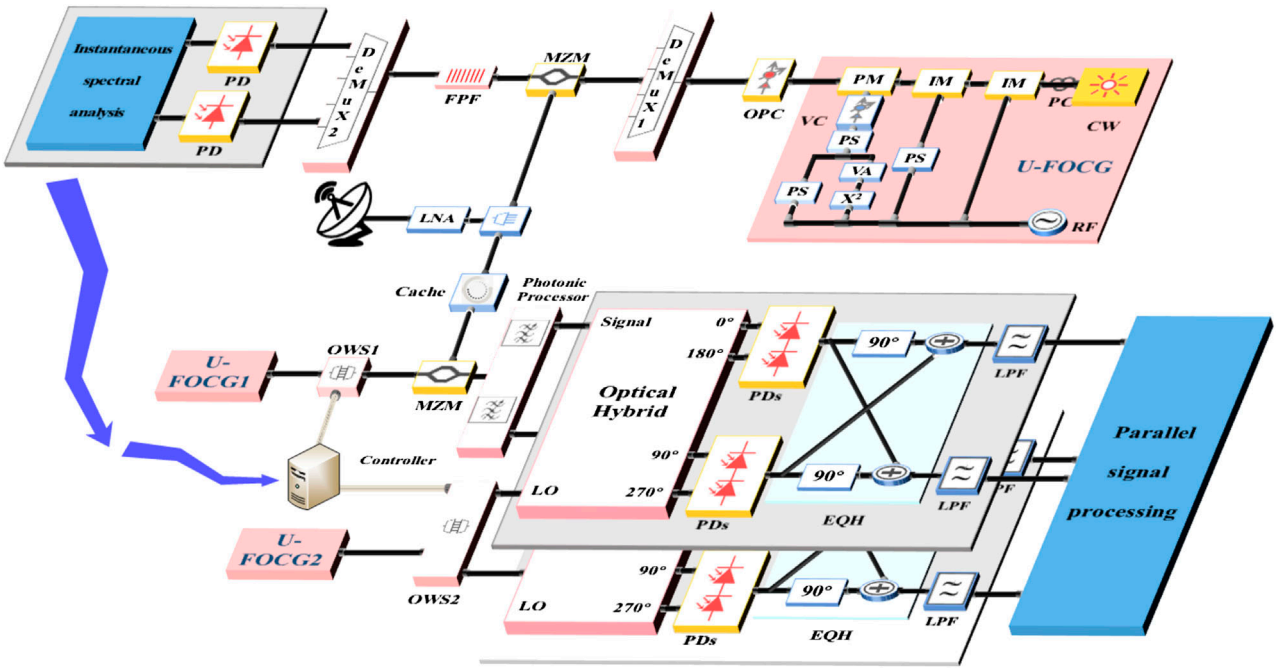

FIGURE 1 | Schematic diagram of the proposed photonics-assisted channelized receiver based on the spectrum analysis and the coherent channelization for multi-band microwave signals. U-FOCG, ultra-flat optical comb generator; OPC, optical power control module; DeMux, demultiplexer; MZM, Mach-Zehnder modulator; FPF, Fabry-Perot filter; LNA, low noise amplifier; PD, photodetector; OWS, optical wavelength selection; EQH, electrical quadrature hybrid; LPF, low-pass filter.

dynamic signal has frequencies between f_L and f_H, $[(f_H - f_L)/2\Delta\delta]$ represents that rounds $(f_H - f_L)/2\Delta\delta$ to the nearest integer, and $\Delta\delta$ is defined as the detuning parameter. The spectra of the selected OCLs are depicted in **Figure 2A**.

The wavelength spacing δ_c is determined by the driven RF source in the U-FOCG. Besides, we assume that the original dynamic signal has frequencies between f_L and f_H. Generally, the bandwidth B of the original signal is much less than the frequency range $(f_H - f_L)$. By setting the driven RF source, the OCLs selected from the U-FOCG satisfy the following spectral relationship [11, 30],

$$f_i^c = f_1^c + 4(i-1) f_{bas} - \Delta\delta, \tag{1}$$

where $f_{bas} = (f_H + f_L)/2$ and $1 \le i \le n$.

Secondly, the optical powers of the selected OCLs are adjusted by an optical power control module, and these OCLs are then injected into a Mach-Zehnder modulator. After double sideband (DSB) modulation, 2n multicast copies are generated and located on the ±1st sidebands of each OCL, as shown in **Figure 2B**. Observe that the original signal is amplified by a low noise amplifier [38, 39] before DSB modulation.

Thirdly, a FPF is taken as a comb filter whose FSR δ_{FPF} is equal to $\delta + \Delta\delta/2$, as plotted in **Figure 2C**, where the detuning parameter $\Delta\delta$ represents the bandwidth of each sub-channel. In other words, the wavelength spacing of those selected OCLs, the FSR of the FPF, and the frequency range of the original signal are optimized as

$$\delta_c = 2\delta_{FPF} - \Delta\delta = 4 f_{bas} - \Delta\delta = 2(f_H + f_L) - \Delta\delta \tag{2}$$

The FPF with $2n$ TPs is introduced to channelize the original signal into a series of narrowband sub-channels, which is termed as *Sub-ch1*, *Sub-ch2* ... *Sub-ch2n*. Specifically, the relationship between the CFs of the first TP and the first OCL satisfies the following equation,

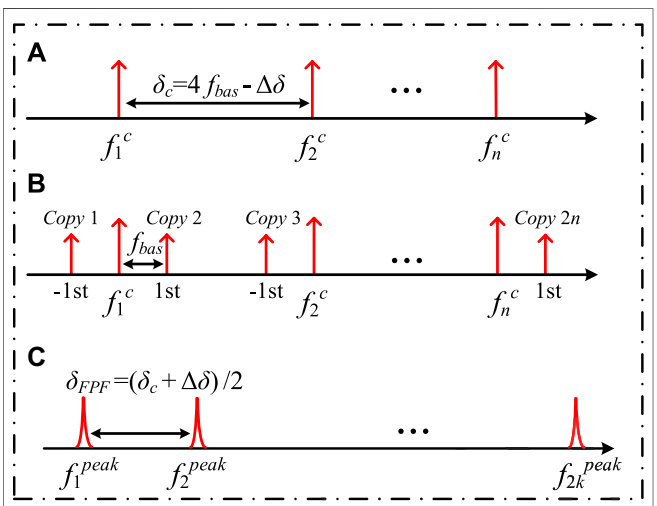

FIGURE 2 | Spectra diagrams of **(A)** the selected optical comb lines, **(B)** modulated optical comb lines after double sideband modulation, and **(C)** transmission peaks of the Fabry-Perot filter.

$$f_1^{peak} = f_1^c - f_H - \frac{\Delta\delta}{2} \tag{3}$$

Moreover, the relationship between the full width half maximum pass-band width δ_{3dB} and the δ_{FPF} has the following form,

$$\frac{\Delta\delta_{FPF}}{\Delta\delta_{3dB}} = \frac{\pi\sqrt{1-t}}{t} \tag{4}$$

where t is the power transmission coefficient of the mirrors in the FPF.

Finally, these $2n$ slices are separated by a demultiplexer termed as DeMux 2, where the channels of DeMux 2 match the FPF's TPs. These slices can be detected by photodetectors (PDs) in parallel, and the frequency range and bandwidth are then determined according to the optical power in each sub-channel. Only a few sub-channels have significant optical power after detecting by PDs because the bandwidth B is much less than the frequency range ($f_H - f_L$). The spectrum information will be employed to control the next coherent channelization.

In the coherent channelization, we equivalently build the multi-band coherent channelizer that can work for microwave signals with a large frequency range ($f_L \le f \le f_H$). We assume that a virtual multi-band microwave signal with frequencies from f_L to f_H is channelized and then received, where the bandwidth and CF of the virtual microwave signal are equal to (f_H–f_L) and ($f_H + f_L$)/2, respectively.

Firstly, we prepare m $\{m = [(f_H - f_L)/4\delta_S]\}$ uniformly-spaced comb lines with specific wavelength spacing as OCs, which can be generated by a U-FOCG termed as U-FOCG 1. An optical wavelength selection (OWS) module such as a wavelength selector switch is introduced to select those OCs from the output of U-FOCG 1, as shown in **Figure 1**. The CF of the rth OC can be express as,

$$f_r^{oc} = f_1^{oc} + (r-1)(4\,f_{bas} - 2\delta_s) \tag{5}$$

where f_1^{oc} represent the CF of the first OC, $1 \le r \le m$, and δ_s represents the bandwidth of each sub-signal after channelization.

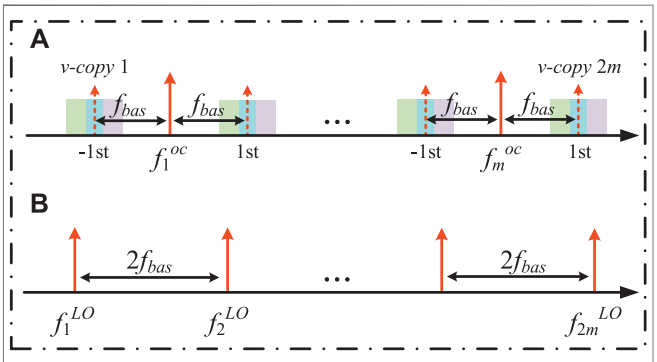

FIGURE 3 | Spectra diagrams of **(A)** the modulated optical carriers s after double sideband modulation and **(B)** local oscillators generated from ultra-flat optical comb generator 2.

Secondly, the virtual wideband signal is modulated on those OCs to generate multicast copies, and the virtual wideband signal is saved in a cache before modulation. Assuming those m OCs are all modulated, $2m$ optical multicast copies of the virtual wideband signal termed as v-copies are obtained after DSB modulation, as illustrated in **Figure3A**.

Thirdly, $2m$ LOs are introduced for coherent detection which can be selected from other U-FOCG (U-FOCG 2), as shown in **Figure 3B**. These $2m$ LOs have a fixed wavelength spacing of $2f_{bas}$. Besides, the CFs of the first LO and the OC have the following relationship that is

$$f_1^{LO} = f_1^{oc} - f_H + \delta_s \tag{6}$$

Next, the photonic processor such as the tunable optical filter (TOF) whose channels match those $2m$ v-copies is introduced to separate them, and an OWS 2 is then employed to separate those $2m$ LOs, as shown in **Figure 1**. When the existing works about the third-category photonic channelization are built for directly channelizing the virtual wideband signal, there are $2m$ optical hybrids (OHs) [40], $8m$ PDs or $4m$ balance photodetectors (BPDs), $2m$ electrical quadrature hybrids (EQHs), and $4m$ electrical low-pass filters (LPFs) at least. Therefore, the coherent channelized receiver will be very complex and redundant, where huge photonic and electrical deceives are wasted. By exploiting the reconfigurability proposed in [11], although $2k$ OHs, $8k$ PDs or $4k$ BPDs, $2k$ EQHs, and $4k$ LPFs $[k = (B/4\delta_S)]$ are enough for channelizing the original wideband signal with dynamic microwave bands, the RF sources must be tunable. In practical application, the bandwidth of a microwave signal is much less than its frequency range, so k is far less than m ($k << m$). Generally, the bandwidth and microwave bands of an original signal are invariant, except the reconfigurable coherent channelizer in [11]. Moreover, in the previous coherent channelized works, it is assumed that the spectrum information is known before the reception.

In this work, we designed a multi-band coherent channelizer which only requires $2k$ OHs, $4k$ PDs, $2k$ EQHs, and $4k$ electrical LPFs, and the CFs of RF sources are fixed. Moreover, for an original signal with dynamic frequencies between f_L and f_H, its bandwidth and microwave bands are given via the instantaneous spectral analysis. So, the spectrum information is not required knowing in advance.

Assuming $2m$ separated v-copies and LOs have been generated. Every *v-copy* consists of different components termed as C_1, C_2 … C_N, where $N = [(f_H - f_L)/\delta_S]$. Most of the components in a *v-copy* are null because the bandwidth B is much less than (f_H–f_L). Taking the first and second v-copies as examples, the spectrum diagrams of the detected signal after the OH and PDs are shown in **Figures 4A, B**. Due to the square-law detection of a PD, the left components of the *v-copy* 1 that have lower frequencies than f_1^{LO} overlap the right components, which is called "Spectrum Aliasing." Moreover, some components are transformed into symmetrical forms after PDs, which are written as C_1^*, C_2^* … C_N^*. This effect can be eliminated by digital signal processing. Similarly, the spectrum diagrams for the first four v-copies and LOs can be obtained and **Figure 4**.

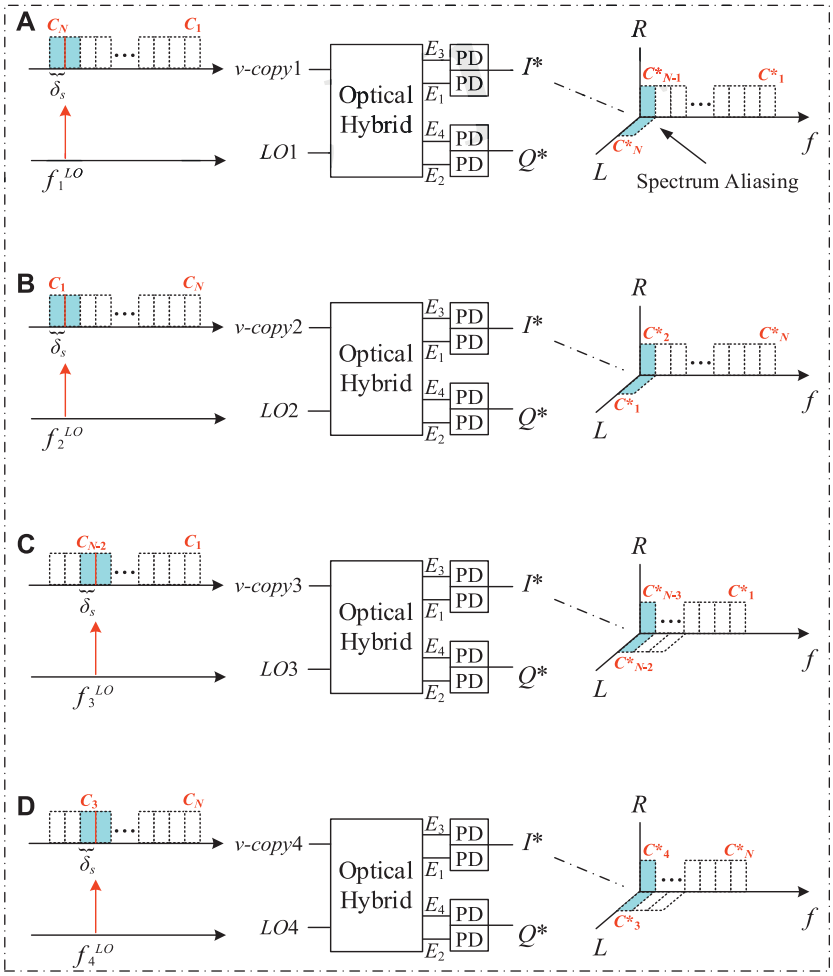

FIGURE 4 | Schematic diagrams of 2 × 4 90° optical hybrid reception for the first four *v-copies* and local oscillators (LOs), where **(A–D)** show the spectra of the *v-copy*1 and the LO 1, the *v-copy*2 and the LO 2, the *v-copy*3 and the LO 3, as well as the *v-copy*4 and the LO 4, and the spectra of them after detecting by photodetectors (PDs), respectively.

To address the spectrum aliasing problem, symmetric EQHs displayed in **Figure 1** are employed. According to the scattering matrix of the EQH [41], the two output signals of the first EQH have the following form,

$$\begin{pmatrix} I \\ Q \end{pmatrix} \propto \begin{pmatrix} 1 & j \\ j & 1 \end{pmatrix} \begin{pmatrix} I^* \\ Q^* \end{pmatrix} = 2 \begin{pmatrix} C_N^* \\ jC_{N-1}^* \cdots jC_1^* \end{pmatrix} \tag{7}$$

In **Eq. (7)**, the I^* and Q^* signals are generated after processing by the OH and PDs which have the following form,

$$\begin{pmatrix} I^* \\ Q^* \end{pmatrix} \propto \Re \sqrt{P_{sig} P_{LO}} \begin{pmatrix} \cos\left[\left(\omega_{sig} - \omega_{LO}\right)t + \left(\varphi_{sig} - \varphi_{LO}\right)\right] + n_1 \\ \sin\left[\left(\omega_{sig} - \omega_{LO}\right)t + \left(\varphi_{sig} - \varphi_{LO}\right)\right] + n_2 \end{pmatrix} \tag{8}$$

where $\Re$ is the responsivity of the PDs, P_{sig} and P_{LO} denote the powers of *v-copy* 1 and LO 1, ω_{sig} and ω_{LO} represent the angular frequencies of *v-copy* 1 and LO 1, φ_{sig} and φ_{LO} are the phases of *v-copy* 1 and LO 1. Moreover, n_1 and n_2 are the additive Gaussian white noise induced by the PDs. From **Eq. (8)**, it is known that the phase and amplitude noise of a laser affect the performance of the proposed channelized receiver. More details about the noise limitations can refer to [11].

Two tunable LPFs with a bandwidth of δ_s are introduced to select the corresponding components from those I and Q signals respectively. Namely, two components $C_N^\star$ and $C_{(N-1)}^\star$ can be obtained from the *v-copy* 1. Similarly, other components can also be selected from other v-copies. Finally, parallel signal processing is performed for those narrowband components, where the electrical processers only require a processing band of δ_s.

The optoelectronic devices such as the OHs, PDs, EQHs, and LPFs in **Figure 4** are transparent for the input. The output components after LPFs are only determined by different v-copies and LOs. Then we can only work effective sub-channels and stop those sub-channels that are corresponding to the null components of the virtual multi-band microwave signal. For example, the virtual multi-band microwave signal only contains components C_1, C_2, C_{N-1}, and C_N, we can channelize and receive this microwave signal by employing *v-copy* 1, *v-copy* 2, LO 1, and LO 2. When the virtual

TABLE 1 | Relationship among multicast copies, local oscillators, and components.

Multicast copies	*v-copy* 1	*v-copy* 2	*v-copy* (2*m*−1)	*v-copy* 2*m*
Local oscillators	*LO* 1	*LO* 2	*LO* (2*m*−1)	*LO* 2*m*
Components	C_N C_{N-1}	C_1 C_2	$C_{(N-2m)}$ $C_{(N-2m+1)}$	$C_{(2m-1)}$ C_{2m}

multi-band microwave signal contains components C_{N-3}, C_{N-2}, C_{N-1}, and C_N at other times, it can be channelized and received by using *v-copy* 2, *v-copy* 3, LO 2, and LO 3. However, the OHs, PDs, EQHs, and LPFs used in the coherent channelization are the same for these two different virtual multi-band microwave signals, because these optoelectronic devices are transparent for the input.

As shown in **Figure 1**, a controller connected to OWS 1 and OWS 2 is introduced to select the necessary OCs (corresponding to specific v-copies) and LOs by the spectrum information from the instantaneous spectral analysis. Therefore, many photonic and electronic devices can be saved. For simplicity, we assume that the $f_H - f_L/4\delta_S$ is a positive integer, so the integer N is equal to $4m$. The relationship among v-copies, LOs, and the channelized components is shown in **Table 1**.

In other words, we equivalently build a coherent channelized receiver for multi-band microwave signals (between f_L and f_H) by using T^*k devices rather than T^*m devices ($k << m$), where T is a positive integer. Compared with the work in [11], the RF sources do not require tuning, and there is no need to know the spectrum information in advance.

Photonics-Assisted Multi-Band Receiver for Dynamic Signals in 2–30 GHz Bands

Most of the initial coherent channelized receivers [23, 26–32] were designed for fixed microwave bands, where the spectrum information must be known in advance. Those coherent channelized receivers can only work for specific microwave bands. When those coherent channelized receivers are deployed for receiving dynamic 4 GHz signals with frequencies between 2 and 30 GHz, they must have an operative bandwidth of 28 GHz regardless of the actual bandwidth. Therefore, lots of active processing devices are wasted. Moreover, the work in [11] has presented a reconfigurable coherent channelized receiver. Although only 4 GHz operative bandwidth is required based on this scheme, the RF sources employed must be tunable.

Based on the proposed scheme, the instantaneous spectral analysis for the dynamic 4 GHz signals has been finished by the following steps. Setting the detuning parameter $\Delta\delta$ to be 1 GHz, 28 sub-channels are required to monitor the dynamic 4 GHz signal. Firstly, 14 OCLs with a wavelength spacing of 63 GHz were generated, where f_1^c = 192.722 THz and f_{bas} = 16 GHz.

These OCLs can be generated from the U-FOCG, and the structure of the U-FOCG is illustrated in **Figure 5A**. First, two IMs and one PM were cascaded in turn, where a continuous-wave laser was connected to the IM1 after a polarization controller. Second, an RF signal source with a frequency of $\delta_c/2$ is separated into three branches. The 1st branch was used to directly drive the IM1. The 2nd branch was employed to drive the IM2 after the first phase shifter (PS1). The 3rd branch was served to drive the PM, where an electric PS (PS2), $a \times 2$ frequency multiplier, and a variable attenuator were employed to eliminate the 1st undesirable term. Third, after processing by the PS3, a voltage control module including a tunable amplifier and attenuator was used to adjust the voltage of the PM's driven signal.

A configuration system used for verifying the feasibility of the U-FOCG was established by VPItransmissionMaker. In the verification system, an RF signal source with a CF of 31.5 GHz and an amplitude of 5 V was employed to drive the modulators, and the CF of the continuous-wave laser was equal to 193.1 THz. Moreover, the phase-shift value of the PS1 was set to $\pi/4$. In order to fully eliminate the 1st undesirable term, the variable attenuator

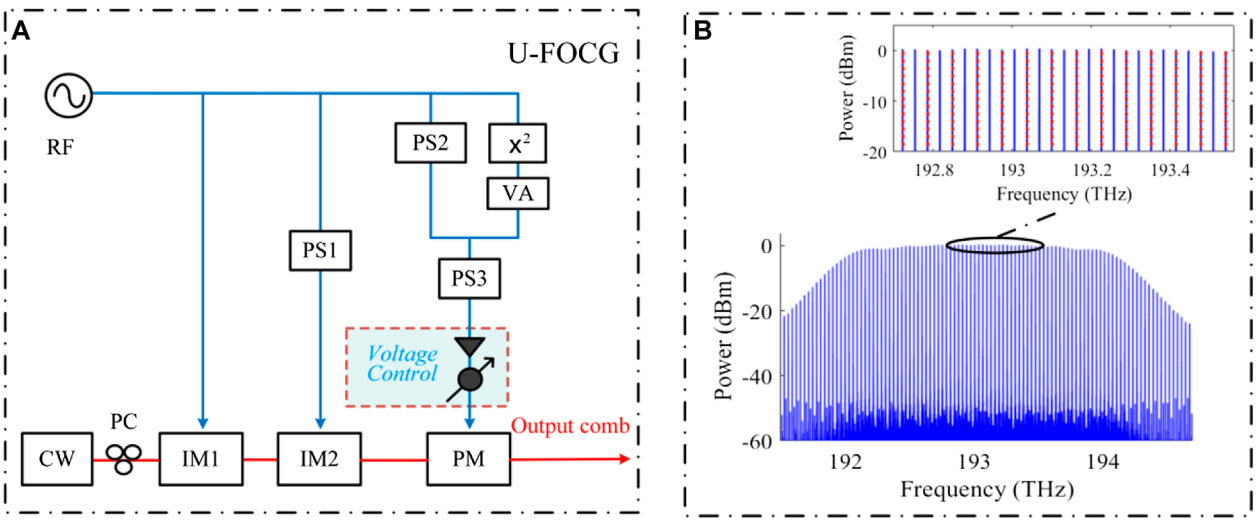

FIGURE 5 | (A) Schematic diagram of the ultra-flat optical comb generator (U-FOCG), **(B)** Spectra of the output of the U-FOCG with 4-dB gain, and the employed 14 optical comb lines with about 0 dBm optical powers. CW, continuous-wave laser; PC, polarization controller; IM, intensity modulator; PS, phase shifter; PM, phase modulator; x^2, *x*2 frequency multiplier; VA, variable attenuator.

and the PS2 were set to $(1/16)^2$ and π, respectively. After setting the PS3 to $\pi/9$, the driven signal of PM was amplified by the voltage control module and then about 4-dB gain has been available for the driven signal. Finally, the output comb is illustrated in **Figure 5B**, where the employed 14 OCLs with about 0 dBm optical powers are also illustrated in the inset in **Figure 5B**.

Secondly, the dynamic 4 GHz signal was modulated on these OCLs. After DSB modulation, 28 optical multicast copies located on the ±1st sidebands of these OCLs were generated. Thirdly, an FPF with an FSR of 32 GHz was introduced to channelize the original signal into 28 narrowband sub-channels, where the CF of the first TP f_1^{peak} must satisfy **(3)**. Finally, by monitoring the optical power in every sub-channel, the spectrum information of the dynamic original signal is then determined. For example, when the optical powers in *Sub-ch.1* and *Sub-ch.2* are significant, the original signal contains frequency bands of 2–3 GHz and 29–30 GHz. Moreover, the relationship among CFs of these OCLs, CFs of these TPs, sub-channels, as well as the monitoring frequency-band range is shown in **Table 2**.

After obtaining the spectrum information, the 4 GHz dynamic signal was received by the proposed multi-band coherent channelizer. Firstly, U-FOCG 1 in **Figure 1** was used to provide 7 OCs, where the RF signal source had a frequency of 31 GHz. The CF of every generated OC satisfies **(5)**, where f_1^{oc} = 192.914 THz and δ_s = 1 GHz. Secondly, 14 *v*-copies were generated after DSB modulation. Thirdly, 14 LOs with wavelength spacing of 32 GHz were generated from U-FOCG 2, where f_1^{LO} = 192.885 THz. It is worth mentioning that CWs used in U-FOCG 1 and U-FOCG 2 should be generated from the same optical laser source [23, 27–31]. For the dynamic 4 GHz signal with microwave bands between 2 GHz and 30 GHz, the relationship among CFs of the OCs, v-copies, CFs of the LOs, and the channelized sub-signals is clearly shown in **Table 3**. Finally, by selecting the corresponding *v-copies* and LOs according to the spectrum information, the 4 GHz dynamic signal was channelized into four 1 GHz sub-signals and then received.

TABLE 2 | Relationship among OCLs, TPs, sub-channels, and monitoring bands for dynamics 4 GHz signal.

CFs of OCLs (THz)	CFs of TPs (THz)	Sub-channels	Monitoring range (GHz)
192.722	192.6925	*Sub-ch.*1	29–30
	192.7245	*Sub-ch.*2	2–3
192.785	192.7565	*Sub-ch.*3	28–29
	192.7885	*Sub-ch.*4	3–4
192.848	192.8205	*Sub-ch.*5	27–28
	192.8525	*Sub-ch.*6	4–5
192.911	192.8845	*Sub-ch.*7	26–27
	192.9165	*Sub-ch.*8	5–6
192.974	192.9485	*Sub-ch.*9	25–26
	192.9805	*Sub-ch.*10	6–7
193.037	193.0125	*Sub-ch.*11	24–25
	193.0445	*Sub-ch.*12	7–8
193.100	193.0765	*Sub-ch.*13	23–24
	193.1085	*Sub-ch.*14	8–9
193.163	193.1405	*Sub-ch.*15	22–23
	193.1725	*Sub-ch.*16	9–10
193.226	193.2045	*Sub-ch.*17	21–22
	193.2365	*Sub-ch.*18	10–11
193.289	193.2685	*Sub-ch.*19	20–21
	193.3005	*Sub-ch.*20	11–12
193.352	193.3325	*Sub-ch.*21	19–20
	193.3645	*Sub-ch.*22	12–13
193.415	193.3965	*Sub-ch.*23	18–19
	193.4285	*Sub-ch.*24	13–14
193.478	193.4605	*Sub-ch.*25	17–18
	193.4925	*Sub-ch.*26	14–15
193.541	193.5245	*Sub-ch.*27	16–17
	193.5565	*Sub-ch.*28	15–16

OCLs, optical comb lines; TPs, transmission peaks; CFs, center frequencies.

TABLE 3 | Relationship among v-copies, LOs, channelized components for dynamics 4 GHz signal.

CFs of OCs (THz)	Multicast copies	CFs of LOs (THz)	Sub-signals (GHz)
192.914	*v-copy* 1	192.885	29–30
			28–29
	v-copy 2	192.917	2–3
			3–4
192.976	*v-copy* 3	192.949	27–28
			26–27
	v-copy 4	192.981	4–5
			5–6
193.038	*v-copy* 5	193.013	25–26
			24–25
	v-copy 6	193.045	6–7
			7–8
193.100	*v-copy* 7	193.077	23–24
			22–23
	v-copy 8	193.109	8–9
			9–10
193.162	*v-copy* 9	193.141	21–22
			20–21
	v-copy 10	193.173	10–11
			11–12
193.224	*v-copy* 11	193.205	19–20
			18–19
	v-copy 12	193.237	12–13
			13–14
193.286	*v-copy* 13	193.269	17–18
			16–17
	v-copy 14	193.301	14–15
			15–16

LOs, local oscillators; CFs, center frequencies, OCs, optical carriers.

RESULTS AND DISCUSSION

To verify the feasibility of the proposed photonics-assisted multi-band microwave receiver, a proof-of-concept system based on the statements in "Photonics-Assisted Multi-Band Receiver for Dynamic Signals in 2–30 GHz Bands" section has been implemented by OptiSystem and VPItransmissionMaker. First, we performed the instantaneous spectral analysis for 4 GHz

TABLE 4 | Considering seven cases and their corresponding frequency ranges.

Seven cases	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
Frequency range (GHz)	2–6	6–10	10–14	14–18	18–22	22–26	26–30

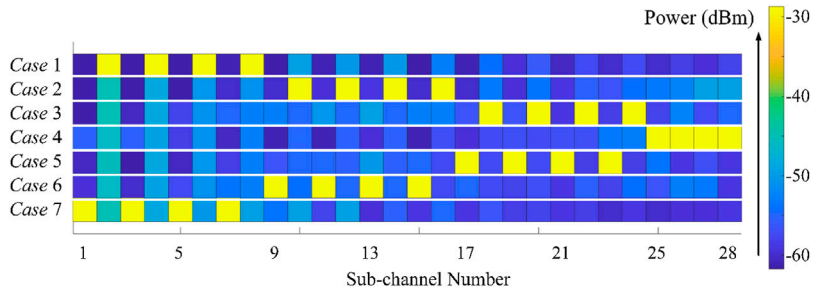

FIGURE 6 | Received optical powers in all 28 sub-channels for seven cases with different frequency ranges, where the spectrum information for each case is obtained by monitoring the optical power in all sub-channels.

dynamic signals. Second, we received the 4 GHz dynamic signals by coherent channelization.

In the instantaneous spectral analysis for the 4 GHz dynamic signals, we considered the following seven cases, as shown in **Table 4**. We employed four-tone signals to equivalently replace the 4 GHz original signals. For case 1, the four-tone signal with frequencies of 2.5, 3.5, 4.5, and 5.5 GHz was employed. After the second-category photonic channelization, the optical powers received from all sub-channels were illustrated in **Figure 6**. We can see that the optical powers in *Sub-ch.2*, *Sub-ch.4*, *Sub-ch.6*, and *Sub-ch.8* are significantly greater than others.

Namely, we can find the sub-channels with significant optical powers and then determine the spectrum information according to the received optical powers and **Table 2**. For all seven cases, the received optical powers after the effective second-category photonic channelizer were shown in **Figure 6**. In practical application, by monitoring the optical power in each sub-channel, the spectrum information is obtained. Based on the spectrum information, different OCs and LOs are selected to implement channelized reception of the original dynamic signal.

For simplicity but without loss of generality, we employed 4 GHz linearly-chirped signals with CFs of 4 and 28 GHz to represent the 4 GHz dynamic signals, as shown in **Figure 7**. After determining the spectrum information of the linearly-chirped signals, the two linearly-chirped signals were received by the coherent channelization.

According to **Table 3**, *v-copy* 2 and *v-copy* 4 were required for receiving the "low-frequency" linearly-chirped signal, where the CFs of the TOFs were set to 192.918 THz and 192.980 THz, and the bandwidths of the TOFs were 5 GHz. The spectra of the EQHs were plotted in **Figure 8A**. Because the TOFs are non-ideal in practice, the OC with the CF of 192.914 THz cannot be effectively suppressed. After heterodyne detection of the OC and the LO with the CF of 192.917 THz, a 3 GHz tone was generated which was shown in Figure 8A**, i**. After filtering by four 1-GHz LPFs, the 4-GHz dynamic signal was channelized into four 1 GHz sub-signals (δ_s = 1 GHz), where we assume that the LPFs had rectangular TPs. Moreover, the waveforms of the channelized sub-signal were depicted in **Figure 8B**, respectively.

Similarly, the "high-frequency" linearly-chirped signal can also be received, where *v-copy* 1 and *v-copy* 3 were required. Moreover, the CFs of the TOFs should be set to 192.886 and 192.948 THz. The obtained spectra after processing by the EQHs and the waveforms of the channelized sub-signals were given in **Figure 9**.

Based on the results shown in **Figures 8, 9**, the proposed channelized receiver can provide a much larger operative bandwidth by using the same number of devices when compared with the existing works in [23, 26–32]. Using two OHs, four BPDs, two EQHs, and four electrical LPFs, the equivalent operative bandwidth of the proposed multi-band channelized receiver can be up to 28 GHz. To the best of our knowledge, it is the first time that a photonics-assisted channelized receiver is built which doesn't require knowing the spectrum information in advance.

CONCLUSIONS

In this work, based on the spectrum analysis and the coherent channelization, a photonics-assisted channelized receiver has been proposed for multi-band microwave signals. After determining the spectrum information by the spectrum analysis, the original wideband signal with dynamic microwave bands was received by the coherent channelization. For seven cases considered, the spectrum information can be obtained by monitoring the optical power in each sub-channel. Moreover, two 4 GHz linearly-chirped signals with CFs of 4 and 28 GHz have been channelized into four 1 GHz sub-channels and received, respectively. We equivalently can build a multi-band coherent

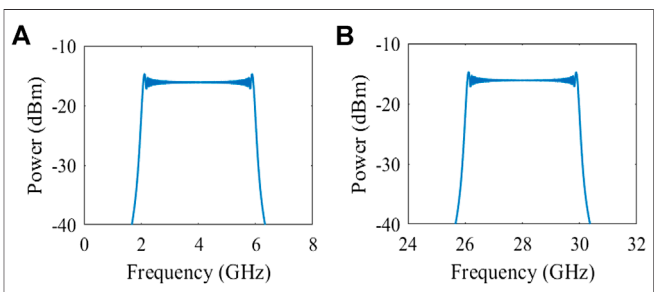

FIGURE 7 | Spectra of original linearly-chirped signals with center frequencies of **(A)** 4 GHz and **(B)** 28 GHz.

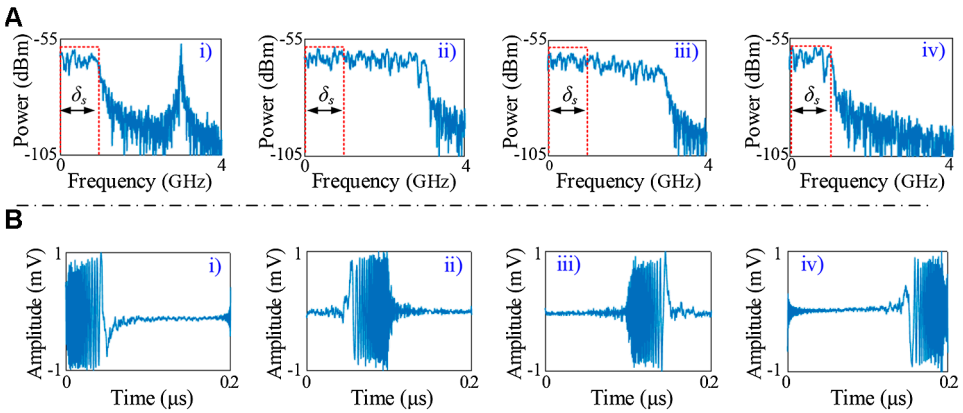

FIGURE 8 | (A) Spectra of the output signals processed by optical hybrids and electrical quadrature hybrids and **(B)** waveforms of channelized signals after low-pass filters, where the 4 GHz linearly-chirped signal has a 4 GHz center frequency.

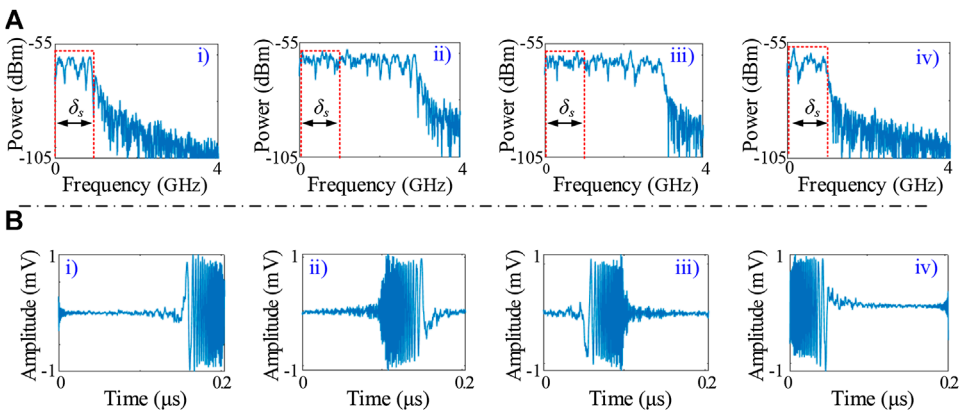

FIGURE 9 | (A) Spectra of the output signals processed by optical hybrids and electrical quadrature hybrids and **(B)** waveforms of channelized signals after low-pass filters, where the 4 GHz linearly-chirped signal has a 28 GHz center frequency.

channelizer that works for dynamic microwave signals by using a few devices. For scenarios requiring multi-band microwave signals such as modern radars, communication satellites, radio over fiber systems, and the 5th generation mobile network, the proposed photonics-assisted channelized receiver provides the low-complexity and low-cost implementation and shows a tremendous advantage and potential.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article.

AUTHOR CONTRIBUTIONS

HH and CF conceived the method and conducted the model. HH, WZ, YC, and ZY contributed to the organization and writing of the manuscript. HH, CF, WZ, YC, HF, ZY, FC, and KQ did the revision and editing. All authors reviewed the manuscript.

FUNDING

Advance Research Project of Common Technology (No. 41418050102), Sichuan Science and Technology Foundation Program (No. 2018JY0145), Project for Innovation Team of Guangdong University (No. 2018KCXTD033), National Key R&D Program of China (no. 2018YFB1801302), Project for Zhongshan Key Social Public Welfare Science and Technology (No. 2019B2007), Zhongshan Government Foundation project for National Laboratory Zhongshan Branch Lab (No. 412S06), Zhongshan Innovative Research Team Program (No. 180809162197886), Zhongshan Institute high-level talent scientific research startup fund project (No. 416YKQ04).

REFERENCES

1. Zhang C, Huang H, Qiu K. Microwave photonics and its optical wireless access systems. In: Proceedings of fiber optic sensing and optical communication; 2018 December; Beijing, China; SPIE 10849. doi:10.1117/12.2505311
2. Li C, Peng Z, Huang T-Y, Fan T, Wang F-K, Horng T-S, et al. A review on recent progress of portable short-range noncontact microwave radar systems. *IEEE Trans Microw Theor Tech* (2017) **65**:1692–706. doi:10.1109/TMTT.2017.2650911
3. Pan S, Zhu D, Liu S, Xu K, Dai Y, Wang T, et al. Satellite payloads pay off. *IEEE Microw Mag* (2015) **16**:61–73. doi:10.1109/MMM.2015.2441619
4. José C, José M, Ivana G, Juan S, Juan L, Salvador S. Microwave photonic signal processing. *J Lightwave Technol* (2013) **31**:571–86. doi:10.1002/9780470744857.ch8
5. Akyildiz IF, Nie S, Lin S-C, Chandrasekaran M. 5G roadmap: 10 key enabling technologies. *Comput Network* (2016) **106**:17–48. doi:10.1016/j.comnet.2016.06.010
6. Zou X, Lu B, Pan W, Yan L, Stöhr A, Yao J. Photonics for microwave measurements. *Laser Photon Rev* (2016) **10**:711–34. doi:10.1002/lpor.201600019
7. Brès C-S, Zlatanovic S, Wiberg AOJ, Radic S. Reconfigurable parametric channelized receiver for instantaneous spectral analysis. *Opt Express* (2011) **19**:3531–41. doi:10.1364/OE.19.003531
8. Xie X, Dai Y, Ji Y, Xu K, Li Y, Wu J, et al. Broadband photonic radio-frequency channelization based on a 39-GHz optical frequency comb. *IEEE Photon Technol Lett* (2012) **24**:661–3. doi:10.1109/lpt.2012.2185787
9. Huang H, Zhang C, Zhou H, Yang H, Yuan W, Qiu K. Double-efficiency photonic channelization enabling optical carrier power suppression. *Opt Lett* (2018) **43**:4073–6. doi:10.1364/OL.43.004073
10. Winnall ST, Lindsay AC, Austin MW, Canning J, Mitchell A. A microwave channelizer and spectroscope based on an integrated optical Bragg-grating Fabry-Perot and integrated hybrid Fresnel lens system. *IEEE Trans Microw Theor Tech* (2006) **54**:868–72. doi:10.1109/TMTT.2005.863052
11. Huang H, Wang R, Zhang C, Chen Y, Yang H, Qiu K. Tunable ultra-flat optical-comb-enabled, reconfigurable, and efficient coherent channelized receiver. *Opt Lett* (2020) **45**:848–51. doi:10.1364/ol.385458
12. Shin H, Harjani R. Low-power wideband analog channelization filter bank using passive polyphase-FFT techniques. *IEEE J Solid State Circ* (2017) **52**: 1753–67. doi:10.1109/JSSC.2017.2700792
13. Nguyen H-N, Kim K-S, Han S-H, Lee J-Y, Kim C, Lee S-G. A low-power interference-tolerance wideband receiver for 802.11af/ah long-range Wi-Fi with post-LNA active N-path filter. *IEEE T Microw Theory* (2018) **66**:2287–98. doi:10.1109/TMTT.2018.2805341
14. Hao W, Dai Y, Yin F, Zhou Y, Li J, Dai J, et al. Chirped-pulse-based broadband RF channelization implemented by a mode-locked laser and dispersion. *Opt Lett* (2017) **42**:5234–7. doi:10.1364/ol.42.005234
15. Khilo A, Spector SJ, Grein ME, Nejadmalayeri AH, Holzwarth CW, Sander MY, et al. Photonic ADC: overcoming the bottleneck of electronic jitter. *Optic Express* (2012) **20**:4454–69. doi:10.1364/OE.20.004454
16. Strutz SJ, Williams KJ. An 8-18-GHz all-optical microwave downconverter with channelization. *IEEE T Microw Theory* (2001) **49**:1992–5. doi:10.1109/22.954819
17. Wang W, Liu J, Sun W, Wang X, Zhu N. Channelization microwave frequency upconversion based on dual-parallel optical frequency comb generators. In: Proceedings of optoelectronic devices and integration; 2015 June; Wuhan, China. **Vol. 2.** (2015). doi:10.1364/OEDI.2015.OW2C.2
18. Zou X, Pan W, Luo B, Yan L. Photonic approach for multiple-frequency component measurement using spectrally sliced incoherent source. *Opt Lett* (2010) **35**:438–41. doi:10.1364/OL.35.000438
19. Xu X, Wu J, Nguyen TG, Chu ST, Little BE, Morandotti R, et al. Broadband RF channelizer based on an integrated optical frequency Kerr comb source. *J Lightwave Technol* (2018) **36**:4519–26. doi:10.1109/JLT.2018.2819172
20. Zhou H, Geng Y, Cui W, Huang S, Zhou Q, Qiu K, et al. Soliton bursts and deterministic dissipative Kerr soliton generation in auxiliary-assisted microcavities. *Light Sci Appl* (2019) **8**:50. doi:10.1038/s41377-019-0161-y
21. Wang W, Davis RL, Jung TJ, Lodenkamper R, Lembo LJ, Brock JC, et al. Characterization of a coherent optical RF channelizer based on a diffraction grating. *IEEE T Microw Theory* (2001) **49**:1996–2001. doi:10.1109/22.954820
22. Rhodes WT Acousto-optic signal processing: convolution and correlation. *Proc IEEE* (1981) **69**: 65–79. doi:10.1109/PROC.1981.11921
23. Tang Z, Zhu D, Pan S. Coherent optical RF channelizer with large instantaneous bandwidth and large in-band interference suppression. *J Lightwave Technol* (2018) **36**:4219–26. doi:10.1109/JLT.2018.2857500
24. Brès CS, Wiberg AOJ, Zlatanovic S, Radic S. Parametric channelized receiver for single-step spectral analysis. In: Optical fiber communication conference and exposition (OFC) and the national fiber optic engineers conference; 2011 March 6–10. Los Angeles, CA: NFOEC (2011). doi:10.1364/OFC.2011.OThC1
25. Brès CS, Wiberg AOJ, Zlatanovic S, Radic S. Characterization of parametric RF channelized receiver through time domain monitoring. In: Conference on lasers and electro-optics (CLEO); May 6–11 2012; San Jose, CA (2012).
26. Wiberg AOJ, Esman DJ, Liu L, Adleman JR, Zlatanovic S, Ataie V, et al. Coherent filterless wideband microwave/millimeter-wave channelizer based on broadband parametric mixers. *J Lightwave Technol* (2014) **32**:3609–17. doi:10.1109/JLT.2014.2320445
27. Yang J, Li R, Dai Y, Dong J, Li W. Wide-band RF receiver based on dual-OFC-based photonic channelization and spectrum stitching technique. *Optic Express* (2019) **27**:33194–204. doi:10.1364/OE.27.033194
28. Jiang W, Zhao S, Tan Q, Liang D, Li X, Gao Y. Wideband photonic microwave channelization and image-reject down-conversion. *Opt Commun* (2019) **445**: 41–9. doi:10.1016/j.optcom.2019.04.013
29. Xie X, Dai Y, Xu K, Niu J, Wang R, Yan L, et al. Broadband photonic RF channelization based on coherent optical frequency combs and I/Q demodulators. *IEEE Photonics J* (2012) **4**:1196–202. doi:10.1109/JPHOT.2012.2207380
30. Chen W, Zhu D, Xie C, Liu J, Pan S. Microwave channelizer based on a photonic dual-output image-reject mixer. *Opt Lett* (2019) **14**:4052–5. doi:10.1364/OL.44.004052
31. Yang J, Li R, Mo Z, Dong J, Li W. Channelized photonic-assisted deramp receiver with an extended detection distance along the range direction for LFM-CW radars. *Opt Express* (2020) **28**:7576–84. doi:10.1364/OE.386819
32. Xie C, Zhu D, Chen W, Pan S. Microwave photonic channelizer based on polarization multiplexing and photonic dual output image reject mixer. *IEEE Access* (2019) **7**:158308–16. doi:10.1109/ACCESS.2019.2947673
33. Li R, Chen H, Yu Y, Chen M, Yang S, Xie S. Multiple-frequency measurement based on serial photonic channelization using optical wavelength scanning. *Opt Lett* (2013) **38**:4781–4. doi:10.1364/OL.38.004781
34. Oleski PJ, Patton RW, Bharj SS, Thaduri M. Transmit receive module for space ground link subsystem (SGLS) and unified S-band (USB) satellite telemetry, tracking and commanding (TT and C), and communications. In: IEEE military communications conference; January 2004. Piscataway, NJ: IEEE (2004).
35. Shambayati S. The struggle for Ka-band: NASA's gradual move towards using 32-GHz Ka-band for deep space missions.In: IEEE aerospace conference; Apirl 2007. Big Sky, MT: Big Sky (2007).
36. Simone L, Salerno N, Maffei M. Frequency-hopping techniques for secure satellite TT&C: system Analysis & trade-offs. In: International workshop on satellite and space communications; October 2006; Madrid, Spain (2006).
37. Nguyen AT, Zhou H, Hadjitheodosiou M, Baras JS. A direct-to-ground architecture for supporting commercial communications from the International Space Station. In: IEEE international conference on communications; 2002 February. Piscataway, NJ: IEEE (2002).
38. Nguyen TK, Oh NJ, Cha CY, Oh YH, Ihm GJ, Lee SG. Image-rejection CMOS low-noise amplifier design optimization techniques. *IEEE T Microw Theory* (2005) **53**:538–47. doi:10.1109/TMTT.2004.840744
39. Lee J, Li YA, Hung MH, Huang SJ. A fully-integrated 77-GHz FMCW radar transceiver in 65-nm CMOS Technology. *IEEE J Solid State Circ* (2010) **45**: 2746–56. doi:10.1109/JSSC.2010.2075250
40. Seimetz M, Weinert C. Options, feasibility, and availability of 2x4 900 hybrids for coherent optical systems. *J Lightwave Technol* (2006) **24**:1317–22. doi:10.1109/JLT.2005.863251
41. Pozar DM. *Microwave engineering*. 4th ed. Chap. 7. New York, NY: John Wiley & Sons (2005).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 12 November 2020
doi: 10.3389/fphy.2020.585707

Optoelectronic Performance Analysis of Low-Energy Proton Irradiation and Post-Thermal Annealing Effects on InGaAs Solar Cell

Y. Zhuang[1], A. Aierken[1], Q. Q. Lei[2,3]*, L. Fang[4], X. B. Shen[2,3], M. Heini[2,3], Q. Guo[2,3], J. Guo[1], X. Yang[1], J. H. Mo[1], R. K. Fan[1], J. Li[1], Q. Y. Chen[1] and S. Y. Zhang[1]*

[1]School of Energy and Environment Science, Yunnan Normal University, Kunming, China, [2]Xinjiang Key Laboratory of Electronic Information Materials and Devices, Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi, China, [3]Key Laboratory of Functional Materials and Devices for Special Environments of Chinese Academy of Sciences, Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi, China, [4]Uniwatt Technology Co. Ltd., Zhongshan, China

OPEN ACCESS

Edited by:
Zichuan Yi,
University of Electronic Science and Technology of China, China

Reviewed by:
Yang Tong,
Oak Ridge National Laboratory (DOE), United States
Wang Xin,
South China Normal University, China

***Correspondence:**
A. Aierken
erkin@ynnu.edu.cn
Q. Q. Lei
1243642152@qq.com

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *21 July 2020*
Accepted: *15 September 2020*
Published: *12 November 2020*

Citation:
Zhuang Y, Aierken A, Lei QQ, Fang L, Shen XB, Heini M, Guo Q, Guo J, Yang X, Mo J H, Fan RK, Li J, Chen QY and Zhang SY (2020) Optoelectronic Performance Analysis of Low-Energy Proton Irradiation and Post-Thermal Annealing Effects on InGaAs Solar Cell. Front. Phys. 8:585707. doi: 10.3389/fphy.2020.585707

The electrical and spectral properties of 150 KeV proton-irradiated MBE-grown $In_{0.53}Ga_{0.47}As$ single junction solar cell and its post-thermal annealing properties were investigated. Both simulation and experimental methods were applied to analyze the irradiation-induced displacement damage and degradation mechanism of cell performance. The results show that most protons would penetrate through the $In_{0.53}Ga_{0.47}As$ emitter and stop in the base region, causing differing extents of electric and spectral degredation. When proton fluence were 1×10^{12} and 5×10^{12} p/cm^2, the remaining factor of I_{sc}, V_{oc}, P_{max}, and FF were degraded to 0.790, 0.767, 0.558, 0.921 and 0.697, 0.500, 0.285, 0.817, respectively. Severer degradation was found in short wave lengths compared to long wave lengths of the solar cell spectral response. After annealing treatments, the normalized I_{sc}, V_{oc}, P_{max}, and FF, significantly recovered from 0.697, 0.500, 0.285, and 0.817 to 0.782, 0.700, 0.499, and 0.912 (fluence: 5×10^{12} p/cm^2), and the irradiation-induced defects in the whole emission area and part of the base area were annihilated, so the observed recovery of the short wavelength of the solar cell was greater than the long wavelength. The performance analysis in this work provided valuable ways to improve the photoelectric efficiency of space solar cells.

Keywords: InGaAs solar cell, proton irradiation, displacement damage, degradation, annealing

INTRODUCTION

III-V-based multi-junction solar cells as a direct energy provider are currently used for different space applications; radiation resistance of the solar cell will directly determine the service life of spacecraft and satellites. In order to accomplish extremely tough and long-term space missions, it is important to develop space solar cells with excellent radiation resistance and high conversion efficiency and stability. Over the past ten years, lattice-matched GaInP/GaAs/Ge three-junction solar cells with relatively high efficiency ($\eta \approx 30\%$, in AM0, 1sun) compared to Si cells have been used as the main power source for space applications. However, the fact that sub-cells of this type of three-junction solar cells are not optimized for solar spectrum hindered the further improvement of photoelectric conversion efficiency [1–3]. So, current researchers are attempting to further improve the efficiency

by selecting dilute nitride materials as the other sub-cell incorporated into the three-junction cell to obtain the GaInP(1.9 eV)/GaAs(1.4 eV)/Ga(In)NAs(Sb) (1.04 eV)/Ge (0.67 eV) four-junction solar cell [4, 5] in different forms, such as wafer bonding solar cells [6], upright metamorphic solar cells [7, 8], and inverted metamorphic solar cells (IMM) [9]. At present, the GaInP/GaAs//InGaAsP/InGaAs four-junction solar cell fabricated by using the wafer bonding technique not only fixed the lattice mismatch issue of multi-junction solar cells, but also obtained high photoelectric conversion efficiency compared to other new-type cells, and so is expected to become the main direction for developing the next generation of space solar cells [10, 11].

There are already many studies about the electrical properties of both InGaAs-based single and multi-junction solar cells. However, the radiation effects for InGaAs sub-cell of GaInP/GaAs//InGaAsP/InGaAs full spectra four-junction solar cells are still limited. Dai et al. [12] studied the 1 MeV electron irradiation effects of wafer bounded GaInP/GaAs//InGaAsP/InGaAs four-junction solar cells and revealed that the degradation of each InGaAs sub-cell and InGaAsP sub-cell are mainly responsible for the deterioration of overall performance. Walters et al. [13] studied the 1 MeV electron-irradiated $In_{0.53}Ga_{0.47}As$ solar cell and found that electron irradiation would induce two trapping centers for the majority carrier, causing degradation of the cell performance. Karlina et al. [14] investigated the 3/10 MeV proton and 1 MeV electron radiation effects of InP/InGaAs solar cells and indicated that InGaAs is an optimum candidate for InP/InGaAs triple-junction solar cells to improve the efficiency. In our previous work, the radiation damage of 3 MeV electron and 1 MeV proton-irradiated $In_{0.53}Ga_{0.47}As$ single cells was well studied and we concluded that the major degradation of spectral response (EQE) usually occurs in the long wave region [15].

Many scholars investigated the effects of proton irradiation with varying energy levels on III-V multi-junction solar cells, and concluded that the degradation caused by low-energy protons is more serious than high-energy protons [2, 16–18]. At present, there is no report about InGaAs sub-cells subjected to low-energy proton irradiation, especially InGaAs sub-cell designed for GaInP/GaAs//InGaAsP/InGaAs four-junction wafer bonded solar cell. So, the low-energy proton irradiated degradation mechanism of this type of solar cell is still unclear. Moreover, space solar cells work at a temperature range of −160~200°C in mission [19, 20]. When the temperature is high enough, solar cells may be annealed. Therefore, the annealing process of solar cells is an important factor that cannot be ignored. In addition, a special feature of the GaAs solar cells found in early research is their ability to recover from radiation damage by annealing at relatively low temperatures under 200°C [21]. So, the annealing effect on InGaAs sub-cells after low-energy proton irradiation is studied in this work, and the annealing temperature is selected as 150°C. Research in this article is part of our ongoing project. The degradation properties of the main solar cell parameters for 150 KeV low-energy proton-irradiated $In_{0.53}Ga_{0.47}As$ single junction solar cells have been clearly discussed and the physical mechanism was well investigated. Moreover, the performance recovery of irradiated $In_{0.53}Ga_{0.47}As$ under different annealing conditions is also discussed. Our experiments could provide evidence for the proper design of annealing cycles for space missions to enhance the performance of solar cells in space.

EXPERIMENTAL DETAILS

$In_{0.53}Ga_{0.47}As$ single junction solar cell was investigated in this work and its detailed structure is shown in **Figure 1A**. The cell was grown on a 350-μm-thick *p*-InP substrate by molecular beam epitaxy (MBE) system, and has a 3-μm-thick base layer with a doping concentration of 1×10^{17} cm^{-3} and a 200-nm-thick emitter layer with a 1×10^{18} cm^{-3} doping concentration. Si and Be were used as the n-type and p-type doping source, respectively. A detailed description for the growth and fabrication process of this solar cell has been reported in our previous work [15]. All tested samples

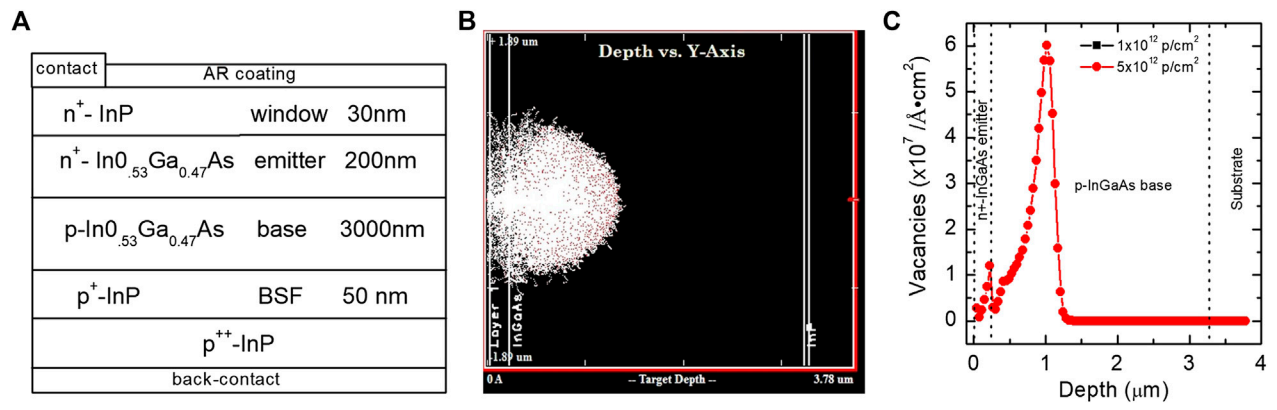

FIGURE 1 | Structure representative of **(A)** InGaAs sub-cells studied in this work; SRIM simulation results of **(A)** 150 KeV proton (number of 10^5) trajectories and **(B)** calculated vacancy densities in $In_{0.53}Ga_{0.47}As$ sub-cells.

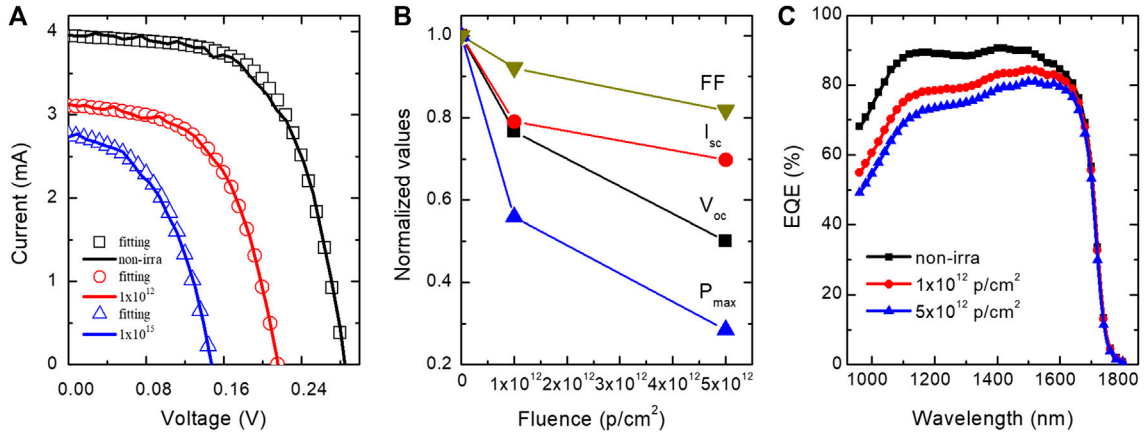

FIGURE 2 | 150 KeV proton with different fluence-irradiated $In_{0.53}Ga_{0.47}As$ solar cells' **(A)** I-V characteristics, solid lines are experimental curves and the theoretical values are denoted by hollow symbols, **(B)** the remaining factor of electrical parameters, and **(C)** EQE spectra.

are 2.5 × 2.5 mm^2 in size and have an initial conversion efficiency of around 11% under the standard AM0 spectra (136.7 mW/cm^2 at 25°C). The 150 KeV proton irradiation experiments were conducted at room temperature with uniform flux, and the total dose was 1.0 × 10^{12} and 5.0 × 10^{12} p/cm^2 for comparison. After proton irradiation, thermal annealing treatments at 150°C for 20, 60, 120, 180, and 360 min were carried out.

RESULTS AND DISCUSSION

Simulation Results

In order to understand proton irradiation-induced displacement damage distribution in $In_{0.53}Ga_{0.47}As$ sub-cells, the SRIM (Stopping and Range of ions in Matter) program was utilized to simulate the trajectories of 10^5 incident protons. **Figure 1B,C** show the trajectories of 150 KeV protons and the corresponding irradiation-induced vacancy densities in $In_{0.53}Ga_{0.47}As$ sub-cells. Results gathered from the simulation indicate that the maximum penetration depth would be at 1.36 μm below the implantation surface, with peak damage occurring at around 1.0 μm, which illustrates that most protons could penetrate through InGaAs emitter region but stop in the base range. This also proves that the proton energy of 150 keV meets the requirement for our experiments.

TABLE 1 | Extracted parameters of $In_{0.53}Ga_{0.47}As$ solar cells from I-V curves.

Fluence (p/cm^2)	I_{sc} (mA)	V_{oc} (V)	FF	P_{max} (mW)	I_{ph} (mA)	I_0 (A) (E)	R_{sh} (Ω)	R_s (Ω)	n
0	3.92	0.280	0.604	0.663	3.96	2–7	1,500	5.5	1.114
1 × 10^{12}	3.13	0.215	0.553	0.371	3.13	1.7–6	1,100	5.6	1.115
5 × 10^{12}	2.74	0.140	0.493	0.189	2.79	1.8–5	900	5.7	1.114

Degradation of Electric and Spectral Properties

Figure 2A shows the I-V characteristics of the $In_{0.53}Ga_{0.47}As$ solar cell before and after proton irradiation with different proton fluence. The basic electrical parameters of solar cells, such as the short-circuit current (I_{sc}), open-circuit voltage (V_{oc}), fill factor (FF), and maximum output power (P_{max}), could be directly derived from the I-V curve. Other parameters, like photo-generated current (I_{ph}), saturation current (I_0), shunt resistance (R_{sh}), series resistance (R_s), and diode ideal factor (n), could be extracted from the I-V curves using a theoretical model [22]. The symbols shown in **Figure 2A** are fitted results and their values are listed in **Table 1**, which were obtained based on the equivalent circuit method described as the following expression:

$$I = I_{ph} - I_0\left[\exp\left(\frac{v + IR_S}{nV_T}\right)\right] - \frac{v + IR_S}{R_{SH}} \quad (1)$$

where V_T is the thermal voltage. It should be noticed from **Figure 2A** that extracted results by theoretical model are highly consistent with the experimental results. The results show that I_0 and R_s increased with the increasing proton fluence and R_{sh} and I_{ph} decreased conversely.

Changes in the remaining factor of I_{sc}, V_{oc}, P_{max}, and FF of $In_{0.53}Ga_{0.47}As$ solar cells before and after differing proton fluence irradiation levels are shown in the curves of **Figure 2B**. When proton fluence was 1 × 10^{12} and 5 × 10^{12} p/cm^2, the remaining factor of I_{sc}, V_{oc}, P_{max}, and FF were 0.790, 0.767, 0.558, 0.921, and 0.697, 0.500, 0.285, 0.817, respectively. All of them degraded seriously when the proton fluence increased, and the P_{max} was the most degraded.

The displacement damage caused by proton irradiation is mainly responsible for the degradation of $In_{0.53}Ga_{0.47}As$ solar cells' performance and electrical parameters. When solar cells are under proton irradiation, the incident proton interacted with atoms on the pristine lattice position and transferred

energy to the atoms by elastic or inelastic collisions, which removed atoms from their initial position, causing large amounts of defects [23, 24]. These irradiation-induced defects introduced five types of new charge-trapping defect levels into the forbidden band, as shown in **Figure 3**. The first mechanism is carrier generation (labeled I), which generates electron-hole pairs assisted by thermal excitation of deep-level defects. Recombination (labeled II) occurs when it is suitable for electron-hole pairs to recombine at the defect sites, and this type of defect usually reduced the minority carrier life. Trapping (labeled III) happens when a carrier is captured by the defect level and emitted to the origin site before recombination or other processes occur. Compensation (labeled IV) occurs when the defect level obtains the free carriers from dopant atoms. The type of defect level IV could reduce the amount of the majority charge carrier. The last influence of defect level is tunneling (labeled V), where carriers can get through the depletion region through trap-assisted tunneling.

The degradation of I_{sc} is mainly due to the displacement damage defects induced by proton irradiation in the active region of solar cells. There are two main mechanisms that cause the degradation of solar cell performance. First, the displacement damage generates non-irradiative recombination centers (defect level II in **Figure 3**) in the band gap to reduce the minority carrier diffusion length, and consequently decreases the collection efficiency of photo-generated carriers. The change of minority carrier diffusion length with respect to the proton irradiation flux can be expressed by the following equation [25]:

$$\frac{1}{L_{\phi}^2} = \frac{1}{L_0^2} + K_L \phi \tag{2}$$

where L_0 and L_φ are the diffusion lengths of minority carriers before and after irradiation, and K_L and φ are the damage coefficient and the proton irradiation fluence. **Equation 2** reveals that L_φ decreases with increasing proton irradiation fluence. Second, the carrier removal effect (defect level III and IV in **Figure 3**) caused by displacement damage is also an important factor resulting in the deduction of I_{sc}. The carrier removal effect would reduce the majority carrier concentration in emission and base regions and weaken the built-in electric field, so as to affect the collection of photo-generated carriers. The correlation between the built-in electric field and majority carrier concentration can been expressed as below:

$$V_D \approx \frac{K_B T}{q} \ln\left(\frac{n_{n0} p_{p0}}{n_i^2}\right) \tag{3}$$

where n_{n0} and p_{p0} are the majority carrier concentrations of emitter and base, T is the temperature, K_B is the Boltzmann constant, and n_i is the intrinsic carrier concentration. It can be seen from **Eq. 3** that the concentration decreased after irradiation and reduced the built-in electric field, consequently lowering the collecting efficiency of photo-generated carriers. The carrier removal effect, which reduced the majority carrier concentration, increased the series resistance R_s of solar cells [26]. Therefore, the R_s increased with increasing proton fluence. The fact that V_{oc} decreased after proton irradiation can be explained by the following equations [27]:

$$V_{OC} = \frac{K_B T}{q} \ln\left(\frac{I_{ph}}{I_0} + 1\right) \tag{4}$$

$$I_0 \propto J_S = \frac{q D_n n_{p0}}{L_n} + \frac{q D_p p_{n0}}{L_p} \tag{5}$$

where I_0 is the reverse saturation current and is proportional to the reverse saturation current density J_S, I_{ph} is the photo current, L_n is the minority carrier diffusion length in p-type base region, and L_p is the minority carrier diffusion length in the n-type emitter region. The variation of the V_{oc} is mainly influenced by I_0 and I_{ph}; both the decrease of I_{sc} and the increase of I_0 would lead to the degradation of V_{oc}. From the above analysis, it can be seen that the displacement damage defect caused by proton irradiation in the active area of $In_{0.53}Ga_{0.47}As$ solar cells combined with the carrier removal effect in emission and base regions leads to the total degradation of I_{sc}. I_0 decreases and R_{sh} increases with the increase of proton irradiation doses. Proton irradiation produces a large number of defects in the junction region of the solar cell, which induce additional deep level defects and increase the recombination rate of electrons and holes at both sides of the junction region, and, as a result, increase the leakage current of the solar cell and decrease the I_0 and R_{sh} at the same time. Therefore, the change of I_{sc} and I_0 is mainly responsible for the degradation of V_{oc}. In addition, **Figure 2A** and **Table 1** show that the FF decreased after proton irradiation due to the decrease of I_{sc} and V_{oc}. Finally, the degradation of V_{oc}, I_{sc}, and FF results in the reduction of P_{max}, which can be described as $P_{max} = FF \times I_{sc} \times V_{oc}$.

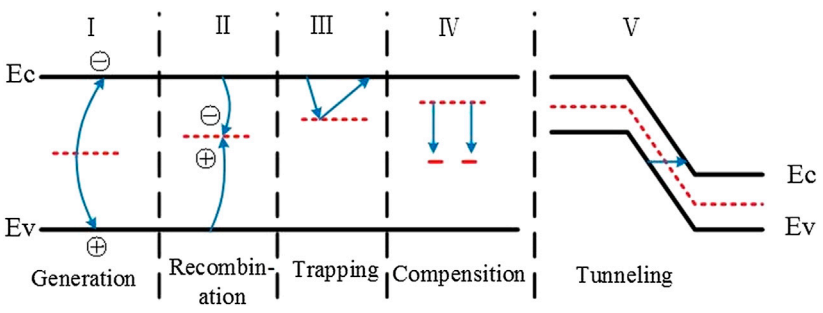

FIGURE 3 | Schematic diagram of irradiation-induced defect levels in band structure.

The EQE spectra of 150 KeV proton-irradiated $In_{0.53}Ga_{0.47}As$ solar cells is shown **Figure 2C**. It can be seen that EQE degraded with the increase of proton irradiation fluence, and the degradation mainly occurred in the short-middle wavelength range. The degradation of the EQE spectral response of solar cells reveals the displacement damage located in different regions, i.e. the displacement damage of the solar cell's emission and base region will lead to the degradation of the short-wavelength and long-wavelength of the EQE spectrum, respectively [17]. According to SRIM simulation results presented in *Simulation Results*, protons are mainly deposited in the whole emission, junction region, and part of the base region. Therefore, 150 KeV proton irradiation has caused displacement damage concentrated upon the whole emission region and junction region, and, consequently, more obvious degradation of EQE spectra occurs in the short-middle wavelength range.

Thermal Annealing of Proton Irradiated $In_{0.53}Ga_{0.47}As$

Thermal annealing experiments were then carried out on $In_{0.53}Ga_{0.47}As$ single junction solar cells irradiated by 150 KeV proton with doses of 5×10^{12} p/cm^2; the annealing temperature was set to 150°C. For the purpose of evaluating the influence of annealing time, the duration of annealing was set to 20, 60, 120, 180, and 360 min. **Figure 4** shows the variation of electrical parameters for irradiated $In_{0.53}Ga_{0.47}As$ sub-cells after different annealing times. The I-V characteristic curves in **Figure 4A** reveal the significant recovery that occurred after annealing for 20 min, but there is no substantial recovery in continuous annealing up to 360 min. It should be noticed from **Figure 4B** that the electrical parameters, such as normalized V_{oc}, I_{sc}, P_{max}, and FF, significantly recovered after annealing for 360 min, changing from 0.5, 0.697, 0.285, and 0.817 to 0.7, 0.782, 0.499, and 0.912, respectively. **Figure 4C** shows the degree of cell recovery as a function of annealing time in terms of the EQE. Obvious recovery occurred at 20 min and was maintained at around the same level during the longer annealing processing. After annealing at 150°C for 360 min, the EQE of $In_{0.53}Ga_{0.47}As$ solar cells recovered by 7.96% and 1.15% at 1050 and 1500 nm wavelengths, respectively. Therefore, the recovery degree of the spectral response of $In_{0.53}Ga_{0.47}As$ solar cells in the short wave region is greater than that in the long wave region after annealing.

After annealing at 150°C, the performance of $In_{0.53}Ga_{0.47}As$ solar cells irradiated by protons was obviously restored. There are two possible reasons for this phenomena. First, during the annealing process, a large number of carriers are generated inside the $In_{0.53}Ga_{0.47}As$ solar cell, which induce a "carrier injection annealing effect", and some proton irradiation-induced defects were annihilated. This consequently increased the minority carrier life and the performance of the $In_{0.53}Ga_{0.47}As$ solar cell could be restored [28–30]. Second, 150 KeV proton irradiation produced a large number of displacement damage defects in the $In_{0.53}Ga_{0.47}As$ solar cell; some studies showed that thermal annealing can remove some of these displacement damage defects [31] and reduce the number of defects and recombination centers to improve the $In_{0.53}Ga_{0.47}As$ solar cell's electrical and optical properties. In our work, when the $In_{0.53}Ga_{0.47}As$ solar cell was annealed at 150°C, a part of the sub-stable displacement damage caused by proton irradiation was annihilated, which resulted in a degree of recovery in cell performance. And the short wavelength recovery of EQE spectrum was more obvious than in the long wavelength. By integrating the SRIM simulation results, which showed that the damage caused by proton irradiation in the solar cell emission area was greater than in the base area, it can be seen that the degradation of the short wavelength of the solar cell spectral response was more serious. After annealing, the defects induced by proton irradiation in the whole emission area and part of the base area were annihilated, so the recovered degree of the short wavelength of the solar cell was greater than in the long wavelength.

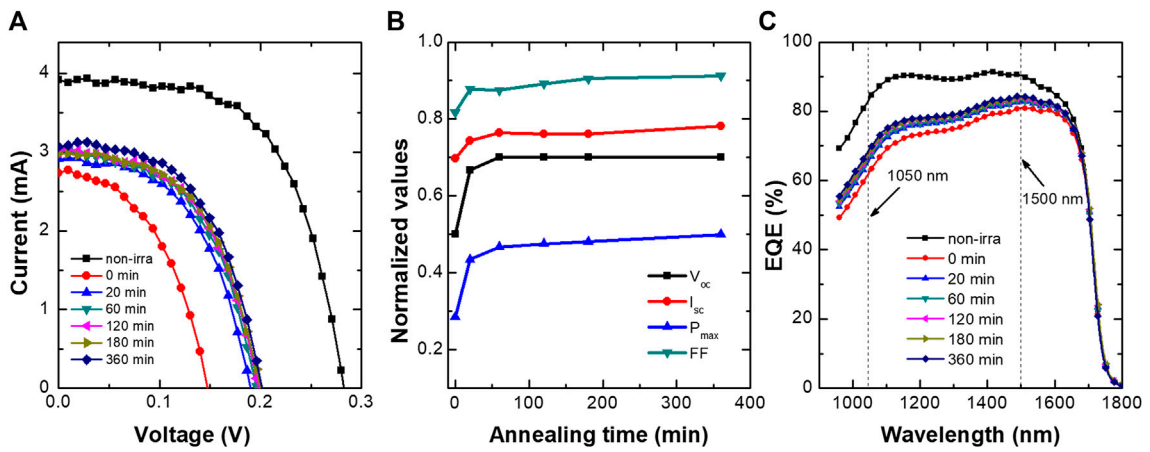

FIGURE 4 | The variation of electrical parameters **(A)** I-V curve and **(B)** normalized V_{oc}, I_{sc}, P_{max}, and FF, and the variation of **(C)** EQE of the solar cell of $In_{0.53}Ga_{0.47}As$ solar cells after annealing at 150°C for different times.

CONCLUSION

In the present article, $In_{0.53}Ga_{0.47}As$ single junction solar cells were subjected to 150 KeV proton radiation with different fluence. In order to understand low-energy proton irradiation-induced displacement damage distribution in $In_{0.53}Ga_{0.47}As$ single junction solar cells, SRIM simulations were firstly carried out to simulate the trajectories of 10^5 incident protons and the induced trap density. And the corresponding electrical and spectral properties after irradiation were well characterized by experimental means. The sequential annealing process from 20 to 360 min were then implemented to reveal the recovery of cell performance correlated with the thermal annealing treatments. The results show that most protons would penetrate through InGaAs emitter and stop in InGaAs base region, causing differing extents of electric and spectral degradation. When proton fluence was 1×10^{12} and 5×10^{12} p/cm^2, the remaining factor of I_{sc}, V_{oc}, P_{max}, and FF were degraded to 0.790, 0.767, 0.558, 0.921 and 0.697, 0.500, 0.285, 0.817, respectively. Severer degradation of short wave lengths then long wave lengths of the solar cell spectral response was observed. After annealing for different time at 150°C, it was shown that significant recovery of cell performance occurred after annealing for 20 min, and there is no obvious substantial recovery in continuous annealing up to 360 min. And the recovered degree of the short wavelength of the solar cell was greater than in the long wavelength, which reveals the fact that the defects induced by proton irradiation in the whole emission area and part of the base area were annihilated. Our performance analysis could provide evidence for the proper design of annealing cycles for space missions to enhance the performance of solar cells in space.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

YZ, AA, and QL designed the research, conducted the experiments and wrote this manuscript. All authors contributed to the discussion of the results and edited the manuscript.

ACKNOWLEDGMENTS

This work was supported by Basic Research Foundation of Yunnan Province(Grant number: 202001AU070090), a key project of the Natural Science Foundation of China (Grant number: 61534008), and Doctoral Start-up Funding of Yunnan Normal University (Gran number: 2019XJLK05/ 01700205020503040).

REFERENCES

1. King RR, Law DC, Edmondson KM, Fetzer CM, Kinsey GS, Yoon H, et al. Advances in high-efficiency III-V multijunction solar cells. *Adv Optoelectron* (2007) **2007**:29523. doi:10.1155/2007/29523
2. Sato S-i, Miyamoto H, Imaizumi M, Shimazaki K, Morioka C, Kawano K, et al. Degradation modeling of InGaP/GaAs/Ge triple-junction solar cells irradiated with various-energy protons. *Sol Energy Mater Sol Cell* (2009) **93**:768–73. doi:10.1016/j.solmat.2008.09.044
3. Shimazaki M. III-V compound multi-junction solar cells: present and future. *Sol Energy Mater Sol Cell* (2003) **75**:261–9. doi:10.1016/s0927-0248(02)00168-x
4. Essig S, Stämmler E, Rönsch S, Oliva E, Schachtner M, Siefer G, et al., Dilute nitrides for 4- and 6- junction space solar cells. The 9th European space power conference; 2011 Jun 6-10; Saint Raphael, France. Noordwijk, Netherlands: European Southern Observatory **Vol. 69**, 978-92-9092-257-5 (2011)
5. Lei QQ, Aierken A, Sailai M, Heini M, Shen XB, Zhao XF, et al. 150 KeV proton irradiation effects on photoluminescence of GaInAsN bulk and quantum well structures. *Opt Mater* (2019) **97**:109375. doi:10.1016/j.optmat.2019.109375
6. Dimroth F, Grave M, Beutel P, Fiedeler U, Karcher C, Tibbits TND, et al. Wafer bonded four-junction GaInP/GaAs//GaInAsP/GaInAs concentrator solar cells with 44.7% efficiency. *Prog Photovoltaics Res Appl* (2014) **22**:277–82. doi:10.1002/pip.2475
7. Liu XQ, Fetzer CM, Rehder E, Cotal S, Mesropian S, Law D, et al. Organometallic vapor phase epitaxy growth of upright metamorphic multijunction solar cells. *J Cryst Growth* (2012) **352**:186–9. doi:10.1016/j.jcrysgro.2011.10.024
8. Aierken A, Fang L, Heini M, Zhang Q, Li ZH, Zhao XF, et al. Effects of proton irradiation on upright metamorphic GaInP/GaInAs/Ge triple junction solar cells. *Sol Energy Mater Sol Cell* (2018) **185**:36–44. doi:10.1016/j.solmat.2018.04.035
9. Yoon H, Haddad M, Mesropian S, Yen J, Edmondson K, Law D, et al. Progress of inverted metamorphic III-V solar cell development at spectrolab. IEEE 33rd photovoltaic specialists conference; 2008 May 11–16; San Diego, CA. Sylmar, CA: IEEE. p. 1–6.
10. Tibbits TND, Beutel P, Grave M, Karcher C, Oliva E, Siefer G, et al. New efficiency frontiers with wafer-bonded multi-junction solar cells. In *Proceedings of the 29th European photovoltaic solar energy conference and exhibition* (2016) p. 1–4.
11. Yamaguchi M, Takamoto T, Araki K, Nobuaki K. Recent results for concentrator photovoltaics in Japan *Jpn J Appl Phys* (2016) **55**:04EA05. doi:10.7567/JJAP.55.04EA05
12. Dai P, Ji L, Tan M, Uchida S, Wu Y, Abuduwayiti A, et al. Electron irradiation study of room-temperature wafer-bonded four-junction solar cell grown by MBE. *Sol Energy Mater Sol Cell* (2017) **171**:118–22. doi:10.1016/j.solmat.2017.06.046
13. Walters RJ, Shaw GJ, Summers GP, Burke EA, Messenger SR. Radiation effects in Ga(0.47)In(0.53)As solar cells. 12th space photovoltaic research and technology conference; Cleveland, OH: SPRAT 12 (1993)
14. Karlina LB, Vozlovskii VV, Solov'ev VA, Shvarts MZ. The effects of irradiation on heteroepitaxial InP/InGaAs solar cells. IEEE photovoltaic specialists conference (1996) Washington, DC, p. 239–42.
15. Shen XB, Aierken A, Heini M, Mo JH, Lei QQ, Zhao XF, et al. Degradation analysis of 1 MeV electron and 3 MeV proton irradiated InGaAs single junction solar cell. *AIP Adv* (2019) **9**:075205. doi:10.1063/1.5094472
16. Wang R, Guo Z, Wang G, Low-energy proton irradiation effects on GaAs/Ge solar cells. *Sol Energy Mater Sol Cell* (2006) **90**:1052–7. doi:10.1016/j.solmat.2005.05.018
17. Hu J-M, Wu Y-Y, Zhang Z, De-Zhuang Y, He S-Y. A study on the degradation of GaAs/Ge solar cells irradiated by <200 keV protons. *Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms* (2008) **266**:267–70. doi:10.1016/j.nimb.2007.11.010

18. Sumita T, Imaizumi M, Matsuda S, Ohshima T, Ohi A, Itoh H. Proton radiation analysis of multi-junction space solar cells. *Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms* (2003) **206**:448–51. doi:10.1016/s0168-583x(03)00791-2
19. Barnes JA, Coswell FN. Thermoplastics for space. *SAMPE Q* (1989) **20**(3): 22–7. doi:10.1016/S0168-583X(03)00791-2
20. Shin K-B, Kim C-G, Hong C-S, Lee H-H. Prediction of failure thermal cycles in graphite/epoxy composite materials under simulated low earth orbit environments. *Compos B Eng* (2000) **31**(3):223–35. doi:10.1016/s1359-8368(99)00073-6
21. Loo RY, Kamath GS, Li SS. Radiation damage and annealing in GaAs solar cells. *IEEE Trans Electron Dev* (1990) **37**(2):485–97. doi:10.1109/16.46387
22. Ciulla G, Brano VL, Dio VD, Cipriani G. A comparison of different one-diode models for the representation of I-V characteristic of a PV cell. *Renew Sustain Energy Rev* (2014) **32**(5):684–96. doi:10.1016/j.rser.2014.01.027
23. Johnston AH. Radation damage of electronic and optoelectronic devices in space. In: 4th international workshop on radiation effects on semiconductor devices for space application; 2000 Oct 11-13; Tsukuba, Japan. Switzerland: Material Science (2000) Fourth international workshop on radiation eflects on semiconductor devices for space application (2000)
24. Luque A, Hegedus S. *Handbook of photovoltaics science and engineering*. 2nd ed. Chap. 8, John Wiley & Sons, Ltd. (2011)
25. Yamaguchi M, Amano C. 60Co γ-ray and electron irradiation damage of GaAs single crystals and solar cells. *J Appl Phys* (1983) **54**(9), 5021–9. doi:10.1063/1.332771
26. Yamaguchi M, Taylor SJ, Matsuda S, Kawasaki O. Mechanism for the anomalous degradation of Si solar cells induced by high fluence 1 MeV electron irradiation. *Appl Phys Lett* (1996) **68**(22):3141–3. doi:10.1063/1.115804
27. Sze SM, Ng KK. *Physics of semiconductor devices*. Section 1.5 and 2.3, 3rd ed., Hoboken, NJ: John Wiley & Sons (2006)
28. Yamaguchi M, Okuda T, Taylor SJ. Minority-carrier injection-enhanced annealing of radiation damage to InGaP solar cells. *Appl Phys Lett* (1997) **70**(16):2180–2. doi:10.1063/1.119034
29. Yamaguchi M, Okuda T, Taylor SJT, Takamoto T. Superior radiation-resistant properties of InGaP/GaAs tandem solar cells. *Appl Phys Lett* (1997) **70**(12): 1566. doi:10.1063/1.118618
30. Ikeda M, Ando K, Yamamoto AUemura C. Injection-enhanced annealing of InP solar-cell radiation damage. *J Appl Phys* (1985) **58**:568. doi:10.1063/1.335664
31. Ahlgren T, Vainonen-Ahlgren E, Likonen J, Li W, Pessa M. Concentration of interstitial and substitutional nitrogen in GaNxAs1−x. *Appl Phys Lett* (2002) **80**:2314. doi:10.1063/1.1465522

Conflict of Interest: Author LF was employed by the company Uniwatt Technology Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 25 November 2020
doi: 10.3389/fphy.2020.616349

Defective Titanium Dioxide-supported Ultrasmall Au Clusters for Photocatalytic Hydrogen Production

Xiaoqin Zhang[1], *Wenna Zhang*[1], *Yuanmei Xu*[1,2,3]* *and Mingliang Jin*[1,2]*

[1]National Center for International Research on Green Optoelectronics, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China, [2]International Academy of Optoelectronics at Zhaoqing, South China Normal University, Guangdong, China, [3]State Key Lab of Silicon Materials, Zhejiang University, Hangzhou, China

Ultrasmall precious metal clusters have attracted extensive attention for providing a very specific surface and promoting electron transfer. In this work, ultrasmall Au clusters based on defective TiO_2 nanosheets (Au/D-TiO_2) were prepared and introduced into photocatalytic hydrogen evolution. Different defects of TiO_2 nanosheets (D-TiO_2) were constructed using a heating process and then loaded with Au clusters. Compared with bare TiO_2, Au clusters established on defective TiO_2 nanosheets with a narrower band gap showed higher light absorption performances, resulting in obviously enhanced photocatalytic hydrogen production performances. The Au/D-TiO_2 displayed the greatly enhanced photocatalytic hydrogen evolution activity of 3,142.33 $\mu mol\ h^{-1}\ g^{-1}$, which was over 45 times than the pure TiO_2. The results showed that the catalysts had good prospects in the field of photocatalytic hydrogen production.

Keywords: Au, cluster, defect TiO_2, H_2 evolution, photocatalytic

OPEN ACCESS

Edited by:
Chongfu Zhang,
University of Electronic Science and Technology of China, China

Reviewed by:
Xiaolei Wang,
University of Alberta, Canada
Yongchao Huang,
Guangzhou University, China

***Correspondence:**
Mingliang Jin
jinml@scnu.edu.cn
Yuanmei Xu
yuanmei.xu@ecs-scnu.org

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *12 October 2020*
Accepted: *02 November 2020*
Published: *25 November 2020*

Citation:
Zhang X, Zhang W, Xu Y and Jin M (2020) Defective Titanium Dioxide-supported Ultrasmall Au Clusters for Photocatalytic Hydrogen Production. Front. Phys. 8:616349. doi: 10.3389/fphy.2020.616349

1. INTRODUCTION

Photocatalytic H_2 evolution, which utilizes water decomposition under solar energy, is a promising route to overcome the growing energy crisis and environment issues. However, in this process, the photocatalyst plays a key role. At present, photocatalysts such as nitrides [1, 2], metal sulfides [3, 4], and metal oxides [5, 6] have been demonstrated in hydrogen evolution systems. Among the traditional photocatalysts, anatase is one of the most extensively studied photocatalysts due to its reasonable price, nontoxicity, and remarkable photochemical stability [7, 8]. However, TiO_2 with a comparatively large band gap (3.2 eV) can absorb only the ultraviolet portion of the solar spectrum, resulting in insufficient photocatalytic activity of TiO_2 for H_2 production.

Photocatalytic efficiency of TiO_2 for water decomposition is limited due to the high recombination rate of photogenerated carriers. Hence, many measures have been taken to solve the problem of low photocatalytic performance of TiO_2, involving the mingling of transition metals [9, 10], adding of the nonmetal ions [11, 12], loading of the precious metals [13, 14], the surface for dye-sensitized [15], and the generation of hybrid semiconductor [16, 17]. The manufacture of metal-semiconductor oxide composite materials is another very dynamic research area that can improve the photocatalytic activity of semiconductor oxide photocatalysts [18]. It has been reported that the transfer of electrons from semiconductor oxides to metals may reduce electron hole recombination events and improve the photocatalytic performances of semiconductor oxide-based catalysts. Thus, the electron hole pairs in the recombined semiconductor oxide produced by photons can be reduced by a large part [19, 20].

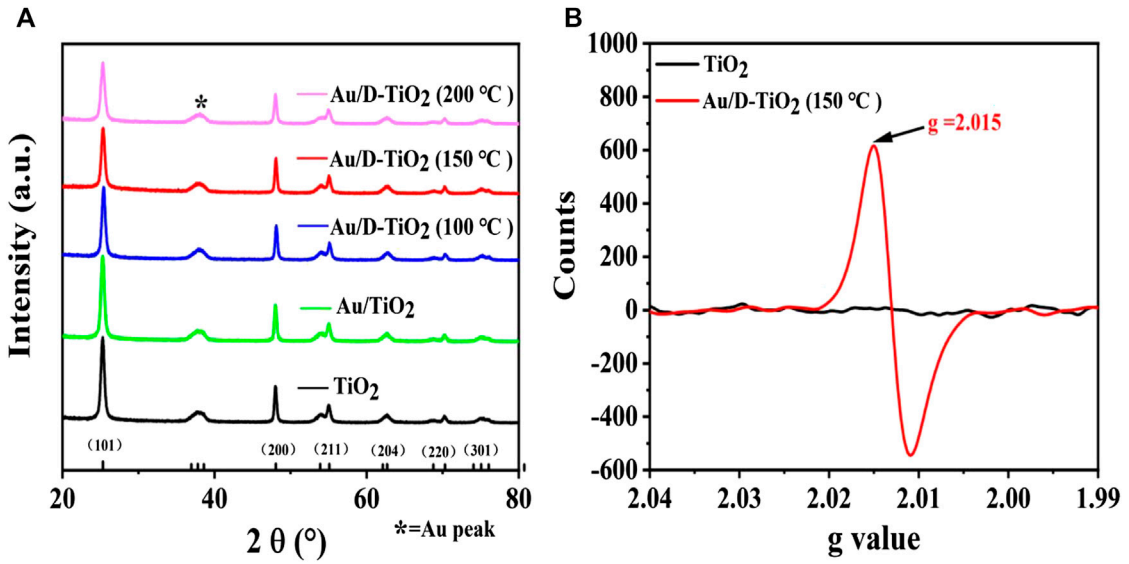

FIGURE 1 | (A) XRD of pure TiO_2, Au/TiO_2, and Au/D-TiO_2 (The D-TiO_2 were treated at 100°, 150°, and 200°C, respectively) **(B)** EPR of TiO_2 and Au/D-TiO_2 (treated at 150°C).

Therefore, in this work, the electron hole recombination rate was reduced, we have demonstrated a method of constructing defects on metal oxides to support Au clusters. Both oxygen vacancies and a Ti-O-Ti structure were formed in this process. Oxygen atoms were lost through the formed oxygen vacancies, and the original oxygen vacancies were occupied by metal clusters, thus effectively reducing the recombination rate of photogenerated carriers. Au clusters built on defective TiO_2 nanosheets (Au/D-TiO_2) were applied to the production of photocatalytic hydrolysis hydrogen and showed significantly enhanced performances. Compared with the traditional TiO_2 catalyst, the catalysts of Au/D-TiO_2 could produce up to 45 times more hydrogen than TiO_2. The results demonstrated that the structural defects on the surface of metal oxide could improve the catalytic performance of Au/D-TiO_2 catalysts. This work laid a foundation for the preparation of catalysts in the future.

2. EXPERIMENTAL SECTION

2.1. Materials

The chemical reagents Tetra-n-butyl Titanate ($Ti(OC_4H_9)_4$, 99.0%, AR grade), hydrofluoric acid solution (HF, 40.0 wt%, AR grade), ethanol (99.7%, AR grade), and ammonium carbonate (40%, AR grade) were bought from Tianjin Damao Chemical Reagent Factory, China. $HAuCl_4{\cdot}3H_2O$ (99.9%, AR grade) was bought from Aladdin, China. All experimental materials were used directly in the experiment after purchase.

2.2. Sample Preparation

2.2.1. Preparation of D-TiO_2 Nanosheets

TiO_2 nanosheets were obtained by hydrothermal method. In a common synthesis, $Ti(OC_4H_9)_4$ (50 ml) and HF (6 ml) were added to the Teflon-lined autoclave. Then, the hydrothermal reaction occurred at 180°C for 24 h. After the reaction, the white sediment of the Teflon-lined autoclave was centrifuged and rinsed three times with water and ethanol, and dried in an oven at a temperature of 80°C for 12 h. Different defects of D-TiO_2 sample were obtained after being calcined at reducing atmosphere (10 vol.% H_2 and 90 vol.% Ar, 2 h) with unequal temperatures (The D-TiO_2 were treated at 100°, 150°, and 200°C, respectively).

2.2.2. Synthesis of Au/D-TiO_2

Au/D-TiO_2 was composite according to the deposition-precipitation method. First, 1 g of pure or defective TiO_2 nanosheets powders were suspended in 50 ml distilled water and stirred for 20 min. Then, the aqueous solution (25 ml) of 2.4 g $(NH_4)_2CO_3$ and 0.01 g $HAuCl_4{\cdot}3H_2O$ were mixed into the above solution drop by drop and stirred for 1 h at room temperature. And then, the samples were collected by centrifugation, and washed with distilled water and ethanol for three times, respectively. Au/D-TiO_2 samples were gained after being dried at 70°C for 6 h and calcined (200°C, 10 vol.% H_2 and 90 vol.% Ar) for 2 h. For comparison, Au/TiO_2 was synthesized with the same method on TiO_2 nanosheets instead of D-TiO_2.

2.2.3. Characterization

The crystalline structures of the acquired samples were analyzed by a powder X-ray diffractometer (XRD, BRUKER D8 ADVANCE) with a scan range of 10°–90° and a step size of 0.02°. The measurement of electron paramagnetic resonance (EPR) was performed at room temperature using a BRUKE A300 EPR spectrometer. The transmission electron microscopy (TEM) and high angle annular dark-field (HAADF) were performed on a JEM-2100HR operating at

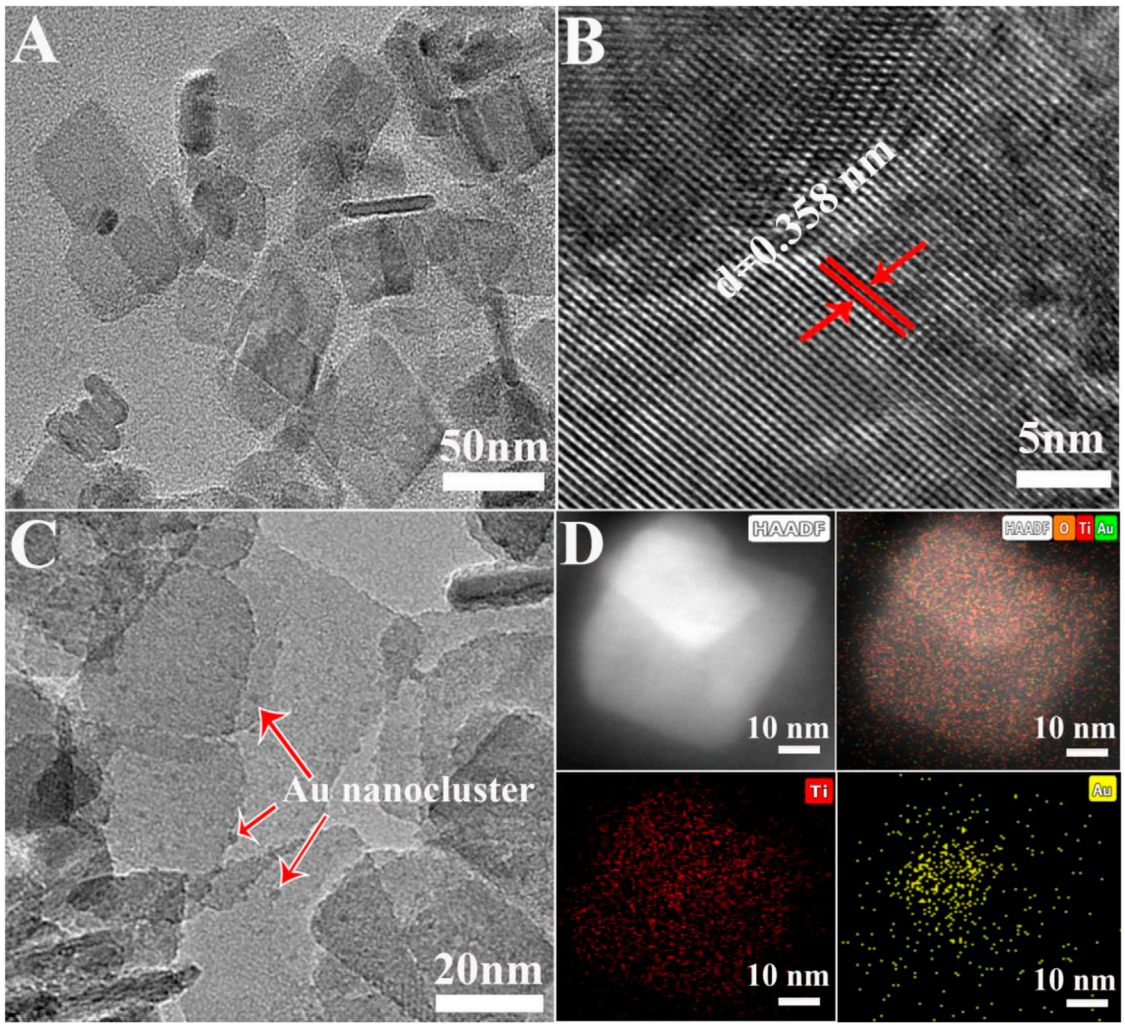

FIGURE 2 | (A) TEM image of TiO_2 nanosheets **(B)** HRTEM image of the selective area **(C)** TEM image, Au clusters high-lighted by red arrows **(D)** elemental mappings show the distribution of Ti (red) and Au (gold).

200 kV. The surface analysis of every sample was examined by X-ray photoelectron spectroscopy (XPS) using a Thermo Fisher Scientific K-Alpha spectrometer. The Metal contents of samples were analyzed by analysis of inductively coupled plasma spectroscopy (ICP, Agilent 700). The UV-vis diffuse reflectance spectra (DRS) recorded on a Lambda 750 spectrophotometer. The photoluminescence (PL) spectra gained by a F-4600 FL Spectrophotometer.

2.2.4. Photocatalytic Hydrogen Evolution

The photocatalytic hydrolysis reaction was carried out in a heat-resistant glass reactor, which was connected to a sealed single channel glass system, a circulating water system, a controller, and a vacuum pump. A 300 w xenon lamp was applied as a light source and fixed with the distance of 1 cm from the glass reactor. The glass reactor was charged with 80 ml of distilled water, 20 ml of methanol solution and 50 mg of photocatalyst. Before the start of the reaction, in order to keep the entire reaction device free of air, the vacuum pump was turned on for 1 h, and the valves of the glass system were rotated while evacuating, and the air of the system was extracted as much as possible. The circulating water and the controller were opened while the vacuum pump was being punched, and the magnetic stirrer placed under the glass reactor was turned at 500 rpm to avoid overheating of the reaction during the experiment and to guarantee uniform dispersion of the photocatalyst in the solution. In the photocatalytic hydrogen production experiment, a gas chromatograph was used to analyze the produced hydrogen gas, and the gas chromatograph was equipped with a packed bed column having a temperature of 70°C and a detector. In the hydrogen production experiment, the xenon lamp irradiation time was 1 h for each time, and the intake air was detected every hour with a total of 4 h [21, 22].

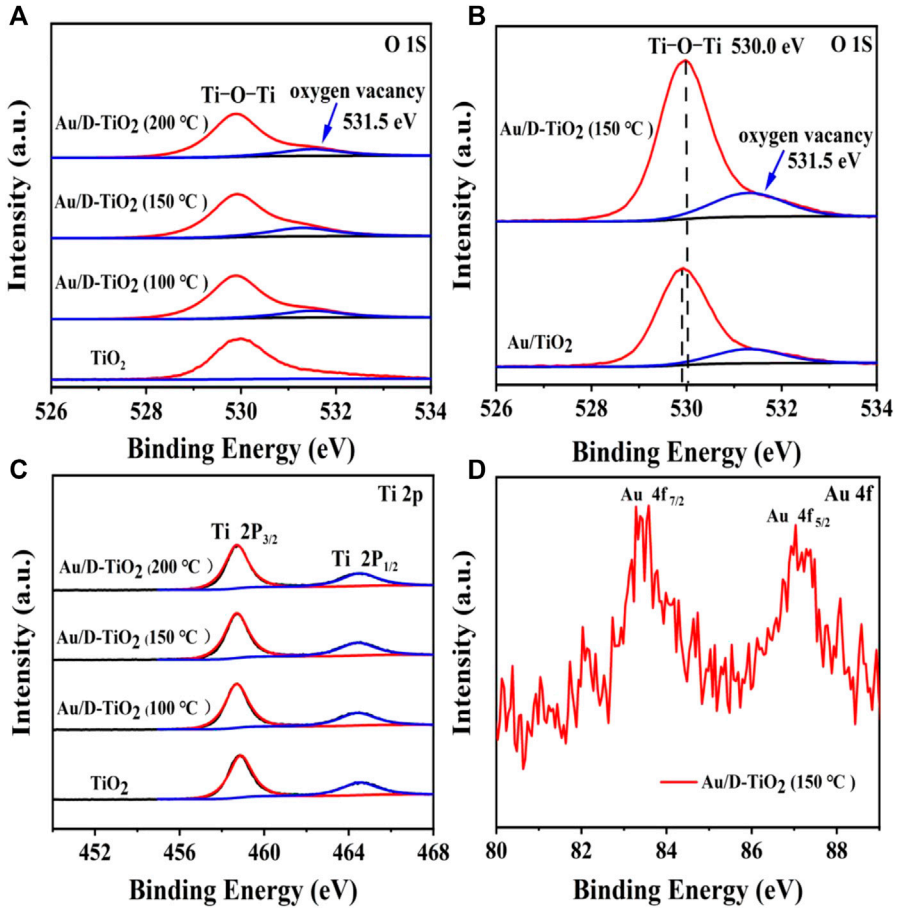

FIGURE 3 | (A) XPS of TiO_2 and D-TiO_2 (The D-TiO_2 were treated at 100˚, 150˚, and 200˚C, respectively) **(B)** XPS of synthesized Au/TiO_2 and Au/D-TiO_2 (treated at 150˚C) **(C)** Ti 2p spectrum **(D)** Au 4f spectrum.

3. RESULTS AND DISCUSSION

The compositions and phase structures of the samples were studied by using XRD (**Figure 1**). It could be seen from the XRD patterns that both pure and defective TiO_2 were with a palpable structure of anatase phase (JCPDS no. 21-1272). In the case of Au/D-TiO_2, the crystallographic peaks of Au could be detected (**Figure 1A**). Defects in D-TiO_2 were tested by electron paramagnetic resonance (EPR) measurements (**Figure 1B**), which indicated an aerobic vacancy defect with a g value of 2.015 [23–25].

The preparation of defective TiO_2 was the first and an important step in the synthesis process of the catalyst Au/D-TiO_2. The transmission electron microscopy (TEM) image (**Figure 2A**) demonstrated the nanosheet structure of D-TiO_2 with a length of about 50 nm. And the high-resolution transmission electron microscopy (HRTEM) showed a lattice spacing of approximate 0.358 nm (**Figure 2B**) corresponding to (101) planes of TiO_2. In addition, **Figure 2C** had many Au clusters distributed and marked by the red arrow. Further, the high angle annular dark-field (HAADF) and the element mapping image (**Figure 2D**) further indicated that the Au clusters were well dispersed on the defective TiO_2 carrier.

In order to further explore the surface defects of the pure TiO_2 and the defective TiO_2, the X-ray photoelectron spectroscopy (XPS) spectra were conducted. The O 1 score-level XPS spectrum manifested two peaks in **Figure 3A**, one of which at 530.0 eV was considered as the oxygen band of Ti-O-Ti, and another one at 531.5 eV could be attributed to oxygen vacancy. The peak area of 531.5 eV of 150˚C treated sample was larger than that of the other samples, indicating the highest oxygen vacancy concentration (**Figure 3A**). **Figure 3B** showed the O 1 s XPS spectrum of Au/TiO_2 and Au/D-TiO_2 (treated at 150˚C). It was found that the binding energy of Au/D-TiO_2 was significantly shifted to higher binding energies compared with Au/TiO_2. And the strength of Ti-O-Ti of Au/D-TiO_2 (treated at 150˚C) was greater than that of Au/TiO_2. Usually, such binding energy transfer is explained by a strong interaction between the two components. **Figure 3C** displayed the Ti 2p XPS spectrum of TiO_2 and Au/D-TiO_2 (The D-TiO_2 were treated at 100˚, 150˚, and 200˚C, respectively) and demonstrated that they

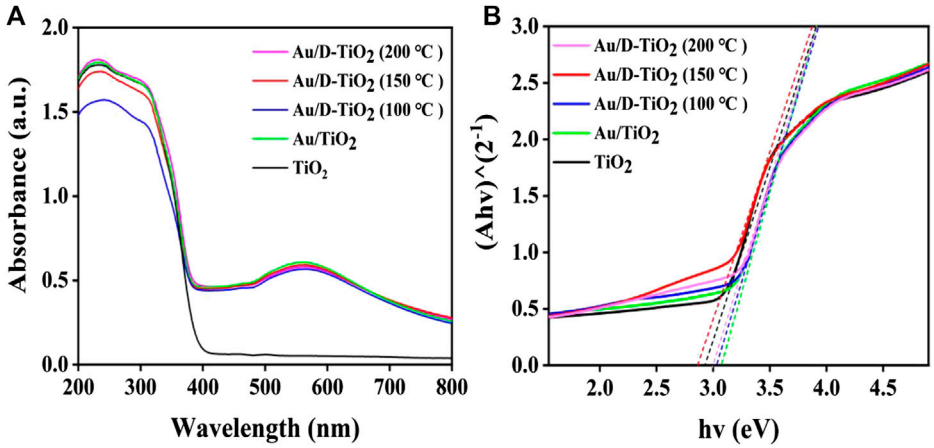

FIGURE 4 | (A) UV-vis DRS of synthesized of TiO_2, Au/TiO_2 and Au/D-TiO_2 (The D-TiO_2 were treated at 100°, 150°, and 200°C, respectively) **(B)** the forbidden band width of synthesized of TiO_2, Au/TiO_2, and Au/D-TiO_2 (The D-TiO_2 were treated at 100°, 150°, and 200°C, respectively).

were not obvious differences. The Au 4f of Au/D-TiO_2 (treated at 150°C) core-level XPS spectrum displayed one special peak at 83.9 eV in **Figure 3D**, and this peak could be attributed to Au clusters [26, 27].

The measurements of UV-visible diffuse-reflectance spectrum (UV-vis DRS) were carried out to investigate the light absorption intensity. As can be seen from **Figure 4A**, the intensity of light absorption for Au/D-TiO_2 was significantly enhanced in the wavelength ranging from 250 to 800 nm compared with the untreated TiO_2 nanosheets. And Au/D-TiO_2 (treated at 150°C) showed the strongest absorption intensity. According to the UV-vis DRS, the optical band gap value E_g of all samples was estimated from these absorption profiles using the Tauc's relation (**Figure 4B**) [28]. As shown in **Figure 4B**, the band gaps of these samples were analogous. However, after combining with Au clusters, the band gap of the samples mildly narrowed, whereby the results of pure TiO_2, Au/TiO_2, Au/D-TiO_2 (treated at 100°C), Au/D-TiO_2 (treated at 150°C) and Au/D-TiO_2 (treated at 200°C) were 2.93, 3.05, 3.03,2.86 and 2.98 ev, respectively [29, 30].

Figure 5 presented a comparison of the PL spectra of synthesized TiO_2 and Au/D-TiO_2 (The D-TiO_2 were treated at 100°, 150°, and 200°C, respectively). PL emission in semiconductors was due to the recombination of free carriers. The PL peak at about 396 nm was attributed to the emission of the bandgap transition. At an excitation wavelength of 230 nm, the light energy was approximately equal to the bandgap energy of anatase (387 nm). As expected, the PL intensity of the prepared Au/D-TiO_2 was significantly reduced as compared with pure TiO_2. This indicated that the recombination rate of electrons and holes of Au/D-TiO_2 sample was low. This might be owing to the fact that electrons were excited from the valence band to the conduction band, and then moved to Au, thereby preventing the direct recombination of electrons and holes. In general, low recombination rate of electrons and holes are often associated with high photocatalytic activity [31, 32].

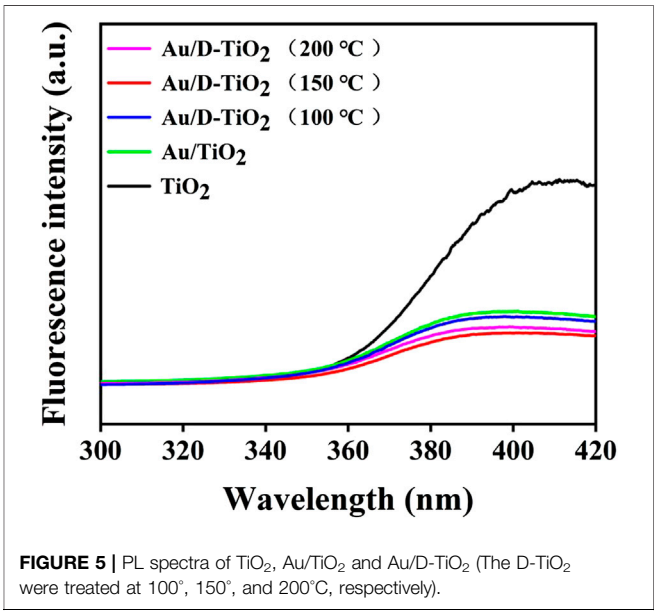

FIGURE 5 | PL spectra of TiO_2, Au/TiO_2 and Au/D-TiO_2 (The D-TiO_2 were treated at 100°, 150°, and 200°C, respectively).

In order to understand the band structure change of TiO_2 nanosheets after Au loading, the Mott-Schottky experiment (MS) was conducted (**Figure 6**).

Based on the MS equation, the capacitance (C) depend on applied potential and could be fitted as follows [33]:

$$\frac{1}{C^2} = + \frac{2}{e\varepsilon\varepsilon_0 N_d}\left(E - E_{FB} - \frac{k_b}{e}\right) \quad (1)$$

Where the slope expresses the type of semiconductor (negative to p-type and positive to n-type). The N_d denotes the carrier density. The *C*, *E* (corrected by the AgCl vs 0.197 eV) denote the space charge capacitance and applied potential. ε denotes the relative permittivity. ε_0 denotes the vacuum permittivity. Where *e*

denotes the electron charge. Besides, the intercept on the x axis denotes the E_{fb} (band potential).

As revealed in **Figure 6**, the positive slope in the Mott–Schottky plot for the samples indicated that TiO_2 nanosheets was typical n-type semiconductors according to the MS equation. Furthermore, the conduction band (CB) of the samples were obtained corrected by the AgCl by calculating, and the flat band potential (E_{fb} (Vs RHE)) of pure TiO_2, Au/TiO_2, Au/D-TiO_2 (treated at 100°C), Au/D-TiO_2 (treated at 150°C) and Au/D-TiO_2 (treated at 200°C) were gained to be −0.16, −0.13, −0.07, −0.04, and −0.01 V, respectively. Moreover, the valence band (VB) was calculated by adding the band gap value to the CB level. whereby the results of pure TiO_2, Au/TiO_2, Au/D-TiO_2 (treated at 100°C), Au/D-TiO_2 (treated at 150°C) and Au/D-TiO_2 (treated at 200°C) were 2.77, 2.92, 2.96, 2.82, and 2.97 V, respectively compared with NHE.

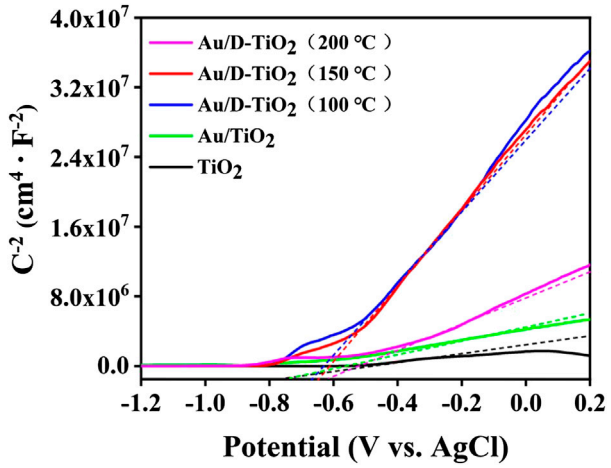

FIGURE 6 | The Mott–Schottky curves of using Ag/AgCl as a reference electrode of TiO_2, Au/TiO_2, and Au/D-TiO_2 (The D-TiO_2 were treated at 100°, 150°, and 200°C, respectively).

The photocatalytic H_2 production activities of the bare TiO_2, Au/TiO_2 and Au/D-TiO_2 were further examined in **Figure 7**. Under the same experimental conditions, the test was performed every 1 h, and a total of 4 h of experiments were performed. Bare TiO_2 and Au/TiO_2 only exhibited a very low photocatalytic H_2 production rate of 69.18 and 138.7 $\mu mol\, h^{-1}\, g^{-1}$, respectively. However, after coupling with Au clusters, the production rate of photocatalytic H_2 was significantly enhanced. Au/D-TiO_2 (treated at 150°C) exhibited the highest photocatalytic H_2 production rate of 3,142.33 $\mu mol\, h^{-1}\, g^{-1}$, suggesting that the construction of Au/D-TiO_2 could effectively boost the production activity of photocatalytic H_2. Besides, the production rate of photocatalytic H_2 first increased and then declined when the treated temperature of D-TiO_2 increased from 100° to 200°C for Au/D-TiO_2 samples, indicating that the treated temperature could efficiently adjust the defects of TiO_2 nanosheets and effect the H_2 production activity [34–36].

The stability test with Au/D-TiO_2 was carried out by a total of 16 h of cycling experiments, which were divided into four groups of 4 h, respectively, and an injection was performed every hour for data recording. The result suggested a good stability of Au/D-TiO_2 at fully photocatalytic hydrogen production experiment process without obvious decrease, which might be due to the Au clusters could be effectively stabilized on TiO_2.

A reaction mechanism for the evolution of hydrogen is shown in **Figure 8**. First, the valence band electrons of TiO_2 are excited to the conduction band, and then immediately transferred to Au through the intimate interface contacts. Since it is well known that H_2O has an ionization balance in the natural state, the water contains trace amounts of H^+ and OH^-. Therefore, the releasing H^+ binding in the e^- of H_2O to produce H_2. H^+ adsorption on the surface of the Au and accepting excited electrons is a key step in the

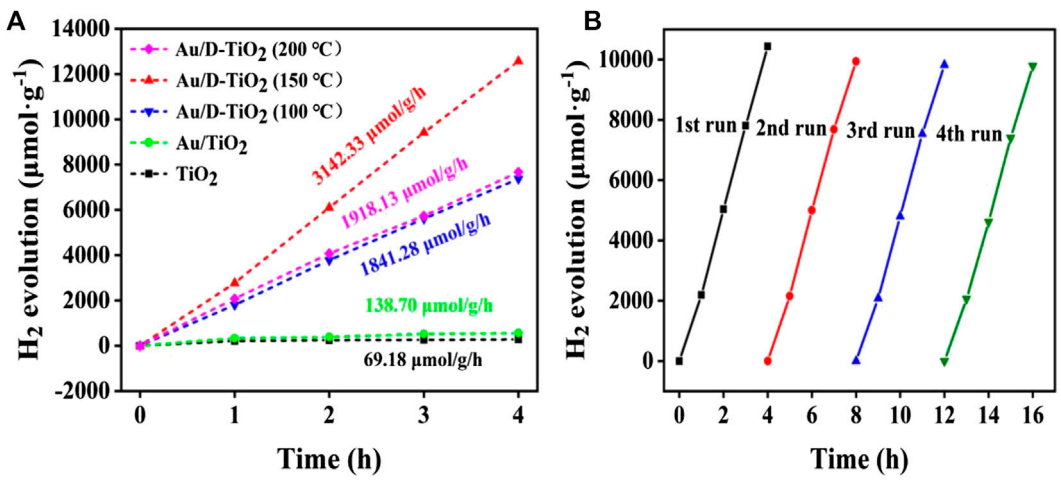

FIGURE 7 | (A) H_2 production of TiO_2, Au/TiO_2, and Au/D-TiO_2 (The D-TiO_2 were treated at 100°, 150°, and 200°C, respectively) under sunlight irradiation **(B)** Circulation experiment of hydrogen evolution of Au/D-TiO_2 (treated at 150°C).

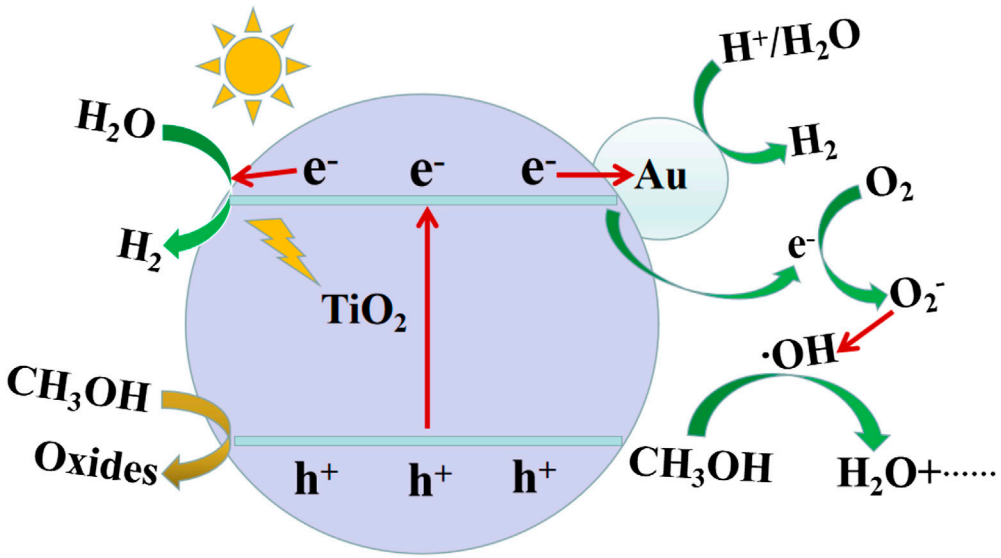

FIGURE 8 | Schematic diagram of charge transfer and H_2 production of Au/D-TiO_2.

hydrogen evolution reaction. On the other hand, H_2O/OH^- reacts with h^+ in Au/D-TiO_2 to produce H_2 and OH. When H_2 is released, O_2 is also formed. Then O_2 reacts with the e^- to form O_2^-. Finally, the O_2^- reacts with CH_3OH to form H_2O. When TiO_2 is coupled with Au clusters, it can provide more active sites for the hydrogen release reaction due to the characteristics of the Au/D-TiO_2 catalyst, thereby increasing photocatalytic activity [37, 38].

CONCLUSION

In conclusion, Au based on defective TiO_2 nanosheets (Au/D-TiO_2) has been successfully prepared and applied to the production of photocatalytic hydrolysis hydrogen. Compared with bare TiO_2 and Au/TiO_2 only exhibit a very low production rate of photocatalytic H_2 of 69.18 and 138.7 μmol h^{-1} g^{-1}, respectively, the production rate of photocatalytic H_2 was significantly enhanced after coupling with Au clusters. The Au/D-TiO_2 treated at 150°C exhibited the highest production rate of photocatalytic H_2 of 3,142.33 μmol h^{-1} g^{-1}. The stability test suggested a good stability of Au/D-TiO_2 at fully photocatalytic hydrogen production experiment process due to the Au clusters could be effectively stabilized on TiO_2. Overall, this could be an effective approach for enhancing photocatalytic hydrogen production efficiency and stability by catalyst.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

XZ: Conducted experiments and wrote manuscript; WZ: Conducted experiments, evaluated data; YX: Designed experiments, evaluated data, and revised manuscript; MJ: Acquired research funding, conceived research and revised manuscript.

FUNDING

This work was financially supported by Guangdong Provincial Grant (2018A050506025), Special Fund Project of Science and Technology Application in Guangdong (2017B020240002), Guangdong Provincial Key Laboratory of Optical Information Materials and Technology (Grant No. 2017B030301007), Guangdong Innovative Research Team Program (No. 2016ZT06C517), Science and Technology Program of Guangzhou (No. 2019050001).

REFERENCES

1. Li L, Yu H, Xu J, Zhao S, Liu Z, Li Y. Rare earth element, Sm, modified graphite phase carbon nitride heterostructure for photocatalytic hydrogen production. *New J Chem* (2019) 43:1716–24. doi:10.1039/C8NJ05619F
2. Ismael M, Wu Y, Wark M. Photocatalytic activity of ZrO_2 composites with graphitic carbon nitride for hydrogen production under visible light. *New J Chem* (2019) 43:4455–62. doi:10.1039/C8NJ06507A
3. Liang Q, Cui S, Xu S, Yao C, Maclachlan MJ, Li Z. A porous triptycene-based covalent polymer stabilized binary metal sulfide for enhanced hydrogen evolution under visible light. *Chem Comm* (2018) 54:3391–4. doi:10.1039/C8CC00665B
4. Zhang K, Guo L. Metal sulphide semiconductors for photocatalytic hydrogen production. *Catal Sci Technol* (2013) 3:1672. doi:10.1039/C3CY00018D
5. Martha S, Sahoo PC, Parida KM. An overview on visible light responsive metal oxide based photocatalysts for hydrogen energy production. *RSC Adv* (2015) 5:61535–53. doi:10.1039/C5RA11682A

6. Tian H, Cui X, Zeng L, Su L, Song Y, Shi J. Oxygen vacancy-assisted hydrogen evolution reaction of the Pt/WO_3 electrocatalyst. *J Mater Chem A* (2019) 7: 6285–93. doi:10.1039/C8TA12219A
7. Kumaravel V, Mathew S, Bartlett J, Pillai SC. Photocatalytic hydrogen production using metal doped TiO_2: a review of recent advances. *Appl Catal B Environ* (2019) 244:1021–64. doi:10.1016/j.apcatb.2018.11.080
8. Salimi M, Behbahani M, Sobhi HR, Gholami M, Jafari AJ, Kalantary RR, et al. A new nano-photocatalyst based on Pt and Bi co-doped TiO_2 for efficient visible-light photo degradation of amoxicillin. *New J Chem* (2019) 43:1562–8. doi:10.1039/C8NJ05020A
9. Yunarti RT, Lee M, Hwang YJ, Choi J, Suh DJ, Lee J, et al. Transition metal-doped TiO_2 nanowire catalysts for the oxidative coupling of methane. *Catal Commun* (2014) 50:54–8. doi:10.1016/j.catcom.2014.02.026
10. Kim J, Kwon G, Lim H, Zhu C, You H, Kim Y. Effects of transition metal doping in Pt/M-TiO_2 (M = V, Cr, and Nb) on oxygen reduction reaction activity. *J Power Sources* (2016) 320:188–95. doi:10.1016/j.jpowsour.2016.04.091
11. Li Y, Xu D, Oh JI, Shen WZ, Li X, Yu Y. Mechanistic study of codoped titania with nonmetal and metal ions: a case of C + Mo codoped TiO_2. *ACS Catal* (2012) 2:391–8. doi:10.1021/cs2006668
12. Yang G, Wang T, Yang B, Yan Z, Ding S, Xiao T. Enhanced visible-light activity of F-N co-doped TiO_2 nanocrystals via nonmetal impurity, Ti^{3+} ions and oxygen vacancies. *Appl Surf Sci* (2013) 287:135–42. doi:10.1016/j.apsusc.2013.09.094
13. Tian F, Zhu R, Ouyang F. Synergistic photocatalytic degradation of pyridine using precious metal supported TiO_2 with $KBrO_3$. *J Environ Sci* (2013) 25: 2299–305. doi:10.1016/S1001-0742(12)60304-0
14. Walsh FC, Bavykin DV, Torrentemurciano L, Lapkin AA, Cressey BA. Synthesis of novel composite materials via the deposition of precious metals onto protonated titanate (TiO_2) nanotubes. *Trans IMF* (2013) 84: 293–9. doi:10.1179/174591906X149077
15. Ooyama Y, Uenaka K, Sato T, Shibayama N, Ohshita J. Effective co-sensitization using D–π–A dyes with a pyridyl group adsorbing at Brønsted acid sites and Lewis acid sites on a TiO_2 surface for dye-sensitized solar cells. *RSC Adv* (2015) 5:2531–5. doi:10.1039/C4RA14190C
16. Fajariah N, Prabowo WA, Fathurrahman F, Melati A, Dipojono HK. The investigation of electronic structure of transition metal doped TiO_2 for diluted magnetic semiconductor applications: a first principle study. *Procedia Eng* (2017) 170:141–7. doi:10.1016/j.proeng.2017.03.032
17. Boga B, Szekely I, Pap Z, Baia L, Baia M. Detailed spectroscopic and structural analysis of TiO_2/WO_3 composite semiconductors. *J Spectrosc* (2018) 2018: 6260458. doi:10.1155/2018/6260458
18. Song S, Cheng B, Wu N, Meng A, Cao S, Yu J. Structure effect of graphene on the photocatalytic performance of plasmonic Ag/Ag_2CO_3-rGO for photocatalytic elimination of pollutants. *Appl Catal B Environ* (2016) 181: 71–8. doi:10.1016/j.apcatb. 2015.07.034
19. Efimkin DK, Burg GW, Tutuc E, MacDonald AH. Tunneling and fluctuating electron-hole Cooper pairs in double bilayer graphene. *Phys Rev B* (2020) 101: 035413. doi:10.1103/PhysRevB.101.035413
20. Prades JD, Hernandezramirez F, Jimenezdiaz R, Manzanares M, Andreu T, Cirera A, et al. The effects of electron–hole separation on the photoconductivity of individual metal oxide nanowires. *Nanotechnol* (2008) 19:465501. doi:10.1088/0957-4484/19/46/465501
21. Qi K, Lv W, Khan I, Liu S. Photocatalytic H_2 generation via CoP quantum-dot-modified g-C_3N_4 synthesized by electroless plating. *Chin J Catal* (2020) 41: 114–21. doi:10.1016/S1872-2067(19)63459-5
22. Reddy NR, Bhargav U, Kumari MM, Cheralathan KK, Shankar MV, Reddy KR, et al. Highly efficient solar light-driven photocatalytic hydrogen production over Cu/FCNTs-titania quantum dots-based heterostructures. *J Environ Manag* (2020) 254:109747. doi:10.1016/j.jenvman.2019.109747
23. Liu Q, Wang F, Lin H, Xie Y, Tong N, Lin J, et al. Surface oxygen vacancy and defect engineering of WO_3 for improved visible light photocatalytic performance. *Catal Sci Technol* (2018) 17:399–406. doi:10.1039/C8CY00994E
24. Alsaad AM, Al-Bataineh QM, Qattan IA, Ahmad AA, Ababne A, Bataineh Z, et al. Measurement and ab-initio investigation of structural, electronic, optical and mechanical properties of sputtered aluminium nitrides thin films. *Front Phys* (2020) 8:115. doi:10.3389/fphy.2020.00115
25. Ye K, Li K, Lu Y, Guo Z, Ni N, Liu H, et al. An overview of advanced methods for the characterization of oxygen vacancies in materials. *Trends Anal Chem* (2019) 116: 102–8. doi:10.1016/j.trac.2019.05.002
26. Zhang Z, Qin J, Shi W, Liu Y, Zhang Y, Liu Y, et al. Enhanced power conversion efficiency of perovskite solar cells with an up-conversion material of Er^{3+}-Yb^{3+}-Li^+ tri-doped TiO_2. *Nanoscale Res Lett* (2018) 13:147. doi:10.1186/s11671-018-2545-y
27. Criado A, Lavela P, Ortiz GF, Tirado JL, Gzouli S, Edfouf Z, et al. CTAB-assisted synthesis of C@$Na_3V_2(PO_4)_2F_3$ with optimized morphology for application as cathode material for Na-ion batteries. *Front Phys* (2019) 7: 207. doi:10.3389/fphy.2019.00207
28. Xu Y, Zhang H, Li X, Wang W, Li J. Ag-encapsulated single-crystalline anatase TiO_2 nanoparticle photoanodes for enhanced dye-sensitized solar cell performance. *J Alloys Compd* (2017) 95:1104–11. doi:10.1016/j.jallcom.2016.10.236
29. Wang S, Li C, Wang T, Zhang P, Li A, Gong J. Controllable synthesis of nanotube-type graphitic C_3N_4 and their visible-light photocatalytic and fluorescent properties. *J Mater Chem A* (2014) 2:2885. doi:10.1039/C3TA14576J
30. Xu Y, Zhang H, Li X, Wu Q, Wang W, Li Z, et al. Investigation of the improved performance with ferrites in TiO_2 dye-sensitized solar cell. *Appl Surf Sci* (2017) 424:245–50. doi:10.1016/j.apsusc.2017.04.210
31. Hisatomi T, Takanabe K, Domen K. Photocatalytic water-splitting reaction from catalytic and kinetic perspectives. *Catal Lett* (2014) 145:95–108. doi:10.1007/s10562-014-1397-z
32. Yu X, Wang Y, Meng X, Yang J. Preparation and characterization of Pd/N codoped TiO_2 photocatalysts with high visible light photocatalytic activity. *Chin J Catal* (2013) 34:1418–28. doi:10.1016/S1872-2067(12)60597-X
33. Martin DJ, Reardon PJ, Moniz SJ, Tang J. Visible light-driven pure water splitting by a nature-inspired organic semiconductor-based system. *J Am Chem Soc* (2014) 136:12568–71. doi:10.1021/ja506386e
34. Akple MS, Low J, Wageh S, Alghamdi AA, Yu J, Zhang J, et al. Enhanced visible light photocatalytic H_2-production of g-C_3N_4/WS_2 composite heterostructures. *Appl Surf Sci* (2015) 358:196–203. doi:10.1016/j.apsusc.2015.08.250
35. Li K, Lu X, Zhang Y, Liu K, Huang Y, Liu H Bi_3TaO_7/Ti_3C_2 heterojunctions for enhanced photocatalytic removal of water-borne contaminants. *Environ Res* (2020) 185:109409. doi:10.1016/j.envres.2020.109409
36. Wei W, Tian Q, Sun H, Liu P, Zheng Y, Fan M, et al.. Efficient visible-light-driven photocatalytic H_2 evolution over MoO_2-C/CdS ternary heterojunction with unique interfacial microstructures. *Appl Catal B Environ* (2020) 260: 118153. doi:10.1016/j.apcatb.2019.118153
37. Xu Y, Wang X, Jin M, Kempa K, Shui L. Water splitting performance enhancement of the TiO_2 nanorod array electrode with ultrathin black phosphorus nanosheets. *Chem Electro Chem* (2019) 7:96–104. doi:10.1002/celc.201901456
38. Huang Y, Guo Z, Liu H, Zhang S, Wang P, Lu J, et al. Heterojunction architecture of N-doped WO_3 nanobundles with Ce_2S_3 nanodots hybridized on a carbon textile enables a highly efficient flexible photocatalyst. *Adv Funct Mater* (2019) 29:45. doi:10.1002/adfm.201903490

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 03 December 2020
doi: 10.3389/fphy.2020.618811

A Multi Waveform Adaptive Driving Scheme for Reducing Hysteresis Effect of Electrowetting Displays

Wei Li[1], *Li Wang*[2]* *and Alex Henzen*[1]

[1]*Guangdong Provincial Key Laboratory of Optical Information Materials and Technology & Institute of Electronic Paper Displays, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China,* [2]*School of Information Engineering, Zhongshan Polytechnic, Zhongshan, China*

Electrowetting display (EWD) is a new reflective display technology, which has the advantages of ultra-low power consumption, high contrast, fast response and full-color. However, due to a hysteresis effect, accurate gray scale display of EWDs cannot be achieved, which seriously restricted the display effect and performance of EWDs. In order to reduce the influence of hysteresis effect, a multi waveform adaptive driving scheme was proposed in this paper. Firstly, a multi waveform driving system was designed and implemented by a STM32 master chip and an AD5304 driver chip. The driving system could automatically select different driving waveforms according to the preset switching conditions. Then, different driving waveforms were designed and implemented according to different driving stages of EWDs. Finally, driving waveforms were mapped with each stage of the driving process one by one to realize the adaptive driving of multiple waveforms. The experimental results showed that, compared with the conventional square wave, the maximum hysteresis difference of hysteresis curve could be reduced by 39.19% with the multi waveform driving scheme.

Keywords: driving waveform, electrowetting display, hysteresis, driving scheme, multi waveform

OPEN ACCESS

Edited by:
Chongfu Zhang,
University of Electronic Science and Technology of China, China

Reviewed by:
Zhixian Lin,
Fuzhou University, China
Jinjia Wang,
Yanshan University, China

***Correspondence:**
Li Wang
wangli_oet@163.com

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *18 October 2020*
Accepted: *11 November 2020*
Published: *03 December 2020*

Citation:
Li W, Wang L and Henzen A (2020) A Multi Waveform Adaptive Driving Scheme for Reducing Hysteresis Effect of Electrowetting Displays. Front. Phys. 8:618811. doi: 10.3389/fphy.2020.618811

INTRODUCTION

As a new optical reflective display device, electrowetting displays (EWDs) have been greatly developed. EWDs have the advantages of ultra-low power consumption, high contrast, fast response and full-color [1]. The concept of EWDs was first proposed by G. Beni [2, 3]. The basic principle of EWDs is to manipulate the liquid transport in a microporous structure by electrowetting effect, so as to change the optical spatial coherence of porous surface and realize an optical display effect of white or transparent switching.

In 2003, a new reflective EWD principle based on oil-water two-phase dielectric wetting system was reported [4, 5]. In 2004, an EWD with high reflectivity was developed [6], which was four times as bright as a reflective liquid crystal display (LCD). At the same time, it also had the advantages of high contrast and fast video response. In 2005, a novel single-layer multi-color EWD which used a variety of dye pastes instead of the conventional color polarizer technology was proposed [7]. In 2010, a three-layer structure color EWD was fabricated by superimposing three primary colors [8]. In 2011, a scheme was proposed to realize a single-layer multi-color EWD by adding different colors of oil into sub-pixel by ink-jet printing (IJP) technology [9].

However, there are still many defects in EWDs, such as oil splitting, oil backflow, charge trapping, hysteresis effect, etc. [10, 11]. In order to solve these defects and further improve the optical

performance of EWDs, many scholars have made a lot of efforts. An asymmetrical driving concept was proposed and a charge trapping suppression method was developed [12], but the variable electrode could lead to the instability of electric field force. And the hysteresis effect would be more significant. Then, a method was proposed to reduce oil splitting by adding a voltage rising gradient in the driving waveform [13], but the rising gradient could increase the response time of EWDs. In addition, an error diffusion algorithm based on pixel neighborhood gray information and direct current (DC) balanced driving waveform was proposed for suppressing the oil backflow [14]. However, the oil backflow caused by the charge trapping cannot be solved effectively. Recently, an amplitude-frequency mixed modulation driving system was proposed to improve the response speed for driving gray scales and enhance the oil stability [15], but the driving scheme was based on a single waveform, which limited the development of driving waveforms.

In order to reduce the influence of hysteresis effect and improve the display performance of EWDs. The reason of hysteresis effect was analyzed, and the hysteresis curve of EWDs was obtained by experiments. Then, a multi waveform adaptive driving system was designed and implemented, and it could effectively reduce the influence of hysteresis effect in EWDs.

PRINCIPLES OF ELECTROWETTING DISPLAYS

Driving Principle of Electrowetting Displays

An EWD is mainly composed of glass substrate, indium tin oxides (ITO) guide electrode, hydrophobic insulation layer (HIL), pixel wall, colored oil, polar liquid, and top plate [16, 17], as shown in **Figure 1A**. Gray scale is realized in EWDs by applying an external voltage to control the movement of the colored oil [18]. The colored oil in a pixel spreads naturally and covers the whole pixel when no external voltage is applied, and the color of oil is displayed, as shown in **Figure 1B**. The colored oil is pushed to a corner in the pixel when the external voltage is applied, and the color of the substrate is displayed, as shown in **Figure 1C**. **Figure 1D** is the top view of oil spreading. **Figure 1E** is the top view of oil contracting.

The different degrees of oil contraction represent different optical states, which are characterized by the aperture ratio. The aperture ratio is a proportion of opening area in a whole pixel. The formula is defined as **Eq. 1** [5].

$$W_A(V) = \left(1 - \frac{S_{oil}(V)}{S_{pix}}\right) \times 100\% \tag{1}$$

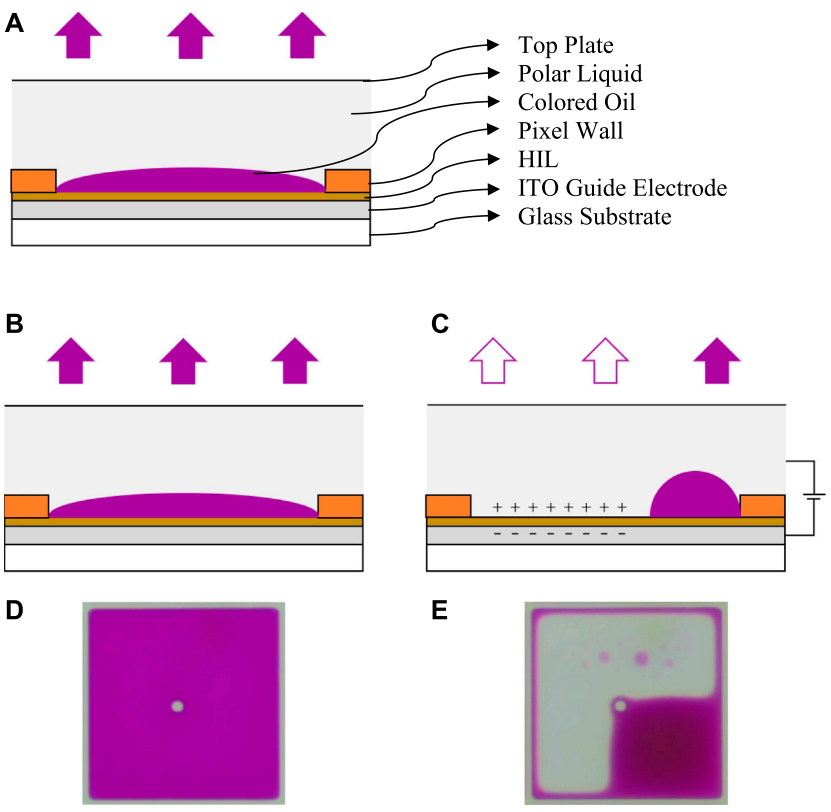

FIGURE 1 | Principle of the EWD. **(A)** Pixel structure of the EWD. **(B)** Without an applied external voltage, the color of oil is displayed in a pixel **(C)** With an applied external voltage, the color of substrate is displayed. **(D)** The top view when the pixel is turned off. **(E)** The top view when the pixel is turned on.

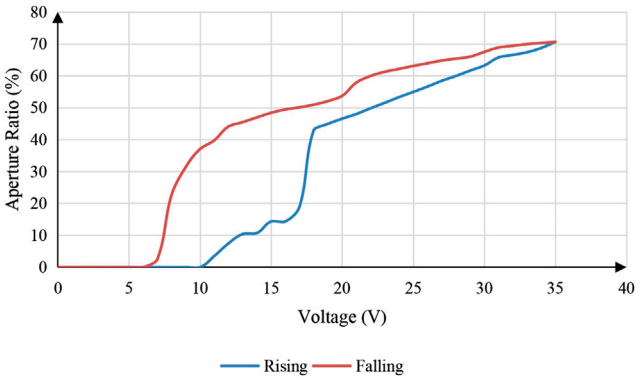

FIGURE 2 | The change of aperture ratio in the process of voltage rising and falling when an EWD is driven by a sine driving waveform.

In **Eq. 1**, $W_A(V)$ represents the aperture ratio, $S_{oil}(V)$and S_{pix} represent the surface area of oil in a single pixel and the surface area of a whole pixel respectively, V represents the driving voltage applied to EWDs, and the area of pixel wall can be ignored in calculating the aperture ratio. The pixel wall is a transparent grid structure which can divide an EWD into several pixels.

Principle of Hysteresis

The conventional sine wave is used to drive an EWD, as shown in **Figure 2**, the change of aperture ratio with driving voltage was tested. The results showed that the aperture ratio values of the same driving voltage value were different in the process of driving voltage rising and falling. This difference in optical response is called hysteresis effect [19].

The hysteresis effect is caused by the contact angle hysteresis in the driving process. It is also the inconsistency between the advancing contact angle and the receding contact angle (oil contracting and spreading) of a droplet. As shown in **Figure 3**, θ_A is the advancing contact angle, θ_R is the receding contact angle [20].

It can be verified by Young's equation and Lippmann-Young equation [21]. Young's equation describes the relationship between the contact angle and the surface tension of droplets. As shown in **Eq. 2**, γ_{LG} is a liquid-gas contact line, $\gamma_{SG}$is a solid-gas contact line, $\gamma_{SL}$is a solid-liquid contact line and $\theta$is a three-phase contact angle.

$$\gamma_{LG}\cos\theta = \gamma_{SG} - \gamma_{SL} \tag{2}$$

The Lippmann-Young equation describes the relationship between the contact angle and the applied voltage in an electrowetting model. As shown in **Eq. 3**, where θ is the contact angle when the driving voltage is V, and θ_0 is the contact angle when the driving voltage is 0, ε_0 and ε_r respectively represent the vacuum dielectric constant and the relative dielectric constant of the dielectric layer. $d$is the thickness of the dielectric layer, $\gamma_{LG}$is the liquid-gas contact line.

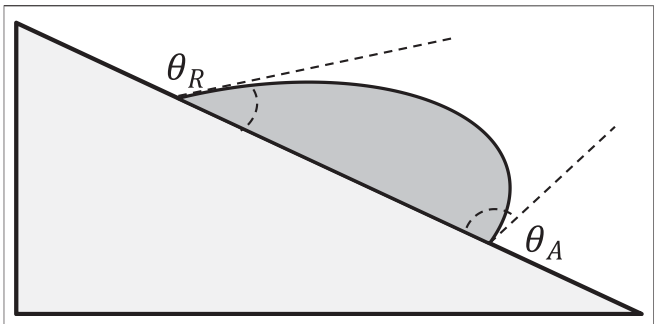

FIGURE 3 | The inconsistency between the advancing contact angle and the receding contact angle.

$$\cos\theta = \cos\theta_0 + \frac{1}{2}\frac{\varepsilon_0\varepsilon_r}{d\gamma_{LG}}V^2 \tag{3}$$

The fundamental reason of contact angle hysteresis is the rough surface of hydrophobic insulation layer and the viscous resistance of two-phase liquid. And this process could be deduced and verified [22], the relationship between the contact angle and driving voltage in a driving process of EWDs can be described as **Eq. 4**, where, θ is the contact angle when the driving voltage is V, θ_0 is the contact angle when the driving voltage is 0, θ_s is the saturation contact angle, θ_m is the contact angle when the driving voltage is V_m, c_{pin} is the maximum pinning force per unit length, $\gamma_{12}$is the surface tension at the interface between insulating fluid and conduction droplet, ε_0 and $\varepsilon_r$respectively represent the vacuum dielectric constant and the relative dielectric constant of the dielectric layer, $d$is the thickness of the dielectric layer, and k_1 is a constant, $C = \varepsilon_0\varepsilon_r/d$is the effective capacitance per unit, $\zeta$is the coefficient of contact line friction, $\eta$is the kinematic viscosity of the liquid, and G is the Langevin's function, V_1, V_2, $V_3$and V_m represent threshold voltage, saturation voltage, turning voltage and maximum voltage respectively.

$$\cos\theta = \begin{cases} \begin{cases} \cos\theta_0 & 0 - V_1 \\ \cos\theta_0 - \frac{c_{pin}}{\gamma_{12}} + \frac{\varepsilon_0\varepsilon_r V^2}{2d\gamma_{12}} - \frac{k_1 C(\zeta+6\eta)}{2\eta\gamma_{12}}V^2 & V_1 - V_2 \\ \cos\theta_0 + (\cos\theta_s - \cos\theta_0)G\left(\frac{\varepsilon_0\varepsilon_r V^2}{2d\gamma_{12}(\cos\theta_s - \cos\theta_0)}\right) & V_2 - V_m \end{cases} & advancing \\ \begin{cases} \cos\theta_m & V_m - V_3 \\ \cos\theta_0 + \frac{c_{pin}}{\gamma_{12}} + \frac{\varepsilon_0\varepsilon_r V^2}{2d\gamma_{12}} + \frac{k_1 C(\zeta+16\eta)}{2\eta\gamma_{12}}V^2 & V_3 - 0 \end{cases} & receding \end{cases} \tag{4}$$

The hysteresis effect seriously affects the gray scale display of EWDs, resulting in an inability of accurate gray scale. Without changing the structure and the surface material of hydrophobic insulating layer, the hysteresis effect can be reduced by driving methods. So, a multi-waveform driving system was designed and implemented, and different driving waveforms were designed for each driving stage.

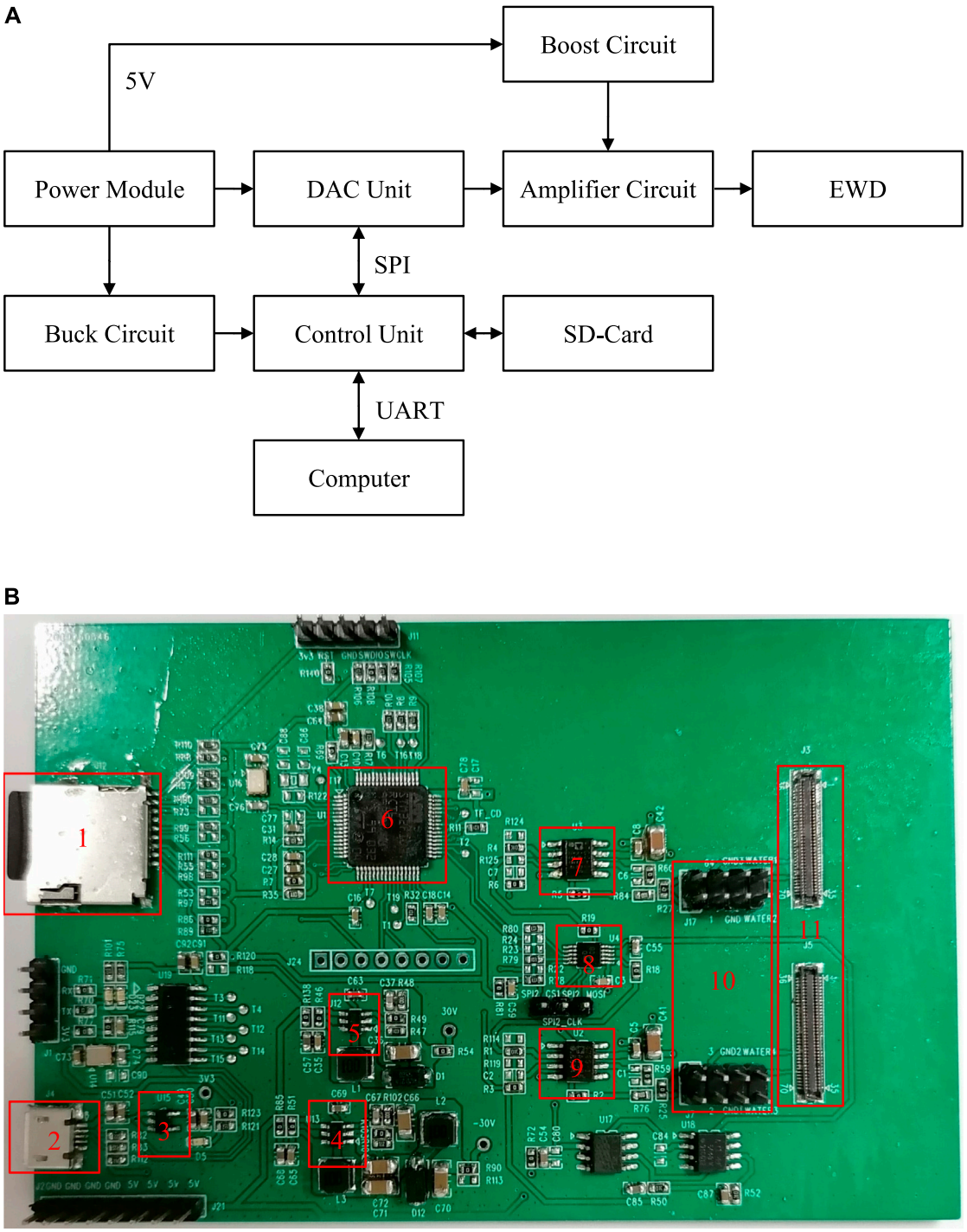

FIGURE 4 | The hardware circuit board consists of a power module, a control unit, a DAC unit, a buck circuit, a boost circuit, an amplifier circuit, a SD-Card and communication interfaces. **(A)** The framework of the hardware circuit board. **(B)** The physical picture of the hardware circuit board. 1) SD-Card. 2) Power Module. 3) Buck Circuit. 4) Boost Circuit. 5) Boost Circuit. 6) Control Unit. 7) Amplifier Circuit. 8) DAC Unit. 9) Amplifier Circuit. 10) Interface A for EWDs. 11) Interface B for EWDs.

DRIVING SYSTEM DESIGN FOR MULTI WAVEFORM

The driving system of EWDs has been reported in many papers. These driving systems were implemented with an FPGA (field programmable gate array) [23–25]. Although the performance was stable and the functions were rich, the multi waveform loading and conditional switching cannot be realized. In this paper, a multi waveform driving system was proposed, which could automatically select different waveforms according to

preset conditions. The whole driving system was mainly composed of a hardware circuit board, a driving program and a waveform editor. The implementation of these three parts is shown as follows.

Hardware Circuit Board

The framework of hardware circuit board is shown in **Figure 4A**. The power module supplied power, and the control unit provided control logic for a whole circuit system, the main control chip was a STM32 (STM32F103RCT6). The control unit could communicate with a PC (personal computer) by a UART (universal asynchronous receiver transmitter) interface to receive command information and data information. In addition, it could communicate with digital-to-analog conversion (DAC) unit by a SPI (serial peripheral interface) interface to send waveform data and receive feedback information. At the same time, a SD-Card was attached to the control unit for the storage of waveform files. An AD5304 chip was used in the DAC unit, which was mainly used to convert the received digital signal of the control unit into analog signals. Finally, the analog signal was amplified to output an appropriate voltage for achieving the control purpose of EWDs. As shown in **Figure 4B**, it is a physical picture of the hardware circuit board.

Driving Program

Driving program is the realization of control logics in a whole driving system. When the program is running, the initialization operation is performed first to reset each pin and communication interface. Then the system enters the command waiting interface. At this time, the control command can be entered. Users can control the system to load waveform files or set waveform switching conditions. Then, the driving program will analyze the waveform files when users give a driving instruction, and send the waveform data to drive EWDs according to switching conditions. In this paper, a STM32 (STM32F103RCT6) was adopted as the main control chip. The driving program was implemented on Windows 10 platform. The integrated development environment was Keil μVision5, and the development language was C programming language. The flow chart of the driving program is shown in **Figure 5**.

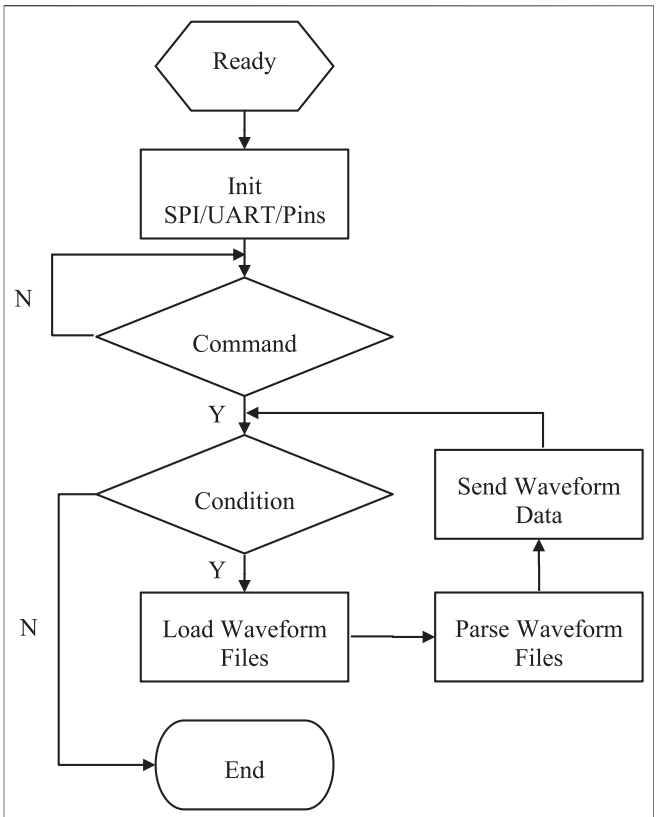

FIGURE 5 | The flow chart of the driving program. When the program starts to run, the initialization operation is performed first. Then the system enters the command waiting interface. Finally, the corresponding operations are executed according to command types.

Waveform Editor

The waveform editor could implement the design of an arbitrary waveform and generate waveform data quickly. The generated waveform data can be saved in a hexadecimal format file, so that the backend hardware system can parse the file and output the corresponding waveform. The design interface of the waveform editor is shown in **Figure 6**. The waveform editor can edit seven waveforms at the same time. In the parameter setting area, the default voltage value of unedited driving waveforms channel can be set, and the waveform repeated interval and repeat times can also be set. In the generated hexadecimal format data, every 11 bytes were in a group. In order to reduce the file size, all channel values were compared in each data interval. If there were any changes, the group of data would be saved, otherwise the group of data could not be saved.

The proposed multi waveform driving system can achieve the purpose of multi waveform loading and condition switching. By using this driving system, the hysteresis effect of EWDs can be reduced. The workflow of the whole system is shown in **Figure 7**.

Firstly, driving waveforms were designed by the waveform editor and saved as waveform files. By the reserved loading interface, the waveform files were loaded and parsed into waveform data. Then, switching conditions of the waveforms can be set. Switching conditions were the setting of voltage values. Finally, the corresponding waveform data was sent to the core driving chip by the driving program according to the current condition. So, EWDs can be driven to display contents by the interface of the hardware.

DRIVING WAVEFORM DESIGN FOR REDUCING HYSTERESIS

For EWDs, if the pixel structure and oil material cannot be improved, the hysteresis effect cannot be completely eliminated. Moreover, due to the variability of driving processes, it is difficult to satisfy different driving processes with a same driving

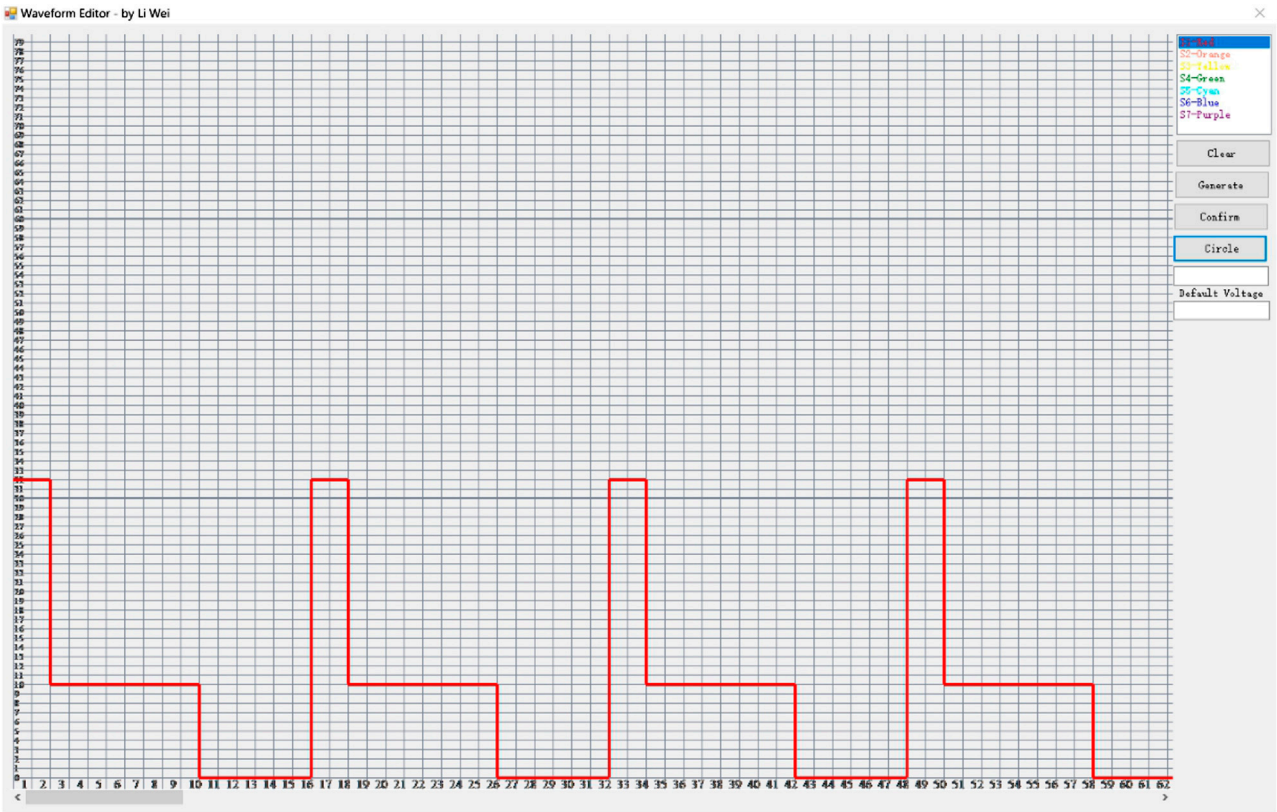

FIGURE 6 | The screenshot of the waveform editor. The waveform editor could implement the design of an arbitrary waveform and generate waveform data quickly.

waveform. Therefore, it is necessary to design different driving waveforms for different driving processes. At the same time, the different driving waveforms can be switched automatically by the multi waveform driving system, so as to realize a better display effect. There are many ways to divide the driving process, and many kinds of multi waveform schemes can be designed.

First scheme, according to **Eq. 4**, the driving process of EWDs can be divided into advancing and receding processes. Between them, the advancing process can be divided into three stages, and the receding process can be divided into two stages. Then driving waveforms can be designed for these five stages. Finally, it is necessary to determine switching conditions of driving waveforms. The starting and ending voltage values of segmented waveforms in this scheme can be used as switching conditions.

Second scheme, the hysteresis curves of EWDs driven by square wave, sine wave and triangle wave are tested by experiments at first. Then three hysteresis curves are superimposed, and the better waveform segment can be used as the driving waveform of the corresponding stage, so as to form a multi waveform. Finally, it is necessary to determine the switching condition of waveform segments, the starting and ending voltage values of selected waveform segments can be set as switching conditions.

In this paper, the second scheme was used to design the multi waveform. As shown in **Figure 8**, the hysteresis curves of EWDs driven by a square wave, a sine wave and a triangle wave were measured by experiments. It can be seen from **Figure 8** that under the same voltage, the square wave has the largest aperture ratio and the triangle wave has the smallest aperture ratio. This is because the overall power of the square wave is larger, so the electric field force is relatively larger. In order to reduce the hysteresis effect as much as possible, the following selections were made according to **Figure 8**. From 0 to 10 V, a square wave was selected as the first segment. From 10 to 18 V, a square wave was selected as the second segment. From 18 to 35 V, a sine wave was selected as the third segment. From 35 to 0 V, a triangle wave was selected as the fourth segment.

Due to the influence of threshold voltages, the aperture ratio of the first segment was zero. But it could be activated by an activation voltage. Therefore, the first segment can be optimized as the driving waveform which is shown in **Figure 9**, EWDs were activated for 2ms with the maximum voltage in each waveform cycle. Finally, the multi waveform can be obtained, as shown in **Figure 10**.

EXPERIMENTAL RESULTS AND DISCUSSION

For the sake of evaluating the display performance of the proposed multi waveform driving scheme, an experimental

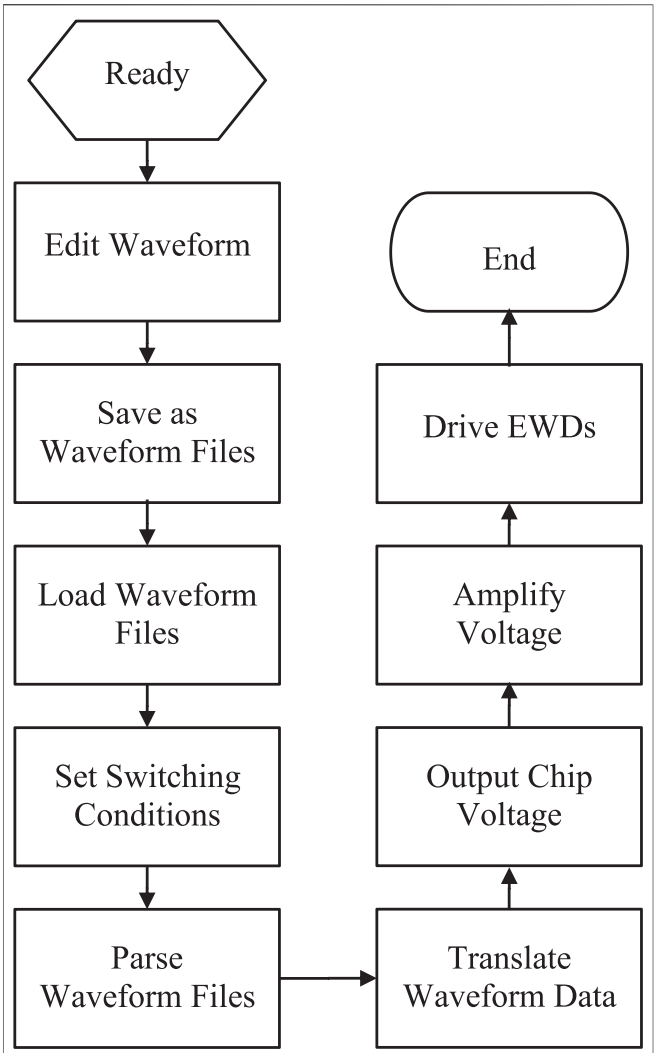

FIGURE 7 | The workflow of the multi waveform driving system. It includes of waveform editing, waveform saving, waveform loading, conditions setting, waveform parsing, waveform translating and voltage amplifying.

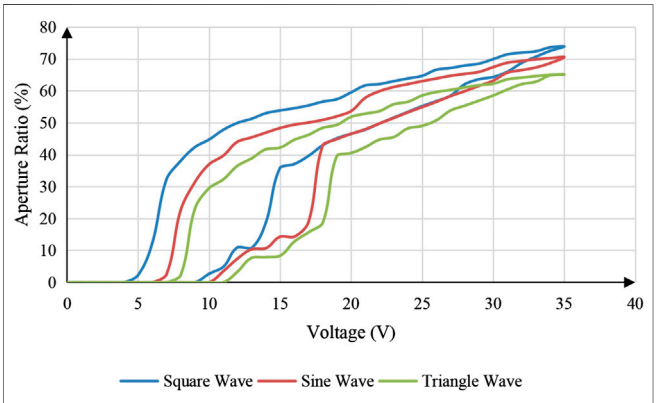

FIGURE 8 | The hysteresis curves of EWDs driven by a square wave, a sine wave and a triangle wave.

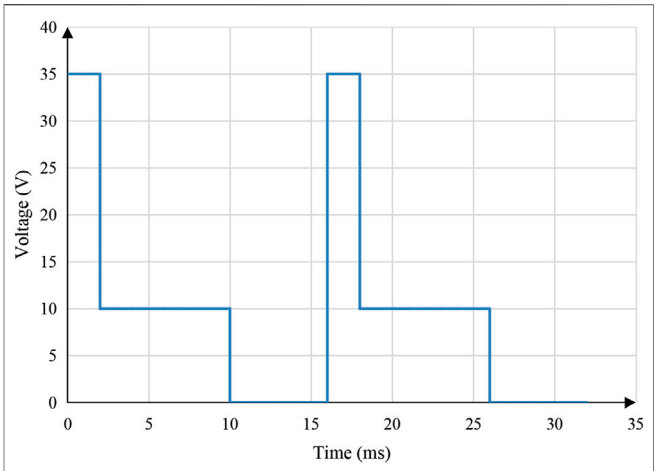

FIGURE 9 | The driving waveform for the first segment of the multi waveform was optimized with an activation voltage. A waveform cycle includes a 2 ms activation phase, an 8 ms driving phase and a 6ms reset phase.

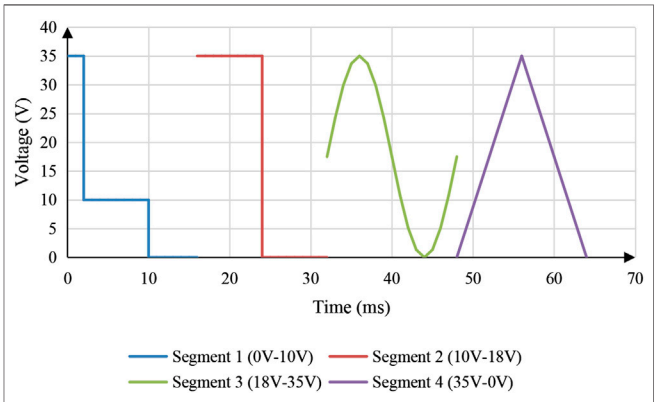

FIGURE 10 | The multi waveform proposed in this paper. There were four segments in total. The first segment was from 0 to 10 V, which was a square wave optimized with an activation voltage. The second segment was from 10 to 18 V, which was a square wave. The third segment was from 18 to 35 V, which was a sine wave. The fourth segment was from 35 to 0 V, which was a triangle wave.

platform was developed to measure the aperture ratio of EWDs, as shown in **Figure 11**. This experimental platform included a conventional driving system, a multi waveform driving system, a testing system, and an EWD panel. The conventional driving system was used to input the conventional driving waveform, which was consisted of a computer, a function generator, and a high-voltage amplifier. The multi waveform driving system was used to load the proposed multi waveform. The testing system was used to measure and record testing results, including a microscope, a high-speed camera, and a computer, and the EWD was the measured object. Its parameters are shown in **Table 1**. By this experimental platform, the conventional driving waveform and the adaptive multi waveform can be tested and compared.

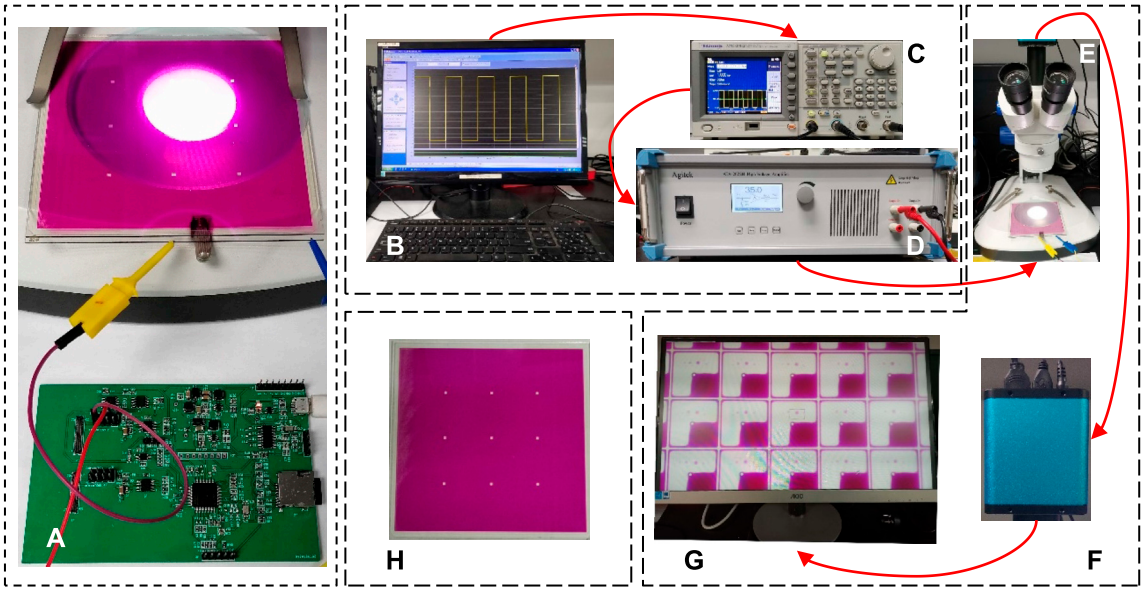

FIGURE 11 | Experimental platform of the driving waveform for EWDs. **(A)** A multi waveform driving system. **(B)** A computer. **(C)** An arbitrary function generator ATA-2022H (Agitek, Xi'an, China). **(D)** An AFG-3052C (Agitek, Xi'an, China) high voltage amplifier. **(E)** A Microscope. **(F)** A high-speed camera. **(G)** A PC for measuring system. **(H)** An EWD panel.

TABLE 1 | Parameters of an EWD panel.

Panel size	Oil color	Resolution	Pixel size	Pixel wall height	Hydrophobic layer thickness	Driving voltage
10 × 10 cm	Magenta	200 × 200	150 μm × 150 μm	5.6 μm	1 μm	0–35 V

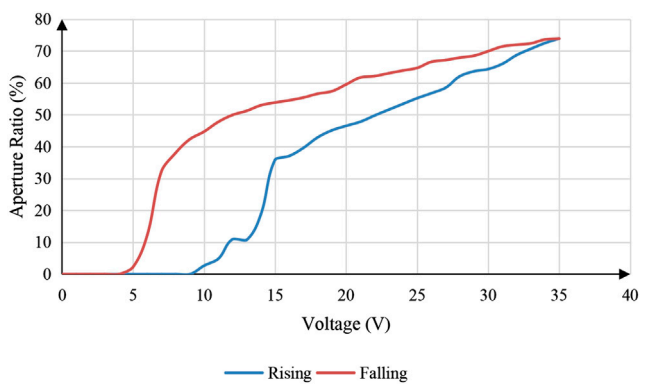

FIGURE 12 | The hysteresis curve of an EWD driven by the square wave.

First, the EWD was driven by the conventional driving system with conventional driving waveforms. Therefore, the hysteresis curves of the EWD driven by conventional driving waveforms could be tested. The experimental process can be divided into two steps. Firstly, the EWD was driven by the conventional driving system. The driving waveform was edited by a computer with ArbexPress software, and it was sent to a function generator by a serial communication, and then, the driving voltage of the driving waveform could be output when it was amplified by a high-voltage amplifier for driving an EWD. Secondly, the experimental data was measured by the testing system. The high-speed camera captured the display state by microscope in real time and transmitted testing data to a computer for calculating the aperture ratio. During the experiment, the temperature was kept a constant to avoid the influence of the external environment. The aperture ratio and driving voltage of the EWD with conventional driving waveforms were measured by the above-mentioned experimental platform. The measurement results are shown in **Figure 12**.

Second, the EWD was driven by the multi waveform driving system with the adaptive multi waveform. So, the hysteresis curve of EWDs driven by an adaptive multi waveform could be tested. Firstly, the multi waveforms were designed with the waveform editor and saved as the corresponding waveform files. Then, the multi waveform files were loaded by the reserved loading interface and switching conditions of the multi waveform were set. In the design process of multi waveforms, switching conditions of waveforms have been given. Finally, EWDs could be driven, and the aperture ratio data was tested and recorded by the testing system. The aperture ratio and driving voltage of the EWD with adaptive multi waveforms was measured. The measurement results are shown in **Figure 13**.

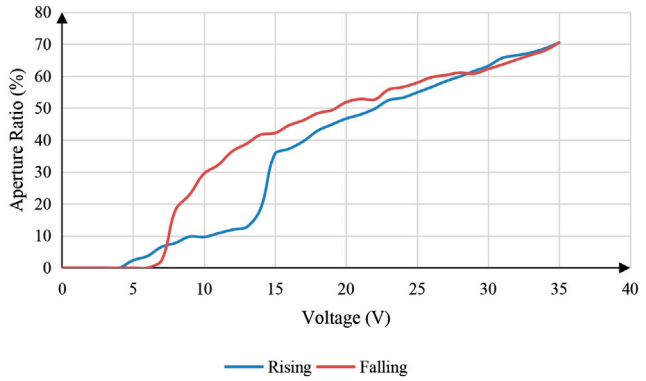

FIGURE 13 | The hysteresis curve of an EWD driven by the multi waveform.

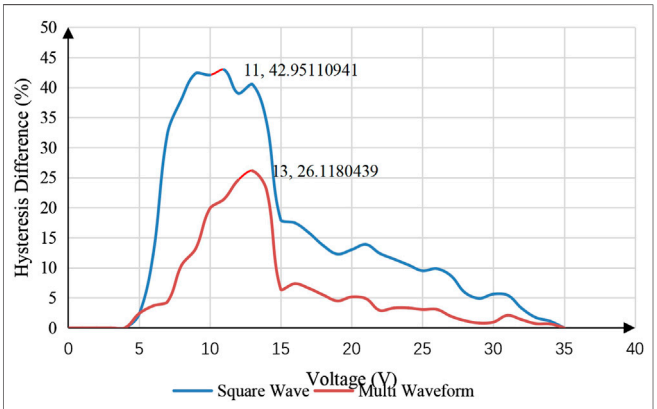

FIGURE 14 | Comparison of the hysteresis difference of an EWD when it was driven by a square wave and a multi waveform.

Compared with **Figures 12**, **13**, it can be seen that the multi waveform driving system and adaptive multi waveform can effectively reduce the hysteresis effect of EWDs. The absolute value of the difference of aperture ratio in the process of driving voltage rising and falling with a same voltage value was defined as hysteresis difference. Then, a comparison of hysteresis difference with a square wave and a multi waveform could be obtained, as shown in **Figure 14**. Compared with the square wave, the maximum hysteresis difference of hysteresis curve could be reduced by 39.19% when the multi waveform driving scheme can be used.

CONCLUSIONS

In this paper, a multi waveform adaptive driving scheme was proposed. This driving scheme includes a driving system and a multi waveform. The driving system could automatically select different driving waveforms according to the preset conditions. The influence of the hysteresis effect of EWDs could be reduced by using the driving scheme. And the driving scheme could provide a strong guarantee for the realization of the accurate gray scale display in EWDs. Although the structural parameters (pixel size, pixel wall height, insulation layer thickness) of the EWD can affect the hysteresis curve, it cannot affect the result that the hysteresis difference of the adaptive multi waveform driving scheme is smaller than that of the conventional driving waveforms. So, the multi waveform adaptive driving scheme proposed in this paper provides a new idea for the design of driving waveforms.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

WL and LW designed this project and carried out most of the experiments and data analysis. AH gave suggestions on the project management and conducted helpful discussion on the experimental results.

FUNDING

Supported by the National Key Research and Development Program of China (2016YFB0401502), Scientific research project of Guangdong Education Department (No. 2020ZDZX3083), Science and Technology Program of Guangzhou (No. 2019050001), Program for Chang Jiang Scholars and Innovative Research Teams in Universities (No. IRT_17R40), Program for Guangdong Innovative and Enterpreneurial Teams (No. 2019BT02C241), Science and technology project of Guangdong Province (No. 2018A050501013), Guangdong Provincial Key Laboratory of Optical Information Materials and Technology (No. 2017B030301007) and the 111 Project.

REFERENCES

1. Yi Z, Huang Z, Lai S, He W, Wang L, Chi F, et al. Driving waveform design of electrowetting displays based on an exponential function for a stable grayscale and a short driving time. *Micromachines* (2020) 11(3):313. doi:10.3390/mi11030313.
2. Beni G, Hackwood S. Electro-wetting displays. *Appl Phys Lett* (1981) 38(4): 207–9. doi:10.1063/1.92322.
3. Beni G, Tenan MA. Dynamics of electrowetting displays. *J Appl Phys* (1981) 52(10):6011–5. doi:10.1063/1.329822.
4. Hayes RA, Feenstra BJ. Video-speed electronic paper based on electrowetting. *Nature* (2003) 425(6956):383–5. doi:10.1038/nature01988.
5. Roques-Carmes T, Hayes RA, Feenstra BJ, Schlangen LJ. Liquid behavior inside a reflective display pixel based on electrowetting. *J Appl Phys* (2004) 95(8): 4389–96. doi:10.1063/1.1667595.
6. Roques-Carmes T, Hayes RA, Schlangen LJ. A physical model describing the electro-optic behavior of switchable optical elements based on electrowetting. *J Appl Phys* (2004) 96(11):6267–71. doi:10.1063/1.1810192.
7. Feenstra BJ. Rapid switching in multiple color active matrix driven electrowetting displays. *IDW Symp Digest* (2005) 2005:861–4.

8. You H, Steckl AJ. Three-color electrowetting display device for electronic paper. *Appl Phys Lett* (2010) 97(2):023514. doi:10.1063/1.3464963.
9. Ku YS, Kuo SW, Huang YS, Chen CY, Lo KL, Cheng WY, et al. Single-layered multi-color electrowetting display by using ink-jet-printing technology and fluid-motion prediction with simulation. *J Soc Inf Disp* (2011) 19(7):488–95. doi:10.1889/JSID19.7.488.
10. Ku YS, Kuo SW, Tsai YH, Cheng PP, Chen JL, Lan KW, et al. The structure and manufacturing process of large area transparent electrowetting display. *SID Symp Digest Techn Papers* (2012) 43(1):850–2. doi:10.1002/j.2168-0159.2012.tb05919.x.
11. Sun B, Heikenfeld J. Observation and optical implications of oil dewetting patterns in electrowetting displays. *J Micromech Microeng* (2008) 18(2): 025027. doi:10.1088/0960-1317/18/2/025027.
12. Chen YC, Chiu YH, Lee WY, Liang CC. A charge trapping suppression method for quick response electrowetting displays. *SID Symp Digest Tech Papers* (2010) 41(1):842–5. doi:10.1889/1.3500607.
13. Zhang XM, Bai PF, Hayes RA, Shui LL, Jin ML, Tang B, et al. Novel driving methods for manipulating oil motion in electrofluidic display pixels. *J Disp Technol* (2016) 12(2):200–5. doi:10.1109/jdt.2015.2477947.
14. Lin S, Zeng S, Qian M, Lin Z, Guo T, Tang B. Improvement of display performance of electrowetting displays by optimized waveforms and error diffusion. *J Soc Inf Disp* (2019) 27(10):619–29. doi:10.1002/jsid.790.
15. Yi Z, Liu L, Wang L, Li W, Shui L, Zhou G. A driving system for fast and precise gray-scale response based on amplitude-frequency mixed modulation in TFT electrowetting displays. *Micromachines* (2019) 10(11):732. doi:10.3390/mi10110732.
16. Dou Y, Tang B, Groenewold J, Li F, Yue Q, Zhou R, et al. Oil motion control by an extra pinning structure in electro-fluidic display. *Sensors* (2018) 18(4):1114. doi:10.3390/s18041114.
17. Li W, Wang L, Zhang T, Lai S, Liu L, He W, et al. Driving waveform design with rising gradient and sawtooth wave of electrowetting displays for ultra-low power consumption. *Micromachines* (2020) 11(2):145. doi:10.3390/mi11020145.
18. Feenstra J. Video-speed electrowetting display technology. *Handbook Visual Display Techn* (2015) 1:1–13. doi:10.1007/978-3-642-35947-7_103-2.
19. Van Dijk R, Feenstra BJ, Hayes RA, Camps IGJ, Boom RGH, Wagemans MMH, et al. Gray scales for video applications on electrowetting displays. *SID Symp Digest Tech Papers* (2006) 37(1):1926–9. doi:10.1889/1.2433427.
20. Yi Z, Feng H, Zhou X, Shui L. Design of an open electrowetting on dielectric device based on printed circuit board by using a parafilm M. *Front Phys* (2020) 8:193. doi:10.3389/fphy.doi:10.3389/fphy.2020.00193.
21. Yi Z, Feng W, Wang L, Liu L, Lin Y, He W, et al. Aperture ratio improvement by optimizing the voltage slope and reverse pulse in the driving waveform for electrowetting displays. *Micromachines* (2019) 10(12):862. doi:10.3390/mi10120862.
22. Rui Z, Qi-Chao L, Ping W, Zhong-Cheng L. Contact angle hysteresis in electrowetting on dielectric. *Chin Phys B* (2015) 24(8):086801. doi:10.1088/1674-1056/24/8/086801.
23. Yi Z, Shui L, Wang L, Jin M, Hayes RA, Zhou G. A novel driver for active matrix electrowetting displays. *Displays* (2015) 37:86–93. doi:10.1016/j.displa.2014.09.004.
24. Chiu YH, Liang CC, Chen YC, Lee WY, Chen HY, Wu SH. Accurate-gray-level and quick-response driving methods for high-performance electrowetting displays. *J Soc Inf Disp* (2011) 19(11):741–8. doi:10.1889/JSID19.11.741.
25. Luo ZJ, Zhang WN, Liu LW, Xie ST, Zhou GF. Portable multi-gray scale video playing scheme for high-performance electrowetting displays. *J Soc Inf Disp* (2016) 24(6):345–54. doi:10.1002/jsid.444.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 09 December 2020
doi: 10.3389/fphy.2020.618917

A Pointer Type Instrument Intelligent Reading System Design Based on Convolutional Neural Networks

Yue Lin, Qinghua Zhong and Hailing Sun*

Guangdong Provincial Key Laboratory of Optical Information Materials and Technology and Institute of Electronic Paper Displays, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China

The pointer instrument has the advantages of being simple, reliable, stable, easy to maintain, having strong anti-interference properties, and so on, which has long occupied the main position of electrical and electric instruments. Though the pointer instrument structure is simple, it is not convenient for real-time reading of measurements. In this paper, a RK3399 microcomputer was used for real-time intelligent reading of a pointer instrument using a camera. Firstly, a histogram normalization transform algorithm was used to optimize the brightness and enhance the contrast of images; then, the feature recognition algorithm You Only Look Once 3rd (YOLOv3) was used to detect and capture the panel area in images; and Convolutional Neural Networks were used to read and predict the characteristic images. Finally, predicted results were uploaded to a server. The system realized automatic identification, numerical reading, an intelligent online reading of pointer data, which has high feasibility and practical value. The experimental results show that the recognition rate of this system was 98.71% and the reading accuracy was 97.42%. What is more, the system can accurately locate the pointer-type instrument area and read corresponding values with simple operating conditions. This achievement meets the demand of real-time readings for analog instruments.

Keywords: pointer instrument, deep learning, You Only Look Once 3rd, convolutional neural networks, automatic reading

OPEN ACCESS

Edited by:
Chongfu Zhang,
University of Electronic Science and Technology of China, China

Reviewed by:
Junsheng Mu,
Beijing University of Posts and Telecommunications (BUPT), China
Yuanquan Hong,
Shaoguan University, China

****Correspondence:***
Qinghua Zhong
zhongqh@163.com

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *19 October 2020*
Accepted: *12 November 2020*
Published: *09 December 2020*

Citation:
Lin Y, Zhong Q and Sun H (2020) A Pointer Type Instrument Intelligent Reading System Design Based on Convolutional Neural Networks. Front. Phys. 8:618917. doi: 10.3389/fphy.2020.618917

INTRODUCTION

The pointer instrument has been widely used in traditional industries, such as industrial production and power transmission because of its properties of being simple, reliable, stable, less affected by temperature, strong anti-interference, easy to maintain, and so on, but it does not reserve any digital interface. Therefore, most of them must be read by humans. This method has disadvantages of low accuracy, poor reliability and low efficiency. So, the pointer position of the instrument should be converted into digital signals by sensors to realize automatic meter reading, which is of great significance in the application of unattended substation [1].

In order to solve this problem, many automatic recognition numerical reading algorithms based on computer readings have emerged in recent years. The existing pointer instrument recognition algorithms can be divided into two kinds, traditional algorithms based on digital image processing technology and modern algorithms based on machine learning or deep learning [2]. Traditional algorithms are the foundation of dilation and corrosion, and noise reduction

filtering and feature matching of image matrixes must be done to perform target recognition. The common traditional algorithms included a binarization threshold segmentation method based on symmetry for dial recognition, and an improved random sample consensus algorithm for pointer reading recognition [3], but this algorithm had high requirements for picture resolution and relatively high definition. Then, a method of accurately positioning the pointer by the Circle-based Regional Cumulative Histogram method was proposed [4]. The adaptability of this method for complex scenes was poor, and the recognition rate was not high. In addition, a visual inspection method was used to detect the transformed image to get the reading of the pointer meter [5]. And a regional growth method was used to locate the dial area and its center, then the improved center projection method was used to perform scale marking and boundary detection by the dial image [6]. Such algorithms based on computer vision and machine learning have poor portability and general versatility. Modern algorithms are machine learning techniques which are used to build and simulate neural networks for analytical learning. From a statistical point of view, this method can predict the distribution of data, learn a model from the data and then use the model to predict new data. So, we can carry out data classification without too much related knowledge [7, 8]. Then, the Mask-RCNN was used to divide the meter dial and the pointer area, the line fitting and angle reading method was used to calculate the pointer reading [9]. The requirement of the instrument placement position in these algorithms was relatively high. At the same time, Faster R-CNN was used to detect the position of the target instrument, and reading by feature correspondence and perspective transformation [10]. However, it had a high requirement for system operation performance. Mask R-CNN was also used to detect key points of tick marks and pointers, then used the intersection of circle and straight line to calculate the reading [11], but the algorithm was complex, and computationally intensive.

Therefore, we proposed a pointer type instrument intelligent reading system based on CNN. The system used a histogram normalization algorithm for image brightness optimization and contrast enhancement, and then YOLOv3 feature recognition algorithm was used to detect a panel area in an image and the corresponding area was saved as a feature image. Then, a multilayer convolution neural network was introduced to read the numerical prediction. The system mainly includes four algorithms: histogram normalization transformation, dial gray-scale transformation, YOLOv3 feature recognition, and convolutional neural network reading prediction, which can accurately locate the panel area and read corresponding numerical values.

SYSTEM DESIGN PRINCIPLE

Histogram Normalization Transformation

The electrician and electric pointer meters are often placed in some special rooms, and the surrounding light environment is complicated so that the quality of the surveillance images is low contrast and has a poor image clarity. Since the gray value distribution of image pixels conforms to the law of probability and statistics distribution, the image is preprocessed by the histogram normalization transformation to realize the balanced distribution of image gray levels for improving image contrast and optimizing image brightness [12]. The row of the input image I is r and the column is c. And $I(r, c)$ is the gray value of input image I in row r and column c, $O(r, c)$ is the gray value of output image O in row r and column c. The minimum gray value of I is I_{min} and the maximum value is I_{max}. *The minimum gray value of O is O_{min} and the maximum value is O_{max}. Then relationship among them is shown in* ***Eq. (1)***.

$$\frac{O(r,c) - O_{min}}{O_{max} - O_{min}} = \frac{I(r,c) - I_{min}}{I_{max} - I_{min}} \tag{1}$$

And it can be converted to **Eq. (2)**.

$$O(r,c) = \frac{O_{max} - O_{min}}{I_{max} - I_{min}} * I(r,c) + O_{min} - \frac{O_{max} - O_{min}}{I_{max} - I_{min}} * I_{min} \tag{2}$$

The process is called histogram normalization transformation. At the same time, $O(r, c)$ can be calculated as **Eq. (3)**, α and β are weight and bias variate.

$$O(r,c) = \alpha * I(r,c) + \beta \tag{3}$$

Therefore, **Eq. (4)** can be obtained.

$$\alpha = \frac{O_{max} - O_{min}}{I_{max} - I_{min}}, \beta = O_{min} - \frac{O_{max} - O_{min}}{I_{max} - I_{min}} * I_{min} \tag{4}$$

As a result, histogram normalization transformation is a linear transformation method to automatically select α and β. General order $O_{min} = 0$, $O_{max} = 255$, then **Eq. (4)** can be converted to **Eq. (5)**.

$$\alpha = \frac{255}{I_{max} - I_{min}}, \beta = -\frac{255}{I_{max} - I_{min}} * I_{min} \tag{5}$$

Dial Gray-Scale Transformation

Most of the panel area of pointer meters is a relatively simple color, either black or white. In order to reduce the amount of computation, image data can be transformed into grayscale. According to the importance of three primary colors, and the sensitivity of human eyes to different colors, the three components of color are given weighted averages by different weights [13], and it can be carried out by **Eq. (6)** to obtain a grayscale image, where i and j represent coordinates of horizontal and vertical of images, $R(i, j)$, $G(i, j)$ and $B(i, j)$ respectively represent components of a points in row i and column j of three primary colors.

$$f(i,j) = 0.299 * R(i,j) + 0.587 * G(i,j) + 0.114 * B(i,j) \tag{6}$$

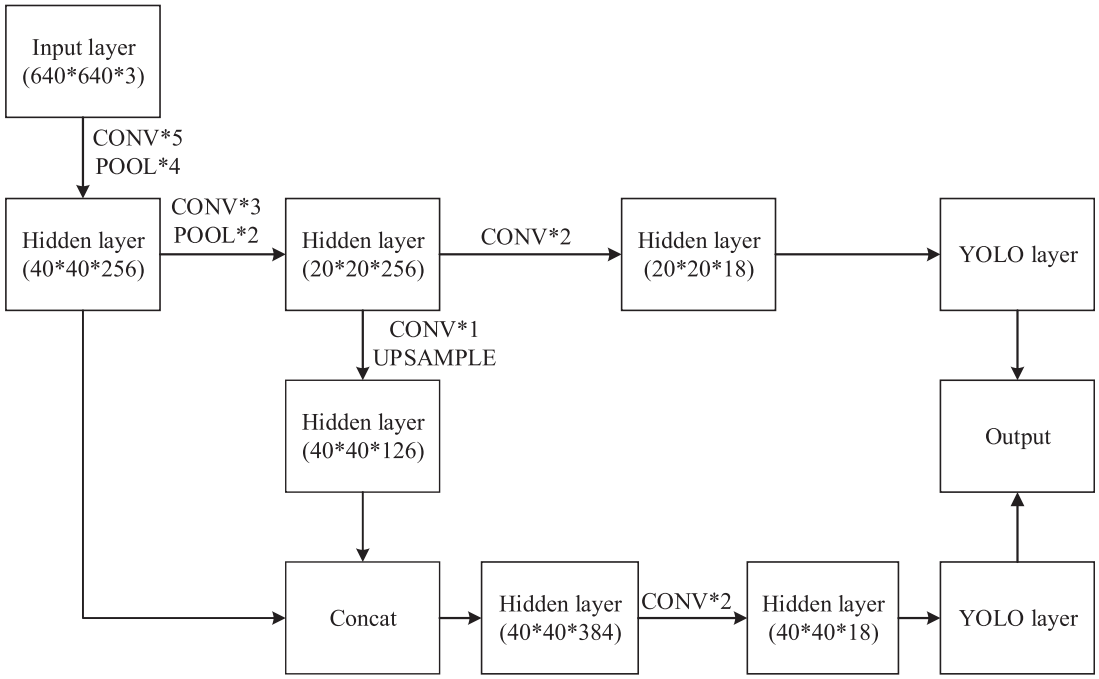

FIGURE 1 | YOLOv3 Tiny Network structure diagram. CONV is a convolution operation, POOL is a pooling operation, and UPSAMPLE is an up sample operation.

You Only Look Once 3rd Feature Recognition

There are two common features in recognition algorithms. One involves candidate regions, and then the region of interest (ROI) is classified, and location coordinates are predicted. This kind of algorithm is called two-stage feature recognition algorithm [14]. Another is one-stage detection algorithm, which needs one network to generate the ROI and predict the category, such as the YOLOv3 feature recognition algorithm [15]. Compared with the two-stage feature recognition algorithm, YOLOv3 uses a single network structure to predict object category and location for generating candidate regions, and each real box of YOLOv3 only corresponds to a correct candidate area [16]. These features provide YOLOv3 with less computation and a faster detection speed, which is more suitable for porting to an embedded computing platform due to weak computing performance.

A standard network structure of YOLOv3 has 107 convolutional layers. The first 74 are based on the Darknet-53 network layer and serve as the main network structure. The 75th to 107th layers are the feature interaction layers, which realizes local feature interaction by means of convolution kernel [17]. Because the target operating platform is an embedded platform, and recognition characteristics of pointer meters are relatively obvious, a simplified version of YOLO network structure YOLOv3 Tiny was used as a network structure by us. Its network structure is shown in **Figure 1**, CONV was a convolution operation, POOL was a pooling operation, and UPSAMPLE was an up sample operation. To reduce computation, the input layer used color images which was resized to a width of 640 pixels and height of 640 pixels, and the YOLO Layer was used as the output layer. Although YOLOv3 Tiny network only retained 24 convolutional layers, it still retained two YOLO network layers, which greatly reduced the amount of computation and still ensured the accuracy of model recognition [18, 19].

Convolutional Neural Network Reading Prediction

As a feedforward neural network, the convolutional neural network has excellent performance in large-scale image processing and it has been widely used in image classification and positioning. Panel image data cannot be linearly classified. To deal with this kind of data, we proposed a multi-layer convolutional neural network, which could perform the classification by mapping the original data to a linearly separable high-dimensional space, and then a specific linear classifier was used [20, 21]. A three-layer neural network model including an input layer, a hidden layer and an output layer was used to train the image data of panel areas by numerical reading [22], as shown in **Eq. (7)**.

$$f_i(x) = act\left(W_i^T x + b_i\right) \tag{7}$$

If input x has m nodes and output $f(x)$ has n nodes, then weight vector W is a matrix with n rows and m columns. Input x is a vector with length m, bias vector b is a vector with length n,

TABLE 1 | | Architecture of proposed CNN algorithm.

Layer	Filters	Size	Output
CONV0	16	9 × 9	80 × 80 × 16
POOL0	—	2 × 2	40 × 40 × 16
CONV1	32	9 × 9	40 × 40 × 32
POOL1	—	2 × 2	20 × 20 × 32
CONV2	64	9 × 9	20 × 20 × 64
POOL2	—	2 × 2	10 × 10 × 64
FC0	1,024	—	1,024
FC1	26	—	26

act is an activation function, and $f(x)$ returns an n-dimensional vector.

The input layer to hidden layer is a fully connected network, each input node is connected to all hidden layer nodes. So, each hidden layer node is equivalent to a neuron, then the output of all nodes of the hidden layer becomes a vector. If the input layer is a vector x, and hidden layer node h has a weight vector W_h, a bias vector b_h, and tanh function is used for the activation function, the output expression of the hidden layer node is shown as **Eq. (8)**.

$$f_h(x) = tanh\left(W_h^T x + b_h\right) = \frac{e^{W_h^T x + b_h} - e^{-(W_h^T x + b_h)}}{e^{W_h^T x + b_h} + e^{-(W_h^T x + b_h)}} \tag{8}$$

The hidden layer to output layer is also a fully connected network, which can be held as a cascade of neurons on the hidden layer. Since it is multi-classified, activation function can use Softmax Regression function, and output formula in output layer is shown in **Eq. (9)**.

$$f_o(x) = softmax\left(W_o^T x + b_o\right) \tag{9}$$

The network structure of CNN used by us is shown in **Table 1**, where CONV was a convolutional layer, POOL was a pooling layer, and FC was fully connected layers. Input image was a grayscale image with a size of 80 × 80 and there were 26 types of output result.

EXPERIMENT AND DISCUSSION

The intelligent reading system consists of two parts: model training and model invocation. Before training, histogram normalization and grayscale transformation were carried out on an image. Then, YOLOv3 feature recognition algorithm was used to train the image to obtain a model of instrument panel. Then, a multi-layer neural network model was built, and numerical reading training was carried out for the instrument panel to obtain a respective model. An RK3399 microcomputer was used to get images by a camera in real time, a histogram normalization algorithm was applied to transform the image brightness optimization and enhance contrast preprocessing. The instrument panel model is called to detect panel area in an image, and extract corresponding regional image, the model of multi-layer neural network model was used for features of an image to read numerical predictions. Finally, we uploaded data by WIFI for subsequent operations. The specific process is shown in **Figure 2**.

System Hardware Architecture

The instrument intelligent reading system designed in this paper was composed of a host and a camera. A microcomputer of an RK3399 hardware platform, which was made by Friendly Arm, was used as the host. The RK3399 processor was a six-core 64-bit CPU newly developed with "big.LITTLE" core architecture had two Cortex-A72 large kernels and four Cortex-A53 small kernels by Rockchip. The

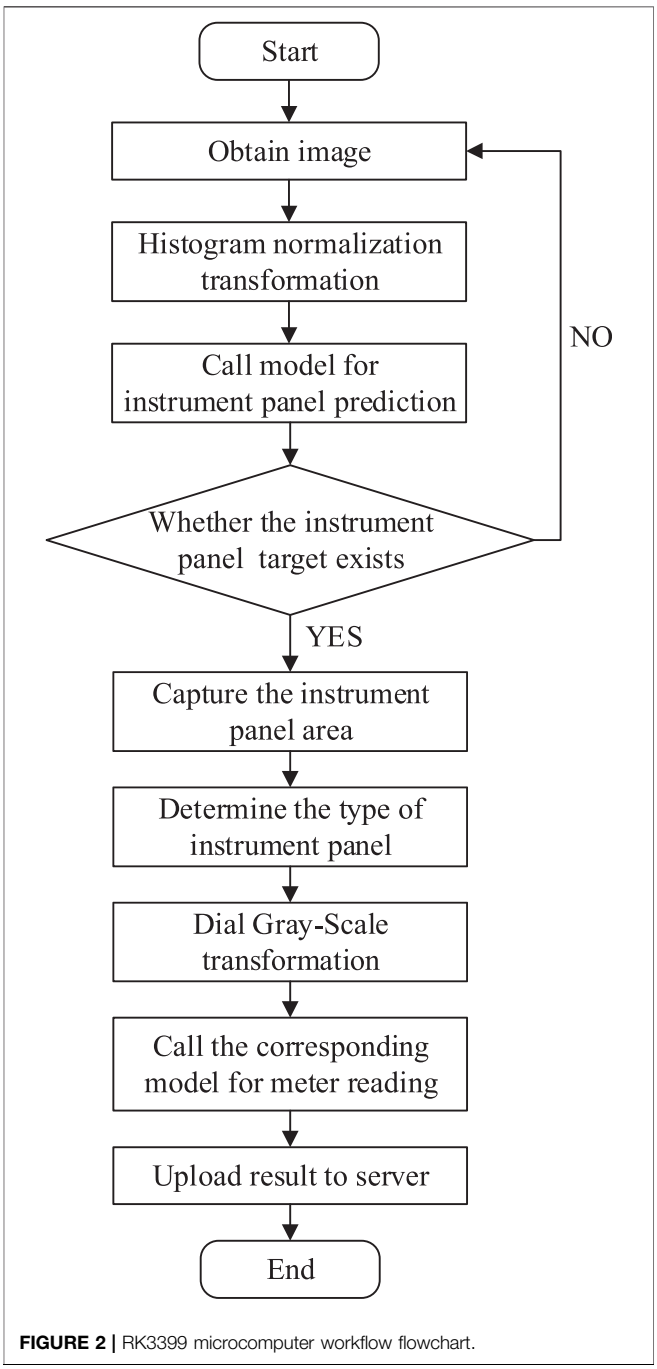

FIGURE 2 | RK3399 microcomputer workflow flowchart.

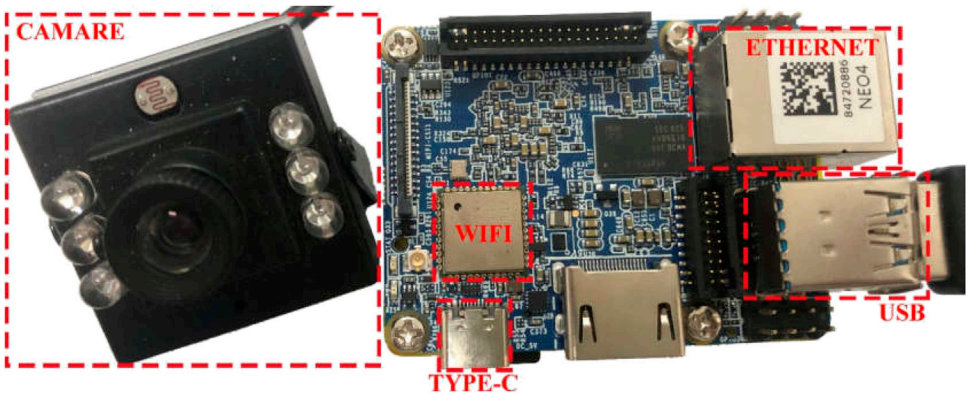

FIGURE 3 | Instruments in our intelligent reading system, in which a camera was connected to the host by a USB interface, the system was connected to the network by a WIFI module or an Ethernet network port, and the power was supplied by a Type-C interface.

TABLE 2 | | Comparison of the running time of the algorithm in different platforms.

CPU in platform	Time of taking photo (ms)	Time of histogram normalization transformation (ms)	Time of calling YOLOv3 model (ms)	Time of calling CNN model (ms)
RK3399	33.3	39.1	7,315	27.18
I5-7200U	33.3	7.9	2,164	12.01

Both had 4G RAM and run Ubuntu 18.04 system, and RK3399 main frequency was locked at 1.8 GHz, I5-7200U was locked at 2.5 GHz.

system had a computing performance close to a CPU whose main frequency is up to 1.8 GHz. It supported WIFI and USB peripherals, and could connect to the Internet and external cameras. The camera used a generic industrial-grade USB camera which had a Built-in Sony Complementary Metal Oxide Semiconductor image sensor. The physical object of the system is shown in **Figure 3**, a camera was connected to the host by an USB interface, the system was connected to the network by a WIFI module or an Ethernet network port, and the power was supplied by a Type-C interface.

Software Architecture

An RK3399 hardware platform was used as the hosting platform. An Ubuntu 18.04 operating system was installed on

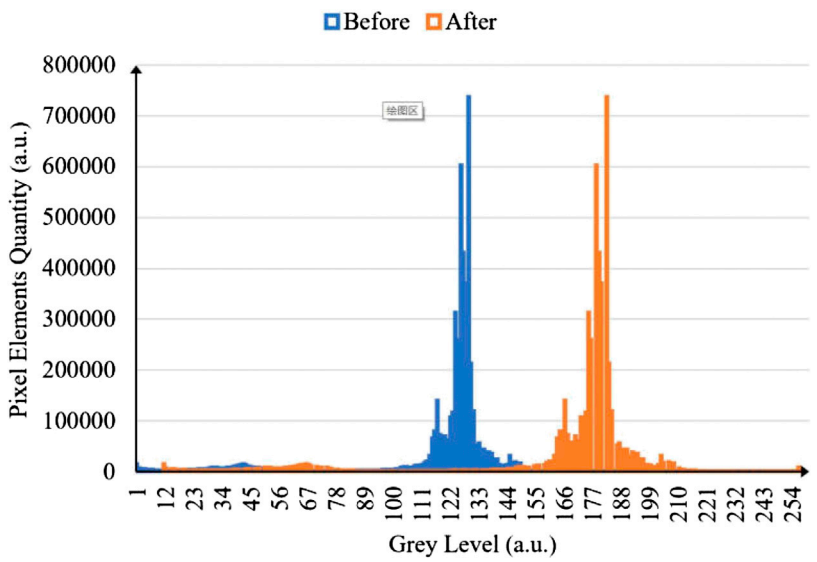

FIGURE 4 | Histogram comparison before and after normalized transformation, blue represents the gray level distribution of an image before the normalized histogram transformation, and yellow represents the transformed gray distribution

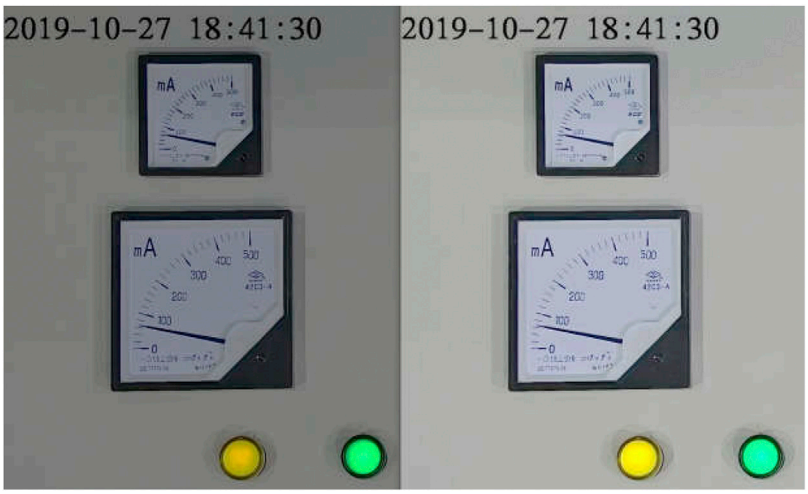

FIGURE 5 | Histogram normalization transformation effect comparison, the left is the pre-processing image, and the right is the finished effect image. The brightness of the image was obviously optimized, while the contrast was increased, which produced clearer details in the darker areas.

TABLE 3 | | The data comparison between our algorithms and others. Epoch was set to 2,000 and batch size to 8. The learning rate was set to 0.01.

Model	Operation time (s)	Accuracy rating (%)	Training time (h)
This paper	6.3	98.71	7.3
Faster-RCNN [10]	8.0	98.04	26.9
Mask-RCNN [23]	9.7	98.25	37.5

the hardware platform, and python3.6, tensorflow1.12 and opencv2.0 operating environments were deployed. **Table 2** is the comparison of the elapsed time. Both had 4G RAM and run Ubuntu 18.04 system, and RK3399 main frequency was locked at 1.8 GHz, I5-7200U was locked at 2.5 GHz. The time for taking photo means how long it requires to take a picture, the time of histogram normalization transformation means how long it can complete histogram normalization transformation, the time of calling YOLOv3 model means how long it takes call a YOLOv3 model, the time of calling CNN model means that how long it takes to call a CNN model. It can be seen that, although the computing performance of the RK3399 hardware platform was worse than that of the ordinary computer platform, it has satisfied the operational requirements of our project. Moreover, it had a high degree of integration and the volume was far smaller than that of ordinary computers, which has greatly improved the portability and flexibility of the system. In addition, it has more advantages in power consumption and cost.

In experiments, a total of 3,745 images were collected as the data set, the original size of images was 1,920 height and 1,080 width. 374 images were randomly chosen from the data, and they were s1 et as validation data, and remaining images were used as training data. In order to ensure the final training result, the data set included images with different tilt angles, different rotation angles, different lighting levels and different degrees of cover.

Histogram Normalization Transformation

Histogram normalization transformation was carried out to automatically adjust image brightness and enhance contrast, so as to make the regional features of pointer instruments more obvious, and the recognition accuracy of the target recognition algorithm was also improved. **Figure 4** is a comparison of image histograms before and after a normalized histogram transformation, in which the horizontal axis x represents the gray level and the vertical axis *Y* represents the number of pixel elements in each gray level, the unit is 10^5 pixels, blue represents the gray distribution of an image before the normalized histogram transformation, and yellow represents the transformed gray distribution. The image processing effect is shown in **Figure 5**. The left is a pre-processing image, and the right is a finished effect image. The brightness of the image was obviously optimized, while the contrast was increased, which produced clearer details in the darker areas.

Pointer Instrument Identification Training

To check the performance superiority of the proposed algorithm, two kinds of two-stage feature recognition algorithms, Faster R-CNN [10] and Mask RCNN [23], were transplanted to the RK3399 hardware platform. Faster R-CNN used Vgg16 as the feature extractor and Mask RCNN used Resnet-18 as the feature extractor. Compared with the algorithm in our system, the average value was calculated after the prediction of all data sets. In order to achieve better results with a shorter training time, we set the epoch to 2,000 and batch size to 8. The learning rate was set to 0.01. The actual prediction results are shown in **Table 3**. Compared with the Fast R-CNN algorithm of a Vgg16 network structure, the operation time used in this system was

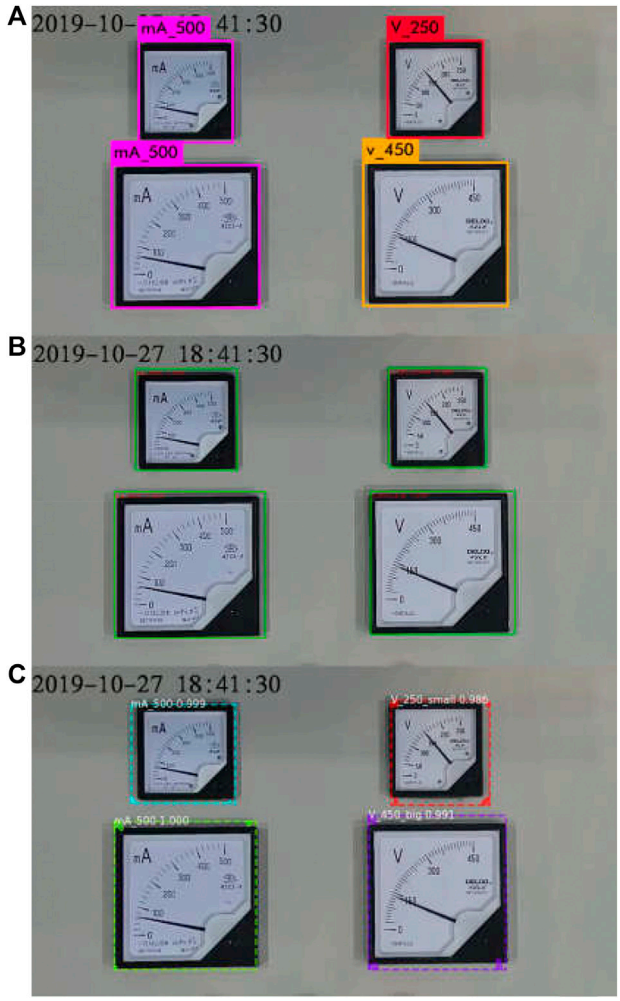

FIGURE 6 | The target region predicted results of the pointer instrument by the three algorithms. **(A)** The predicted result which used the algorithms by this paper. **(B)** The predicted result which used Faster R-CNN algorithms. **(C)** The predicted result which used Mask RCNN algorithms.

shortened by 1.7 s, the recognition accuracy was improved by 0.67%, and the time spent in model training was shortened by 19.6 h. Compared with the Mask RCNN algorithm with a resnet-18 network structure, the operation time was reduced by 3.4 s, the recognition accuracy was improved by 0.46%, and the time spent in model training was reduced by 30.2 h. Three algorithms were used to predict the target image, and the target region prediction results of pointer instrument are shown in **Figure 6**. The experimental results showed that compared with the Mask RCNN and Faster R-CNN algorithms, the proposed algorithm not only had a slight improvement in the recognition accuracy, but also had a great advantage in the call time of the model, which was more suitable for transplantation to an embedded platform. Moreover, it greatly shortened the length of time consumed by the training model.

Pointer Meter Reading Training

Compared with the traditional machine vision algorithm, the multi-layer neural network had the advantages of high universality, wide application range, and high prediction accuracy in the recognition application, especially in the scene with complex environmental illumination. We used the algorithm proposed by Lai et al [24] to predict the readings of the data set. In complex situations, and in some lights, the recognition accuracy rate was very low, when the images' angles were changed, either tilted, rotated around, or the microscopic angle changed, the identification accuracy could be reduced. As shown in **Figure 7**, the identification result is very inaccurate, or even unable to be identified. The image in **Figure7A** can only identify the scale in the clockwise direction, and the identification result of the scale value was unstable, as shown in **figure7B**. Because the dial area in **Figure7C** was fuzzy, only the pointer could be identified, the scale and scale value cannot be identified. As shown in **Figure7D**, because of the uneven dial scale values, special treatment was required. So, in the four cases, the algorithm in this paper can carry out

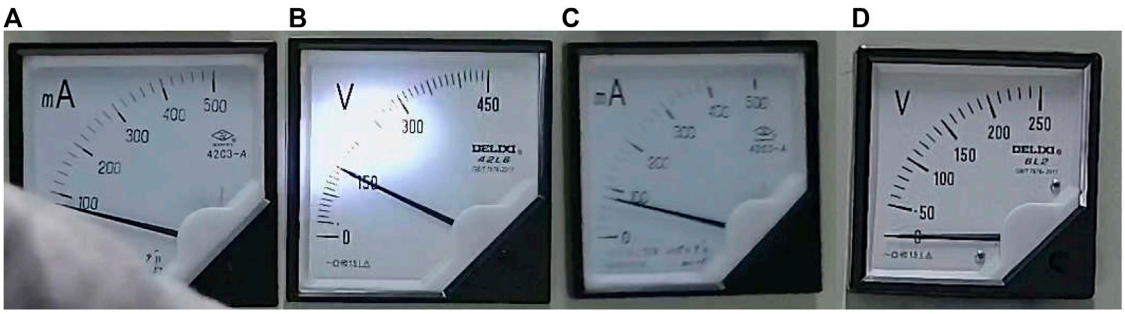

FIGURE 7 | The situation of inaccurate identification results. **(A)** It can only identify the scale in the clockwise direction, and the identification result of the scale value was unstable, while the contrary was the case in **(B)**, **(C)** was fuzzy, only the pointer can be identified, and the scale and scale value cannot be identified. **(D)** Because of the uneven dial scale values, special treatment was required.

FIGURE 8 | The reading result which exceeds the accuracy of a scale value. It could have read 70 or 80 V, but reading results showed that the voltage was 75 V

perfect recognition without interference. In particular, the trained model can even predict the reading result which exceeded the accuracy of a scale value when the data set was well processed. As shown in **Figure8**, it could have read 70 or 80 V, but reading results showed that the voltage was 75 V. **Table 4** is the comparison of the verification results of reading prediction for the data set between the algorithm used in this system and a traditional algorithm [24]. The accuracy rating is the average verification accuracy found by training all data sets. The operation time means the average time consumed by training all data sets. Although the algorithm used in this system was time-consuming, its accuracy rate was significantly increased, which greatly improved the reliability of system readings.

CONCLUSION

An instrumental intelligent reading system was designed in this paper, which had a recognition rate of 98.71% for pointer instrument panel and a reading accuracy rate of 97.42% for numerical values. It can accurately locate panel area and read corresponding values to meet the demand of real-time detection for pointer instruments. In addition, identification results can be uploaded to a server synchronously for monitoring online, which was convenient for real-time detection of on-site measurement problems. At the same time, it was convenient to collect and analyze historical data for system optimizations. Moreover, the system had simple operating conditions, compact size and convenient use. There was no need to modify the instrument itself, and it retained the advantages of pointer instruments, which has a high feasibility and a practical value.

TABLE 4 | | Comparison of verification results of the data set between the algorithm used in this system and a traditional algorithm [24].

Instrument panel	Accuracy rating (%)			Operation time (s)		
	250 V	450 V	500 mA	250 V	450 V	500 mA
CNN	97.06	97.98	97.22	2.93	2.87	2.91
Algorithm proposed by Lai et al. [24]	78.35	80.24	79.08	1.79	1.78	1.79

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YL and QZ designed this project. YL carried out most of the experiments and data analysis. HS contributed to the data analysis and correction. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by National Key Research and Development Program of China (2016YFB0401501), Science and Technology Program of Guangzhou (No. 2019050001), Program for Chang Jiang Scholars and Innovative Research Teams in Universities (No. IRT_17R40), Program for Guangdong Innovative and Enterpreneurial Teams (No. 2019BT02C241), Guangdong Provincial Key Laboratory of Optical Information Materials and Technology (No. 2017B030301007), Guangzhou Key Laboratory of Electronic Paper Displays Materials and Devices (No. 201705030007) and the 111 Project.

REFERENCES

1. Zhang T, Zhang S, Wang P. Automatic recognition of instrument dial readings with multiple reference points. In: IEEE 2017 Chinese Automation Congress. Jinan (2018).
2. Yi ZC, Chen ZB, Peng B, Li SX, Bai PF, Shui LL, et al. Vehicle lighting recognition system based on erosion algorithm and effective area separation in 5G vehicular communication networks. *IEEE Access* (2019) 7: 111074–83. doi:10.1109/ACCESS.2019.2927731
3. Ma YF, Jiang Q. A robust and high-precision automatic reading algorithm of pointer meters based on machine vision. *Meas Sci Technol* (2019) 30(1):21–7. doi:10.1088/1361-6501/aaed0a
4. Yang Z, Niu W, Peng X, Gao Y, Dai Y. An image-based intelligent system for pointer instrument reading. In: IEEE International Conference on Information Science and Technology. Shenzhen (2014).

5. Bao HJ, Tan QC, Liu ST, Miao JW. Computer vision measurement of pointer meter readings based on inverse perspective mapping. *Appl Sci Basel* (2019) 9(18):13–9. doi:10.3390/app9183729
6. Chi J, Liu L, Liu J, Jiang Z, Zhang G. Machine vision based automatic detection method of indicating values of a pointer gauge. *Math Probl Eng*. (2015) 12:1–19. doi:10.1155/2015/283629
7. Jorge N. Optimization methods for large-scale machine learning. *SIAM Rev* (2016)60(2):33–7. doi:10.1137/16M1080173
8. Zhang Q, Yang LT, Chen Z, Li P. A survey on deep learning for big data. *Inf Fusion* (2018) 42:146–57. doi:10.1016/j.inffus.2017.10.006
9. He P, Zuo L, Zhang C, Zhang Z. A value recognition algorithm for pointer meter based on improved mask-RCNN. In: International Conference on Information Science and technology. Hulunbuir (2019).
10. Liu Y, Liu J, Ke Y. A detection and recognition system of pointer meters in substations based on computer vision. *Measurement*. (2020) 152:122–6. doi:10.1016/j.measurement.2019.107333
11. Perez FJ, Signol F, Perez-Cortes JC, Fuster A, Pollan M, Perez B, et al. A deep learning system to obtain the optimal parameters for A threshold-based breast and dense tissue segmentation. *Comput Methods Progr Biomed* (2020) 14:195. doi:10.1016/j.cmpb.2020.105668
12. Yadav J, Rajpal N, Mehta R. An improved illumination normalization and robust feature extraction technique for face recognition under varying illuminations. *Arab J Sci Eng* (2019) 44(11):9067–86. doi:10.1007/s13369-019-03729-6
13. Wen K, Li D, Zhao X, Fan A, Mao Y, Zheng S. Lightning arrester monitor pointer meter and digits reading recognition based on image processing. In: Proceedings of 2018 IEEE 3rd advanced Information technology, electronic and automation Control Conference. Chongqing (2018).
14. Ahmed B, Gulliver TA, Saif AS. Image splicing detection using mask-rcnn image splicing detection using mask-rcnn. *Signal Image Video Process* (2020) 7: 7. doi:10.1007/s11760-020-01636-0
15. Liu G, Nouaze JC, Touko PL, Kim JH. YOLO-tomato: a robust algorithm for tomato detection based on YOLOv3. *Sensors* (2020) 20(7):27. doi:10.3390/s20072145
16. Tian Y, Yang G, Wang Z, Li E, Liang Z. Detection of apple lesions in orchards based on deep learning methods of CycleGAN and YOLOV3-dense. *J Sensors* (2019) 17:121–7. doi:10.1155/2019/7630926
17. Redmon J, Farhadi A. *YOLOv3: an incremental improvement*. Arxiv (2018).
18. Fu L, Feng Y, Wu J, Liu Z, Gao F, Majeed Y, Al-Mallahi A. Fast and accurate detection of kiwifruit in orchard using improved Yolov3-tiny model. *Precis Agric* (2020) 7:144–65. doi:10.1007/s11119-020-09754-y
19. Zhang Y, Shen YL, Zhang J. An improved tiny-Yolov3 pedestrian detection algorithm. *OPTIK Int J Light Electron Opt* (2019) 183:17–23. doi:10.1016/j.ijleo.2019.02.038
20. Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. *Adv Neural Inform Process Syst* (2012) 25(2):22–39. doi:10.1145/3065386
21. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imag* (2016) 35(5): 1285–98. doi:10.1109/TMI.2016.2528162
22. Liu W, Wang Z, Liu X, Zeng N, Lin Y, Alsaadi FE. A survey of deep neural network architectures and their applications. *Neurocomputing* (2017) 234: 11–26. doi:10.1016/j.neucom.2016.12.038
23. Zuo L, He P, Zhang C, Zhang ZA. Mask RCNN based automatic reading method for pointer meter. *Neurocomputing* (2020) 388:90–101. doi:10.1016/j.neucom.2020.01.032
24. Lai HW, Kang Q, Pan L, Cui C. A novel scale recognition method for pointer meters adapted to different types and shapes. In: IEEE international conference on automation science and engineering. Vancouver (2019).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 10 December 2020
doi: 10.3389/fphy.2020.618752

Design of an AC Driving Waveform Based on Characteristics of Electrowetting Stability for Electrowetting Displays

*Linwei Liu[1], Zhuoyu Wu[1], Li Wang[2], Taiyuan Zhang[1], Wei Li[1], Shufa Lai[1] and Pengfei Bai[1]**

[1]Guangdong Provincial Key Laboratory of Optical Information Materials and Technology and Institute of Electronic Paper Displays, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China, [2]School of Information Engineering, Zhongshan Polytechnic, Zhongshan, China

In traditional electrowetting display (EWD) drivers, direct current (DC) voltage and pulse width modulation are often used, which easily caused an electrowetting charge trapping phenomenon in a hydrophobic insulating layer. Therefore, the driving voltage must be increased for driving EWDs, and oil backflow cannot be solved. Aqueous solutions are often used as polar liquids for EWDs, and the reverse voltage of alternating current (AC) driving can cause chemical reactions between water and indium tin oxide (ITO). So, a driving waveform was proposed, which included a DC waveform and an AC waveform, to separately drive EWDs for oil rupture and open state. Firstly, a DC waveform was used when the oil was broken, and the response time was reduced by designing the DC voltage and duration. Secondly, an AC waveform was used when the oil required to be stable. Oil backflow could be suppressed by the AC waveform. The main parameters of AC waveform include reverse voltage, frequency and duty cycle. The reverse voltage of EWDs could be obtained by voltammetry. The frequency could be obtained by analyzing the rising and falling edges of the capacitance voltage curve. The experimental results showed that the proposed waveform can effectively suppress oil backflow and shorten the response time. The response time was about 86% lower than the conventional driving waveforms, and oil backflow was about 72% slower than the DC driving waveform.

Keywords: electrowetting display, low energy consumption, oil backflow, response time, electrochemical reaction, voltammetry

OPEN ACCESS

Edited by:
Chongfu Zhang,
University of Electronic Science and Technology of China, China

Reviewed by:
Jitesh Barman,
Banaras Hindu University, India
Lelun Jiang,
Sun Yat-sen University, China
Zontao Li,
South China University of Technology, China

****Correspondence:***
Pengfei Bai
baipf@scnu.edu.cn

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *18 October 2020*
Accepted: *18 November 2020*
Published: *10 December 2020*

Citation:
Liu L, Wu Z, Wang L, Zhang T, Li W, Lai S and Bai P (2020) Design of an AC Driving Waveform Based on Characteristics of Electrowetting Stability for Electrowetting Displays. Front. Phys. 8:618752. doi: 10.3389/fphy.2020.618752

INTRODUCTION

With the rapid development of display technology in the world, liquid crystal display (LCD) and electrophoretic display (EPD) have become mainstream display technologies [1–3]. Electrowetting display (EWD) technology is an emerging technology to enhance the user experience of electronic paper [4, 5]. It has the characteristics of fast response time, low energy consumption and so on, which can meet the requirements of high energy saving in current network terminal devices [5–7].

At present, driving chips for EPDs is used for driving EWDs, and PWM modulation is still the main control scheme for displaying grayscales in EWDs [8–10]. Aqueous solutions are often used as polar liquids for EWDs, and the chemical reactions between water and ITO is caused by a reverse voltage. So, EWDs are often damaged. The reverse voltage is not designed according to the structure and material of EWDs, resulting in a short life of EWDs. By using an unipolar PWM waveform, the

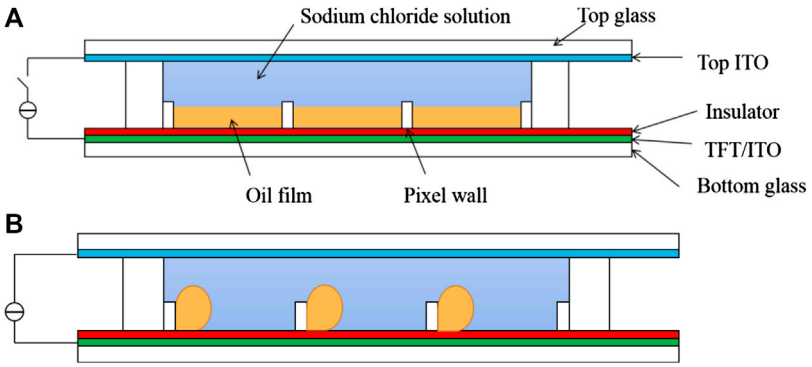

FIGURE 1 | The switching mechanism and the structure of an EWD. **(A)** The pixel is turned off when no voltage is applied, the oil is tiled in the pixel, and the color of the oil is displayed. **(B)** The pixel is turned on when voltage is applied, the oil is pushed to a corner, and the reflected light is the color of the pixel substrate and the color of a small portion of the oil.

backflow can be suspended, but the response time is extended [11]. Therefore, a driving waveform which can suppress backflow and shorten response time must be designed for EWDs.

The EWD technology belongs to the micro-flow control technology, which makes it difficult to accurately control the pixel aperture in EWDs [12–14]. The driving voltage and characteristics of EWDs are closely related to the process and materials, the contrast and color gamut are affected by the concentration of dye [15–18]. The main purpose of driving waveform for EWDs is to control the pixel aperture and oil movement accurately [19]. Because of the hysteresis effect, which can make charge capture and oil backflow, the driving waveform cannot be designed with EPD driver chips according to the characteristics of EWDs [20, 21]. Therefore, an AC-common driving waveform was proposed for more gray scales and a fast response speed [22]. The oil splitting problem was improved with an AC-common driving waveform. And the oil backflow effect was suppressed effectively. AC-common driving waveform provided a way to continuously improve the response time of EWDs. Then, an active thin film transistor (TFT) EWD driving system was proposed based on EPD ICs [23]. PWM waveform was used in the driving system to display images and videos. The driving system used an unipolar PWM waveform modulation to realize the display of gray scales in EWDs. A reset frame was designed to release the trapped charge and inhibit oil backflow, and the pixel can remain stable in an open state for a longer time. In 2016, a platform was proposed to display 16 gray scales dynamic video based on a TFT EWD [24]. The driving waveform was constructed by seven sub-frames and a dynamic reset frame. The trapped charge was released by a dynamic reset frame. In 2019, a frequency-amplitude mixed modulation driving system was proposed to improve the response speed for driving gray-scale and enhancing oil stability based on an active TFT EWD [8]. The oil was pushed by a high voltage to close to a target reflectivity, and then, the driving voltage was decreased to stabilize the oil to a target reflectivity. The response time was improved, but the second stage cannot solve the problem of oil backflow.

So, we proposed a mixed driving waveform which included a direct current (DC) driving waveform and an alternating current (AC) driving waveform. And a voltammetry methodology was proposed for the reverse voltage of AC driving waveform. The response time and oil stability were improved by the DC driving waveform. The oil backflow was optimized by the parameter setting of the AC driving waveform. Then, the driving waveform can realize a shorter response time and a slower backflow by adjusting duty cycle and reverse voltage.

PRINCIPLE OF EWDS

The structure of EWDs is mainly composed of upper and lower substrates, polar liquid, pixel wall, oil, sealant and insulator [25, 26]. The EWD device was reported by Hayes and Feenstra in 2003 [27]. The switching mechanism and structure of EWDs are shown in **Figure 1**. Indium tin oxide (ITO) glass is generally used as the common electrode on upper substrate. The lower substrate can be made by ITO or TFT. ITO is generally used to control fewer pixels, such as segment display. TFT glass is used to control more pixels, such as dot matrix displays. Polar liquids are usually aqueous solutions, such as sodium chloride solutions. Decane (C10H22) is used as oil. The color oil is made by dissolving the colored dye in decane (C10H22, ε_r = 2.2). The insulator material is Teflon af1600 and FC-43. The pixel wall material is transparent polyimide.

The driving mechanism of EWDs is to control contact angle between the insulator and oil by applying voltage. The reflectivity of EWDs is determined by the contact angle. In 1805, Young proposed an equation about interfacial tension and equilibrium contact angle, as shown in **Eq. 1**.

$$\gamma_{LG}\cos\theta_0 = \gamma_{SG} - \gamma_{SL} \tag{1}$$

Where, γ_{sg} is surface tension of solid-gas, γ_{sl} is surface tension of solid-liquid, γ_{lg} is surface tension of liquid-gas, θ_0 is the initial contact angle of the droplet.

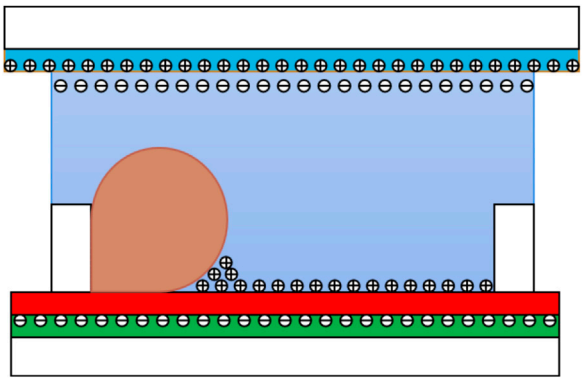

FIGURE 2 | Charge distribution in electric field of a EWD pixel. When voltage is applied, ions squeeze the oil with the electric force, ions in the aqueous solution can push the oil to a corner.

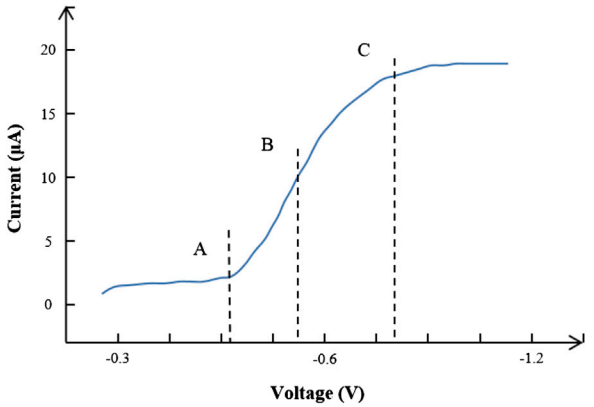

FIGURE 3 | The polarographic diagram. The curve represents the change of the current with the voltage changes, and the analysis of the current can be used to determine whether a chemical reaction occurs. The voltage at point A is the starting voltage for chemical reactions.

The relationship between interfacial tensions and electromotive force V is described by Lippmann equation, as shown in **Eq. 2**.

$$\gamma_{sl}(V) = \gamma_{sl} - \frac{\varepsilon_0 \varepsilon_r V^2}{2d} \tag{2}$$

Where, $\gamma_{sl}(V)$ and γ_{sl} are surface tensions of the solid-liquid interface when the voltage is applied and when the voltage is canceled. ε_0 and ε_r are dielectric constant in vacuum and the effective dielectric constant of the dielectric layer, d is the thickness of the dielectric layer.

The relationship between the contact angle and electromotive force V can be derived from **Eqs 1** and **2**, as shown in **Eq. 3**.

$$\cos\theta = \cos\theta_0 + \frac{\varepsilon_0 \varepsilon_r V^2}{2d\gamma_{LV}} \tag{3}$$

V can directly control the contact angle θ. With an electric field, the water can move to the insulator in EWDs. The oil film tends to rupture at the thinnest area where the electric field is the highest in a pixel. When water contacts the insulator, three-phase contact line and contact angles are formed. The relationship between contact angle of aqueous solution and applying voltage can be described by **Eq. 3**. With the applying voltage, the charge distribution and oil status are shown in **Figure 2**. When the electromotive force V is applied, the oil film is broken and driven to a corner by a force from aqueous solution. The force of the aqueous solution can be obtained by the Korteweg-Helmholtz equation, as shown in **Eq. 4**

$$f_e = \sigma \mathrm{E} - \frac{\varepsilon_0}{2}\mathrm{E}^2\nabla\varepsilon + \nabla\left[\frac{\varepsilon_0}{2}\mathrm{E}^2\frac{\partial\varepsilon}{\partial\rho}\rho\right] \tag{4}$$

Where, σ is the density of free charges, E is the intensity of electric field, ε_0 is a dielectric constant, ρ is the density of liquid, and the second term is a mass dynamic, the third term is an electrostrictive force.

The number of trapped charge in the insulator directly affects the backflow. When voltage is applied to the pixel, positive ions are gathered at the three-phase contact line, as shown in **Figure 2**. The electric field near the three-phase contact line will be distorted. The imbalance between Maxwell pressure and Laplace pressure at the three-phase contact line will cause backflow.

DESIGN PRINCIPLE OF ALTERNATING CURRENT WAVEFORM

The capacitance of pixels have been widely used in EWDs, such as oil distribution, design of driving waveform, threshold voltage. The driving waveform is a superposition of a staircase wave and a square wave. The AC driving frequency was obtained by measuring EWD capacitance voltage (CV) Curve. When the

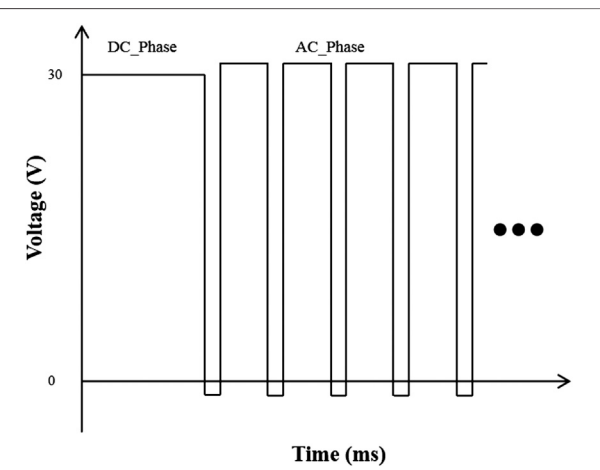

FIGURE 4 | The mixed DC and AC waveform. In DC phase, the main parameters that affect the oil movement are the voltage value and driving time. In AC phase, the main parameters that affect the oil backflow are the reverse voltage, duty cycle and frequency.

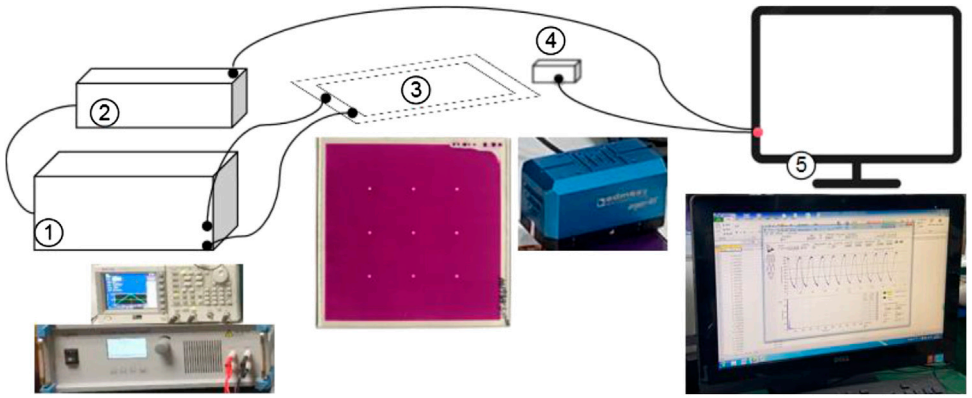

FIGURE 5 | Experimental platform for EWDs. ① High voltage amplifier. ② waveform generator. ③ EWD. ④ Admesy colormeter. ⑤ Computer.

TABLE 1 | Specifications of EWDs.

Panel size (cm^2)	Oil color	Resolution	Pixel size (μm^2)	Pixel wall height (μm)	Dielectric (μm)	Hydrophobic layer thickness (μm)
3.5 × 3.5	Magenta	50 × 50	300 × 300	5	1	1

rising edge curve and the falling edge curve were partly coincided, we used this frequency as a driving frequency to design the driving waveform. In the initial stage of oil rupturing, we used different driving time in a DC waveform to drive a pixel for reaching an aperture ratio, and then switch to a AC driving waveform. Reverse voltage is the key point of waveform parameters in AC driving waveform. Aqueous solution is used as the polar liquid of EWDs. The upper substrate electrode is made of ITO. The electrochemical reaction between water and ITO material is produced by the reverse voltage, the electrochemical reaction equation of ITO and water is shown in **Eq. 5**.

$$SnO_2 + 2H_2O + 4e = Sn + 4OH(-) \quad (5)$$

The oxidation-reduction reaction needs to reach half-wave potential for displaying in EWDs. So, the reverse voltage was designed lower than the voltage of the oxidation-reduction reaction to avoid damage to the EWD.

The relationship between the driving voltage and current was measured by voltammetry [28], the potential without oxidation-reduction reaction and half-wave potential can be obtained [29]. The half-wave potential is shown in **Eq. 6**. **Figure 3** shows the polarographic diagram. Point A is the potential without oxidation-reduction reaction, point B is the half-wave

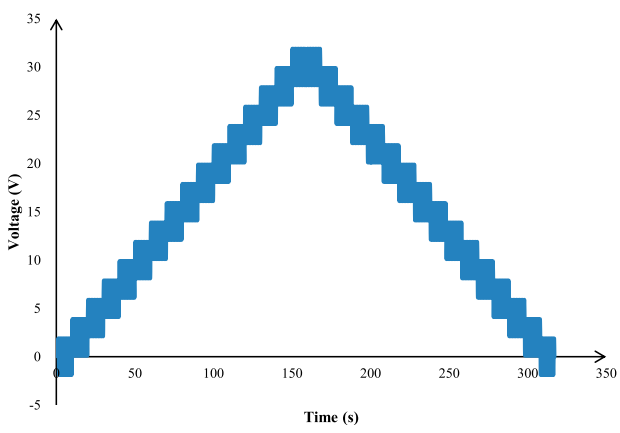

FIGURE 6 | The waveform for measuring the hysteresis curve, the frequency of AC signal was 500 Hz, the voltage peak to peak was 1 V, step voltage was 2 V, step frequency was 0.2 Hz.

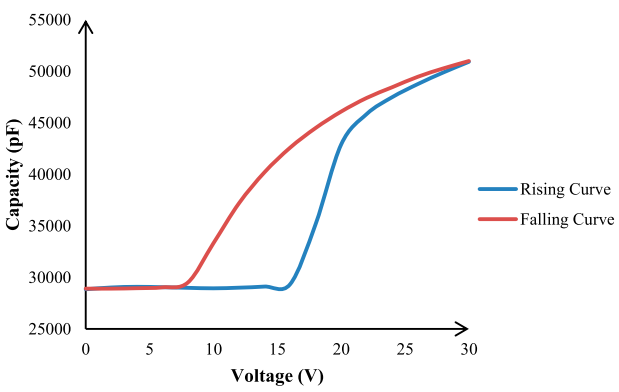

FIGURE 7 | EWDs CV curve. In the rising curve, the threshold voltage was about 16 V. In the falling curve, the switch off voltage was about 7 V. The CV curve can also represent the curve of aperture and reflectivity. When the *x*-axis was less than 7, the rising edge curve was coincided with the falling edge curve.

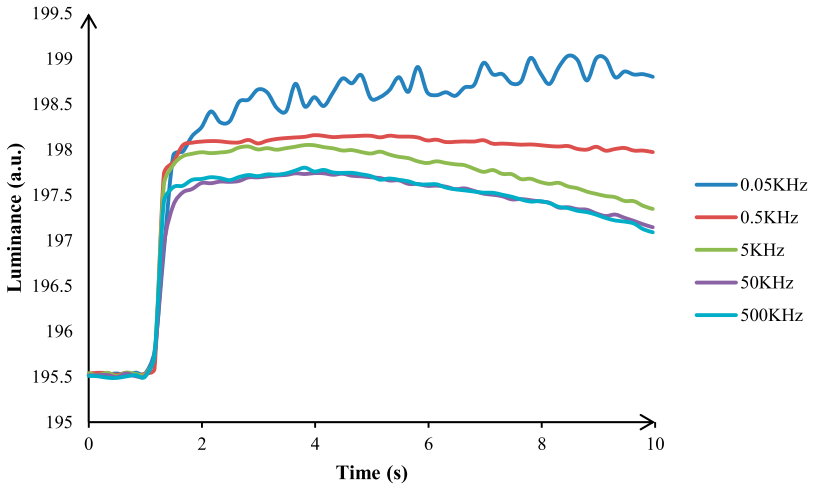

FIGURE 8 | Oil backflow with different driving frequencies. The oil could maintain stability at the frequency of 0.05 KHz. The driving waveform at frequency of 0.5 KHz had a good backflow phenomenon, and the aperture could not be increased. The driving waveform at frequency of 5 KHz had a serious backflow phenomenon. The driving waveform at frequency of 50 KHz had a backflow phenomenon and could reduce the aperture. The driving waveform at frequency of 500 KHz had the same driving effect as 50 KHz.

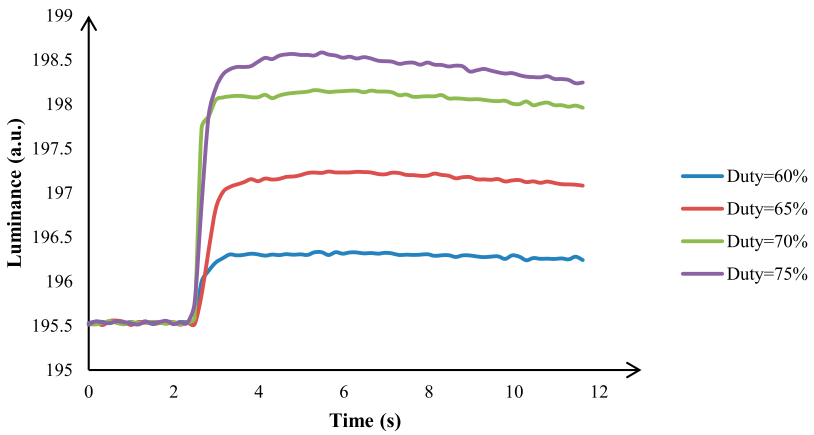

FIGURE 9 | Oil backflow and response time with the different duty cycles. The waveform had the least response time when Duty = 70%. The smaller the duty cycle, the smaller the effective voltage and the smaller the brightness value of EWDs.

FIGURE 10 | The electrochemical reaction between ITO and water. After the reaction, part of ITO was reduced to metal Sn, which gradually reduced the conductivity of ITO until it was completely non-conductive.

potential, and point C is the potential when the current reaches saturation. The maximum of the reverse voltage is the voltage value at point A.

$$E = E_{1/2} - \frac{0.059}{n} \ln \frac{i}{i_d - i} \tag{6}$$

Where, i is the electrolysis current, i_d is the diffusion current value, the temperature is 25°C.

The proposed driving waveform is shown in **Figure 4**. This waveform contains two phases. The first phase is a DC phase, so that the oil has a greater driving force. The second phase is an oil stabilization stage, reverse voltage is used to suspend the oil backflow.

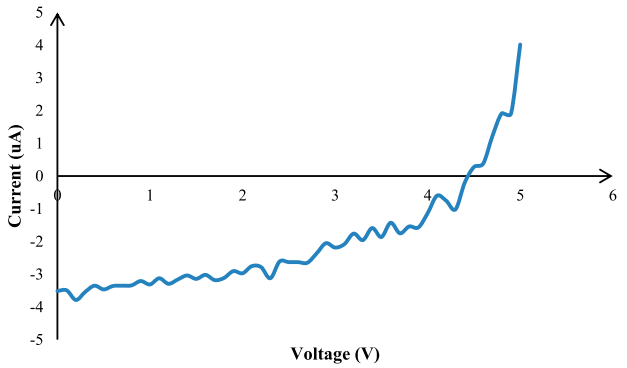

FIGURE 11 | The drop-down curve of voltage and current, voltage step amplitude was –0.1 V, step time was 1 s. As the voltage gradually decreased, the current of the device gradually decreased.

The driving waveform for measuring the hysteresis curve is shown in **Figure 6**. The driving waveform was a superposition of a staircase wave and a square wave. The AC driving frequency was obtained by measuring CV Curve. When the frequency of AC square wave was 500 Hz, a clear CV curve can be obtained, as shown in **Figure 7**. We used this frequency as the driving frequency of the driving waveform.

By the analysis and comparison of the experimental results with different frequencies, it was found that the oil backflow phenomenon become slower when the driving frequency was 0.5 Hz. As shown in **Figure 8**, we used five different driving frequencies for comparative experiments. From the experimental data, it can be seen that oil backflow was more serious than the frequency of 0.05 KHz. And the fluctuations in luminance were visible at the frequency of 0.05 KHz.

The zero voltage helps to delay the phenomenon of charge trapping, and the phenomenon of charge trapping gradually increases as the driving frequency gradually increases. Because the time at zero voltage is too short, the trapped charge cannot be released.

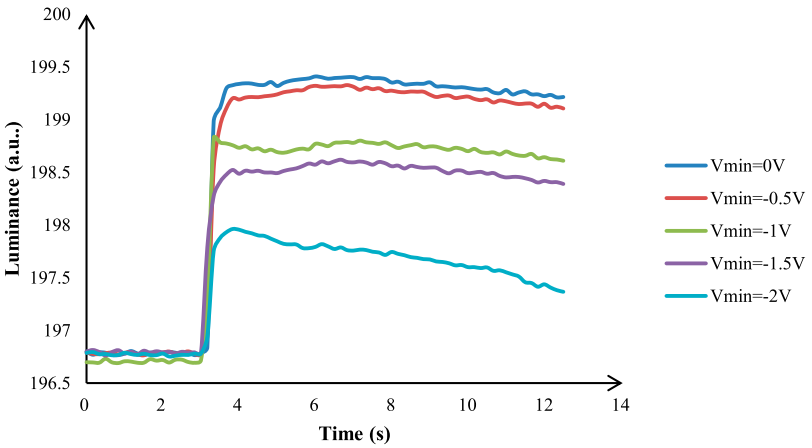

FIGURE 12 | Oil backflow with different reverse voltages. A bigger reverse voltage could reduce the initial aperture, and the backflow phenomenon was enhanced by a high reverse voltage.

EXPERIMENTAL RESULTS AND DISCUSSION

The luminance measurement platform was buided for the experiment, as shown in **Figure 5**. This experimental platform included a waveform generator, a high voltage amplifier, an EWD cell, a colormeter and a computer. And we designed software to control the waveform generator in order to quickly switch the waveform from the waveform generator. The main measuring device is Admesy arg-45, which is a colorimeter developed by Admesy. It has the characteristics of fast measuring speed and high measuring accuracy. The specification of EWDs is shown in **Table 1**.

The colorimeter emits light with an angle of 45° to irradiate the pixels in an area. After the light is absorbed and transmitted by the pixels, it reaches the substrate for reflection. Then the colorimeter can monitor the intensity of the reflected light.

Duty cycle is an important parameter in the AC waveform design. The effective voltage value is also different when the duty cycle is different. According to the Lippman equation, the contact angle can be controlled by the duty cycle. We used 30 V DC voltage for 100 ms to open pixels for obtaining a same aperture, and then, the influence of different duty cycles on oil backflow and response time were studied by the AC waveform, as shown in **Figure 9**.

TABLE 2 | Straight line slope table at different reverse voltages.

Reverse voltage (V)	0	–0.5	–1	–1.5	–2
Straight line slope	–41	–44	–26	–39	–97

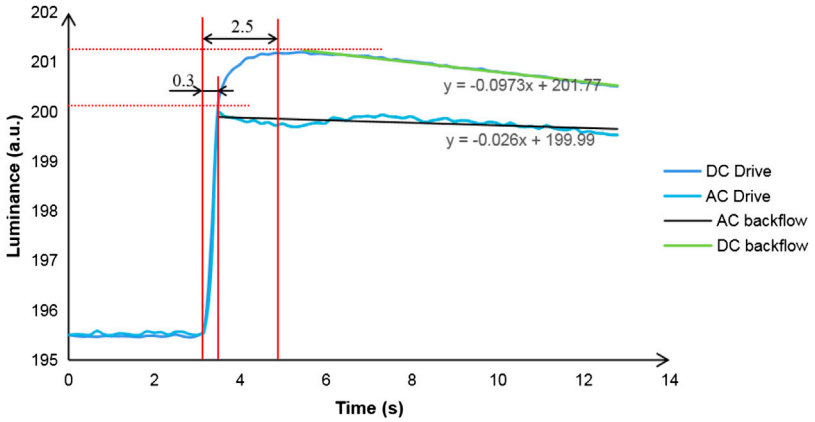

FIGURE 13 | Comparison of response time and backflow phenomenon with different driving waveforms. The AC waveform can inhibit the slow growth of aperture and suspend the backflow phenomenon.

The duty cycle directly affects the release of trapped charges. The duty cycle can effectively control the amount of trapped charge and the amount of released trapped charge. The amount of trapped charge is zero by adjusting duty cycle. So, the backflow can be suppressed.

The ITO electrode is oxidized and reduced to a metal element. As a result, the electric field cannot be formed due to the disconnection of top ITO, as shown in **Figure 10**. The maximum value of the reverse voltage was judged by the drop-down curve in the voltammetry. As shown in **Figure 11**, the measurement method uses a voltage range of −4 to 0 V, and the current value is recorded 2 s after the output voltage.

It can be seen in **Figure 12** that the greater the reverse voltage, the faster the oil backflow. In order to analyze the speed of backflow more intuitively, we simplified the backflow curve to a straight line. We can intuitively see the trend of oil backflow, as shown in **Table 2**. It had the best performance when the reverse voltage was −1 V.

The PWM was often used to drive EWDs. But the response time was too long in PWM waveform, so we compared the response time of a DC waveform, as shown in **Figure 13**. The mixed DC and AC waveform response time was less than 300 ms, and the DC waveform response time was about 2.5 s. In a traditional PWM waveform, pixels were periodically switched on and off, which was the main energy for pushing oil and maintaining state. PWM waveform had a certain inhibition effect on oil backflow. The traditional PWM waveform was similar to the AC waveform with a reverse voltage of 0 V, and the linear slope was −0.0162. In the mixed DC and AC waveform, the oil was pushed by DC waveform, and the oil maintaining state depended on AC waveform. After adding a reverse voltage (−1 V) and using 70% duty cycle to the AC waveform, the oil backflow phenomenon was better than the PWM waveform, and the linear slope was −0.0043.

CONCLUSION

This paper analyzed the driving principle of EWDs. Then, compared the oil backflow and response time among existing driving waveforms. According to the polar solution of water, a design method of adding a safe reverse voltage was proposed. We divided the driving waveform into DC and AC according to different functions, which improved the response time and the oil backflow phenomenon. The driving waveform had better performance for static and dynamic display in EWDs.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

ZW and TZ carried out most of the experiments and data analysis. WL and SL assisted in writing this paper. LW and PB provided theoretical guidance for this paper.

FUNDING

Supported by the National Key Research and Development Program of China (2016YFB0401502), Scientific research project of Guangdong Education Department (No. 2020ZDZX3083), Key-Area Research and Development Program of Guangdong Province (no. 2019B010924002) and Program for Guangdong Innovative and Enterpreneurial Teams (No. 2019BT02C241), Guangdong Provincial Key Laboratory of Optical Information Materials and Technology (No. 2017B030301007) and the 111 Project.

REFERENCES

1. Bai PF, Hayes RA, Jin ML, Shui LL, Yi ZC, Wang L, et al. Review of paper-like display technologies (invited review). *Prog Electromagn Res* (2014) 147:95–116. doi:10.2528/PIER13120405
2. Yi ZC, Bai PF, Wang L, Zhang X An electrophoretic display driving waveform based on improvement of activation pattern. *J Cent South Univ* (2014) 21(8): 3133–7. doi:10.1007/s11771-014-2285-9
3. Wang L, Yi ZC, Jin ML, Shui LL, Zhou GF. Improvement of video playback performance of electrophoretic displays by optimized waveforms with shortened refresh time. *Displays* (2017) 49:95–100. doi:10.1016/j.displa.2017.07.007
4. Roques CT, Hayes RA, Feenstra BJ, Schlangen L. Liquid behavior inside a reflective display pixel based on electrowetting. *J Appl Phys* (2004) 95(8): 4389–96. doi:10.1063/1.1667595
5. Feenstra BJ, Hayes RA, Dijk R, Boom RGH, Wagemans MMH, Camps IGJ, et al. Electrowetting-based displays: bringing microfluidics alive on-screen. In: 19th IEEE international conference on micro electro mechanical systems; 2006 Jan 22–26; Istanbul, Turkey. IEEE (2006). p. 48–53. doi:10.1109/MEMSYS.2006.1627733
6. Yi ZC, Feng HQ, Zhou X, Shui LL. Design of an open electrowetting on dielectric device based on printed circuit board by using a parafilm M. *Front Phys* (2020) 8:193. doi:10.3389/fphy.2020.00193
7. Li W, Wang L, Zhang TY, Lai SF, Liu LW, He WY, et al. Driving waveform design with rising gradient and sawtooth wave of electrowetting displays for ultra-low power consumption. *Micromachines* (2020) 11:145. doi:10.3390/mi11020145
8. Yi ZC, Liu L, Wang L, Li W, Shui LL, Zhou GF. A driving system for fast and precise gray-scale response based on amplitude-frequency mixed modulation in tft electrowetting displays. *Micromachines* (2019) 10:732. doi:10.3390/mi10110732
9. He WY, Yi ZC, Shen ST, Huang ZY, Liu LW, Zhang TY, et al. Driving waveform design of electrophoretic display based on optimized particle activation for a rapid response speed. *Micromachines* (2020) 11:498. doi:10.3390/mi11050498
10. Wang L, Yi ZC, Peng B, Zhou GF. An improved driving waveform reference grayscale of electrophoretic displays. *Proc SPIE* (2015) 9672:967204. doi:10.1117/12.2199212
11. Jiang CD, Tang B, Xu BJ, Groenewold J, Zhou GF. Oil conductivity, electric-field-induced interfacial charge effects, and their influence on the electro-optical response of electrowetting display devices. *Micromachines* (2020) 11: 702. doi:10.3390/mi11070702
12. Shen ST, Gong YX, Jin ML, Yan ZB, Xu C, Yi ZC, et al. Improving electrophoretic particle motion control in electrophoretic displays by eliminating the fringing effect via driving waveform design. *Micromachines* (2018) 9(4):143. doi:10.3390/mi9040143
13. Giraldo A, Massard R, Mans J, Derchx E, Aubert J, Mennen J, 10.3: ultra low-power electrowetting-based displays using dynamic frame rate driving (2011). Oxford UK: Blackwell Publishing Ltd. SIDSymposium Digest of Technical Papers 42. p. 114–7. doi:10.1889/1.3621027
14. Jiang CD, Tang B, Xu BJ, Groenewold J, Zhou GF. Oil conductivity, electric-field-induced interfacial charge effects, and their influence on the electro-optical response of electrowetting display devices. *Micromachines* (2020) 11(7):702. doi:10.3390/mi11070702
15. Yi ZC, Huang ZY, Lai SF, He WY, Wang L, Chi F, et al. Driving waveform design of electrowetting displays based on an exponential function for a stable grayscale and a short driving time. *Micromachines* (2020) 11(3):313. doi:10.3390/mi11030313
16. You H, Steckl AJ. Lightweight electrowetting display on ultra-thin glass substrate. *J Soc Inf Disp* (2013) 21(5):192–7. doi:10.1002/jsid.169
17. Roques CT, Palmier S, Hayes RA, Schlangen L. The effect of the oil/water interfacial tension on electrowetting driven fluid motion. *Colloid Surface Physicochem Eng Aspect* (2005) 267(1–3):56–63. doi:10.1016/j.colsurfa.2005.06.056
18. Moon H, Cho SK, Garrell R. Low voltage electrowetting-on-dielectric. *J Appl Phys* (2002) 92(7):4080–7. doi:10.1063/1.1504171
19. Murade CU, Oh JM, Ende D, Mugele F. Electrowetting driven optical switch and tunable aperture. *Optic Express* (2011) 19(16):15525–31. doi:10.1364/OE.19.015525
20. Barman J, Pant R, Nagarajan AK, Khare K. Electrowetting on dielectrics on lubricating fluid-infused smooth/rough surfaces with negligible hysteresis. *J Adhes Sci Technol* (2017) 31(2):159–70. doi:10.1080/01694243.2016.1205245
21. Massard R, Mans J, Adityaputra A, Leguijt R, Staats C, Giraldo A. Colored oil for electrowetting displays. *J Inf Disp* (2013) 14(1):1–6. doi:10.1080/15980316.2012.751939
22. Chiu YH, Liang CC, Chen YC, Lee WY, Chen HY, Wu SH. Accurate-gray-level and quick-response driving methods for high-performance electrowetting displays. *J Soc Inf Disp* (2011) 19(11):741–8. doi:10.1889/JSID19.11.741
23. Yi ZC, Shui LL, Wang L, Jin ML, Hayes RA, Zhou GF. A novel driver for active matrix electrowetting displays. *Displays* (2015) 37:86–93. doi:10.1016/j.displa.2014.09.004
24. Luo ZJ, Zhang WN, Liu LW, Xie ST, Zhou GF. Portable multi-gray scale video playing scheme for high-performance electrowetting displays. *J Soc Inf Disp* (2016) 24(6):345–54. doi:10.1002/jsid.444
25. Yi ZC, Feng WY, Wang L, Liu LM, Lin Y, He WY, et al. Aperture ratio improvement by optimizing the voltage slope and reverse pulse in the driving waveform for electrowetting displays. *Micromachines* (2019) 10(12):862. doi:10.3390/mi10120862
26. Jin ML, Shen ST, Yi ZC, Zhou GF, Shui LL. Optofluid-based reflective displays. *Micromachines* (2018) 9(4):159. doi:10.3390/mi9040159
27. Hayes RA, Feenstra BJ. Video-speed electronic paper based on electrowetting [J]. *Nature* (2003) 425(6956):383–5. doi:10.1038/nature01988
28. Zeng AP, Liu EQ, Tan SN, Zhang S, Gao J. Cyclic voltammetry studies of sputtered nitrogen doped diamond-like carbon film electrodes. *Electroanalysis* (2002) 14(15–16):1110–5. doi:10.1002/1521-4109(200208)14:15/16<1110::AID-ELAN1110>3.0.CO;2-E
29. Shao YH, Stewart AA, Girault HH. Determination of the half-wave potential of the species limiting the potential window. Measurement of Gibbs transfer energies at the water/1, 2-dichloroethane interface. *J Chem Soc Faraday Trans* (1991) 87(16):2593–7. doi:10.1039/FT9918702593

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fphy.2020.616930

Specular Highlight Detection Based on Color Distribution for Endoscopic Images

Baoxian Yu[1,2,3]*, *Wanbing Chen*[1,2], *Qinghua Zhong*[1,2] *and Han Zhang*[1,2,3]*

[1]School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou, China, [2]Guangdong Provincial Engineering Technology Research Center of Cardiovascular Individual Medicine and Big Data, South China Normal University, Guangzhou, China, [3]SCNU Qingyuan Institute of Science and Technology Innovation Co., Ltd., Qingyuan, China

Endoscopic imaging systems have been widely used in disease diagnosis and minimally invasive surgery. Practically, specular reflection (a.k.a. highlight) always exists in endoscopic images and significantly affects surgeons' observation and judgment. Motivated by the fact that the values of the red channel in nonhighlight area of endoscopic images are higher than those of the green and blue ones, this paper proposes an adaptive specular highlight detection method for endoscopic images. Specifically, for each pixel, we design a criterion for specular highlight detection based on the ratio of the red channel to both the green and blue channels. With the designed criteria, we take advantage of image segmentation and then develop an adaptive threshold with respect to the differences between the red channel and the other ones of neighboring pixels. To validate the proposed method, we conduct experiments on clinical data and CVC-ClinicSpec open database. The experimental results demonstrate that the proposed method yields an averaged *Precision*, *Accuracy*, and *F*1-*score* rate of 88.76%, 99.60% and 72.56%, respectively, and outperforms the state-of-the-art approaches based on color distribution reported for endoscopic highlight detection.

Keywords: adaptive detection, color distribution, endoscopic images, RGB color space, specular highlight detection

OPEN ACCESS

Edited by:
Zichuan Yi,
University of Electronic Science and Technology of China, China

Reviewed by:
Dong Li,
Macau University of Science and Technology, Macau
Xiao Zhang,
Lamar University, United States
Zhiming Dai,
Sun Yat-sen University, China

***Correspondence:**
Baoxian Yu
yubx@m.scnu.edu.cn
Han Zhang
zhanghan@scnu.edu.cn

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *13 October 2020*
Accepted: *11 November 2020*
Published: *12 January 2021*

Citation:
Yu B, Chen W, Zhong Q and Zhang H (2021) Specular Highlight Detection Based on Color Distribution for Endoscopic Images. Front. Phys. 8:616930. doi: 10.3389/fphy.2020.616930

1. INTRODUCTION

Endoscopic imaging systems have been widely used in disease diagnosis and minimally invasive surgery, which, compared to traditional surgery, takes shorter recovery time. Since the angles of lighting source and camera are almost identical, specular reflection always exists in endoscopic images and may conceal valuable information such as vessels and lesions, which is essential for surgeons' observation and judgment. Therefore, specular detection and removal are significant for endoscopic techniques, where the former is the prerequisite of the latter.

For natural images, most of the pioneer works are based on various color spaces. Xia et al. [1] proposed a joint hue, saturation, value (HSV) and red, green, blue (RGB) space threshold detection set using gradient magnitude in dark channel in RGB space. In [2], based on the global brightness of the images, the threshold of HSV color space is automatically set to separate specular reflections. Besides, the dichromatic reflection model is widely used for specular reflection detection of natural images [3]. Specifically, this method uses intensity ratio to extract specular and diffuse components from images [4, 5]. Owing to the different characteristics of natural images, such as uneven color

distribution and lighting, oversaturated specular highlight, the above methods designed for natural images are not applicable for endoscopic images ones.

For endoscopic images, the highlight detection methods can be mainly classified into methods based on different color spaces and the ones with classifier. Considering the real-time detection on highlight of endoscopic images, methods based on color spaces entail lower computations in comparison with those using machine learning techniques [6, 7] and thus have advantages in practice, even though the latter can achieve higher accuracy. At present, the most commonly used color spaces are gray-level [8, 9], HSV [10, 11], and RGB [12, 13]. Specifically, preset thresholds on different color spaces are employed to determine specular highlight. To reduce the impact of halo effect of highlight, Shen et al. [8] proposed a specular detection method by adopting a morphological dilation operation to enlarge the specular reflection regions which is obtained by preset threshold on gray-level image. To address the problem that some pixels in specular reflection areas have lower intensity than nonspecular reflection areas, Oh et al. [10] defined specular reflection area into absolute bright area and relative bright area, which is determined by outlier detection. However, the detected relative bright area might include not only specular highlight, but also white tissues. Zimmerman-Moreno et al. [11] used probabilistic modeling for precise extraction of highlight from segmentation on hue and saturation components for specular regions detection inside the coarse regions. In order to detect the less intense specular highlight, Arnold et al. [13] compared the original image and median filtered image, which is modified by filling each possible specular region with surrounding information. In addition to the application of traditional color spaces, some innovative approaches are proposed. Akbari et al. [14] applied nonlinear SVM classifier trained by statistical features including the mean and standard deviation extracted from each channel of both RGB and HSV color spaces to evaluate the specular detection methods with adaptive threshold. Meslouhi et al. [15] applied the luminance and normalized chrominance of the CIE-XYZ color space to identify specular regions by thresholding.

However, specular reflection detection based on color space mainly suffers from two challenges. On one hand, empirical thresholds are set in advance and, thus, are unable to be adjusted adaptively in different scenes. On the other hand, large specular reflection regions with high intensity cannot be well detected due to the limited difference between highlights and the neighboring areas [16].

To address the above issues, we, in this study, propose an adaptive detection method based on the color distribution characteristics of endoscopic images. The main contributions are summarized in the following two points.

(1) We take advantage of color distribution and propose a criterion characterized by the ratio of the red channel to both the green and blue ones, which provides a more distinguishable feature of specular reflection of endoscopic images.
(2) We propose an adaptive threshold for specular reflection detection. By taking advantage of the difference between the red channel and the other ones, and integrating with overlapped windowing, the proposed adaptive threshold is applicable for large highlight regions with high intensity.

2. PROPOSED METHOD

Without loss of generality, given that the intensity of highlight pixel is higher in the surrounding area, this pixel in an image is defined as

$$I(x) > \alpha * I_{mean}, \tag{1}$$

where $I(x)$ is the intensity of pixel x, I_{mean} is the mean intensity of small region to which pixel x belongs, and α represents a constant parameter. The above criterion is suitable for small specular region with high intensity which is more sensitive to surrounding information. However, pixels within large specular reflection regions are hard to be recognized due to higher I_{mean}.

To address the above issue, we propose a criterion based on the distribution of RGB color space. Basically, in nonhighlight cases, most of endoscopic images are reddish due to the presence of hemoglobin; the value of red channel could be higher than those of green and blue ones. In highlight cases, however, the values of all three channels are nearly identical and saturated, especially for large highlight regions. Therefore, we introduce the criterion for saturated specular pixels by using the ratio between red channel and the other ones as follows:

$$R = \frac{I_R(x)}{\frac{1}{2}(I_G(x) + I_B(x))}, \tag{2}$$

where $I_R(x)$, $I_G(x)$, and $I_B(x)$ are the intensity of RGB channels of pixel x, respectively. From **(2)**, R of nonhighlight pixels is higher than that of highlight ones. Therefore, a threshold t is introduced to discriminate highlight and nonhighlight pixels, which is defined as

$$t = \mu_1 * d^2 + \mu_2 * d + \mu_3, \tag{3}$$

where d is the difference between the red channel and green and blue ones of each small region,

$$d = mean\left(I_R(x) - \frac{1}{2}(I_G(x) + I_B(x))\right) \; x \in \mathbf{A}, \tag{4}$$

where $\mathbf{A}$ represents the set of all pixels in each patch. With the defined threshold, pixels with R lower than t are marked as specular highlight pixels,

$$R < t, \tag{5}$$

However, for darker pixels, the values of all three channels are nearly identical, like highlight pixels. To avoid the confusion between the dark and the highlight cases, we apply an empirical value of threshold in grayscale of 200 as [8], i.e.,

$$I(x) > 200. \tag{6}$$

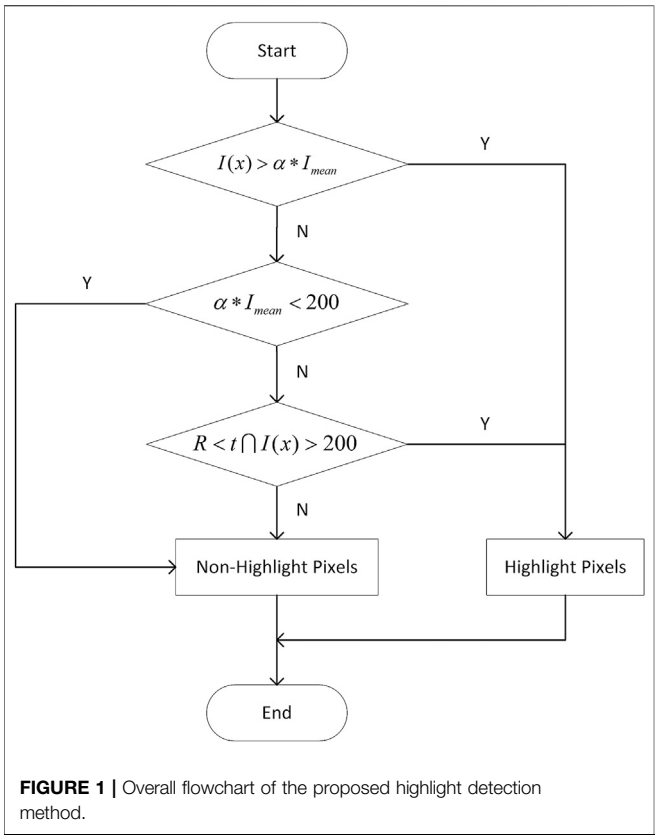

FIGURE 1 | Overall flowchart of the proposed highlight detection method.

Using the above designed criterion, the pixels of an image satisfying **(5)** and **(6)** are detected as highlight. In summary, the overall flowchart of the proposed criteria with adaptive threshold for highlight detection is shown in **Figure 1**.

To further improve the adaptation of the criteria of threshold for highlight detection, we apply the preprocess of segmentation by splitting an image into overlapped equal-sized windows. Therefore, the problem of uneven illumination caused by single light source can be solved through setting adaptive threshold for each patch. Specifically, in each patch, the I_{mean} in **(1)** is defined as

$$I_{mean} = mean(I(x)) \quad x \in \mathbf{A}. \tag{7}$$

Additionally, as shown in **(3)** and **(4)**, d is calculated from the information of each patch, which means t in **(5)** computed by d is adaptively set for each patch.

3. RESULTS AND DISCUSSION

To evaluate the proposed method, we conduct experiments with clinical data provided by Guangzhou Improve Medical Instruments Co., Ltd., and open dataset, namely, CVC-ClinicSpec, which annotated specular highlight ground truth of endoscopic images [6]. The size of images from clinical data and CVC-ClinicSpec is 768 × 576 and 384 × 288, respectively. According to extensive experiments on clinical data using quadratic polynomial fitting, the parameter α in **(1)** is 2.4 and the parameters μ_1, μ_2, and μ_3 in **(3)** are set to $-2.151*10^{-5}$, $2.031*10^{-3}$, and 1.221, respectively.

For measurement, we assess the performance of the proposed method with visual inspection and quantitative evaluation. Specifically, true positive pixels (TPPs) and false positive pixels (FPPs) are the pixels detected as highlights within and not in the ground truth, respectively. Similarly, true negative pixels (TNPs) and false negative pixels (FNPs) are the pixels marked as nonhighlights not in and within the ground truth, respectively. In this study, we employ *Precision*, *Recall*, *Accuracy*, and *F1-score* to quantify the performance of the proposed highlight detection method, which can be written as

$$\begin{aligned} Precision &= \frac{TPPs}{TPPs + FPPs} \\ Recall &= \frac{TPPs}{TPPs + FNPs} \\ Accuracy &= \frac{TPPs + TNPs}{TPPs + TNPs + FPPs + FNPs} \\ F1-score &= 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}. \end{aligned} \tag{8}$$

In the results of the experiment shown below, for the detection results generated from clinical data, the white denotes the detected highlight. As for the results from CVC-ClinicSpec, the white, black, red, and blue denote TPPs, TNPs, FPPs, and FNPs, respectively.

Firstly, we evaluate the detected results of highlight on clinical data and CVC-ClinicSpec database, respectively. To be specific, we conduct experiments on **Figures 2A,B,C** using different criteria to evaluate the performance of each procedure of the proposed detection method. **Figures 2A,B** are the original images from clinical data, and **Figure 2C** is from CVC-ClinicSpec. **Figures 2D,E,F** show the results using the detection procedure in **(1)**. Clearly, only sporadic specular pixels can be detected due to the limitation of using only local information for comparison. By using the proposed criteria in **(1)** and **(5)**, highlight can be well detected, especially for the case of large region, as shown in **Figures 2G,H,I**. Besides, from the comparison between **Figure 2I** and **L**, the effectiveness of distinguishing highlight from dark regions using **(6)** is verified. We can learn from **Figures 2J–L** that the proposed method can detect highlight precisely and achieve visually satisfying results.

Next, we examine the performance improvement by using the preprocess of image segmentation. As elaborated in *Proposed Method*, in order to improve the adaptation of the thresholds in **(1)** and **(5)** for highlight detection, we apply image segmentation to the images from CVC-ClinicSpec database (**Figure 3A,B,C**) and divide each image into nine equisized subgraphs with 50% overlapping with each other. In **Figures 3G,H,I**, it can be observed that there is misjudgment at the boundaries of each window due to the inability to effectively use the boundary information. As shown in **Figure 3**, the highlight detection with image segmentation provides a superior solution in comparison with that without image segmentation. The result is expected, since the proposed adaptive thresholds in **(1)** and **(5)**

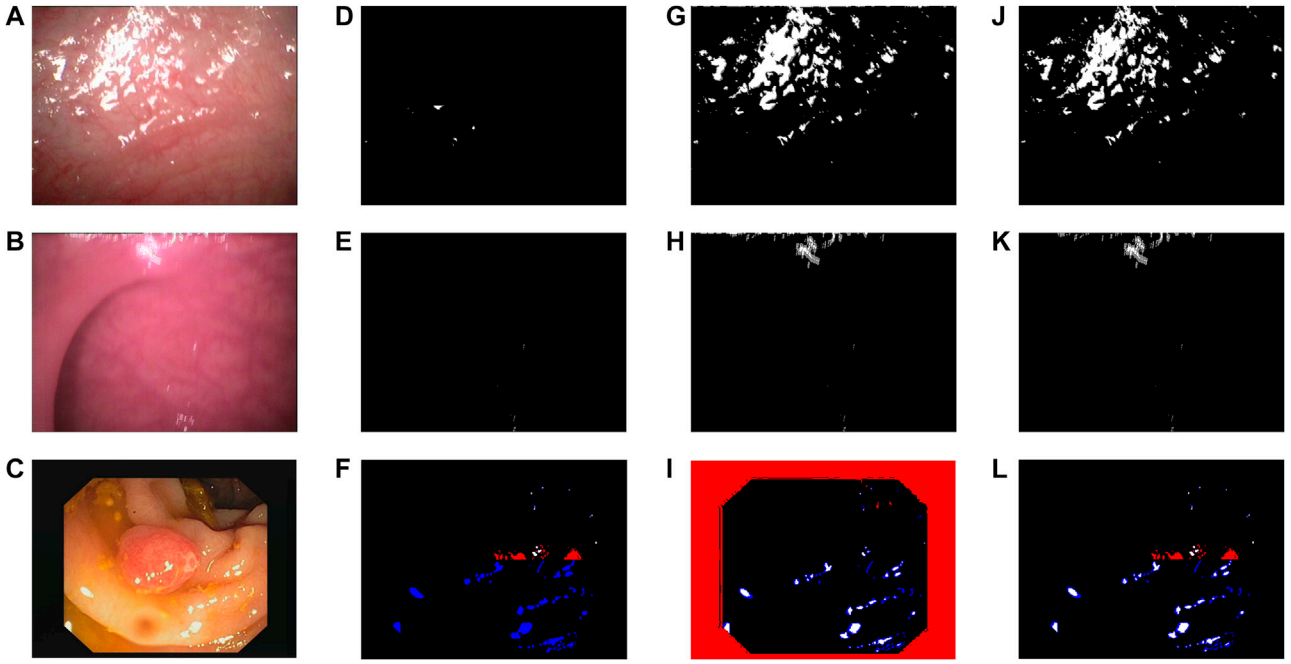

FIGURE 2 | Performance of the proposed highlight detection method. **(A)**–**(B)** Original images from clinical data. **(C)** Original image from CVC-ClinicSpec. **(D)**–**(F)** Applying (1). **(G)**–**(I)** Applying (1) and (5). **(J)**–**(L)** Proposed method.

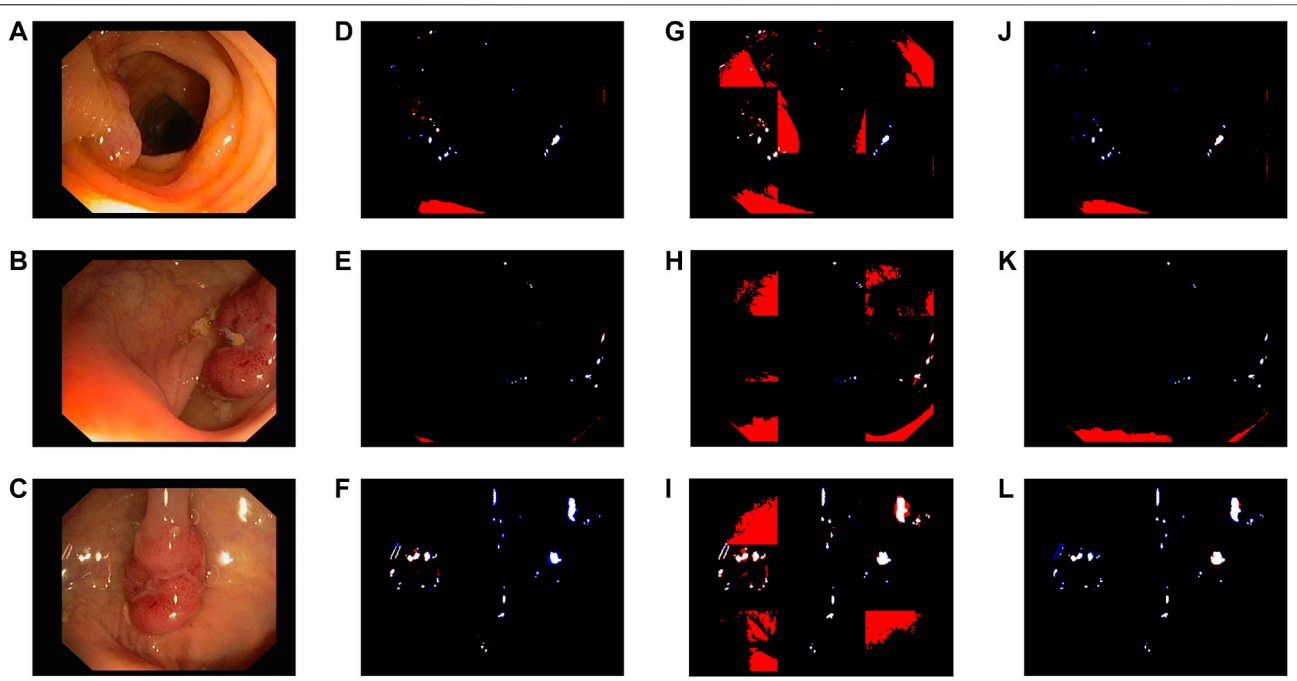

FIGURE 3 | Performance of the proposed highlight detection method with different traversal methods. **(A)**–**(C)** Original images from CVC-ClinicSpec. **(D)**–**(F)** 50% overlapped windows one-fourth the size of the original images. **(G)**–**(I)** Nonoverlapped windows one-ninth the size of the original images. **(J)**–**(L)** Using global information.

are sensitive to uneven lighting, which leads to errors for pixels with high intensity of highlight or dark regions, as shown in **Figures 3J–L**. Comparatively, the image segmentation with overlapping can effectively reduce the effect of uneven lighting and enables a superior detection performance to that with nonoverlapped windows and without segmentation.

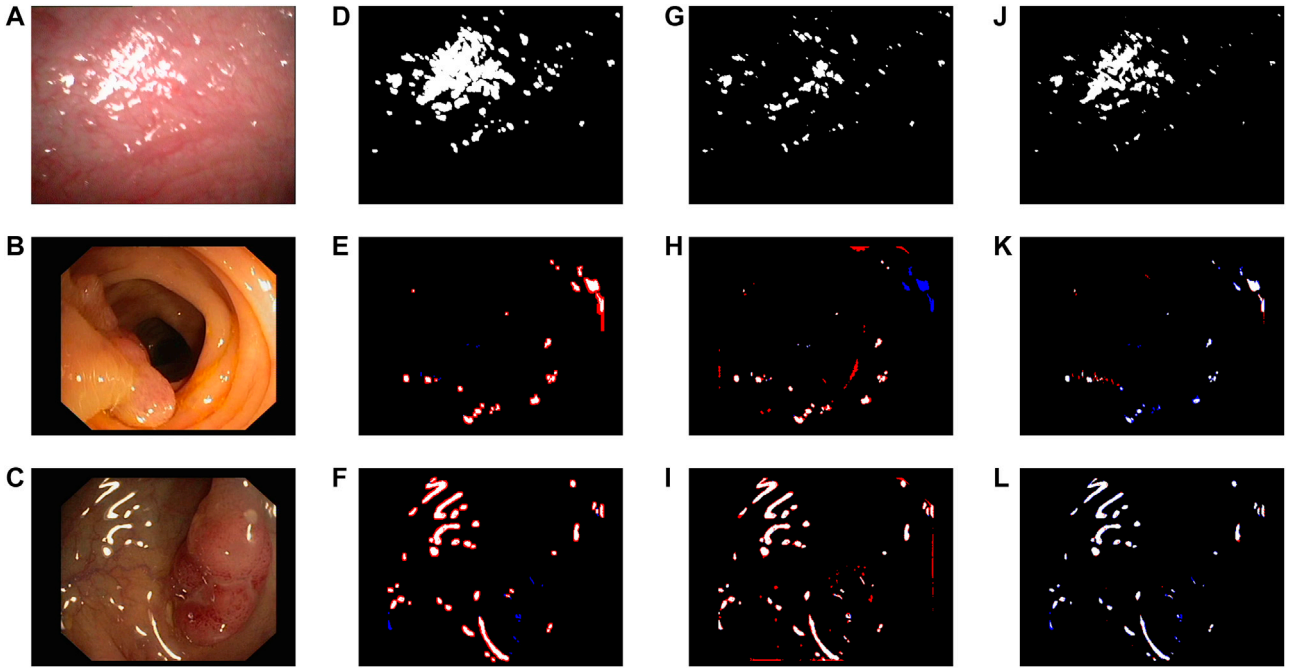

FIGURE 4 | Comparison between different methods. **(A)** Original image from clinical data. **(B)–(C)** Original images from CVC-ClinicSpec. **(D)–(F)** [8]. **(G)–(I)** [13]. **(J)–(L)** Proposed method.

TABLE 1 | Comparison between TPPs, FPPs, *Precision*, *Recall*, *Accuracy*, and *F1-score*.

Image	Method	TPPs	FPPs	*Precision* (%)	*Recall* (%)	*Accuracy* (%)	*F1-score* (%)
Figure 4B	[8]	1,031	1,461	41.37	**96.45**	98.64	57.91
	[13]	612	788	43.71	57.25	98.87	49.57
	Proposed	715	153	**82.37**	66.88	**99.54**	**73.83**
Figure 4C	[8]	2,880	2,504	53.49	94.15	97.57	68.22
	[13]	3,014	1,619	65.06	**98.53**	98.50	78.37
	Proposed	2,580	129	**95.24**	84.34	**99.45**	**89.46**

To further validate the superiority of the proposed method, we compare with the benchmark specular detection methods in [8, 13], as shown in **Figure 4**. **Figures 4A** and **B-C** are the original images from clinical data and CVC-ClinicSpec, respectively. As it illustrated in **Figures 4D,E,F**, it is seen that method of [8] tends to detect a larger area than the actual highlight because of the dilation operation. **Figure 4G** shows the method in [13] is unable to detect large specular highlight region due to the application of local surrounding information. From **Figures 4J–L**, we can observe that the proposed method can obtain accurate highlight detection results and can take both large and small highlight regions into account. To further compare the performance of different methods, we perform quantitative evaluation on images of **Figure 4**. As shown in **Table 1**, although the proposed method performs inferior to that of [8, 13] in terms of *Recall* due to missing specular pixels, which are marked in blue in **Figures 4K,L**, our method obtains the best performance in terms of *Precision*, *Accuracy*, and *F1-score*.

Finally, to validate the robustness of the proposed method, we test 24 images from CVC-ClinicSpec and compare the proposed method with the benchmark methods from different perspective. **Figure 5** illustrates the statistical comparison on *Precision*, *Recall*, *Accuracy*, and *F1-score* using boxplots. It can be observed that the proposed method outperforms the benchmarks in terms of *Precision*, *Accuracy*, and *F1-score*. Theoretically, *Precision* and *Recall* are negatively correlated. Although the proposed method yields a lower averaged *Recall* value, the efficiency of the method can be verified by the highest *F1-score*, which is the harmonic mean of *Precision* and *Recall*.

4. CONCLUSION

In this paper, an adaptive specular highlight detection method for endoscopic images is proposed. Taking the color distribution characteristics of endoscopic images into account, the criteria with adaptive thresholds for highlight detection are developed. The experimental results demonstrate that the proposed method outperforms conventional schemes based on color space and can achieve averaged *Precision*, *Accuracy*, and *F1-score* of 88.76%, 99.60% and 72.56%, respectively.

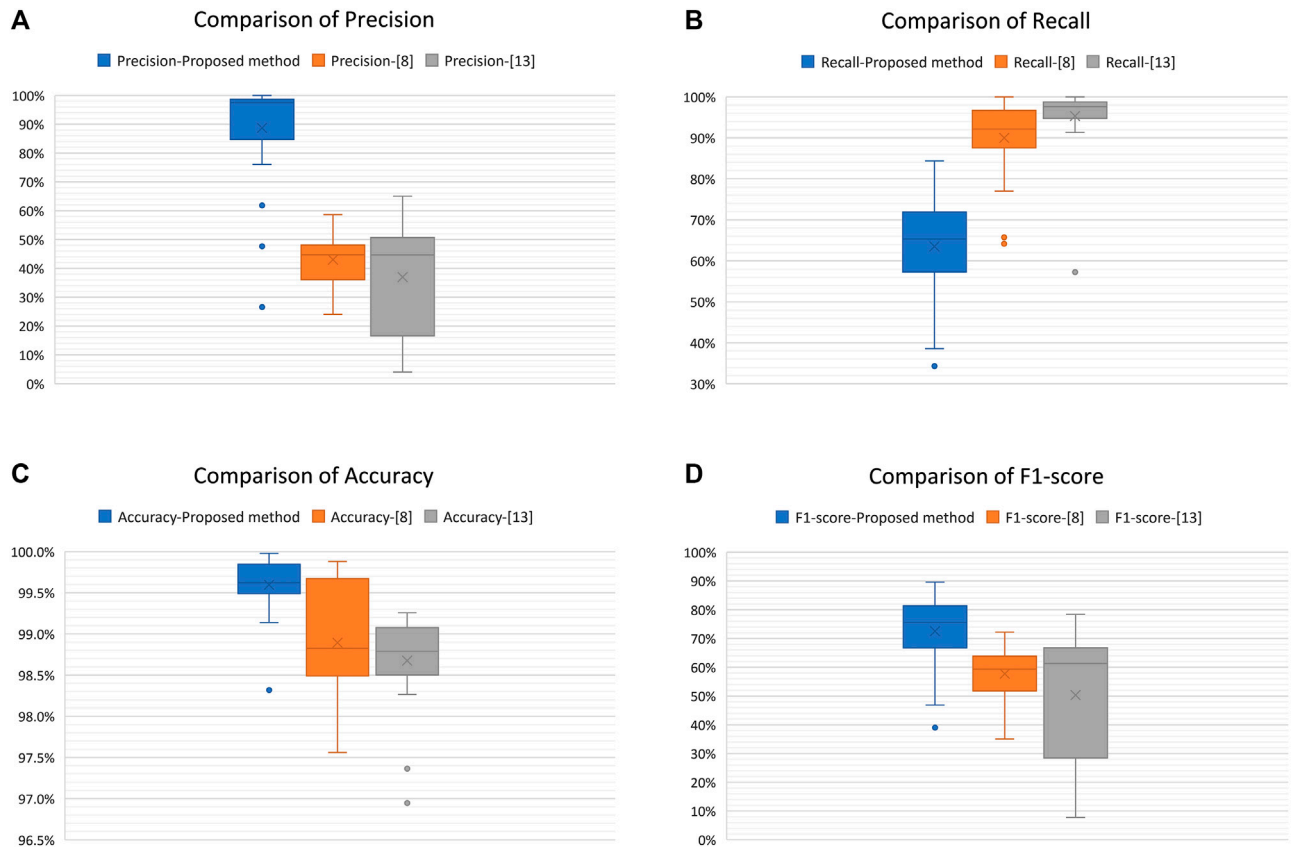

FIGURE 5 | Comparison of *Precision*, *Recall*, *Accuracy*, and *F1-score*, where the crosses mark averages. **(A)** Comparison of *Precision*. **(B)** Comparison of *Recall*. **(C)** Comparison of *Accuracy*. **(D)** Comparison of *F1-score*.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

BY, WC, and HZ contributed the ideas. WC performed the experiments and drafted the paper. BY, WC, QZ, and HZ all did the revision.

FUNDING

This work was supported by the Natural Science Foundation of Guangdong Province, Grant No. 2019A1515011940; Science and Technology Program of Guangzhou, Grant Nos. 2019050001 and 202002030353; Science and Technology Planning Project of Guangdong Province, Grant Nos. 2017B030308009, 2017KZ010101, and 2018A030313990; Special Project for Youth Top-notch Scholars of Guangdong Province, Grant No. 2016TQ03X100; and also Scientific Research Cultivation Project for Young Scholars of South China Normal University, Grant No. 19KJ16.

REFERENCES

1. Xia W, Chen ECS, Pautler SE, Peters TM. A global optimization method for specular highlight removal from a single image. *IEEE Access* (2019) 7: 125976–90. doi:10.1109/ACCESS.2019.2939229
2. Morgand A, Tamaazousti M. Generic and real-time detection of specular reflections in images. In: 2014 International conference on computer vision theory and applications (VISAPP); 2014 January 5–8; Lisbon, Portugal (2014). p. 274–82.
3. Shafer SA. Using color to separate reflection components. *Color Res Appl* (1985) 10:210–8. doi:10.1002/col.5080100409
4. Shen HL, Zheng ZH. Real-time highlight removal using intensity ratio. *Appl Optic* (2013) 52 4483–93. doi:10.1364/AO.52.004483
5. Suo J, An D, Ji X, Wang H, Dai Q. Fast and high quality highlight removal from a single image. *IEEE Trans Image Process* (2016) 25:5441–54. doi:10.1109/TIP.2016.2605002
6. Sánchez F, Bernal J, Sánchez-Montes C, Rodríguez de Miguel C, Fernández-Esparrach G. Bright spot regions segmentation and classification for specular highlights detection in colonoscopy videos. *Mach Vis Appl* (2017) 28:917–36. doi:10.1007/s00138-017-0864-0
7. Gao Y, Yang J, Ma S, Ai D, Lin T, Tang S, et al. Dynamic searching and classification for highlight removal on endoscopic image. *Procedia Comp Sci* (2017) 107:762–7. doi:10.1016/j.procs.2017.03.161
8. Shen DF, Guo JJ, Lin GS, Lin JY. Content-aware specular reflection suppression based on adaptive image inpainting and neural network for endoscopic images. *Comput Methods Progr Biomed* (2020) 192 105414. doi:10.1016/j.cmpb.2020.105414

9. Chu Y, Li H, Li X, Ding Y, Yang X, Ai D, et al. Endoscopic image feature matching via motion consensus and global bilateral regression. *Comput Methods Progr Biomed* (2020) 190 105370. doi:10.1016/j.cmpb.2020.105370
10. Oh J, Hwang S, Lee J, Tavanapong W, Wong J, de Groen PC. Informative frame classification for endoscopy video. *Med Image Anal* (2007) 11 110–27. doi:10.1016/j.media.2006.10.003
11. Zimmerman-Moreno G, Greenspan H. Automatic detection of specular reflections in uterine cervix images. *SPIE Med Imaging* (2006) 6144: 2037–45. doi:10.1117/12.653089
12. Alsaleh SM, Aviles-Rivero AI, Hahn JK. Retouchimg: fusioning from-local-to-global context detection and graph data structures for fully-automatic specular reflection removal for endoscopic images. *Comput Med Imag Graph* (2019) 73 39–48. doi:10.1016/j.compmedimag.2019.02.002
13. Arnold M, Ghosh A, Ameling S, Lacey G. Automatic segmentation and inpainting of specular highlights for endoscopic imaging. *EURASIP Journal on Image and Video Processing* (2010) 1–12. doi:10.1155/2010/814319
14. Akbari M, Mohrekesh M, Najariani K, Karimi N, Samavi S, Soroushmehr SMR. Adaptive specular reflection detection and inpainting in colonoscopy video frames. In: 2018 25th IEEE International conference on image processing; 2018 October 7–10; Athens, Greece. (ICIP) (2018). p. 3134–8.
15. Meslouhi O, Kardouchi M, Allali H, Gadi T, Benkaddour Y. Automatic detection and inpainting of specular reflections for colposcopic images. *Open Computer Science* (2011) 1:341–54. doi:10.2478/s13537-011-0020-2
16. Voronin V, Semenishchev E, Zelensky A, Agaian S. Specular reflection detection algorithm for endoscopic images. *Electron Imag* (2019) 2019: 222–1–222–6. doi:10.2352/ISSN.2470-1173.2019.11.IPAS-222

Conflict of Interest: Authors BY and HZ were employed by the company SCNU Qingyuan Institute of Science and Technology Innovation Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 18 January 2021
doi: 10.3389/fphy.2020.631031

Quantum Interference Effects in Quantum Dot Molecular With Majorana Bound States

Feng Chi[1], Jing Wang[1], Tian-Yu He[1], Zhen-Guo Fu[2]*, Ping Zhang[2], Xiao-Wen Zhang[3], Lihui Wang[4] and Zongliu Lu[4]*

[1]School of Electronic and Information Engineering, Zhongshan Institute, University of Electronic Science and Technology of China, Zhongshan, China, [2]Institute of Applied Physics and Computational Mathematics, Beijing, China, [3]School of Materials Science and Engineering and Guangxi Key Laboratory of Information Materials, Guilin University of Electronic Technology, Guilin, China, [4]Guilin Key Laboratory of Microelectronic Electrode Materials and Biological Nanomaterials and National Special Mineral Materials Engineering Technology Research Center and Guangxi Key Laboratory of Superhard Materials, China Monferrous Metal (Guilin) Geology and Mining Co., Ltd, Guilin, China

Non-invasive detection of the Majorana bound state (MBSs), a kind of quasiparticle without charge and mass, is one of the core issues in current condensed matter physics. Here we study in theory the quantum interference effect in parallel-coupled double quantum dots which are connected either by Majorana bound states (MBSs) or regular fermions. We find that the zero-energy conductance develops a sharp peak when the dots are connected by the MBSs, whereas that in the case of the dots are coupled via regular fermions shows a valley. By varying the coupling strength between the dots and the electron reservoirs, the conductance in the two structures changes in different ways. By comparing the properties of the linear conductance in the two systems, the information of the MBSs formed at the two ends of a topological superconductor nanowire then can be inferred. We also find that the MBSs in the present structure also induces the Fano effect, and is favorable in quantum information processing.

Keywords: quantum dots, Majorana bound states, quantum interference, Fano effect, antiresonance

OPEN ACCESS

Edited by:
Guofu Zhou,
South China Normal University, China

Reviewed by:
Yingjie Chen,
Qufu Normal University, China
Xiu Qing Wang,
Inner Mongolia University for Nationalities, China

****Correspondence:***
Feng Chi
chifeng@semi.ac.cn
Zhen-Guo Fu
fu_zhenguo@iapcm.ac.cn

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *19 November 2020*
Accepted: *09 December 2020*
Published: *18 January 2021*

Citation:
Chi F, Wang J, He T-Y, Fu Z-G, Zhang P, Zhang X-W, Wang L and Lu Z (2021) Quantum Interference Effects in Quantum Dot Molecular With Majorana Bound States. Front. Phys. 8:631031. doi: 10.3389/fphy.2020.631031

INTRODUCTION

In recent years, the preparation and detection of the exotic self-conjugate Majorana fermions have been extensively studied both experimentally and theoretically [1–3]. In solid platforms, they are realized in the form of zero-energy Majorana bound states (MBSs), and are promising in fault-tolerant quantum computing over the last two decades because of their non-Abelian statistics and robustness against decoherence [4, 5]. Up to now, MBSs have been prepared in various setups, in which the simplest and the most frequently studied one is perhaps the topological superconducting nanowires (TSNW). It involves the use of a nanowire having strong spin-orbit interactions under a uniform magnetic field in the proximity of an *s*-wave superconductor [5–7]. The magnetic field breaks the time-reversal symmetry in the TSNW and produces spinless electrons with *p*-wave superconducting pairing [5–7]. With the rapid developments of the nano-fabrications, the MBSs have also be successfully prepared in semiconducting nanowires [8, 9], magnetic atom chains [10], and planar Josephson junctions [11, 12].

The detection of the MBSs is quite challenging due to its charge neutrality and massless nature. One of the current strong evidence of the existence of MBSs is the zero-bias anomaly of the conductance peak in tunnel-spectroscopy experiments [7, 11], which means that under the condition

zero bias, the conductance develops a peak other than the usual zero value. The zero-bias anomaly of the conductance arises from the hybridization between the quantum dot and the Majorana bound state, and then can be used for deducing the existence of the latter. However, such a phenomenon can also arise from some other mechanisms, such as the Kondo effect [13–15]. The Josephson effect has also been extensively investigated due to its phase-sensitive nature as the MBSs in a Josephson junction enable a nontrivial Josephson current having 4π periodicity in the Josephson phase [1, 11, 12]. This is quite different from the Josephson current with 2π periodicity in the usual junctions. Meanwhile, there were also some proposes of hybridizing a quantum dot (QD) to one end of a TSNW [16–20] in that the information of the MBSs will "leak" into the QD and then can be detected by spectroscopy transport measurements in a non-invasive way [21, 22]. Some other proposals such as the thermoelectric effect [23] and optical schemes [24, 25] based on QDs were also put forward. For example in Ref. 24, they argued that the MBSs will absorb and emit photons leading photon-assisted side band peaks. Such processes result in non-zero MBSs mode and then indicate a novel detection scheme for MBSs. The height of the photon-induced side band peaks is proportional to the magnitude of the photon field, and the time-dependent conductance is negative for a certain period of time in the presence of the MBSs [24].

Electronic transport through QD structures has been an active research subject during the past several decades [26, 27]. Due to the confinement of electrons in both the three directions, QD is characterized by the quantized energy levels, and is usually called "artificial atoms" [26]. Coherence of electrons tunneling through QD is greatly preserved, which has been demonstrated in phenomena such as the Aharonov-Bohm oscillations in closed ring-shape geometries [28], the subtle Kondo effect in QD connected to leads [13, 14, 29], and Fano resonances in systems with multiple channels [26–29]. In fact, two or more QDs coupled to each other will form an "artificial molecule" and result in the formation of bonding and antibonding states that are interesting in potential applications in quantum computing devices [26–29]. A lot of research works then have been carried out on the quantum interference effects in diverse parallel and *T*-shaped setups. In the systems of hybridized QD-TSNW, there is usually more than one electron transport channels and the quantum interference effect is an efficient mean to detect the existence of the MBSs formed at the ends of the TSNW [30–35]. For example, it was shown that when the MBSs is coupled to the QD in multiple channel systems, antiresonance in the conductance emerges and its location depends on the direct hybridization between the MBSs [32, 33]. If the coupling strengths between the QD and the two modes of the MBSs are the same, the Fano effect is induced as the Fano lineshape in the conductance or spectrum function [35]. Such interesting results may provide new information for the detection of the MBSs or find practical applications in design of high-efficiency quantum devices.

In the present paper, we study Fano effect in parallel double QDs in the presence of MBSs. We consider the cases of the two dots are simultaneously coupled to the left and right leads with different strengths, and to each other by the TNSW hosting MBSs (QD-MBSs). Our numerical results show that the zero-energy conductance is sensitive to the dot-lead coupling. We also compare the results in our QD-MBSs structure with those in the QDs connected to each other by the usual tunnel junction or regular fermions (another QD), and show that the central (zero-energy) peak in the two cases undergoes a peak to valley evolution, which can be used for detecting the MBSs.

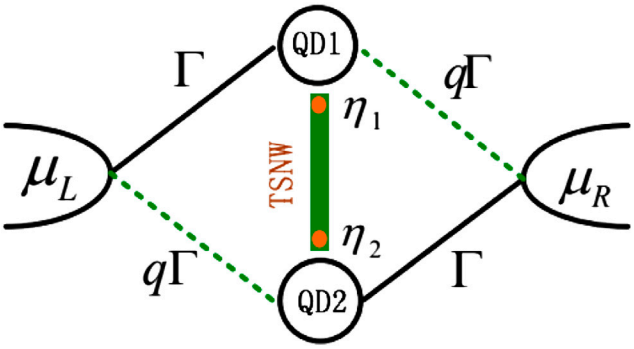

FIGURE 1 | Schematic plot of the parallel double QDs coupled to the left and right leads with coupling strength $\Gamma_{ij}^{L/R}$, and to each other by a TSNW hosting MBSs. Dot 1(2) couples to the mode $\eta_{1(2)}$ of the MBηSs with hybridization strength $\lambda_{1(2)}$. The chemical potentials of the leads are μ_L and μ_R, which are set to be $\mu_L = \mu_R = 0$ in the linear response regime.

MODEL AND METHODS

The schematic plot of the present structure is shown in **Figure 1**, in which two QDs are simultaneously coupled to the left and right leads with different coupling strengths, and connected to each other by a TNSW with MBSs. The Hamiltonian can be written as the following form Refs. 36–40,

$$H = \sum_{k\alpha}\varepsilon_{k\alpha}\varepsilon_{k\alpha}^{\dagger}\varepsilon_{k\alpha} + \sum_{i=1,2}\varepsilon_i d_i^{\dagger}d_i + t_c\left(d_1^{\dagger}d_2 + h.c\right) + \sum_{k\alpha i}\left(V_{k\alpha i}d_i^{\dagger}c_{k\alpha} + h.c.\right) + H_{MBSs}, \quad (1)$$

where the creation (annihilation) operator $c_{k\alpha}^{\dagger}(c_{k\alpha})$ is for electrons having wave vector k, energy $\varepsilon_{k\alpha}$ in the αth ($\alpha = L/R$) normal metal lead. The second term in the right side of **Eq. 1** is for electrons on the two QDs with energy level $\varepsilon_i$with creation (annihilation) operator $d_i^{\dagger}(d_i)$. The third term in the right side of **Eq. 1** stands for possible hopping between the two QDs through a tunnel barrier with strength t_c. The forth term in the right side of **Eq. 1**corresponds to tunneling between the QDs and the leads with amplitude $V_{k\alpha i}$. The last term in right side of **Eq. 1** denotes the MBSs formed at opposite ends of the TSNW and their couplings to the QDs [35–37].

$$H_{MBSs} = \frac{\lambda_1}{\sqrt{2}}\left(d_1 - d_1^{\dagger}\right)\gamma_1 + i\frac{\lambda_2}{\sqrt{2}}\left(d_2 + d_2^{\dagger}\right)\gamma_2 + i\delta_M\gamma_1\gamma_2, \quad (2)$$

in which λ_i is the hybridization amplitude between the QD-*i* and the mode-*i* of the MBSs, and the self-conjugate nature of the

MBSs requires $\gamma_{1(2)} = \gamma_{1(2)}^{\dagger}$ and $\{\gamma_i, \gamma_j\} = 2\delta_{ij}$. Following previous work [17], we transform the Majorana fermion operators $\gamma_{1(2)}$ into the regular fermion operators f and $f^{\dagger}$ by defining $\gamma_1 = \frac{1}{\sqrt{2}}(f^{+} + f), \gamma_2 = \frac{i}{\sqrt{2}}(f^{+} - f)$, and then **Eq. 2** becomes into the following form,

$$H_{MBSs} = \delta_M\left(f^{\dagger}f - \frac{1}{2}\right) + \frac{\lambda_1}{\sqrt{2}}\left(d_1 - d_1^{\dagger}\right)\left(f + f^{\dagger}\right) + \frac{\lambda_2}{\sqrt{2}}\left(d_2 + d_2^{\dagger}\right)\left(f - f^{\dagger}\right). \tag{3}$$

At ultra-low temperature favorable for the preparation of the MBSs, the linear conductance G is obtained with the help of the transmission $T(\varepsilon)$ of an electron by the Landauer formula $G = (2e^2/h)T(\varepsilon)$. The transmission coefficient is calculated in terms of the Green's functions as $T(\varepsilon) = Tr\{\widehat{G}^{a}(\varepsilon)\widehat{\Gamma}^{R}\widehat{G}^{r}(\varepsilon)\widehat{\Gamma}^{L}\}$, in which $\widehat{G}^{r(a)}(\varepsilon)$ is the retarded (advanced) Green's function and $\widehat{\Gamma}^{L(R)}$ the line-width function denoting coupling between the dots and the leads. For the present structure, they are all 6×6matrixes. We calculate the retarded Green's function $\widehat{G}^{r}(\varepsilon) = [\widehat{G}^{a}(\varepsilon)]^{\dagger}$ by using the Dyson's equation method combined with the equation of motion technique. After some straightforward calculations, the Green's function is given by Ref. 14 $\widehat{G}^{r} = (\widehat{g}^{r-1} - \widehat{T}_c - \widehat{K} - i(\widehat{\Gamma}^{L} + \widehat{\Gamma}^{R})/2)^{-1}$, in which $\widehat{g}^{r}$ is the retarded Green's function of the central region free from couplings whose diagonal elements are $g^{r}_{11(22)}(\varepsilon) = 1/(\varepsilon \mp \varepsilon_1)$, $g^{r}_{33(44)}(\varepsilon) = 1/(\varepsilon \mp \varepsilon_2)$, and $g^{r}_{55(66)}(\varepsilon) = 1/(\varepsilon \mp \delta_M)$. Matrix $\widehat{T}_c$ denotes the tunnel-coupling between the dots whose elements are $\widehat{T}_{c,13} = \widehat{T}_{c,31} = -\widehat{T}_{c,24} = -\widehat{T}_{c,42} = t_c$. The matrix elements of are $\widehat{K}_{15(6)} = \widehat{K}_{5(6)1} = -\widehat{K}_{25(6)} = -\widehat{K}_{5(6)} = \xi_1$ and $\widehat{K}_{3(4)5} = \widehat{K}_{53(4)} = -\widehat{K}_{3(4)6} = -\widehat{K}_{63(4)} = \xi_2$, where $\xi_{1(2)} = \pi\rho_i\lambda^2_{1(2)}/2$ being the dot-MBSs coupling strength and ρ_i the density of the states in QDs i. In the present paper, we assume that the MBSs are coupled to the two QDs with equal strength, and fix $\xi_1 = \xi_2 = \lambda$. The line-width function is given by

$$\widehat{\Gamma}^{L/R} = \begin{pmatrix} \Gamma_{11}^{L/R} & 0 & \sqrt{\Gamma_{11}^{L/R}\Gamma_{22}^{L/R}} & 0 & 0 & 0 \\ 0 & \Gamma_{11}^{L/R} & 0 & \sqrt{\Gamma_{11}^{L/R}\Gamma_{22}^{L/R}} & 0 & 0 \\ \sqrt{\Gamma_{11}^{L/R}\Gamma_{22}^{L/R}} & 0 & \Gamma_{22}^{L/R} & 0 & 0 & 0 \\ 0 & \sqrt{\Gamma_{11}^{L/R}\Gamma_{22}^{L/R}} & 0 & \Gamma_{22}^{L/R} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{4}$$

in which the $\Gamma_{ij}^{\alpha} = 2\pi V_{k\alpha i}V^{*}_{k\alpha j}\rho_{\alpha}$ with ρ_{α} being the local density of states in lead α.

RESULTS AND DISCUSSION

In the following numerical calculations, we fix the line-width function $\Gamma_{11}^{L} = \Gamma_{22}^{R} = \Gamma \equiv 1$ as the energy unit, and assume $\Gamma_{22}^{L} = \Gamma_{11}^{R} = q\Gamma$ with q a dimensionless parameter throughout the paper. We also fix the dots' levels at $\varepsilon_1 = \varepsilon_2 = 0$ in that the MBSs exert significant impacts on electronic transport around zero energy.

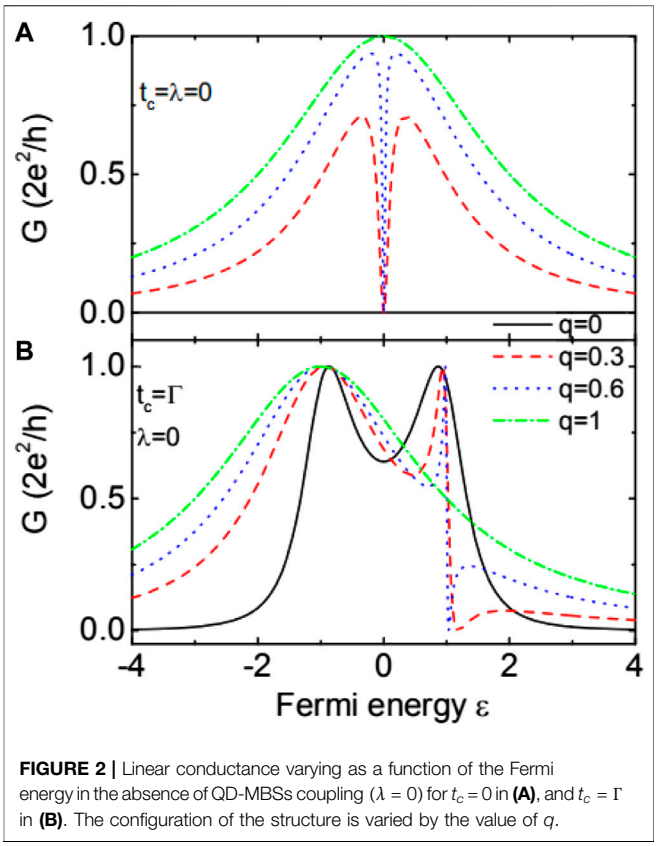

FIGURE 2 | Linear conductance varying as a function of the Fermi energy in the absence of QD-MBSs coupling ($\lambda = 0$) for $t_c = 0$ in **(A)**, and $t_c = \Gamma$ in **(B)**. The configuration of the structure is varied by the value of q.

For $0 < q \leq 1$, the two dots are parallel-coupled to the two leads, and $q = 0$ corresponds to the case of the serially coupled configuration. **Figure 2** presents the results of linear conductance through the double QDs for different values of q. For $q = 0$ and $t_c = 0$, the system is an open circuit and the conductance is zero $(G = 0)$ as indicated by the solid black line in **Figure 2A**. For $0 < q < 1$, the two dots are connected in parallel configuration with asymmetrical left-right couplings, and the analytical expression of the linear conductance is Refs. 28 and 31

$$G = \frac{2e^2}{h}\frac{4q\varepsilon^2\Gamma^2}{\left[\varepsilon^2 - \left(\frac{1-q}{2}\Gamma\right)^2\right]^2 + \left[\varepsilon\Gamma(1+q)\right]^2}. \tag{5}$$

From above equation, one can see that G develops a dip at $\varepsilon = 0$, and has two resonances located at $\varepsilon = \pm(1-q)\Gamma/2$ as shown by the red dashed and blue dotted lines in **Figure 2A**. For the two dots are coupled to the dots symmetrically $q = 1$, the conductance exhibits a single wide peak centered at as indicated by the green dash-dotted line, which can also be seen from **Eq. 5**. Under the condition of $t_c \neq 0$, the conductance is composed by one Breit-Wigner resonance at $\varepsilon = t_c$ and a Fano one at $\varepsilon = t_c$, which respectively correspond to the bonding and antibonding states in this QDs molecular [28, 31]. The conductance now is given by

$$G = \frac{2e^2}{h}\frac{4\Gamma^2\left(\frac{1+q}{2}t_c - \sqrt{q}\,\varepsilon\right)^2}{\left[\varepsilon^2 - t_c^2 + \left(\frac{1-q}{2}\Gamma\right)^2\right]^2 + 4\Gamma^2\left[\frac{1+q}{2}\varepsilon + \sqrt{q}\,t_c\right]^2}. \tag{6}$$

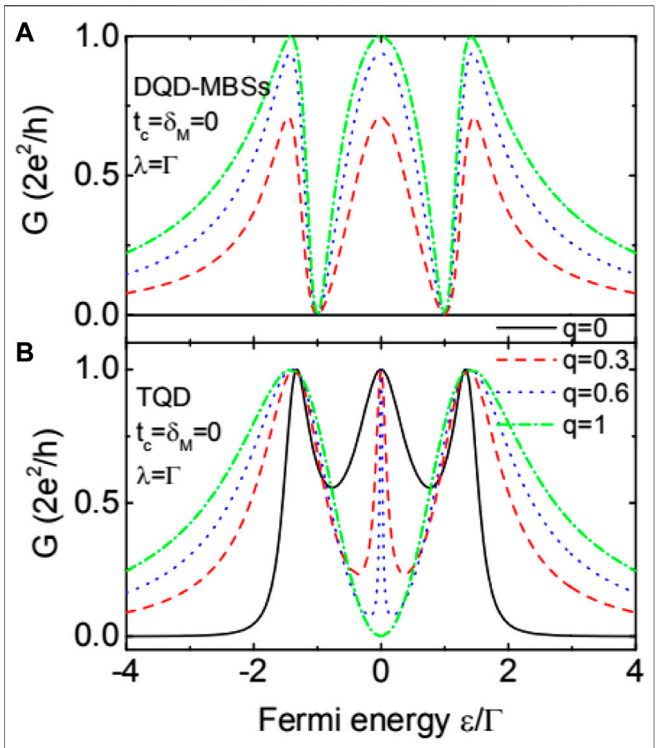

FIGURE 3 | Dependence of the linear conductance on the Fermi energy for the case of the double QDs are coupled to each other by the MBSs (DQD-MBSs) in **(A)** and the case of the two QDs connected by regular fermions, i.e., another QD in **(B)**. The latter corresponds to the triple-QD (TQD) configuration.

For $q = 0$ (serially connected double QDs), the conductance shows two Lorentz peaks at the bonding and antibonding states, respectively. Whereas for $0 < q < 1$, the resonance at the antibonding state $\varepsilon = t_c$ shows the typical asymmetrical Fano line-shape, which has been systematically investigated in Refs. 28 and 31 and we do not discuss them in more detail.

We now study the case of the two QDs are connected by the MBSs (DQD-MBSs configuration) in **Figure 3A**, and also shows the results when the two dots are connected by regular fermions (TQD configuration). For $q = 0$, the conductance remains at zero in that the electrons can not transport through the TSNW for $t_c = \delta_M = 0$ as shown by the black solid line in **Figure 3A**. The conductance develops triple peaks for nonzero $q \neq 0$ which locate about at $\varepsilon = 0$ and $\varepsilon = \pm(\lambda + \Gamma/2)$, respectively. Meanwhile, there are two antiresonances $(G = 0)$ at $\varepsilon = \pm\lambda$. With increasing q, the height of the three peaks increases with unchanged configuration, and reaches its quantum value $2e^2/h$ for $q = 1$. In structure of single QD hybridized with the MBSs, the zero energy conductance is half of its quantum value $G(\varepsilon = 0) \equiv e^2/h$, which is believed to be the evidence of the MBSs [19]. In the present DQD-MBSs systems, however, this criterion will not hold true, and the existence of the MBSs can only be deduced from the properties of the resonances and antiresonances in the conductance. For this reason, we show the results when the MBSs connecting the two QDs is replaced by another QD (dot-3) serving as a pair of regular fermions. For the convenience of comparison, we use similar symbols of the MBSs to denote dot-3, and then Hamiltonian in **Eq. 1** is changed into

$$H = \sum_{k\alpha}\varepsilon_{k\alpha}\varepsilon^{\dagger}_{k\alpha}\varepsilon_{k\alpha} + \sum_{i=1,2}\varepsilon_i d_i^{\dagger}d_i + t_c\left(d_1^{\dagger}d_2 + h.c\right) + \delta_M d_3^{\dagger}d_3 + \left(\lambda d_1^{\dagger}d_3 + \lambda d_3^{\dagger}d_1 + h.c\right) + \sum_{k\alpha i}\left(V_{k\alpha i}d_i^{\dagger}c_{k\alpha} + h.c.\right), \tag{7}$$

from which the retarded Green's function and the conductance can be easily obtained by the equation of motion method in the absence of Coulomb interaction between the electrons [41].

As shown in **Figure 3B**, the conductance develops triple peaks for $0 \leq q < 1$ with unchanged peaks' value $2e^2/h$. With increasing

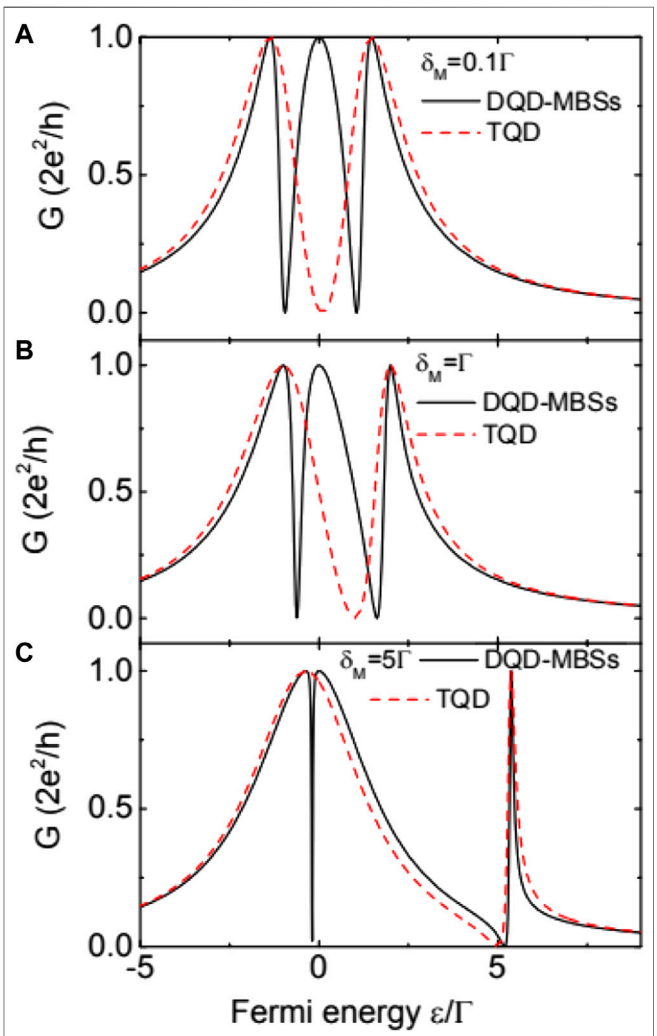

FIGURE 4 | Linear conductance as a function of the Fermi energy for DQD-MBSs and TQD configurations and different values of direct overlap amplitude δ_M and fixed $\lambda = 1$ and $q = 1$. In **(A)** we set $\delta_M = 0.1\Gamma$, and $\delta_M = \Gamma$ in **(B)**, $\delta_M = 5\Gamma$ in **(C)**. The conductance in **(C)** displays typical asymmetric Fano line-shape in higher energy regimes.

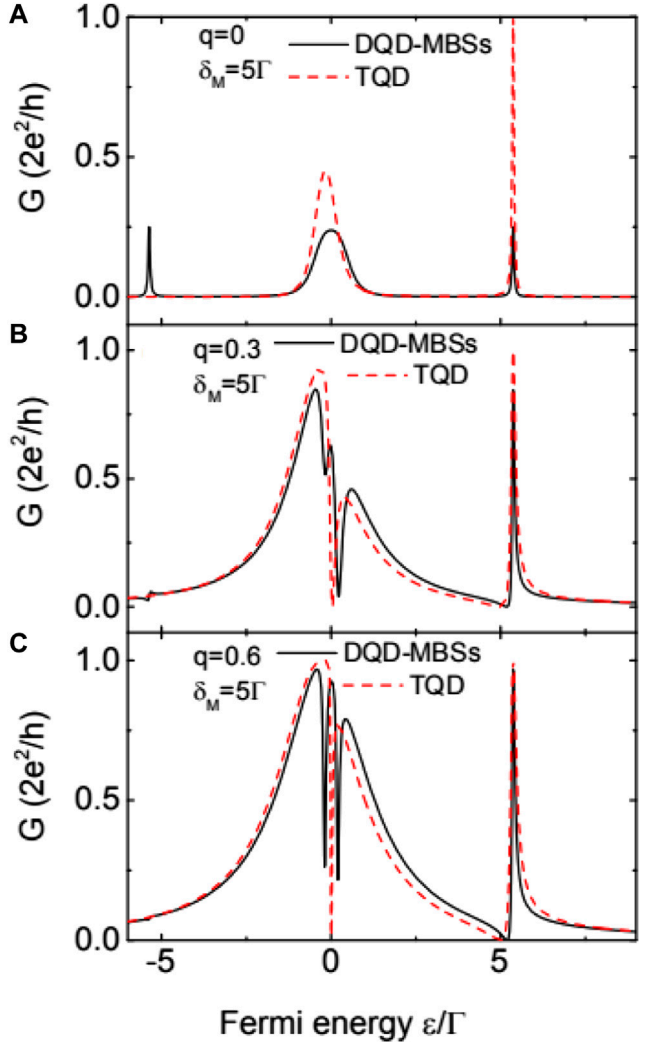

FIGURE 5 | Linear conductance as a function of the Fermi energy for DQD-MBSs and TQD configurations for $q = 0$ in **(A)**, $q = 0.3$ in **(B)**, $q = 0.6$ in **(C)**. Except the parameters given in the figure, others are as in **Figure 4**.

q, the two dips move toward $\varepsilon = 0$ with the narrowing of the central peak. For, the central peak involves into an antiresonance $G(\varepsilon = 0) \equiv 0$ and the conductance reduces to a double-peak configuration. By comparing **Figures 3A,B**, one can find that the properties of the conductance in DQD-MBSs are quite different from those in TQD, and then the information of the MBSs can be then inferred. For example, the zero-energy conductance develops a peak when the two dots are symmetrically coupled to the leads ($q = 1$) in the DQD-MBSs, whereas it becomes to be zero when the two dots are connected by another QD (TQD). In fact, this is one of the interesting effects induced by the interaction between the MBSs and the electrons on the QDs that the behaviors of the conductance near the zero-energy point are mainly influenced. The reason is that the MBSs are zero energy like the electron-hole pair, and then induces a peak at zero Fermi energy, or a peak in the differential conductance when the bias voltage is zero (zero bias anomaly in the conductance found in Ref. 13). In higher energy regime, the impacts of the MBSs then are weak and the transport processes are dominated by the electrons from different channels. For this reason, the typical asymmetrical Fano line-shape in the conductance in higher energy regimes emerges, but vanishes around the zero-energy.

Figure 4 presents the calculation results in structures of DQD-MBSs (black solid line) and TQD (red dashed line) with varied inter-MBSs coupling strength δ_M (or energy level of dot-3). For small value $\delta_M = 0.1\Gamma$ as shown in **Figure 4A**, the conductance's configuration in both of the two structures remains almost unchanged, which is also quite different from the result in single dot coupled to MBSs [19], in which the value of the zero-energy conductance recovers $2e^2/h$ as long as $\delta_M \neq 0$. With increasing δ_M as shown in **Figure 4B**, the conductance peak in negative (positive) energy regimes moves toward (away from) the zero-energy state of $\varepsilon = 0$ for the case of DQD-MBSs (the solid black line). The conductance in the TQD, however, both the positions of the peaks and the dip move to higher energy level, as indicated by the red dashed line. When the hybridization between the MBSs is $\delta_M > \Gamma$, the conductance in high energy regime exhibits typical Fano line-shape, which holds true in both the DQD-MBSs and TQD configurations. Around the zero-energy state, however, the conductance in DQD-MBSs remains the double-peak configuration with an antiresonance. The conductance in TQD around $\varepsilon = 0$ shows the single wide peak configuration, which is completely different from that in DQD-MBSs.

Finally, we study the conductance in **Figure 5** for different values of q and fixed $\delta_M = 5\Gamma$. For nonzero δ_M, the conductance in serially coupled double QDs has three peaks individually abut at $\varepsilon = 0$ and $\varepsilon = \pm\sqrt{\lambda^2 + \delta_M^2}$ in DQD-MBSs, as shown by the solid black line in **Figure 5A**. The peaks are also equal in height but the width of the central peak is much wider than that of the other two. With increasing q as indicated by the sold lines in **Figures 5B, C**, the peak in the negative energy regime is destroyed with the splitting of the central one, exhibiting the Fano line-shape. The peak in higher energy regime becomes higher and then also displays the asymmetric Fano line-shape which is similar to the case in **Figure 4**. The conductance in TQD has only two peaks for $q = 0$, and then changes in similar way as that of DQD-MBSs. Interestingly, in the case of $q = 1$, we find in **Figure 5A** that the zero-energy conductance of DQD-MBSs develops a peak whereas that of TQD is a dip with the value of $G = 0$. Such an evolution from peak to valley is a good evidence of the MBSs. This difference between the two cases can be attributed to the fact that there is a resonant state at zero-energy induced by the MBSs.

SUMMARY

In summary, we have studied the properties of the conductance in double QDs coupled to the left and right leads in parallel configuration within the framework of the Green's function method. We find that when the two dots are connected by a TSNW with MBSs at its two ends, the conductance develops three

peaks, whose height is enhanced with increasing left-right symmetry. Between the peaks, there are two dips with zero conductance. If the dots are connected by regular fermions, however, the dips is nonzero, and the central peaks will evolve into a dip if the system is totally symmetric in left-right coupling strength. Such a peak to valley change of the conductance provides a non-invasive detection of the MBSs. When the MBSs formed at the ends of the TSNW is overlapped, the resonances at zero and positive energy regimes will display the typical Fano line-shape, which is useful in quantum information processing or design of novel quantum devices.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

FC and Z-GF contributed the ideas equally and derived the formulae. JW and T-YH performed the numerical calculations. FC, Z-GF, and PZ wrote the original manuscript. X-WZ, LW, and ZL helped to discussed the results and polished the manuscript.

FUNDING

This research was supported by the NSFC (Grant No. 62075035), the Initial Project of UEST of China, Zhongshan Institute (415YKQ02), Project for Innovation Team of Guangdong University (No. 2018KCXTD033), Science and Technology Bureau of Zhongshan (Grant No. 2017B1116, 180809162197886), and the Research Funds for Beijing Universities (NO.KM201910009002). X-W Zhang acknowledges support from Key Research and Development Program of Guangxi (GuiKeAB19110032).

REFERENCES

1. Kitaev AY. Unpaired majorana fermions in quantum wires. *Phys Usp* (2001) 44(10):131–6. doi:10.1070/1063-7869/44/10S/S29
2. Nayak C, Simon SH, Stern A, Freedman M, Das Sarma S. Non-Abelian anyons and topological quantum computation. *Rev Mod Phys* (2008) 80(3):1083–1159. doi:10.1103/revmodphys.80.1083
3. Alicea J. New directions in the pursuit of Majorana fermions in solid state systems. *Rep Prog Phys* (2012) 75(7):076501. doi:10.1088/0034-4885/75/7/076501
4. Sau JD, Lutchyn RM, Tewari S, Das Sarma S. Generic new platform for topological quantum computation using semiconductor heterostructures. *Phys Rev Lett* (2010) 104(4):040502. doi:10.1103/PhysRevLett.104.040502
5. Karzig T, Knapp C, Lutchyn RM, Bonderson P, Hastings MB, Nayak C, et al.. Scalable designs for quasiparticle-poisoning-protected topological quantum computation with Majorana zero modes. *Phys Rev B* (2017) 95(23):235305. doi:10.1103/physrevb.95.235305
6. Oreg Y, Refael G, von Oppen F. Helical liquids and Majorana bound states in quantum wires. *Phys Rev Lett* (2010) 105(17):177002. doi:10.1103/PhysRevLett.105.177002
7. Mourik V, Zuo K, Frolov SM, Plissard SR, Bakkers EP, Kouwenhoven LP. Signatures of Majorana fermions in hybrid superconductor-semiconductor nanowire devices. *Science* (2012) 336(6084):1003–7. doi:10.1126/science.1222360
8. Deng MT, Vaitiekėnas S, Hansen EB, Danon J, Leijnse M, Flensberg K, et al.. Majorana bound state in a coupled quantum-dot hybrid-nanowire system. *Science* (2016) 354(6319):1557–62. doi:10.1126/science.aaf3961
9. Zhang H, Liu CX, Gazibegovic S, Xu D, Logan JA, Wang G, et al.. Quantized Majorana conductance. *Nature* (2018) 556(7699):74–9. doi:10.1038/nature26142
10. KouwenhovenXu S, Xie Y, Li J, Wang Z, Bernevig BA, Yazdani A. Distinguishing a Majorana zero mode using spin-resolved measurements. *Science* (2017) 358(6364):772–6. doi:10.1126/science.aan3670
11. Fornieri A, Whiticar AM, Setiawan F, Portolés E., Drachmann ACC, Keselman A, et al. Evidence of topological superconductivity in planar Josephson junctions. *Nature* (2019) 569(7754):89–92. doi:10.1038/s41586-019-1068-8
12. NichelePortolés H, Pientka F, Hart S, Pierce A, Kosowsky M, Lunczerau L, et al.. Topological superconductivity in a phase-controlled Josephson junction. *Nature* (2019) 569(7754):93–8. doi:10.1038/s41586-019-1148-9
13. YacobyPierce A, Das A, Ronen Y, Most Y, Oreg Y, Heiblum M, et al.. Zero-bias peaks and splitting in an Al-InAs nanowire topological superconductor as a signature of Majorana fermions. *Nat Phys* (2012) 8(12):887–95. doi:10.1038/NPHYS2479
14. Górski G, Kucab K. The spin-dependent coupling in the hybrid quantum dot–Majorana wire system. *Phys Status Solidi B-Basic Solid State Phys* (2019) 256(3):1800492. doi:10.1002/pssb.201800492
15. Weymann I, Wojcik KP, Majek P. Majorana-Kondo interplay in T-shaped double quantum dots. *Phys Rev B* (2020) 101(23):235404. doi:10.1103/PhysRevB.101.235404
16. Yi GY, Jiang C, Zhang LL, Zhong SR, Chu H, Gong WJ, et al.. Manipulability of the Kondo effect in a T-shaped triple-quantum-dot structure. *Phys Rev B* (2020) 102(8):085418. doi:10.1103/PhysRevB.102.085418
17. Feng JJ, Huang Z, Wang Z, Niu Q. Josephson radiation from nonlinear dynamics of Majorana zero modes. *Phys Rev B* (2020) 101(18):180504. doi:10.1103/PhysRevB.101.180504
18. Liu DE, Cheng M, Lutchyn RM. Probing Majorana physics in quantum-dot shot-noise experiments. *Phys Rev B* (2015) 91(8):081405. doi:10.1103/PhysRevB.91.081405
19. Liu DE, Baranger HU. Detecting a Majorana-fermion zero mode using a quantum dot. *Phys Rev B* (2011) 84(20):201308. doi:10.1103/PhysRevB.84.201308
20. Sherman D, Yodh JS, Albrecht SM, Nygård J, Krogstrup P, Marcus CM. Normal, superconducting and topological regimes of hybrid double quantum dots. *Nat Nanotechnol* (2016) 12(3):212–7. doi:10.1038/nnano.2016.227
21. Ricco LS, Campo VL, Shelykh IA, Seridonio AC. Majorana oscillations modulated by Fano interference and degree of nonlocality in a topological superconducting-nanowire–quantum-dot system. *Phys Rev B* (2018) 98(7):075142. doi:10.1103/PhysRevB.98.075142
22. Ricco LS, de Souza M, Figueira MS, Shelykh IA, Seridonio AC. Spin-dependent zero-bias peak in a hybrid nanowire-quantum dot system: distinguishing isolated Majorana fermions from Andreev bound states. *Phys Rev B* (2019) 99(15):155159. doi:10.1103/PhysRevB.99.155159
23. López R, Lee M, Serra L, Lim JS. Thermoelectrical detection of Majorana states. *Phys Rev B* (2014) 89(20):205418. doi:10.1103/PhysRevB.89.205418
24. Tang HZ, Zhang YT, Liu JJ. Photon-assisted tunneling through a topological superconductor with Majorana bound states. *AIP Adv* (2015) 5(12):127129. doi:10.1063/1.4939096
25. Chi F, He TY, Wang J, Fu ZG, Liu LM, Liu P, et al.. Photon-assisted transport through a quantum dot side-coupled to Majorana bound states. *Front Phys* (2020) 8:254. doi:10.3389/fphy.2020.00254
26. Van der Wiel WG, De Franceschi S, Elzerman JM, Fujisawa T, Tarucha S, Kouwenhoven LP. Electron transport through double quantum dots. *Rev Mod Phys* (2002) 75(1):1–22. doi:10.1103/RevModPhys.75.1
27. Hanson R, Kouwenhoven LP, Petta JR, Tarucha S, Vandersypen LMK. Spins in few-electron quantum dots. *Rev Mod Phys* (2007) 79(4):1217–65. doi:10.1103/RevModPhys.79.1217

28. Chi F, Liu JL, Sun LL. Fano-Rashba effect in a double quantum dot Aharonov-Bohm interferometer. *J Appl Phys* (2007) 101(9):093704. doi:10.1063/1.2720097
29. Hofstetter W, König J, Schoeller H. Kondo correlations and the Fano effect in closed Aharonov-Bohm interferometers. *Phys Rev Lett* (2001) 87(15):156803. doi:10.1103/PhysRevLett.87.156803
30. Miroshnichenko AE, Flach S, Kivshar YS. Fano resonances in nanoscale structures. *Rev Mod Phys* (2010) 82(3):2257–98. doi:10.1103/RevModPhys.82.2257
31. de Guevara MLL, Claro F, Orellana PA. Ghost Fano resonance in a double quantum dot molecule attached to leads. *Phys Rev B* (2003) 67(19):195335. doi:10.1103/PhysRevB.67.195335
32. Li YX, Bai ZM. Tunneling transport through multi-quantum-dot with Majorana bound states. *J Appl Phys* (2013) 114(3):033703. doi:10.1063/1.4813229
33. Wang N, Lv SH, Li YX. Quantum transport through the system of parallel quantum dots with Majorana bound states. *J Appl Phys* (2014) 115(8):083706. doi:10.1063/1.4867040
34. Xia JJ, Duan SQ, Zhang W. Detection of Majorana fermions by Fano resonance in hybrid nanostructures. *Nanoscale Res Lett* (2015) 10:223. doi:10.1186/s11671-015-0914-3
35. Jiang C, Zheng YS. Fano effect in the Andreev reflection of the Aharonov-Bohm-Fano ring with Majorana bound states. *Solid State Commun* (2015) 212:14–18. doi:10.1016/j.ssc.2015.04.001
36. Ueda A, Yokoyama T. Anomalous interference in Aharonov-Bohm rings with two Majorana bound states. *Phys Rev B* (2014) 90(8):081405. doi:10.1103/PhysRevB.90.081405
37. Zeng QB, Chen S, Rong L. Fano effect in an AB interferometer with a quantum dot side-coupled to a single Majorana bound state. *Phys Lett* (2016) 380:951–7. doi:10.1016/j.physleta.2015.12.026
38. Ivanov TI. Coherent tunneling through a double quantum dot coupled to Majorana bound states. *Phys Rev B* (2017) 96(3):035417. doi:10.1103/PhysRevB.96.035417
39. Cifuentes JD, Dias da Silva LGGV. Manipulating Majorana zero modes in double quantum dots. *Phys Rev B* (2019) 100(8):085429. doi:10.1103/PhysRevB.100.085429
40. Tang LW, Mao WG. Detection of Majorana bound states by sign change of the tunnel magnetoresistance in a quantum dot coupled to ferromagnetic electrodes. *Front Phys* (2020) 8(5):147. doi:10.3389/fphy.2020.00147
41. Liu LM, Chi F, Fu ZG, Yu SC, Chen HW. Spin-polarized transport and spin Seebeck effect in triple quantum dots with spin-dependent interdot couplings. *Nanoscale Res Lett* (2018) 13(1):358. doi:10.1186/s11671-018-2773-1

Conflict of Interest: Authors LW and ZL are employed by China Monferrous Metal (Guilin) Geology and Mining Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 18 January 2021
doi: 10.3389/fphy.2020.629620

Differential Entropy Feature Signal Extraction Based on Activation Mode and Its Recognition in Convolutional Gated Recurrent Unit Network

Yongsheng Zhu[1] *and Qinghua Zhong*[1,2]*

[1]School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou, China, [2]South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China

In brain-computer-interface (BCI) devices, signal acquisition via reducing the electrode channels can reduce the computational complexity of models and filter out the irrelevant noise. Differential entropy (DE) plays an important role in emotional components of signals, which can reflect the area activity differences. Therefore, to extract distinctive feature signals and improve the recognition accuracy based on feature signals, a method of DE feature signal recognition based on a Convolutional Gated Recurrent Unit network was proposed in this paper. Firstly, the DE and power spectral density (PSD) of each original signal were mapped to two topographic maps, and the activated channels could be selected in activation modes. Secondly, according to the position of original electrodes, 1D feature signal sequences with four bands were reconstructed into a 3D feature signal matrix, and a radial basis function interpolation was used to fill in zero values. Then, the 3D feature signal matrices were fed into a 2D Convolutional Neural Network (2DCNN) for spatial feature extraction, and the 1D feature signal sequences were fed into a bidirectional Gated Recurrent Unit (BiGRU) network for temporal feature extraction. Finally, the spatial-temporal features were fused by a fully connected layer, and recognition experiments based on DE feature signals at the different time scales were carried out on a DEAP dataset. The experimental results showed that there were different activation modes at different time scales, and the reduction of the electrode channel could achieve a similar accuracy with all channels. The proposed method achieved 87.89% on arousal and 88.69% on valence.

Keywords: differential entropy, signal extraction, activation mode, convolutional neural network, bidirectional gated recurrent unit network

OPEN ACCESS

Edited by:
Zichuan Yi,
University of Electronic Science and Technology of China, China

Reviewed by:
Congcong Ma,
Nanyang Institute of Technology, China
Junsheng Mu,
Beijing University of Posts and Telecommunications (BUPT), China
Guodong Liang,
Jinan University, China

***Correspondence:**
Qinghua Zhong
zhongqinghua@m.scnu.edu.cn

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *15 November 2020*
Accepted: *04 December 2020*
Published: *18 January 2021*

Citation:
Zhu Y and Zhong Q (2021) Differential Entropy Feature Signal Extraction Based on Activation Mode and Its Recognition in Convolutional Gated Recurrent Unit Network. Front. Phys. 8:629620. doi: 10.3389/fphy.2020.629620

INTRODUCTION

Signal recognition plays an important role in BCI devices [1]. The ability of perceived robots for expressing similar human behaviors is considered to be more approachable and humanized, which can obtain higher participation and more pleasant interaction in reality [2]. In recent years, an increasing number of researchers are attracted to the research of signal recognitions by computers. Electroencephalogram (EEG) signals can avoid the camouflage and subjectivity of human behaviors [3]. Therefore, feature signal recognition based on BCI devices is becoming a research focus.

At present, there are two technical problems in the process of feature signal recognition based on BCI devices. One is how to extract distinctive feature signals from original signals, and the other is how to establish a more effective feature recognition calculation model [4, 5]. Fast Fourier transform (FFT) was a common method to extract feature signals from the original signal [6]. However, FFT cannot reflect temporal information in frequency signal, so a short-time Fourier transform was used to extract time-frequency domain features which were recognized as a feature signal [7]. Human brain is a nonlinear dynamic system. It is difficult to analyze the original signal by traditional time-frequency feature extraction and analysis methods. So, by calculating the DE of the original signal, the differential asymmetry and rational asymmetry signals of the symmetrical electrodes in the left and right hemispheres of a brain were used for feature signal recognition, which achieved an average recognition accuracy of 69.67% on the DEAP dataset [8]. However, this could only explore the relationship between symmetric electrodes, not connect all electrodes in a spatial position. Recent research has shown that distinctive feature signals were closely related to multiple areas of the cerebral cortex in BCI [9]. The weights of brain areas were calculated by attention mechanism and the sum of weights was taken as the contribution value of brain areas, which showed that frontal lobe areas play an important role in feature signal recognition experiments [5]. The feature signals of different activation areas were extracted by DE and PSD topographic distribution, which found that prefrontal and temporal lobes of the cerebral cortex were related to feature signal states [10]. However, they did not use the relevant brain areas to improve the recognition rate of feature signals. Hence, a feature extraction method of multivariate empirical mode decomposition (MEMD) was used to select feature signal of appropriate channels, which achieved 75.00% on arousal and 72.87% on valence for feature signal recognition based on an Artificial Neural Network (ANN) classifier [11]. However, the traditional machine learning model is unable to extract more subtle feature signals, which could lead to a low performance of feature signal recognition. In recent years, feature signal recognition methods based on deep learning have developed rapidly. Especially, CNN model has become a leading method to improve recognition performance. A method of feeding time-frequency features of each channel into a 2DCNN model for feature signal state recognition was proposed, which achieved 78.12% on arousal and 81.25% on valence [12]. The original signal was decomposed into time frames, and the multi-channel time frame signals were used as inputs of a 3DCNN model, which achieved a recognition accuracy of 88.49% on arousal and 87.44% on valence [13]. The frequency domain feature, spatial feature and frequency band features of fusion multi-channel signals were fed into a Capsule Network (CapsNet) based on CNN, which achieved 68.28% on arousal and 66.73% on valence [14]. Although the CNN model can effectively extract the spatial information from feature signals, it cannot effectively extract the temporal information. Therefore, a hybrid neural network model combined CNN and Recurrent Neural Network (RNN) was proposed [15]. They used CNN model to extract the correlation of signals in physical adjacent channels, and used RNN model to mine the context information of feature signal sequences, which achieved 74.12% on arousal and 72.06% on valence. A Stack AutoEncoder (SAE) was used to establish a linear mixed model, and a long-short-term memory recurrent neural network (LSTM-RNN) was used for feature signal recognition, which could achieve 81.10% on arousal and 74.38% on valence. However, the unidirectional RNN and LSTM cannot backward learn the feature signal sequences, which was the reason of a low recognition rate.

For solving existing problems in previous studies, firstly, considering that different areas played different roles in feature signal recognition, activation pattern was introduced to reflect the weight of region contribution. So, a method of the DE feature signal extraction based on an activation mode was proposed. Secondly, a 1D and 3D feature signal representation method of considering the spatial-temporal information were also proposed, which could improve the recognition rate of feature signals by utilizing the temporal information of different areas and spatial connection of electrode positions. Lastly, a recognition framework based on Convolutional Gated Recurrent Unit network were proposed in this paper. The recognition framework was composed of 2DCNN and BiGRU in parallel, which could not only learn more distinctive and robust feature signals but also improve the recognition rate.

METHODLOGY

DE Feature Signal Extraction

Original signals collected by the BCI include rhythm signals, event-related potentials, and spontaneous potential activity signals [4]. A Butterworth filter [16] is used to decompose the original signal (X) into four frequency band signals: X_θ, X_α, X_β, and, X_γ where θ is 4–7 Hz, α is 8–13 Hz, β is 14–30 Hz, and γ is 31–45 Hz.

DE Algorithm

DE is suitable for decoding characteristic signals [7, 10]. Each frequency band signal is divided into X_i/τ equal parts by a time window τ, and then analyzed by a DE algorithm. DE can discretize the value of continuous random variables. The signal sequence values are divided into small parts with Δx. According to the mean value theorem, there is always a value x_i in each part to make **Eq. 1** true.

$$\int_{i\triangle x}^{(i+1)\triangle x} p(x)dx = p(x_i)\triangle x \tag{1}$$

where $p(x_i)$ is a probability density function of discrete signals. Each point at i is assigned to x_i, and then the **Eq. 1** is substituted into the Shannon formula for the discrete variables. The process is shown in **Eq. 2**.

$$\begin{aligned} H(\triangle x) &= -\sum_{i=1}^{n} p(x_i)\triangle x \ln\left[p(x_i)\triangle x\right] \\ &= -\sum_{i=1}^{n} p(x_i)\triangle x \ln p(x_i) - \sum_{i=1}^{n} p(x_i)\triangle x \ln\triangle x \end{aligned} \tag{2}$$

When $\triangle x$ approaches 0, $\sum_{i=1}^{n} p(x_i)\triangle x$ approaches 1 and $\ln\triangle x$ approaches $-\infty$. So, the right side of the **Eq. 2** approaches ∞, and the left side of **Eq. 2** is seen as the DE of a continuous signal in **Eq. 2**. The DE can be defined as **Eq. 3**.

$$H(X) = -\int_{X} f(X)\log[f(x)]dx \tag{3}$$

where X is a random variable, $f(x)$ is a probability density function of X. Assuming that the original signal X obeys normal distribution $N(\mu, \sigma^2)$, the DE can be solved as **Eq. 4**.

$$\begin{aligned} H(X) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}} log\left[\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{\sigma^2}}\right] dx \\ &= \frac{1}{2} log\left(2\pi e\sigma^2\right) \end{aligned} \tag{4}$$

where μ is a mean of X, and σ^2 is a variance of X. In **Eq. 4**, the DE of signal source X_i can be calculated as long as σ^2 is known, and the variance of normal distribution $N(\mu, \sigma^2)$ can be calculated via **Eq. 5**.

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{n} x_i^2 \tag{5}$$

The spectral energy of the discrete signal is defined as $P = \int_{-\infty}^{+\infty} f^2(t)\,dt$. According to **Eq. 5**, the variance of signal source X_i is an average spectral energy value P. From **Eq. 4**, the variance of X_i is a constant multiple (P_i/N^2) of the spectral energy in each frequency band. So, the DE of a specific frequency band can be defined as **Eq. 6**.

$$H_i(X) = \frac{1}{2} log\left(2\pi e\sigma_i^2\right) = \frac{1}{2} log\left(P_i\right) + \frac{1}{2} log\left(\frac{2\pi e}{N}\right) \tag{6}$$

where $H_i(X)$ is the DE of X_i, P_i is a spectral energy of X_i, ${\sigma_i}^2$ is a variance of X_i, and N is a constant.

DE Feature Signal Vector

A distinctive feature vector is constructed by using $H_i(X)$, the processing process for a baseline signal of a specific frequency band can be expressed as **Eq. 7**.

$$v_j^i = v_{trail}^i(j) - \frac{\sum_{i=1}^{m} v_{base}^i(k)}{m},\ j \in \left(N = \frac{t}{\tau}\right) \tag{7}$$

Where t is a total signal time, τ is a time sliding window, v_j^i is the final DE feature vector at *jth* segment of *ith* band, $v_{trail}^i(j)$ is a DE feature vector at *jth* segment of *ith* frequency band, $v_{base}^i(k)$ is a DE feature vector of a baseline signal in *jth* segment of *ith* band, and m is a number of segments in baseline signals. So, a 1D DE feature signal vector can be expressed as $v_j^i = [c_\tau^1, c_\tau^2, \ldots, c_\tau^n]^T \in R^n$, where n is the number of electrode channels, c_j^n is the pre-processed signal at *nth* channel of *jth* segment.

The 1D data of n channels are filled into the space electrode position of $d \times d$, and the unused electrode position is filled with zero value. Then, a 2D matrix ($\mathbf{f}_\tau$) can be obtained. In order to make the matrix denser, a radial basis function (RBF) interpolation of Gaussian kernel function [17] is used to fill in zero values. This process can be expressed as **Eq. 8**.

$$\hat{\mathbf{f}}_\tau(c) = \sum_{i=1}^{n} e^{\left(-\frac{\left\|x-c_\tau^i\right\|^2}{2\sigma^2}\right)} \tag{8}$$

Where σ is an extension constant of the RBF function, x is a center point, c is an electrode channel point, and $\|\cdot\|$ is 2-norm.

Convolutional Gated Recurrent Unit Network

The convolutional gated recurrent unit network is composed of 2DCNN and BiGRU in parallel, as shown in **Figure 1**.

Structural Principle of 2DCNN

CNN is a kind of forward feedback neural network. The model structure mainly includes input layers, hidden layers and output layers. The network structure of 2DCNN is shown in **Figure 2**. The feature signal matrix $\hat{\mathbf{f}}_\tau(c)$ is used as the input of 2DCNN. The abstract feature extraction of the DE feature signal is completed by setting the size of the 2D filters, the process can be defined as **Eq. 9**.

$$s(i, j) = \left(\hat{\mathbf{f}}_\tau * \mathbf{W}\right) = \sum_m \sum_n f(i+m,\ j+n) w(m, n) \tag{9}$$

where $\mathbf{W}$ is the convolution kernel, (m, n) is the size of the convolution kernel $\mathbf{W}$, $\hat{\mathbf{f}}_\tau$ is the input matrix, (i, j) is the matrix coordinate. After each convolution operation, the feature data of each layer is batch-normalized (BN), and a RELU activation function is added to make the model have nonlinear feature transformation capability. The RELU function is expressed as **Eq. 10**.

$$RELU(x) = max(x, 0) = \begin{cases} x, & if\ x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where *max* is the maximum function, x is the inputs. The feature matrix $\mathbf{S}$ is fed into the fully connected layer to make it more expressive in space. The process is shown as **Eq. 11**.

$$\mathbf{FS} = FC(\mathbf{S}),\ \mathbf{FS} \in R^{1024} \tag{11}$$

where R^{1014} represents a dimension of 1,024, *FC* is the fully connection layer, and **FS** is a 1,024-dimensional vector.

Structural Principle of BiGRU

GRU [18] is an improvement of LSTM [19]. Compared with LSTM, GRU is capacity of dealing with a smaller amount of data, which has a faster calculation speed and can better solve the problem of gradient disappearance. The schematic diagram of GRU is shown in **Figure 3A**. The GRU processes sequence information by resetting gate r_z and updating gate z_t, and its parameter update equation is shown in **Eqs. 12–15**.

$$r_t = \sigma(w_r x_t \oplus U_r h_{t-1}) \tag{12}$$

$$z_t = \sigma(w_z x_t \oplus U_z h_{t-1}) \tag{13}$$

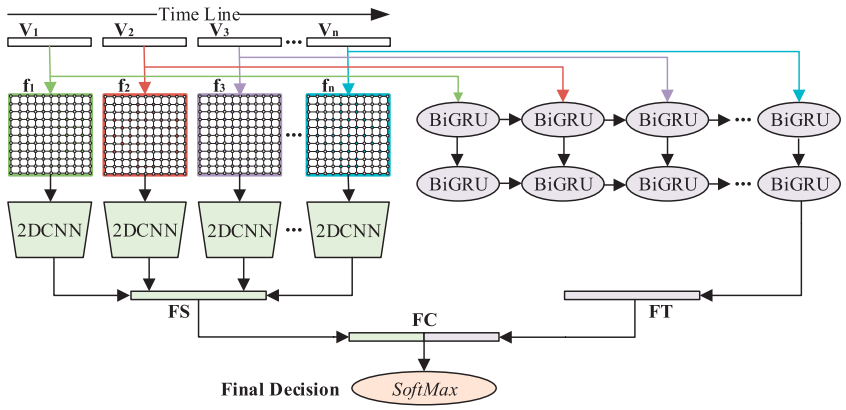

FIGURE 1 | Schematic diagram of DE feature signal recognition model based on 2DCNN-BiGRU. The 1D feature sequences of each time step are fed into BiGRU to extract the time information of feature signals. The 2D feature matrices of each time step are fed into 2DCNN to extract the spatial information of feature signals. Finally, the decision layer gives the recognition result.

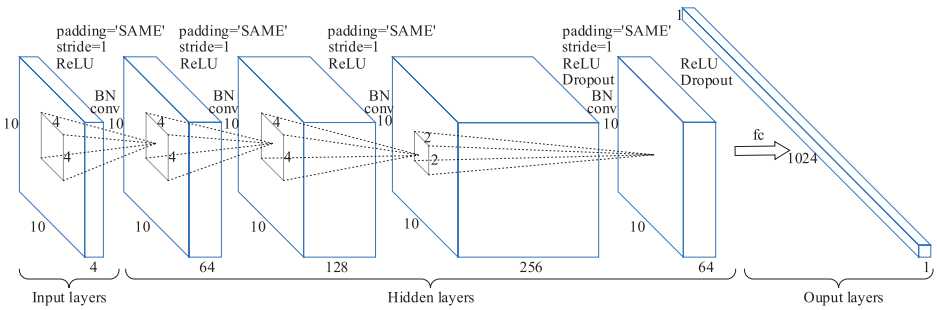

FIGURE 2 | Schematic diagram of the 2DCNN model. The inputs are a 3D DE feature matrix of 10 × 10 × 4, and its construction process is shown in *Construction of 3D DE Feature Matrix* section. Padding = 'SAME' represents the edge information is filled with zeros, Stride = 1 means that the step size of each convolution operation is 1, and Dropout means that hidden neurons are deleted randomly.

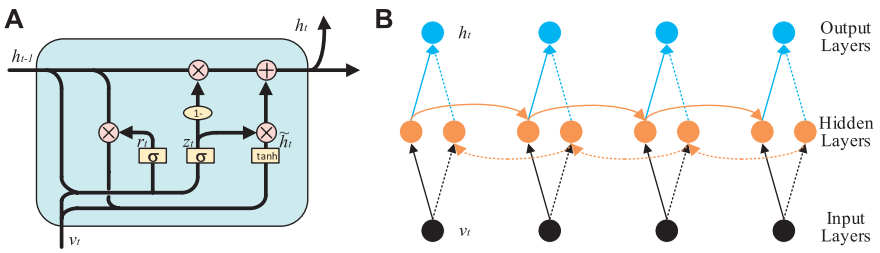

FIGURE 3 | (A) Schematic diagram of GRU network. **(B)** Schematic diagram of BiGRU network.

$$\tilde{h}_t = \tanh[w_h x_t \oplus U_h(r_t \otimes h_{t-1})] \quad (14)$$

$$h_t = (1 - z_t) \otimes h_{t-1} \oplus z_t \otimes \tilde{h}_t \quad (15)$$

where w_r, w_z, w_h, U_r, U_z, and U_h are the weight parameters of the BiGRU network, r_t is reset gate, z_t is update gate, $\tilde{h}_t$ is candidate activation unit, h_t is the hidden unit at time t, h_{t-1} is the hidden unit at time $t-1$, σ is the activation function, V_t is the GRU input at time t, ⊗ represents multiplying by elements, and ⊕ represents adding by elements.

DE feature matrix **V** is exploited to be the original input of BiGRU network. The BiGRU network is composed of forward GRU, backward GRU, and forward-backward output state connection layers. The structure of BiGRU network is shown in **Figure 3B**, which mainly includes input layers, hidden layers and output layers.

TABLE 1 | Specific experimental environment.

Name	Version
CPU	Intel Core i7-9750H @2.60 GHz
GPU	NVIDIA GeForce RTX 2060 6 GB
RAM	DDR4 16 GB
OS	Windows 10
Frameworks	Tensorflow-GPU 1.14.0, MATLAB 2019b

EXPERIMENTAL RESULTS AND DISCUSSION

In this part, the experimental processes would be introduced and our method would be compared with other methods. Then, the effectiveness of our framework was evaluated on the DEAP dataset. To achieve a more reliable emotion recognition process, the emotion recognition performance of the EEG access was analyzed by a 5-fold cross-validation technology.

Experimental Environment and Experimental Dataset

Table 1 Shows the specific experimental environment for experiments.

In the DEAP dataset [4], EEG signals of 32 subjects who watched 40 1-minute music videos were recorded, and each subject contained 63s EEG data of 32 electrode channels. Among them, the first 3s was the baseline signal recorded in the relaxed state, and the last 60s was the trial signal recorded when watching the videos.

According to the level of arousal and valence, the distinctive categories of DE feature signal states were obtained. In our experiment, the DE feature signal recognition was divided into two binary classifications. If scores of the arousal or valence were less than or equal to 5, the label was marked as low. If scores were greater than 5, the label was marked as high. Thus, there were four labels on arousal and valence: high arousal (HA), low arousal (LA), high valence (HV) and low valence (LV).

Data Preprocessing

In order to improve the accuracy of recognition, the influence of baseline signals on trial signals needs to be considered. Before extracting the DE feature signal of original signals, the original signals are usually divided into short time frames [15, 19, 20]. The baseline signal was divided into three segments with a 1s sliding window and the trial signal into $n = 60/\tau$ segments with a τwindow. As shown in **Figure 4**, a channel signal of the original data was taken out, and the original signal of each second is decomposed into θwave, αwave, βwave, and γwave through the Butterworth filters. The DE vectors Base_Vector (i) of the three baseline signals and the DE vectors Trail_Vector (n) of the trial signals in each time window were calculated by the DE algorithm. Then, n-feature signals ($DE^1_{i=\theta,\alpha,\beta,\gamma}$) in this channel could be obtained by Trail_Vector (n) minus Mean$\sum_{i=1,2,3}$ Base_Vector (i) in turn. Finally, the 1D vector $DE_i = [DE_i^1, DE_i^1, \ldots, DE_i^{32}]$ of 32 electrode channels could be obtained, where i is the frequency band.

Construction of 3D DE Feature Matrix.

The DE feature signal value of 32 channels was filled to the orange position in **Figure 5B**, and the gray point was filled with zero values. The electrodes circled in orange were the test points used in the DEAP dataset, as shown in **Figure 5A**. The electrodes of the international 10–20 system [10] were connected with the test electrodes of the DEAP dataset, which could construct a square matrix $N \times N$ (N is the maximum number of points between the horizontal test points and the vertical test points). In

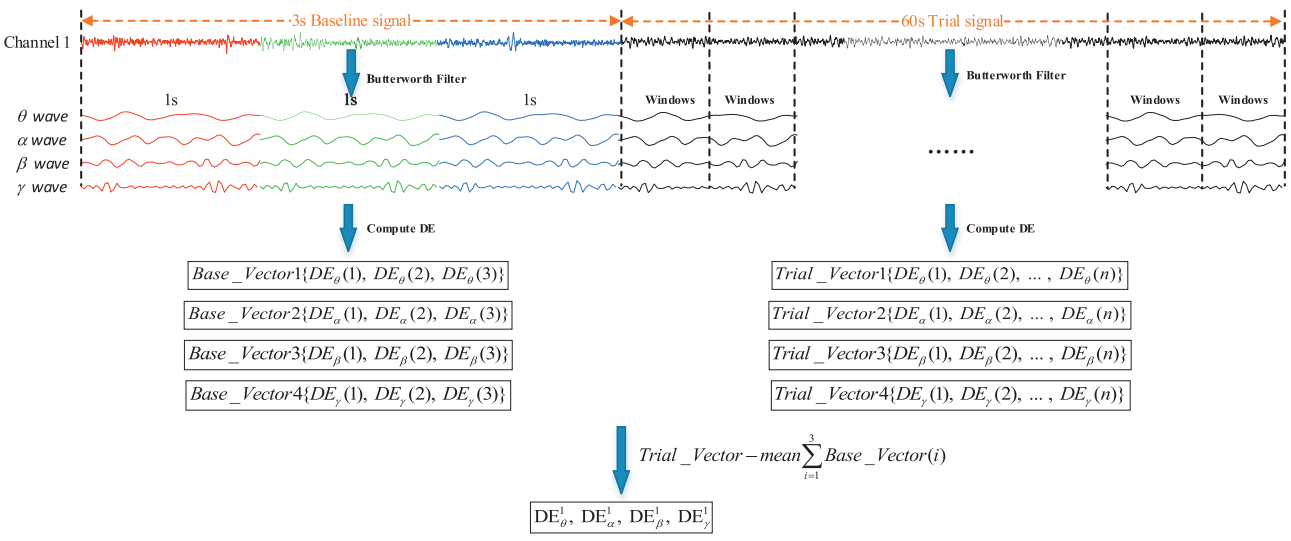

FIGURE 4 | Schematic diagram of the original signal preprocessing flow. Among them, Trail_Vector (n) is the baseline signal feature vector. Trail_Vector (n) is the trial signal feature vector. $DE_i = [DE_i^1, DE_i^1, \ldots, DE_i^{32}]$ is the DE feature signal vector of 32 channels. Mean ∑ Base_Vector (i) is the sum average DE of three baseline segmentations.

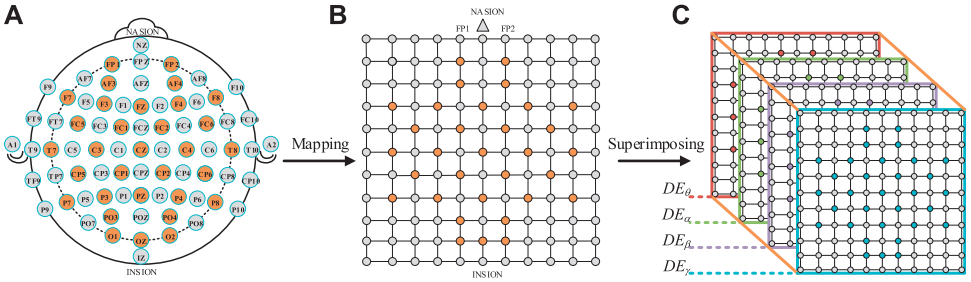

FIGURE 5 | Schematic diagram of a 3D feature signal matrix construction. **(A)** International 10–20 system [10]. **(B)** A 2D square matrix of 32 EEG channels. **(C)** A 3D feature matrix of DE_θ, DE_α, DE_β and DE_γ.

TABLE 2 | The number of feature signal samples of each frequency band signal at 1, 10, 30, and 60s time windows.

Windows (s)	Arousal			Valence		
	LA	HA	Total	LV	HV	Total
1	34,320	42,480	76,800	32,580	44,220	76,800
10	3,432	4,248	7,680	3,258	4,422	7,680
30	1,144	1,416	2,560	1,086	1,474	2,560
60	572	708	1,280	543	737	1,280

addition, in order to avoid the loss of edge information, a layer of gray unused points was added to the outer layer of the matrix, as shown in **Figure 5B**. In order to make the matrix denser, the RBF interpolation was used to fill in the zero values [17]. Finally, a 3D feature matrix was obtained by stacking the 2D feature matrices of four frequency bands, as shown in **Figure 5C**.

The sliding windows of 1, 10, 30 and 60s were used to divide the original signals, and the number of DE feature signal samples obtained is shown in **Table 2**. Notably, the time step window of 60s was the original signal length. The total samples of each frequency band were $32 \times 40 \times n$, where 32 was the number of subjects, 40 was the number of experiments of each subject, and n was the number of signals divided by the time window. Finally, the same number of samples of the 1D feature signal vectors and 2D feature signal matrices of each frequency band were obtained.

2DCNN-BiGRU Model Training and Parameter Setting

The 1D feature signal vectors and 3D feature signal matrices were fed into BiGRU model and 2DCNN model respectively. The proposed model was implemented with Tensorflow framework and trained on an NVIDIA GeForce RTX 2060 GPU. The Adam optimizer was adopted to minimize the cross-entropy loss function. The keep probability of dropout operation was 0.5. The penalty strength of L2 was 0.5. The hidden sates of the GRU cell is the number of channels. The learning rate was initialized to 0.001. When the verification errors of the model stopped dropping, the learning rate was divided by 10 until the iteration stopped.

In the 2DCNN model of the first three convolutional layers, 64, 256, and 512 convolution kernels with a size of 4×4 were used respectively. In order to reduce the amount of calculation, 64 convolution cores with a size of 2×2 were used in the fourth convolution layer, which added a dropout operation. In addition, in each convolutional layer, stride was set to 1, padding was set to SAME, and zero padding was used to prevent information from being lost at the edge of the inputs. A fully connected layer was used to convert input features into spatial abstract features. In order to avoid learning overfitting and improve the generalization ability of the model, L2 regularization was added to the network. And then, two layers BiGRU were used to fused the temporal features obtained by the BiGRU model with the spatial features obtained by the 2DCNN model. Finally, the DE feature signal recognition result was obtained through a *SoftMax* classifier.

Results of 2DCNN-BiGRU in All Electrode Channels

The 5-fold cross-validation technology was used to validate all subjects and the recognition results of θ frequency band, α band signal, β frequency band signal, γ band signal and four band signal combinations were counted at four-time windows. The recognition results were shown in **Table 3**. In the dimensions of arousal and valence, the high frequency band (β and γ) had higher average recognition accuracy than the low frequency band (θ and α), which showed that the high frequency band had more abundant DE feature signal information. It also could be observed that the accuracy of all band combinations was higher than a single band. The 2DCNN-BiGRU model achieved the highest average recognition accuracy of 87.20 and 87.90% on arousal and valence at the 10s sliding window.

Results of DE Feature Signal Recognition in Activation Mode

In order to explore the influence of electrode channels on the recognition rate of DE feature signals, the activation model of DE feature signals and PSD feature signals were studied [19]. The DE feature signals were classified by reducing the electrode channels

TABLE 3 | When the inputs of the 2DCNN-BiGRU model were data of 32 electrodes, the DE feature signal recognition results of each frequency band signal and all frequency band combinations under the time window of 1, 10, 30, and 60s, respectively.

Windows (s)	Arousal accuracy (%)					Valence accuracy (%)				
	θ	α	β	γ	All	θ	α	β	γ	All
1	69.85	70.38	70.08	73.72	81.69	70.48	70.43	69.51	74.17	82.13
10	73.07	77.97	78.41	78.34	**87.20**	72.32	74.56	77.59	76.65	**87.90**
30	67.95	67.94	70.86	69.85	78.66	68.15	69.52	72.97	69.92	79.35
60	59.41	57.73	59.14	61.25	62.43	58.73	61.10	61.33	61.66	63.52

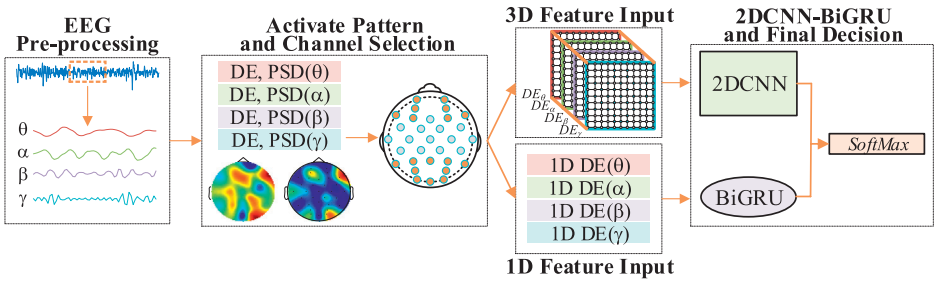

FIGURE 6 | A brief frame diagram of DE feature signal recognition of convolutional gated recurrent network based on activation mode.

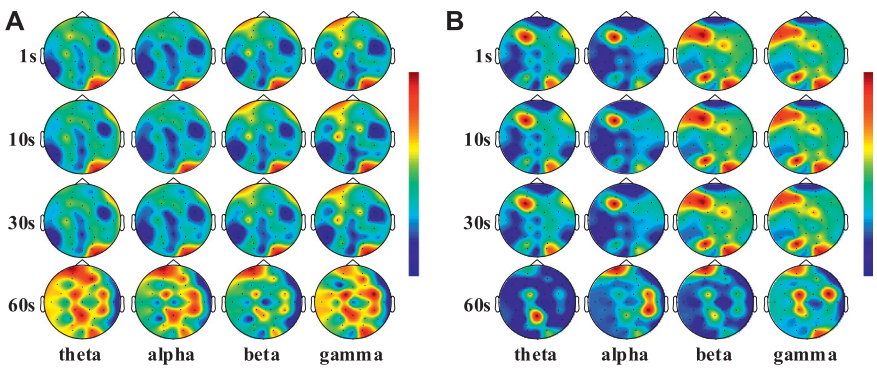

FIGURE 7 | (A) Scalp distribution of the DE feature signal at 1, 10, 30, and 60s windows of different frequency bands. **(B)** Scalp distribution of the PSD feature signal at 1, 10, 30, and 60s windows of different frequency bands.

under the activation model. A brief framework for the recognition process is shown in **Figure 6**.

In the DEAP dataset, the average value of DE and PSD were calculated, which were from the 32 electrode channels of all subjects at different time windows. **Figure 7A** and **Figure 7B** showed the averaged PSD and DE distribution, where four frequency bands (theta, alpha, beta and gamma) represented four activation models. It was found that the electrode channels located in the frontal and occipital lobes had a higher activation capacity. However, different time windows have similar activation patterns on different frequency bands, which is the reason for the lower recognition accuracy of DE feature signals. The activation ability of high frequency bands (beta and gamma) is greater than that of low frequency band (theta and alpha), which also explains that beta and gamma bands have better recognition effect than theta and alpha bands. According to the spatial locations of the electrodes, the 32 electrodes used in the DEAP dataset were divided into five clusters, namely, five brain areas, as shown in **Figure 8A**. **Table 4** summaries the electrode channels in each brain region, where the frontal lobe represents the electrodes of FP1, AF3, F7, F3, FP2, AF4, F8, F4, and FZ, the central lobe represents the electrodes of FC1, CP1, C3, FC2, CP2, C4, and CZ, the temporal lobe represents the electrodes of FC5, T7, CP5, FC6, T8, and CP6, the parietal lobe represents the electrodes of P7, P3, P8, P4, and PZ, and the occipital lobe represents the electrodes of PO3, O1, PO4, O2, and OZ.

According to the activation areas of each time window, the combinations of different areas were selected. As shown in **Figure 8B**, the frontal, parietal, and occipital areas were considered as the DE feature signal activation areas at 1, 10, and 30s windows. The number of electrodes were reduced from 32 to 19, where the selected electrodes were FP1, AF3, F7, F3, FP2,

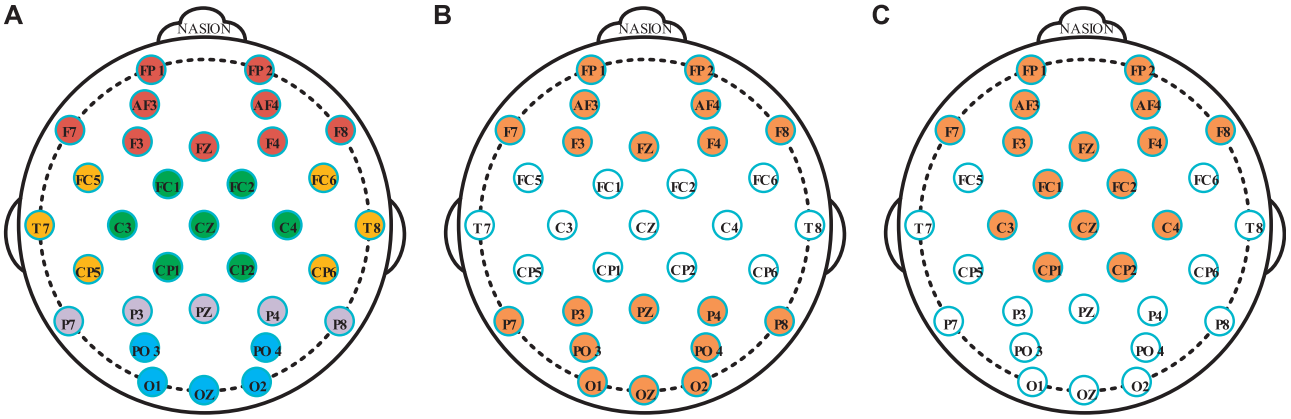

FIGURE 8 | (A) The 32 electrodes of the DEAP dataset are divided into five groups, and the same color represents the same group of areas. **(B)** The 19 electrodes are selected by the activation modes at the 1, 10, and 30s windows. **(C)** The 16 electrodes are selected by the activation modes at the 60s windows.

TABLE 4 | The 32 electrodes in the DEAP dataset are divided into five areas and the electrodes represents by each brain area.

Brain areas	Electrodes name
Frontal	FP1, AF3, F7, F3, FP2, AF4, F8, F4, FZ
Central	FC1, CP1, C3, FC2, CP2, C4, CZ
Temporal	FC5, T7, CP5, FC6, T8, CP6
Parietal	P7, P3, P8, P4, PZ
Occipital	PO3, O1, PO4, O2, OZ

AF4, F8, F4, FZ, P7, P3, P8, P4, PZ, PO3, O1, PO4, O2, and OZ. Under the 60s-time step window, the frontal lobe and central area were used as the activation areas of the DE feature signals, as shown in **Figure 8C**. The number of electrodes were reduced from 32 to 16, and the selected electrodes were FP1, AF3, F7, F3, FP2, AF4, F8, F4, FZ, FC1, CP1, C3, FC2, CP2, C4, and CZ.

The DE feature signals of the four frequency bands were used as the inputs of the 2DCNN-BiGRU model, and the DE feature signal recognition experiments with the selected electrode were performed at each time window. The recognition results were shown in **Figure 9**. At 1s window, the recognition rate of 19 electrodes improved by 0.06% on valence compared with 32 electrodes, while decreased by 0.69% on arousal. At 10s window, the recognition rate of 19 electrodes improved by 0.79% on arousal and 0.06% on valence compared with 32 electrodes. At 30 s window, the recognition rate of 19 electrodes decreased by 0.77% on arousal and 1.21% on valence compared with 32 electrodes. At 60 s window, the recognition rate of 16 electrodes improved by 0.04% on arousal and 0.19% on valence compared with 32 electrodes. Notably, when the time window was 10 s, the 2DCNN-BiGRU model achieves the highest accuracy. Experimental results showed that there were different activation modes at different time scales. By reducing the number of electrodes in the activation mode, not only could achieve the recognition rate which was similar to all electrodes of DE feature signal recognition, but also the performance and robustness of the recognition models could be improved.

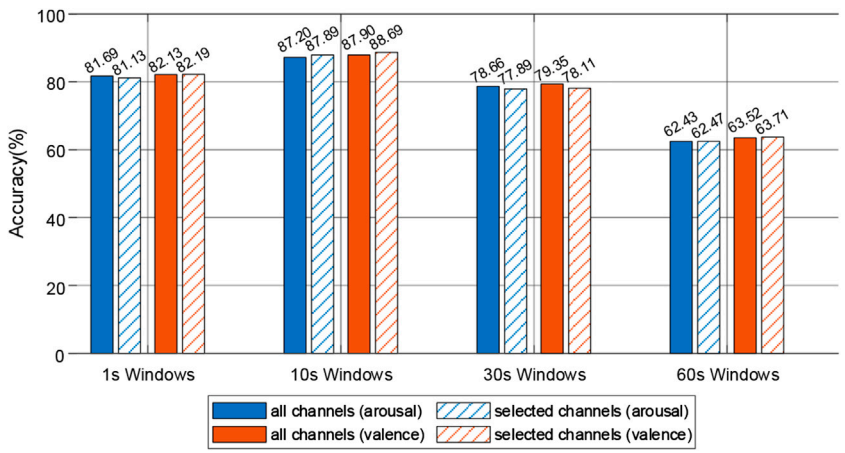

FIGURE 9 | DE feature signal recognition results of 2DCNN-BiGRU in activation mode.

TABLE 5 | The structure and inputs of 2DCNN, BiLSTM, BiGRU, 2DCNN-BiLSTM, and 2DCNN-BiGRU models.

Models	Data inputs	Network structures
2DCNN	3D matrix	4 × (conv2D) + FC + SoftMax
BiGRU	1D vector	2BiGRU + FC + SoftMax
BiLSTM	1D vector	2BiLSTM + FC + SoftMax
2DCNN-BiLSTM	2DCNN: 3D input BiLSTM: 1D input	4 × (conv2D) + FC + FC + SoftMax 2BiLSTM + FC
2DCNN-BiGRU	2DCNN: 3D input BiGRU: 1D input	4 × (conv2D) + FC + FC + SoftMax 2BiGRU + FC

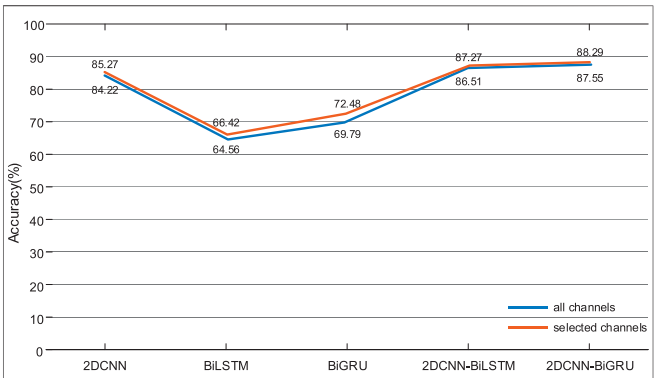

FIGURE 10 | At the 10s windows, the DE feature signal recognition results of 2DCNN, BiLSTM, BiGRU, 2DCNN-BiLSTM, and 2DCNN-BiGRU in the activation mode.

In order to further verify the reduction of electrode channels could achieve similar accuracy to all electrodes, and to verify that the hybrid model is better than the single model, four models of 2DCNN, BiLSTM, BiGRU, and 2DCNN-BiLSTM are compared with the 2DCNN-BiGRU. **Table 5** shows the structure and inputs of different models.

In the experiment, the data of the 10s window were used as the inputs of the models, and the sum average of arousal and valence as the final results. In order to make the experiment comparable, the convolutional kernels of each model and the fully connected layer parameter settings were consistent in experiments. The DE feature signal recognition rate of each model was shown in **Figure 10**. The recognition rate of the selected electrodes was slightly higher than that of all electrodes in different models. The recognition rate of 2DCNN-BiLSTM and 2DCNN-BiGRU is higher than that of 2DCNN, BiLSTM and BiGRU, which indicated that the hybrid models could effectively extract the spatial-temporal features of DE feature signals. The recognition rate of the 2DCNN-BiGRU model was slightly higher than that of the 2DCNN-BiLSTM model, which indicated that the GRU unit was superior to the LSTM unit in handling small samples.

Comparison of the results of different experimental methods.

The proposed method was compared with the current recognition methods based on feature signals, which were applied to the DEAP dataset. As shown in **Table 6**, the binary classification experiments of valence and arousal were carried out, and the similar methods were followed to evaluate the recognition accuracy.

TABLE 6 | Comparison of the different experimental methods in the DEAP dataset.

Studies	Models	Feature signals	Evaluation methods	Accuracy (%)	
				Arousal	Valence
Chen et al. [23]	HMM	Fusion feature	5-Fold	73.00	75.63
Zhuang et al. [22]	SVM	Intrinsic mode functions	leave-one-trail-out	69.10	71.99
Mert et al. [11]	ANN	MEMD-based features	leave-one-trail-out	75.00	72.87
Li et al. [15]	CNN-LSTM	Wavelet transform	5-Fold	74.12	72.06
Alhagry et al. [24]	LSTM	Raw	4-Fold	85.65	85.65
Kwon et al. [12]	2DCNN	Wavelet transform	10-Fold	78.12	81.25
Xing et al. [16]	LSTM	Frequency band power	10-Fold	74.38	81.10
Our proposed method	2DCNN-BiGRU	Differential entropy	5-Fold	**87.89**	**88.69**

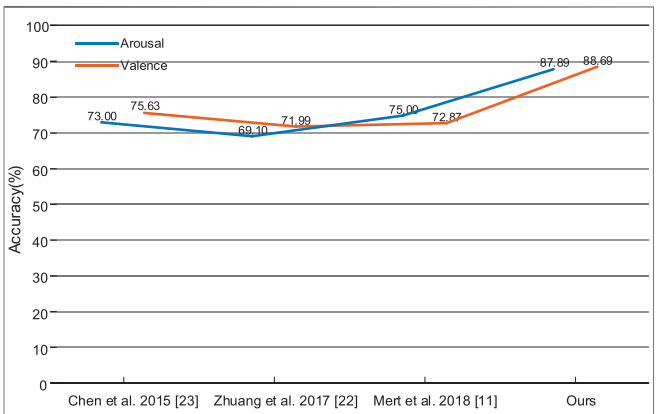

FIGURE 11 | The proposed method is compared with the traditional machine learning methods.

Our model was compared with traditional machine learning models of HMM [21], SVM [22] and ANN [11], as shown in **Figure 11**. The accuracy of our method improved by 12.89% on arousal and 13.06% on valence, which showed that the DE feature signal recognition based on deep learning method could deeply extract more subtle abstract features and achieve higher recognition rate.

In order to further verify the effectiveness of the proposed method, the 2DCNN-BiGRU model was compared with the latest deep learning methods, such as 2DCNN [12], LSTM [19, 23] and CNN-LSTM [15], as shown in **Figure 12**. Compared with the 2DCNN model, our model improved by 9.77% on arousal and 7.44% on valence, which indicated that BiGRU could handle the

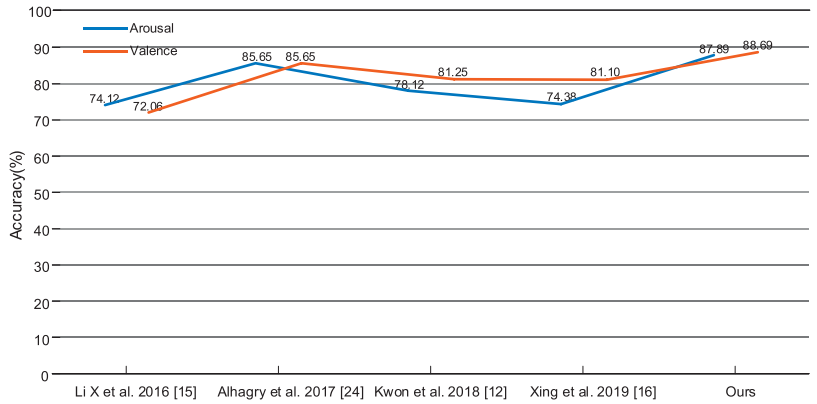

FIGURE 12 | The proposed method is compared with other deep learning methods.

dynamic information of deep feature signal sequences. Compared with the LSTM model, our model improved by 2.24% on arousal and 3.04% on valence, which indicated that the 2DCNN model could effectively extract the spatial feature. Compared with the CNN-LSTM model, our model improved by 13.77% on arousal and 16.63% on valence, which indicated that our spatial-temporal features based on the activation modes are more effective. In addition, the DE feature signal was compared with wavelet transform (WT) [12, 15] power spectral density (PSD) [19] and raw signals [24], which showed that the DE feature signals are more effective in our model. However, on one hand, the hybrid 2DCNN-BiGRU model contains massive amounts of parameters, which is necessarily unfriendly to hardware devices. On the other hand, a more advanced Graph Convolution Network (GCN) [23, 25] can be considered to further explain the relationship between the electrodes.

CONCLUSION

In this paper, a DE feature signal extraction method based on an activation mode and its recognition in a Convolutional Gated Recurrent Unit network were proposed. The DE and PSD feature signals were used to mine activation patterns at different time scales to reduce electrode channels. The 1D temporal and 3D spatial feature signals were respectively fed into 2DCNN and BiGRU models, which achieved a recognition accuracy of 87.89% on arousal and 88.69% on valence of the DEAP dataset. It was found that DE feature signals of reducing electrode channels could achieve similar recognition accuracy to all electrode channels, which was of great significance to develop a recognition device based on BCI system.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://www.eecs.qmul.ac.uk/mmv/datasets/deap/index.html.

AUTHOR CONTRIBUTIONS

YZ designed the framework, conducted experiments and wrote the manuscript. QZ carried out experiments and analyzed the results and presented the discussion and conclusion parts.

FUNDING

This research was funded by the National Natural Science Foundation of China (No. 61871433), the Natural Science Foundation of Guangdong Province (No. 2019A1515011940), the Science and Technology Program of Guangzhou (Nos. 202002030353 and 2019050001), the Science and Technology Planning Project of Guangdong Province (Nos. 2017B030308009 and 2017KZ010101), the Guangdong Provincial Key Laboratory of Optical Information Materials and Technology (No. 2017B030301007), and the Guangzhou Key Laboratory of Electronic Paper Displays Materials and Devices.

REFERENCES

1. Korovesis N, Kandris D, Koulouras G, Alexandridis A. Robot motion control via an EEG-based brain–computer interface by using neural networks and alpha brainwaves. *Electronics* (2019) 8(12):1387–1402. doi:10.3390/electronics8121387
2. Xiao G, Ma Y, Liu C, Jiang D. A machine emotion transfer model for intelligent human-machine interaction based on group division. *Mech Syst Signal Process* (2020) 142:106736–850. doi:10.1016/j.ymssp.2020.106736
3. Hu B, Li X, Sun S, Ratcliffe M. Attention recognition in EEG-based affective learning research using CFS+KNN algorithm. *IEEE/ACM Trans Comput Biol Bioinf* (2018) 15(1):38–45. doi:10.1109/tcbb.2016.2616395
4. Koelstra S, Muhl C, Soleymani M, Lee J-S, Yazdani A, Ebrahimi T, et al. DEAP: a database for emotion analysis; using physiological signals. *IEEE Trans Affective Comput* (2012) 3(1):18–31. doi:10.1109/t-affc.2011.15
5. Li Y, Zheng W, Wang L, Zong Y, Cui Z. From regional to global brain: a novel hierarchical spatial-temporal neural network model for EEG emotion recognition. *IEEE Trans Affective Comput* (2019) 99:1. doi:10.1109/taffc.2019.2922912

6. Yin Z, Wang Y, Liu L, Zhang W, Zhang J. Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination. *Front Neurorobot* (2017) 11:19. doi:10.3389/fnbot.2017.00019
7. Liu YJ, Yu M, Zhao G, Song J, Ge Y, Shi Y. Real-time movie-induced discrete emotion recognition from EEG signals. *IEEE Trans Affective Comput* (2018) 9(4):550–62. doi:10.1109/taffc.2017.2660485
8. Zheng WL, Zhu JY, Lu BL. Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans Affective Comput* (2019) 10(3):417–29. doi:10.1109/taffc.2017.2712143
9. Zhong Q, Zhu Y, Cai D. Electroencephalogram access for emotion recognition based on deep hybrid network. *Front Hum Neurosci* (2020) 14:1–13. doi:10.3389/fnhum.2020.589001
10. Li P, Liu H, Si Y, Li C, Li F, Zhu X, et al. EEG based emotion recognition by combining functional connectivity network and local activations. *IEEE Trans Biomed Eng* (2019) 66(10):2869–81. doi:10.1109/tbme.2019.2897651
11. Mert A, Akan A. Emotion recognition from EEG signals by using multivariate empirical mode decomposition. *Pattern Anal Appl* (2016) 21(1):81–9. doi:10.1007/s10044-016-0567-6
12. Kwon YH, Shin SB, Kim SD. Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system. *Sensors* (2018) 18(5):1383–95. doi:10.3390/s18051383
13. Salama ES, El-Khoribi RA, Shoman ME, Shalaby MAW. EEG-based emotion recognition using 3D convolutional neural networks. *Int J Adv Comput Sci Appl* (2018) 9(8). doi:10.14569/IJACSA.2018.090843
14. Chao H, Dong L, Liu Y, Lu B. Emotion recognition from multiband EEG signals using CapsNet. *Sensors* (2019) 19(9):2212. doi:10.3390/s19092212
15. Li X, Song D, Zhang P, Yu G, Hou Y, Hu B. Emotion recognition from multi-channel EEG data through convolutional recurrent neural network. *IEEE Int Conf Bioinform Biomed* (2016) 1:352–9. doi:10.1109/bibm.2016.7822545
16. Mahata S, Saha SK, Kar R, Mandal D. Optimal design of fractional order low pass Butterworth filter with accurate magnitude response. *Digital Signal Process* (2018) 72:96–114. doi:10.1016/j.dsp.2017.10.001
17. Cho J, Hwang H. Spatio-temporal representation of an electoencephalogram for emotion recognition using a three-dimensional convolutional neural network. *Sensors* (2020) 20(12):3491–508. doi:10.3390/s20123491
18. Cho K, Van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *Comput Sci* (2014) 2014: 1724–34. doi:10.3115/v1/D14-1179
19. Xing X, Li Z, Xu T, Shu L, Hu B, Xu X. SAE+LSTM: a new framework for emotion recognition from multi-channel EEG. *Front Neurorobot* (2019) 13:37. doi:10.3389/fnbot.2019.00037
20. Wang XW, Nie D, Lu BL. Emotional state classification from EEG data using machine learning approach. *Neurocomputing* (2014) 129:94–106. doi:10.1016/j.neucom.2013.06.046
21. Chen J, Hu B, Xu L, Moore P, Su Y. Feature-level fusion of multimodal physiological signals for emotion recognition. *IEEE Int Conf Bioinform Biomed* (2015) 1:395–99. doi:10.1109/bibm.2015.7359713
22. Zhuang N, Zeng Y, Tong L, Zhang C, Zhang H, Yan B. Emotion recognition from EEG signals using multidimensional information in EMD domain. *Biomed Res Int* (2017) 2017:83173579. doi:10.1155/2017/8317357
23. Xiao L, Hu X, Chen Y, Xue Y, Gu D, Chen B, et al. Targeted sentiment classification based on attentional encoding and graph convolutional networks. *Appl Sci* (2020) 10(3):957–73. doi:10.3390/app10030957
24. Alhagry S, Fahmy AA, El-Khoribi RA. Emotion recognition based on EEG using LSTM recurrent neural network. *Int J Adv Comput Sci Appl* (2017) 8(10):355–8. doi:10.14569/IJACSA.2017.081046
25. Xiao L, Hu X, Chen Y, Xue Y, Chen B, Gu D, et al. Multi-head self-attention based gated graph convolutional networks for aspect-based sentiment classification. *Multimedia Tools Appl* (2020) 2020:1–12. doi:10.1007/s11042-020-10107-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 25 January 2021
doi: 10.3389/fphy.2020.619363

A Frequency Modulation Fingerprint-Based Positioning Algorithm for Indoor Mobile Localization of Photoelectric Modules

Chi Duan[1], *Lixia Tian*[1]*, *Pengfei Bai*[1] *and Bao Peng*[2]

[1]Guangdong Provincial Key Laboratory of Optical Information Materials and Technology, Institute of Electronic Paper Displays, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China, [2]Shenzhen Institute and Information Technology, Shenzhen, China

Optoelectronic modules have a wide range of applications in the field of wireless communication. However, the function of mobile localization has not been realized in optoelectronic modules. In this paper, an indoor positioning algorithm, which was based on frequency modulation (FM) signals, was realized in optoelectronic modules. Firstly, FM monitoring receiver DB4004 was used to collect FM signals; Secondly, FM signals were preprocessed and analyzed to build a FM dataset. Finally, weighted centroid k-nearest neighbors (WC-KNN) precise positioning algorithm was proposed to obtain the position information of the photoelectric module. Experimental results showed that the median location error of the WC-KNN algorithm can reach 0.8 m and additional hardware equipment was not required. The research results provided the feasibility for the practical application of equipment based on optoelectronic devices in various fields.

Keywords: indoor localization, optoelectronic module, fingerprint-based positioning, frequency modulation (FM), the weighted centroid k-nearest neighbors

OPEN ACCESS

Edited by:
Chongfu Zhang,
University of Electronic Science and Technology of China, China

Reviewed by:
Xiaolong Yang,
Brunel University London, United Kingdom
Xie Xianming Xie,
Guangxi University of Science and Technology, China

***Correspondence:**
Lixia Tian
2019010238@m.scnu.edu.cn

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *20 October 2020*
Accepted: *23 November 2020*
Published: *25 January 2021*

Citation:
Duan C, Tian L, Bai P and Peng B (2021) A Frequency Modulation Fingerprint-Based Positioning Algorithm for Indoor Mobile Localization of Photoelectric Modules. Front. Phys. 8:619363. doi: 10.3389/fphy.2020.619363

INTRODUCTION

With the rapid development of optoelectronic technologies in recent years, various optoelectronic devices were widely used in indoor mobile robots [1], Internet of Things [2] and other industries. And they were related to professional fields, such as visual analysis [3] and display technology [4]. Optoelectronic technologies are a basic technology in the current industrial application fields. Because of exponential growth in the use of optoelectronic modules, it is necessary to deploy thousands of intelligent terminals which can integrate with optoelectronic modules in indoor applications [5]. Especially, the location information of these modules is a basic information for fault detection and other various application services in a mobile application environment [6]. It is the core data for supporting specific applications. So, the research of indoor positioning algorithm model has a great significance [7].

In recent years, global positioning system (GPS) technology had been a mainstream choice for global positioning systems [8]. Assisted global positioning system (A-GPS) technology had the advantages of high speed and high accuracy in outdoor positioning applications [9]. However, it showed obvious shortcomings in indoor positioning. A-GPS localization required multiple network communications with a server, which took up a lot of communication resources. Especially, A-GPS localization was vulnerable to network congestion in areas where mobile phones were used intensively. In addition, Wi-Fi positioning was a more popular positioning technology in the application of indoor

positioning [10]. Wi-Fi positioning accuracy could reach a meter level, but Wi-Fi signals had a short wavelength, severe transmission attenuation, which could limit the effective positioning range. Moreover, the Wi-Fi signals have not been deployed in many areas. So, FM signals were proposed for indoor positioning. Compared with A-GPS positioning and Wi-Fi positioning, there was no need to communicate with the server. Network congestion problems were avoided when FM signals were used to position [11]. What's more, FM signals had a long wavelength, a long transmission distance, and it was less affected by multipath effects. It could form a better stability and robustness in a complex indoor environment [12]. In addition, FM signals based stations were widely available [13]. So, no additional hardware equipment was required. It reduced the cost of popularization greatly.

In this paper, a fingerprint location algorithm WC-KNN based on FM signals was proposed to solve the indoor location of photoelectric modules. And a practical fingerprint positioning model was proposed by simulation analysis. The model only used a public FM transmission information and floorplans to predict the distribution of RSS. So, it can achieve a high-precision positioning.

ALGORITHM PRINCIPLE DESCRIPTION

The positioning process can be divided into two stages, offline and online phase [14]. In the offline phase, the positioning area is discretized, the received signal strength (RSS) information of each discrete point is collected. Then, the RSS feature vector of each discrete point is extracted as a fingerprint information. Finally, a location fingerprint library is constructed based on feature vectors of all discrete points [15]. In the online calculation process, the matching algorithm WC-KNN is used to find the point with the highest RSS feature similarity in the location fingerprint dataset as an estimated location.

Offline Phase

The main work of the offline phase can be divided into two points. Firstly, the appropriate reference points (RPs) density for the positioning area are selected after selecting experimental environment, and the continuous positioning area is discretized. Then, the RSS signals of each channel in the FM broadcast are collected and entered into the dataset by a special testing equipment [16]. Secondly, the collected RSS information is preprocessed, outliers are eliminated, and the processed data for feature extraction are used to build a location fingerprint dataset, which is shown in **Eq. 1**.

$$\left\{RSSI_{rq} = \left[RSSI_{r1}, RSSI_{r2}, ..., RSSI_{rp}\right],\ r = 1, 2, ..., R,\ q = 1, 2, .., Q\right\} \tag{1}$$

$$\overline{RSSI} = \frac{1}{Q}\sum_{q=1}^{Q} RSSI_{rq} \tag{2}$$

$RSSI_{rq}$ is a R×Q matrix which contains collected RSS information. According to **Eq. 2**, the average value of measured values in each FM channel is calculated and recorded as the reference data of a RP.

Online Phase

In the fingerprint-based localization stage, deterministic and probabilistic methods are used to estimate the position mainly. Three different algorithms are compared and analyzed. The first algorithm is the nearest neighbor algorithm (NN) [17, 18]. The position of the nearest RP to the unknown node is regarded as the estimated position. The nearest RP is decided with the shortest distance from an unknown point. This distance calculation is based on the Manhattan distance or Euclidean distance between the observed fingerprint and the fingerprint recorded in the database. The second algorithm is K nearest neighbor method (KNN) [19, 20]. The basic principle of the algorithm is that the fingerprint $r_i, i = 1, 2, ..., L$ in the fingerprint library and the fingerprint $\overline{r}$ collected online are both the mean value of each FM broadcast signal sample. During the execution process, the Euclidean distance between $\overline{r}$ and the fingerprint RP_i stored in the fingerprint library are solved, and it is sorted from small to large [21]. Then, K coordinate positions l_i corresponding to RP_i with the smallest Euclidean distance are selected for weighted calculation, which is showed in **Eq. 3**. Finally, the weighted coordinate positions are the coordinate positions of the terminal.

$$r = \underset{r_i, i=1,2,...,K}{argmin} \sum_{i=1}^{K} \left\| r_i - \overline{r}^2 \right\| \tag{3}$$

Where, $r = [r_1, r_2, ..., r_k]^T$ represents the fingerprints dataset at K RPs with the smallest Euclidean distance. The relative coordinate position dataset is $l = [l_1, l_2, ..., l_k]^T$. Then the coordinate position is estimated to be $\overline{l}$ by the KNN algorithm, which is shown in **Eq. 4**.

$$\overline{l} = \frac{1}{k}\sum_{i=1}^{K} l_i \tag{4}$$

The position estimation $\overline{l}$ of WC- KNN algorithm is optimized by **Eq. 5**.

$$\overline{l} = \frac{1}{\sum_{i=1}^{K} w_i}\sum_{i=1}^{K} w_i l_i \tag{5}$$

Where $w_i = \frac{1}{\varsigma + \|r_i - \overline{r}\|}$, ς is a constant approximately zero. From **Eq. 5**, it can be seen that WC-KNN algorithm is a special case of the weighting factor in the KNN algorithm to remove the mean value.

In order to improve the estimation accuracy, more RPs are set when collecting data when data is collected. The dense distance between RPs often leads to close proximity of fingerprint data in neighboring RPs, which forms a fuzzy judgment for the traditional KNN algorithm. So, the positioning accurate is affected ultimately. So, the graphic area enclosed between RPs is reduced in order to avoid overfitting and fuzzy judgment. Centroid method is proposed to optimize the algorithm model as follows.

An intelligent terminal A has its real coordinates in a 2D experimental environment. The coordinates of RP_i are (x_i, y_i). The $RSSI_p, i = 1, 2, ..., p$ received from point A is calculated for each RP according to **Eq. 5**. And the K RPs are selected with the highest geometric spatial correlation. In this paper, the distance d_n is used to represent the geometric spatial correlation. The d_n is obtained in **Eq. 6**.

$$d_n = \left\| \overline{rssi} - \overline{RSSI_r}^2 \right\|, r = 1, 2, ..., R \tag{6}$$

The area which is composed of selected RPs is considered, the coordinates of centroid O_b is (x_{Ob}, y_{Ob}), which is obtained in **Eq. 7**.

$$x_{ob} = \frac{1}{K}\sum_{k=1}^{K} x_k, y_{ob} = \frac{1}{K}\sum_{k=1}^{K} y_k \tag{7}$$

Then, the distance d_{Ob} between the centroid O_b and A is obtained in **Eq. 8**.

$$\begin{aligned} d_{Ob} &= \sqrt{(x_{Ob} - x)^2 + (y_{Ob} - y)^2} \\ &= \sqrt{\left(\frac{1}{k}\sum_{k=1}^{K} x_k - x\right)^2 + \left(\frac{1}{k}\sum_{k=1}^{K} y_k - y\right)^2} \\ &= \frac{1}{K}\sqrt{\left[\sum_{k=1}^{K}(x_k - x)\right]^2 + \left[\sum_{k=1}^{K}(y_k - y)\right]^2} \\ &= \frac{1}{K}\left[\sum_{k=1}^{K}(x_k - x)^2 + \sum_{k=1}^{K}(y_k - y)^2 + 2\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(x_i - x)(x_j - x)\right. \\ &\quad \left. + 2\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(y_i - y)(y_j - y)\right]^{1/2} \end{aligned} \tag{8}$$

Where, the coordinates of RP_i are (x_i, y_i). The relationship between d_{Ob} and d_n is obtained in **Eq. 9**.

$$\begin{aligned} d_{Ob} = \frac{1}{K}\Bigg[& K\sum_{k=1}^{K}(d_n)^2 - \sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(x_i - x_j)^2 \\ & - \sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(y_i - y_j)^2 \Bigg]^{1/2} \end{aligned} \tag{9}$$

Therefore, coordinates (x_k, y_k) of K RPs are obtained, and the distance d_n between them and the intelligent terminal A are calculated. And coordinates of the space centroid O_b enclosed by the selected RP and the distance between O_b and A can be obtained by **Eq. 7** and **Eq. 9**. **Eq. 9** is simplified further, **Eq. 10** can be obtained.

$$\begin{aligned} d_{Ob}^2 = \frac{1}{K}\sum_{k=1}^{K}(d_n)^2 - \frac{1}{K^2}\Bigg[& \sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(x_i - x_j)^2 \\ & + \sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(y_i - y_j)^2 \Bigg] = F_1 - F_2 \end{aligned} \tag{10}$$

Where, $F_1 = \frac{1}{K}\sum_{k=1}^{K}(d_n)^2$, $F_2 = \frac{1}{K^2}\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(d_{ij})^2$. d_{ij} indicates the distance between RP_i and RP_j. Without loss of generality, it is assumed that the distance between the selected K RPs and the intelligent terminal A satisfies **Eq. 11**.

$$0 < d_1 \le d_2 \le d_3 \cdots \le d_{K-1} \le d_K \tag{11}$$

Substitute **Eq. 11** into **Eq. 10**, F_1 and F_2 can be obtained in **Eq. 12**.

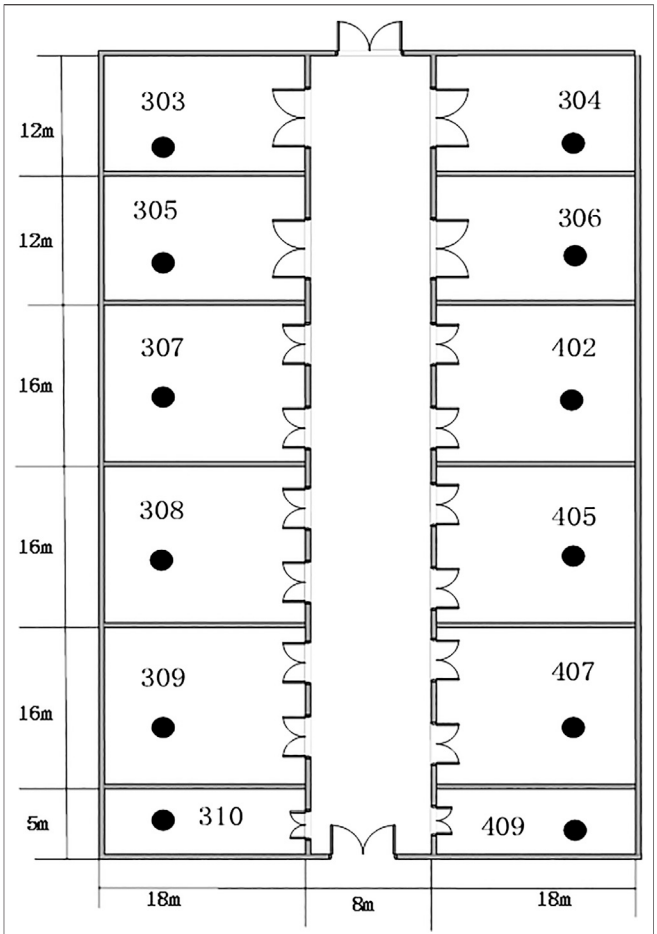

FIGURE 1 | The experimental environment was located on a fourth floor in a building. The bird's-eye view of the fourth floor is shown in the figure. Black dots represent the layout of wireless FM signal data.

$$\begin{cases} F_1 = \frac{1}{K}\sum_{k=1}^{K}(d_n)^2 \le d_K^2 \\ F_2 = \frac{1}{K^2}\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(d_{ij})^2 > 0 \end{cases} \tag{12}$$

Eq. 10 and **Eq. 12** are combined, the conclusion can be drawn in **Eq. 13**.

$$d_{Ob} < d_K \tag{13}$$

Where, d_N is distance between RP and the intelligent terminal A. There is at least one RP in the known K RPs. d_N must be greater than the obtained distance d_{Ob}.

Therefore, the current centroid O_b is used to replace the RP which is farthest from the intelligent terminal A. At this time, the plane enclosed by the new K RPs must be smaller than the plane enclosed by the original K RPs. The area where the intelligent terminal A can be further reduced and the positioning performance is improved by multiple iterations.

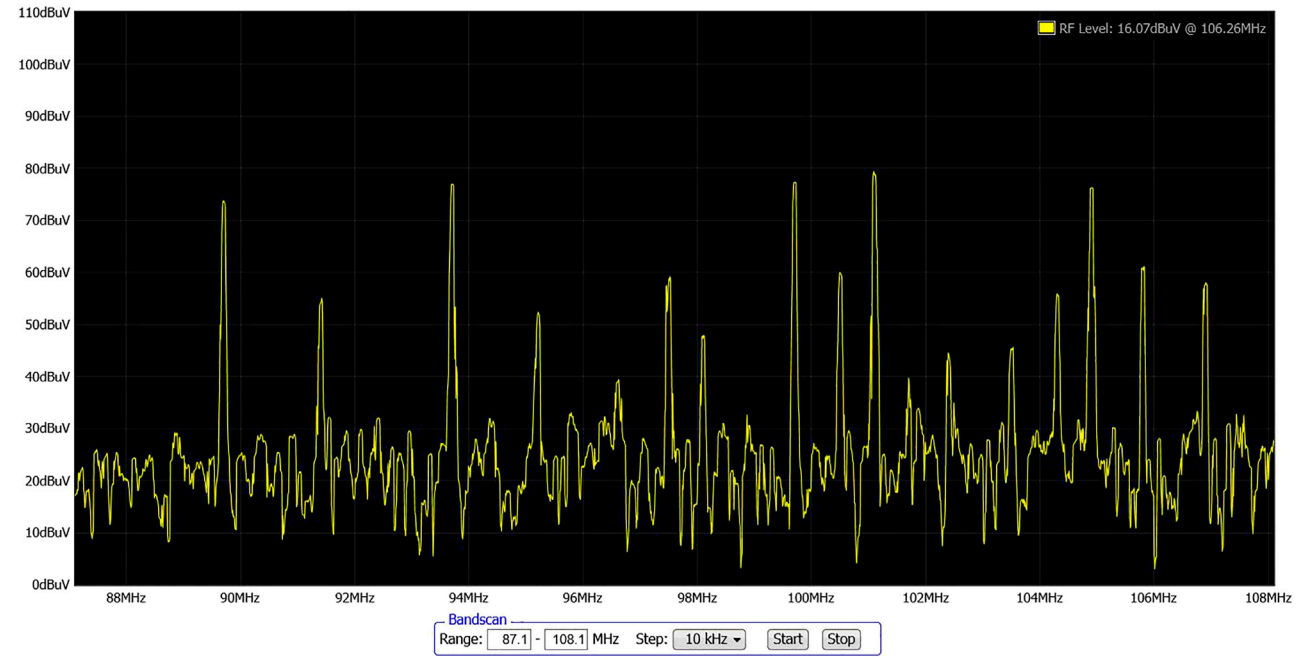

FIGURE 2 | Amplitude diagram of full-band FM signals, which can cover the entire FM bandwidth from 88 to 108 MHz. Then, the stronger the signal, the better the stability, so the frequency band with a larger amplitude is selected as a collection channel.

EXPERIMENTAL ENVIRONMENT

In this part, experimental environment was configured, experimental process was designed, and FM signals of each channel were collected.

The experimental environment was located on a fourth floor of an office building and a layout of the experiment environment was shown in **Figure 1**. It was a typical indoor office environment, including 12 rooms and corridors. Firstly, we selected a test point (TP) in each room. Then, FM broadcast signals with strong nearby signals were searched by a FM test equipment. **Figure 2** was a full-band FM signal RSSI diagram. Each TP could search 15 FM channels (p = 15), which could cover an entire FM bandwidth from 88 to 108MHz, as shown in **Figure 2**. Finally, FM monitoring receiver DB4004 was used to open the FM full-band to sweep FM signals of all frequencies. There were 15 sets of FM programs with frequencies of 89.7–106.9 MHz.

ILLUSTRATIVE EXPERIMENTAL RESULTS

In this part, performance of several algorithms was evaluated, and changes in performance caused by parameter changes were analyzed to verify the effectiveness of the algorithm.

Because our FM-based positioning was two-dimensional positioning, the height of the FM antenna remained a constant in all measurements. Signals were collected three times a day and measured 10 times each time in two weeks. So, scene of human presence was avoided effectively. At the same time, room 308 was selected as a testing platform for fingerprint positioning. **Figure 3** showed the indoor environment layout and RPs setting diagram of room 308.

FM fingerprint was generated by MS9801 field intensity meter to collect FM broadcast signal. Because of office furniture, 42 data collection points were set up in one room with a size of $16m \times 18m$. The collection point size was $3m \times 3m$. Each FM fingerprint included 15 channels, and each channel collected 10 RSSI samples at a time. In the simulation analysis, interpolation and increased coverage area were used to increase sampling data and provided positioning accuracy. The RPs were expanded to 280, and the size of collection points was $1m \times 1m$.

A path which was consisted of 100 points was selected as the real path randomly, which was showed as a blue line in **Figure 4**. And the position information was used to estimate by the algorithm in this chapter as the predicted path, showed in the red line in **Figure 4**. The predicted location was compared with the actual location, the average positioning error of the system was 0.8 m. The overall positioning results were basically consistent with the real trajectory, although some points had large positioning errors. It indicated that the algorithm could better meet the location service.

WC-KNN algorithm was compared with three different positioning algorithms, which were NN, KNN, and Bayesian algorithms [22], respectively. **Figure 5** showed the performance results of the FM positioning system by using all broadcast FM stations in simulation. The median location error

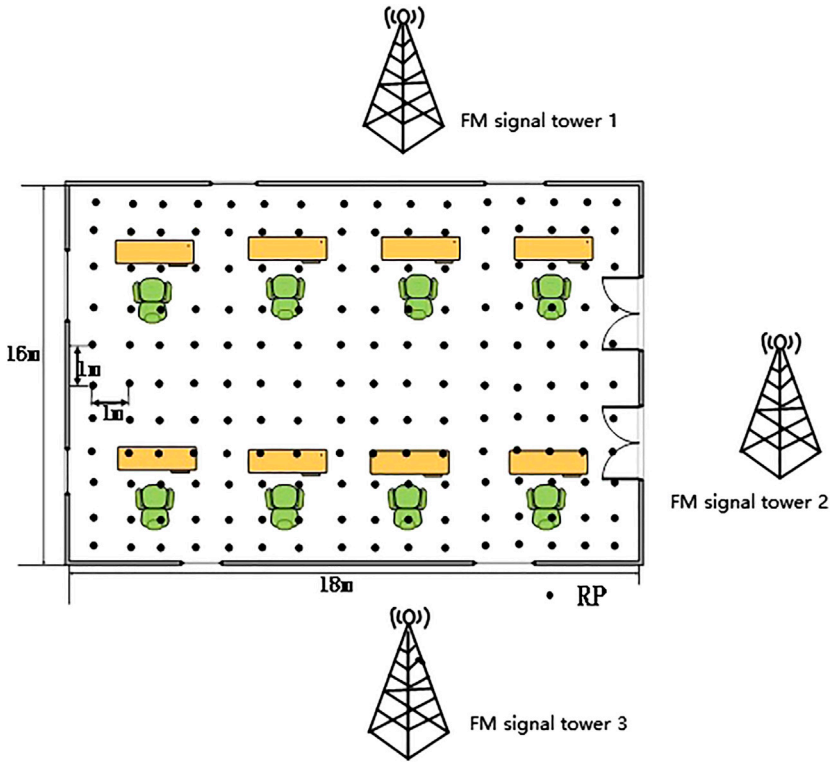

FIGURE 3 | Room 308 was 18 m in length and 16 m in width. Each RP was set in every meter distance.

of the WC-KNN algorithm was the smallest, which could reach 0.8 m. The median location error of the KNN algorithm was 1.2 m, the median location error of the Bayes algorithm was 1.8 m, and the median location error of the NN algorithm was 2.1 m. Positioning errors were increased by 33%, 55% and 62%, respectively. Obviously, such a small positioning error was due to the effectiveness of fingerprint library modeling. So, the performance of WC-KNN algorithm in indoor positioning had an improved significantly.

INFLUENCE OF PARMETERS ON LOCALIZATION

In order to verify the effectiveness of the algorithm, MATLAB was used to compare the real path and the estimated path. And the average positioning error of algorithms was calculated. Because the estimation error of the WC-KNN algorithm was greatly affected by factors such as K value, access point (AP) number. Controlling variable method was used to analyze the impact of parameter changes effectively.

The Influence of K Value on Localization

One motion node was selected for algorithm simulation. **Figure 6** showed that different K values in the algorithm correspond to different positioning errors. When K < 20, the positioning error dropped rapidly, and it entered the relative convergence stage when K = 20. As K increased, the positioning error generally showed a downward trend, then it began to rebound. Finally, it rose after reaching a certain value. The overall positioning performance was the best when the value of K was around 30. When farther points were selected, it would also affect the classification of positioning nodes. And it would lead to the situation of under-fitting, which can lead to a large error in this node.

The Influence of AP Number on Localization

Signals from all FM frequency points were used for positioning calculation. However, wider FM fingerprint data required more time for data collection, and it would also increase the practical cost of a positioning system and the dimension of positioning calculation significantly. In addition, the calculation amount of the algorithm was increase significantly. Therefore, not all FM frequency points could play the same role in the positioning calculation, so it was necessary to find an optimal balance point between the FM fingerprint width (the number of APs) and the positioning accurate.

FM signals with 15 different frequency points were received from three different broadcast transmitters. Among them, broadcast transmitter 1 received 6 broadcast signals, broadcast transmitter 2 received 5 broadcast signals, and broadcast transmitter 3 received 4 broadcast signals. The

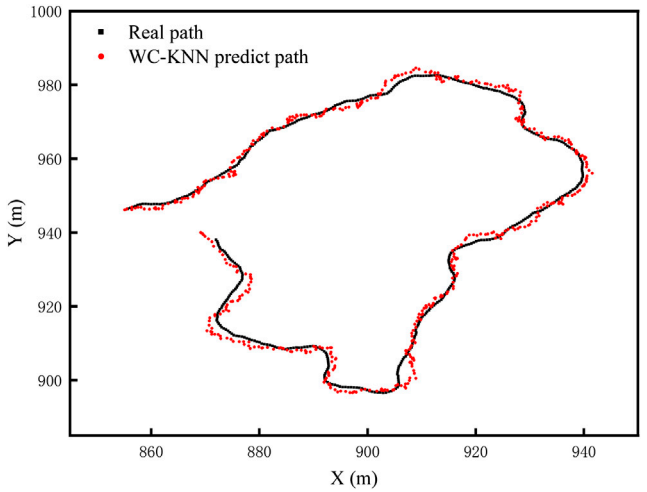

FIGURE 4 | Blue line was a path which generate randomly. It was selected as the real path. Red line was a path which is predicted by WC-KNN algorithm. They all contained 100 points.

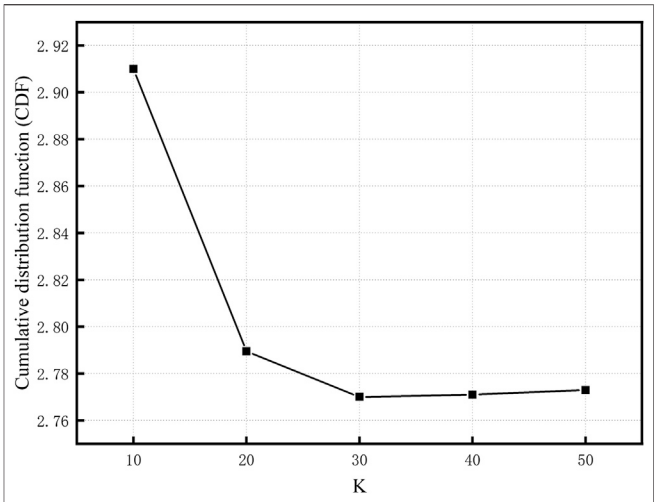

FIGURE 6 | When K < 20, the positioning error dropped rapidly, and when K = 20, it entered a relative convergence stage. When K = 30, the overall positioning performance was the best. When K continued to increase, the classification of positioning nodes was affected.

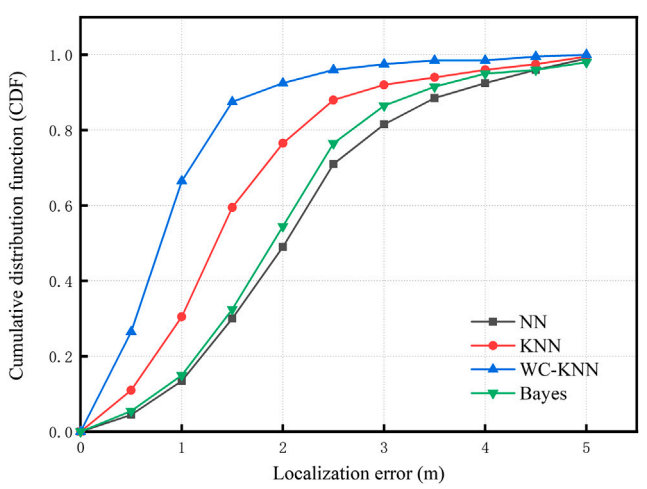

FIGURE 5 | Cumulative error distribution function diagram of four different algorithms. The performance of WC-KNN algorithm was the best, its median location error was 0.8 m. The median location errors of KNN, Bayes, NN were 1.2, 1.8, 2.1 m respectively.

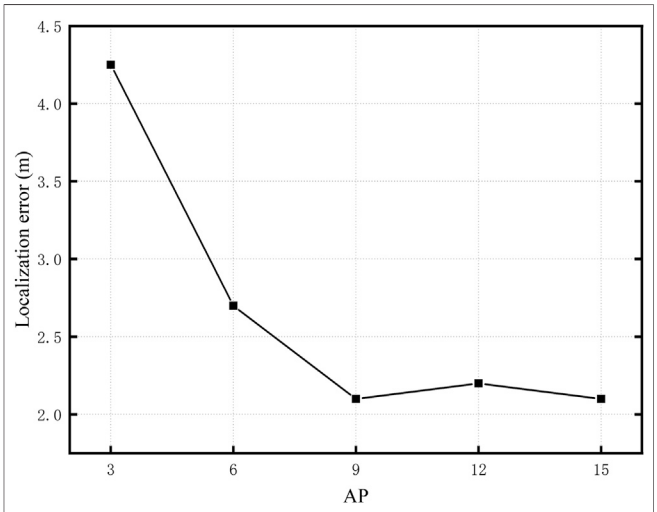

FIGURE 7 | The relationship between frequency AP and error. When AP = 9, the positioning performance was the best, which was similar to AP = 15. When AP = 12, positioning error slightly increased because of accidental factors.

same number of AP fingerprint data was selected from different broadcast transmitting stations. The number of APs were 3, 6, 9, 12, and 15, respectively, showed in **Figure 7**. The K value was 30.

Figure 7 showed that when 9 APs were selected, the positioning error was 2.1 m, which was equivalent to the positioning error of all APs. The analysis results showed that when the FM broadcast AP was effectively selected, not only a lower positioning error could be achieved, but also the fingerprint database involved in the calculation was also reduced. It had the effect of reducing the computational complexity and improving the operating efficiency of system. A large number of test and analysis results showed that an effective selection of FM broadcast AP could reduce the width of the fingerprint by about 40% with less impact on the positioning error.

The Influence of Different Receiving Equipment on Localization

The receiving equipment used for the signal acquisition was a field intensity meter. The resolution of the equipment was

FIGURE 8 | FM receiving equipment. **(A)** Field strength meter. **(B)** FM monitoring receiver DB4004.

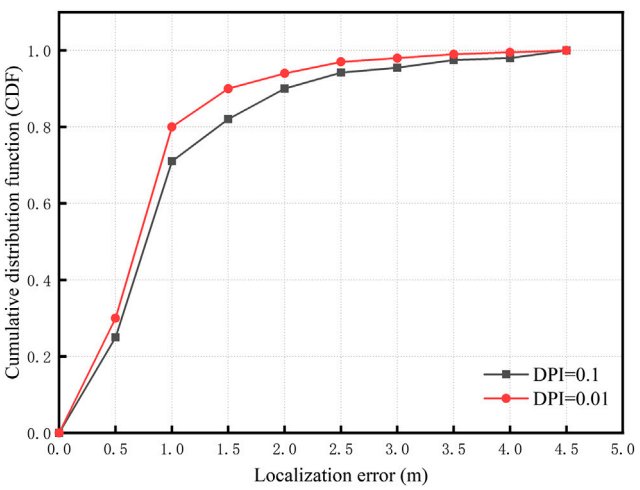

FIGURE 9 | The cumulative error distribution of two different devices. The median positioning error of DB4004 was 0.7 m. The median positioning error of the field strength meter was 0.8 m. Smaller dots Per Inch (DPI) device, smaller positioning error.

0.1 dBm. And FM monitoring receiver DB4004 was used for data acquisition. The resolution of the equipment was 0.01 dBm. As shown in **Figure 8**.

Figure 9 showed that the data was collected by two different resolution receivers. When the WC-KNN algorithm was used, the median positioning error of the FM monitoring was 0.7 m. Compared with the field strength meter, the median positioning error was smaller than 0.1 m. The difference in positioning error between the two was about 0.5 m with 95% confidence. Therefore, the resolution of the receiving device had a greater impact on the indoor positioning error.

CONCLUSION

In this paper, an improved WC-KNN algorithm based on FM signals was proposed as a positioning algorithm. The fingerprint dataset was established according to the geometric layout of the indoor environment. And fluctuation characteristics of signals were quantified and added as a weight to the distance calculation of the positioning algorithm, which solved the fuzzy judgment problem of the traditional KNN algorithm. The FM positioning experimental environment was built in the school laboratory building, and RSS dataset which was corresponded to the physical space was established. The experimental data verified that positioning accuracy was improved by the WC-KNN algorithm compared to the three traditional algorithms. Then, the influence of different parameters on the positioning accuracy was discussed, the effectiveness of the WC-KNN algorithm under different conditions was verified. In the 16 × 18 m positioning area, the average positioning error was 0.8 m, the maximum was 2.3 m, and the positioning error of 90% in the test points was below 2.0 m. The impact of complex indoor environment was reduced effectively and positioning accuracy was improved, which had a certain practical value.

DATA AVAILABITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

CD and BP designed the projects. LT and CD carried out most of the experiments and data analysis. BP and PB performed part of experiments and helped with discussions during manuscript preparation. LT and CD contributed to the data analysis and correction. BP and PB gave suggestions on project management and provides helpful discussions on the experimental results. All authors have read and agreed to the published version of the manuscript.

FUNDING

Supported by Program for Chang Jiang Scholars and Innovative Research Teams in Universities (No. IRT_17R40), Program for Guangdong Innovative and Enterpreneurial Teams (No. 2019BT02C241), Science and Technology Program of Guangzhou (No. 2019050001), Guangdong Provincial Key Laboratory of Optical Information Materials and Technology (No. 2017B030301007), Guangzhou Key Laboratory of Electronic Paper Displays Materials and Devices (201705030007) and the 111 Project.

REFERENCES

1. Xiaohua T, Sujie Z, Sijie X. Performance analysis of Wi-Fi indoor localization with channel state information. *IEEE Trans Mobile Comput* (2019) 18:1870–84. doi:10.1109/TMC.2018.2868680
2. Zeger L, Pei C, Kobayashi H. Effects of multipath interference on an FM data subcarrier. IEEE pacific rim conference on communications. *Comp Signal Process* (1997) 10:40–4. doi:10.1109/PACRIM.1997.619898
3. Pan C, Jianga S, Fuqiang G. Learning RSSI feature via ranking model for Wi-Fi fingerprinting localization. *IEEE Trans Veh Technol* (2020) 69:1695–705. doi:10.1109/TVT.2019.2959308
4. Mukherjee T, Kumar P, Pati D, Blasch E, Pasiliao E, Liqin X. LoSI: large scale location inference through FM signal integration and estimation. *Big Data Mining Analy* (2019) 2:319–48. doi:10.26599/BDMA.2019.9020013
5. Ren L, TungJung H. Indoor localization system based on hybrid Wi-Fi/BLE and hierarchical topological fingerprinting approach. *IEEE Trans Veh Technol* (2019) 68:10791–806. doi:10.1109/TVT.2019.2938893
6. Xiansheng G, Nkrow Raphael E, Nirwan A. Robust WiFi localization by fusing derivative fingerprints of RSS and multiple classifiers. *IEEE Trans Indust Inform* (2020) 16:3177–86. doi:10.1109/TII.2019.2910664
7. Xin L, Min J, Xueyan Z, Weidang L. A novel multichannel internet of things based on dynamic spectrum sharing in 5G communication. *IEEE Inter Things J* (2019) 6:5962–70. doi:10.1109/JIOT.2018.2847731
8. Shengcheng Y, Wanghsin H, Wenyen L, Wu Y. Study on an indoor positioning system using earth magnetic field. *IEEE Trans Instrum Meas* (2020) 69:865–72. doi:10.1109/TIM.2019.2905750
9. Yang X, Cao R, Zhou M, Xie L, Temporal-frequency attention-based human activity recognition using commercial wi-fi devices. *IEEE Access* (2020) 8: 137758–69. doi:10.1109/ACCESS.2020.3012021
10. Guoliang S, Yongxin G. An optimal design for passive magnetic localization system based on SNR evaluation. *IEEE Trans Instrum Meas* (2020) 69:4324–33. doi:10.1109/TIM.2019.2947173
11. Capraro F, Segura M, Sisterna C. Human real time localization system in underground mines using UWB. *IEEE Latin Am Trans* (2020) 18: 392–9. doi:10.1109/TLA.2019.9082253
12. Fan Q, Sun B, Sun Y, Zhuang X. Performance enhancement of MEMS-based INS/UWB integration for indoor navigation applications. *IEEE Sensor J* (2017) 17:3116–30. doi:10.1109/JSEN.2017.2689802
13. Yang X, Liu Z, Nie W, He W, Pu Q. AP optimization for Wi-Fi indoor positioning-based on RSS feature fuzzy mapping and clustering. *IEEE Access* (2020) 8:153599–609. doi:10.1109/ACCESS.2020.3018147
14. Xin L, Xueyan Z. NOMA-based resource allocation for cluster-based cognitive industrial internet of things. *IEEE Trans Indust Inform* (2020) 16:5379–88. doi:10.1109/TII.2019.2947435
15. Vahideh M, Andrew G. Indoor location fingerprinting using FM radio signals. *IEEE Trans Broadcast* (2014) 60: 336–46. doi:10.1109/TBC.2014.2322771
16. Li C, Jingnan T, Honglei Q. A practical floor localization algorithm based on multifeature motion mode recognition utilizing FM radio signals and inertial sensors. *IEEE Sensor J* (2020) 20:8806–19. doi:10.1109/JSEN.2020.2985934
17. Ye X, Yin X, Cai X. Neural-network-assisted UE localization using radio-channel fingerprints in LTE networks. *IEEE Access* (2017) 5:12071–87. doi:10.1109/ACCESS.2017.2712131
18. Lingfei M, Chenyang L. Passive UHF-RFID localization based on the similarity measurement of virtual reference tags. *IEEE Transa Instrum Meas* (2019) 8: 2926–33. doi:10.1109/TIM.2018.2869408
19. Wang Y, Ren A, Zhou M, Wang W, Yang X. A novel detection and recognition method for continuous hand gesture using FMCW radar. *IEEE Access* (2020) 8: 167264–75. doi:10.1109/ACCESS.2020.3023187
20. Fang S, Wang C, Huang T-Y. An enhanced ZigBee indoor positioning system with an ensemble approach. *IEEE Commun Lett* (2012) 16:564–7. doi:10.1109/LCOMM.2012.022112.120131
21. Yu Z, Guo G. Improvement of positioning technology based on RSSI in ZigBee networks. *Wireless Pers Commun* (2017) 95:1943–62. doi:10.1007/s11277-016-3860-1
22. Andrei P. Improving ambient FM indoor localization using multipath induced amplitude modulation effect: a year-long experiment. *Pervasive Mob Comput* (2019) 58:1–10. doi:10.1016/j.pmcj.2019.05.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 16 February 2021
doi: 10.3389/fphy.2020.616066

First-Principles Investigation on Electrochemical Performance of Na-Doped $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$

Yumei Gao[1], Kaixiang Shen[2], Ping Liu[1], Liming Liu[1], Feng Chi[1], Xianhua Hou[2] and Wenxin Yang[1]*

[1]College of Electron and Information, University of Electronic Science and Technology of China, Zhongshan Institute, Zhongshan, China, [2]Guangdong Provincial Key Laboratory of Quantum Engineering and Quantum Materials, Guangdong Engineering Technology Research Center of Efficient Green Energy and Environment Protection Materials, School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou, China

The cathode material $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ for lithium-ion battery has a better electrochemical property than $LiCoO_2$. In order to improve its electrochemical performance, Na-doped $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ is one of the effective modifications. In this article, based on the density functional theory of the first-principles, the conductivity and the potential energy of the Na-doped $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ are calculated with Materials Studio and Nanodcal, respectively. The calculation results of the band gap, partial density of states, formation energy of intercalation of Li^+, electron density difference, and potential energy of electrons show that the new cathode material $Li_{1-x}Na_x$ $Ni_{1/3}Co_{1/3}Mn_{1/3}O_2$ has a better conductivity when the Na-doping amount is x = 0.05 mol. The 3D and 2D potential maps of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ can be obtained from Nanodcal. The maps demonstrate that Na-doping can reduce the potential well and increase the removal rate of lithium-ion. The theoretical calculation results match well with experimental results. Our method and analysis can provide some theoretical proposals for the electrochemical performance study of doping. This method can also be applied to the performance study of new optoelectronic devices.

Keywords: first-principles, lithium-ion battery, cathode material, $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$, Na-doped

OPEN ACCESS

Edited by:
Guofu Zhou,
South China Normal University, China

Reviewed by:
Jun Liu,
South China University of Technology, China
Liangzhong Xiang,
University of California, Irvine, United States

****Correspondence:***
Yumei Gao
yumeigao5697@163.com

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *10 October 2020*
Accepted: *16 November 2020*
Published: *16 February 2021*

Citation:
Gao Y, Shen K, Liu P, Liu L, Chi F, Hou X and Yang W (2021) First-Principles Investigation on Electrochemical Performance of Na-Doped $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$. Front. Phys. 8:616066. doi: 10.3389/fphy.2020.616066

INTRODUCTION

Since Sony introduced commercial lithium-ion batteries (LIBs) that used $LiCoO_2$ as cathode, LIBs have been used in electric vehicles (EVs), hybrid electric vehicles (HEVs), and mobile electronic devices for their good cycle performance, long life, less self-discharge, high specific capacity, and working voltage [1]. In the current era of LIBs, there is an ever-growing demand for even higher energy densities to power mobile devices with increased power consumption.

Cathode materials are one of the key materials to determine the electrochemical performance of LIBs. Recently, the ternary cathode material $LiNi_xCo_yMn_zO_2$ (NCM, $0 < x, y, z < 1$) has been paid much attention for its lower price and better electrochemical performance. $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$, $LiNi_{0.5}Co_{0.2}Mn_{0.3}O_2$, and $LiNi_{0.4}Co_{0.2}Mn_{0.4}O_2$ [2, 3] are often studied by scholars and experts. Their precursors are made by hydroxide coprecipitation reaction combined with solid-phase sintering step [4]. $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ is considered a promising cathode material for commercial LIBs due to its low price, high conductivity, excellent cycle performance, high reversibility, and stable structure [5]. However, the commercial $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ has already reached its theoretical limit, with poor

cycling performance at high rate and thermal stability at high cutoff voltage, which limit its further application in higher-power systems such as EVs and HEVs [6].

Surface modification [6, 7] and doping [8, 9] are analyzed experimentally to improve the electrochemical performance of $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ for high-power electric devices. The discharging capacity and rate capacity can dramatically be improved by coating with Li_2MoO_4 [7]. The lithium difluorophosphate ($LiPO_2F_2$) electrolyte additive drops are applied to significantly improve the electrochemical performance of $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ cell at room and low temperature [10]. Some cations [11], anions [12], and polyanions [13] are doped to substitute part of Ni, Co, Mn, or oxygen sites to improve the electrochemical performance of $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$. The K-doped $LiNi_{0.5}Co_{0.2}Mn_{0.3}O_2$ has a higher column efficiency, larger reversible discharging capacity, and rate capacity. Because bigger metal ions K^+ can extend the space of Li-O, decrease the cation mixing (Li^+/Ni^{2+}), and enhance the structure stability [14]. The Na-doped $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ had been extensively investigated by Li et al. [15]; their experimental investigations showed that $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ has a more excellent electrochemical performance.

Samples and raw materials needed in the NCM ternary cathode material experiments are of high cost, and some experiments may be conducted several times; thus, the entire experiment will need a long time. However, based on first-principles [16–18], the physical and electrochemical properties can be calculated and analyzed by Materials Studio, Nanodcal, and MATLAB; the results can give some theoretical advice or investigation directions about the relevant experiments, and the theoretical calculation can effectively shorten the whole investigation period and reduce the cost [16]. In this article, the Na-doped layer-structure $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ is studied theoretically with density functional theory of first-principles; the results show that the electrochemical performance of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ is affected by the proportion of Na-substitution and the best proportion of Na-substitution in $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ is in agreement with that of experiments [15].

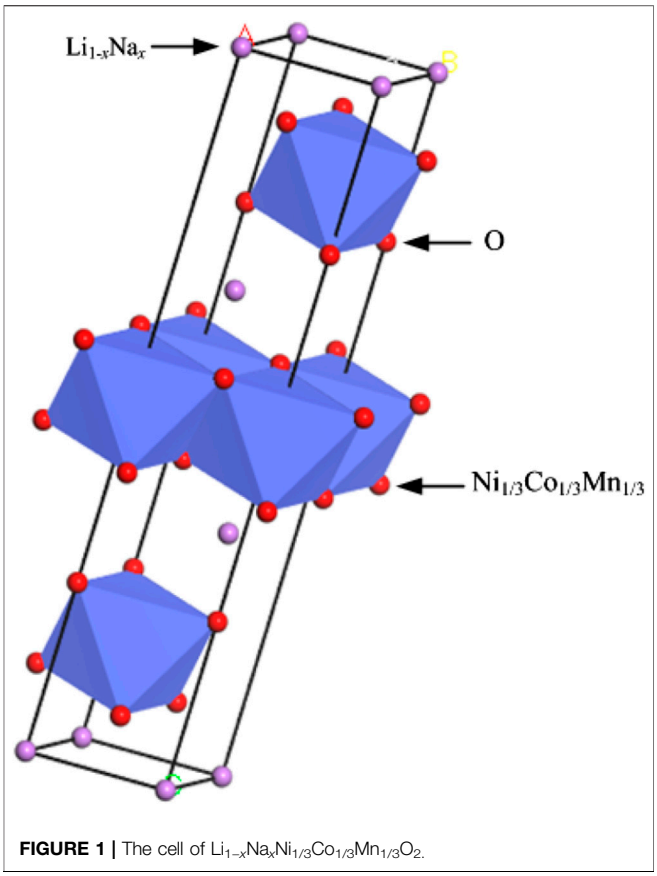

FIGURE 1 | The cell of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$.

PRINCIPLE

The Theory Foundation of the First-Principles

Quantum mechanics is about the properties of molecules and atoms, which is closer to the nature of all things in the universe, so it is called the first principle. In fact, according to the nuclear charge number of atoms and some simulated environmental parameters (such as electron quantity and static mass), the first principle means to solve the Schrödinger equation as shown in **Equation (1)**.

$$\hat{H}\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}), \tag{1}$$

where $\hat{H}$ is the Hamiltonian and E is the system's energy. The solution of the **Equation (1)** is the eigenvalue (energy value) and eigenfunction (wave function), which are used to analyze the related properties of particles. However, for the multiparticle system, it is impossible to solve the Schrödinger equation exactly because of a complex electronic interaction in the system, so it is necessary to simplify and approximate it reasonably. The mass of the nucleus is much larger than that of the electron, and the velocity of the nucleus is much smaller than that of the electron. For the electron, the nucleus can be assumed to be stationary, and then the Schrödinger equation can be approximated adiabatically, as shown in **Equation (2)**.

$$\hat{H}(\boldsymbol{R})\psi(\boldsymbol{r}) = E(\boldsymbol{R})\psi(\boldsymbol{r}), \tag{2}$$

where $\hat{H}(\boldsymbol{R})$ is the adiabatical Hamiltonian and $E(\boldsymbol{R})$ is the system's adiabatical energy.

Density Functional Theory

The density functional theory originates from the Thomas–Fermi model proposed by Thomas [19] and Fermi [20]. This model ignores the exchange-correlation between electrons. On this assumption, the Schrödinger equation can be converted into a simpler wave equation, and then the function of the total energy about the electronic system is determined only by the electron density function $\boldsymbol{r}(\boldsymbol{r})$, which is called density functional theory (DFT). This model is too rough to be used directly. Based on the Thomas–Fermi model, under the Born–Oppenheimer approximation,

TABLE 1 | Band gap value of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ with different Na-doping amount *x*.

x (mol)	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
Band gap (eV)	1.603	1.503	1.445	1.377	1.312	1.249	1.186	1.123	1.063	1.004	0.941

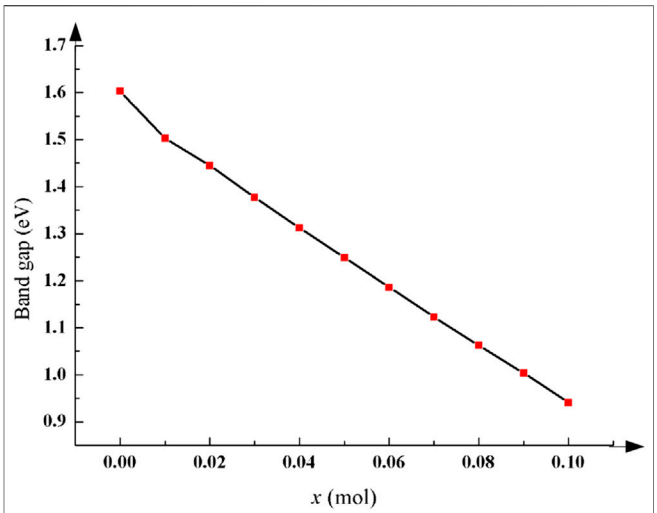

FIGURE 2 | The band gap of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ declines after Na-doping and the conductivity is improved significantly.

nonrelativistic adiabatic approximation, and single-electron approximation, Hohenberg and Kohn proposed the more exact density functional method (HK theorems) [21, 22]. The physical properties of matter can be analyzed by self-consistent calculation. According to HK theorems, properties of particles' ground state can be expressed by particle number density $\rho(\boldsymbol{r})$ instead of the wave function, the system degrees turning 3N into 3, so a lot of computing resources can be saved.

However, HK theorems still have the following three problems: firstly, how to determine the particle number density function $\rho(\boldsymbol{r})$; secondly, how to determine the kinetic energy functional $\boldsymbol{T}[\rho(\boldsymbol{r})]$; lastly, how to determine the exchange-correlation energy functional $E_{\mathrm{xc}}[\rho(\boldsymbol{r})]$.

Kohn–Sham Equation and Exchange-Correlation Functional

The first and the second problems can be solved by the Kohn–Sham equation [23], as shown in **Equation (3)**.

$$\left\{-\nabla^2 + V_{\mathrm{KS}}[\rho(\boldsymbol{r})]\right\}\varphi_i(\boldsymbol{r}) = E_i\varphi_i(\boldsymbol{r}), \tag{3}$$

where $V_{\mathrm{KS}}[\rho(\boldsymbol{r})] = V_{\mathrm{ne}}[\rho(\boldsymbol{r})] + V_{\mathrm{coul}}[\rho(\boldsymbol{r})] + V_{\mathrm{xc}}[\rho(\boldsymbol{r})]$, $V_{\mathrm{ne}}[\rho(\boldsymbol{r})]$ is the attraction potential between nuclei, $V_{\mathrm{coul}}[\rho(\boldsymbol{r})]$ is the Coulomb potential between electrons, and $V_{\mathrm{xc}}[\rho(\boldsymbol{r})]$ is the exchange-correlation potential. In **Equation (3)**, the Hamiltonian with the interacting particles is replaced with that of the noninteracting particles, and the interacting particles are described by the exchange-correlation functional $E_{\mathrm{xc}}[\rho(\boldsymbol{r})]$; then, the single-electron equation can be obtained.

The last problem is generally solved by the local density approximation (LDA) and the general gradient approximation (GGA). LDA can give more accurate results for the system with small change of electron density relative in the space. Considering the ununiformity of the electron density, GGA can give more exact results than LDA [24].

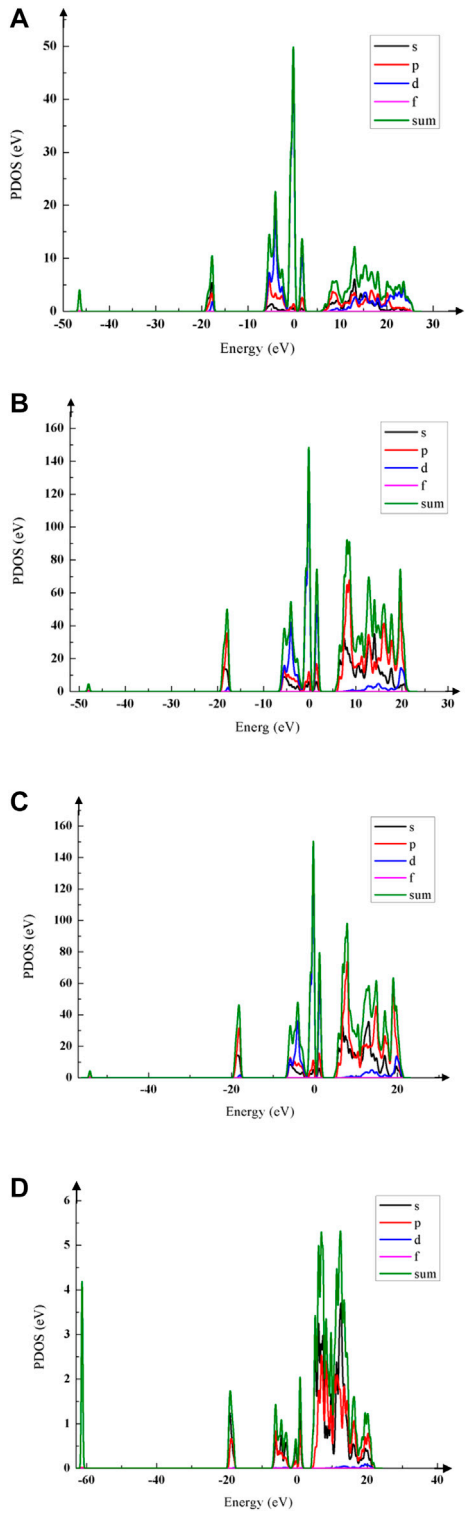

FIGURE 3 | PDOS schemas of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ with different Na-doping amount *x*. **(A)** When *x* = 0 mol, $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ is not doped. **(B)** When *x* = 0.01 mol, PDOS peak of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ is higher than that of the pristine. **(C)** When *x* = 0.05 mol, the PDOS peak rises to about 157 eV, its conductivity is best. **(D)** When *x* = 0.10 mol, the PDOS peak declines dramatically to 5.5 eV, which is much lower than that of the pristine.

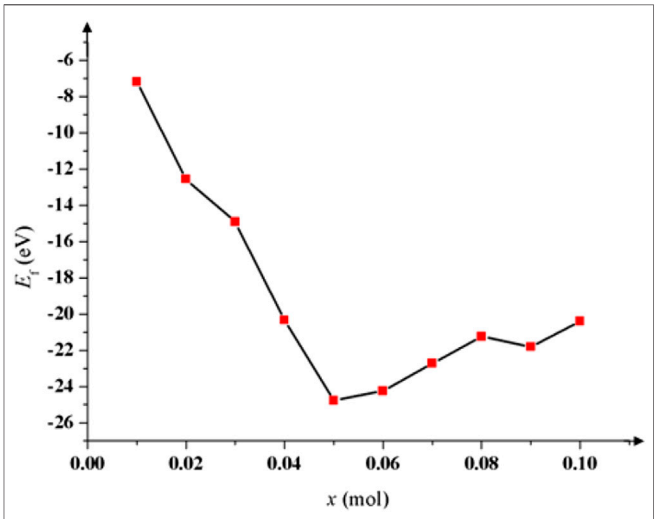

FIGURE 4 | Formation energy of intercalation of Li^+ with different doping amount *x* of Na. When *x* = 0.05 mol, E_f is lowest and the cycling performance is best.

With the development of computer technology, computing power has been improved dramatically, and programs based on DFT have gradually become a universal means to study the properties of materials. Multielectron system is the research object of DFT, and HK theorems are the theoretical basis of this method.

METHOD AND MODEL

Based on PW91 (Perdew-Wang) [25] method with the PBE (Perdew-Burke-Ernzenhof) [26] exchange-correlation functional, DFT calculations have been carried out using GGA in the CAmbridge Serial Total Energy Package (CASTEP) program of Materials Studio 8.0. The CASTEP is the quantum mechanical procedure of the plane wave pseudopotential method. An ultrasoft pseudopotential is used to describe the Coulomb attraction potential between the inner layer electrons around the nucleus and that of the outer layer [27], A plane wave cutoff of 340 eV and a k-point mesh of 4 × 4 × 1 in the Monkhorst–Pack [28] sampling scheme are used. The parameter about the self-consistent calculation is 1×10^{-6} eV/atom. To obtain the local stable structure of material, the structural geometry is optimized until the average force on every atom falls less than 0.3 eV/nm, the maximum displacement tolerance is less than 1×10^{-4} nm, and the maximum stress tolerance less than 0.05 GPa.

For this theoretical calculation, the electronic conductivity of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ is calculated by the CASTEP program with the virtual mixed atom method. The chemical formula assumes Li and Na as 1 mol, the doping amount of Na is *x* mol, and then the doping amount of Li is 1-*x* mol. The $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ cell model is shown in **Figure 1**. Li and Na occupy 3a, $Ni_{1/3}Co_{1/3}Mn_{1/3}$ occupies 3b, O atoms occupy 6c, and the space group is $R\bar{3}m$. Based on the model

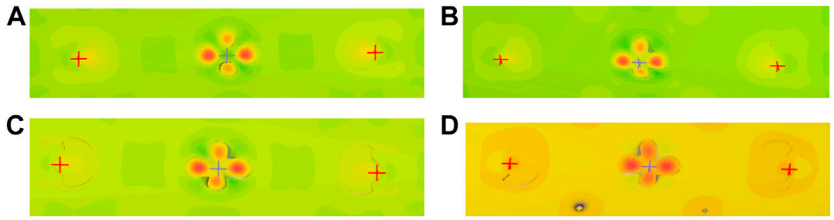

FIGURE 5 | Electron density difference of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$. **(A)** When $x = 0$, the color and coverage of the electron cloud around atoms are shown. **(B)** When $x = 0.01$ mol, there is no difference for the color between $x = 0.01$ mol and $x = 0$ mol, and the coverage of the electron cloud expands. **(C)** When $x = 0.05$ mol, its color gets darker. **(D)** When $x = 0.10$, the orange color indicates that its conductivity is improved extremely.

of **Figure 1**, 4 × 3 × 2 supercell model is built. Virtual atomic mixing can speed up the calculation and meet the precision requirements. The results of calculation of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ (x = 0.01, 0.02, 0.03 ... 0.10) are analyzed in this article.

RESULTS AND DISCUSSION

Band Gap and Partial Density of States

The band gap and the partial density of states (PDOS) of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ were calculated when Na-doping amount x = 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, and 0.10 mol. The electronic conductivity of the material depends on the number of electrons contained in the conduction band. The wider the band gap is, the less easily the valence band electrons can leap to the conduction band and the worse the conductivity of the material is. After Na-doping, the whole band structure of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ is not affected, but the energy gap has changed significantly. The relevant data of the band gap is shown in **Table 1**, and the diagram of the relationship between x and the band gap is shown in **Figure 2**. Compared with the pristine, the change of the band gap is most obvious when x = 0.01 mol. The results show that its band gap is decreasing when x increases, and Na-doping can effectively improve the conductivity of $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$.

The peak of the PDOS reflects the number of electrons in this energy. The bonding and the density of states near the Fermi level can be obtained by PDOS, and the influence of Na-doping on the conductivity of $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ can be analyzed. Their PDOS are shown in **Figure 3** when x = 0, 0.01 and 0.05 mol. The colored lines represent the density of states of different orbitals.

Figures 3A–D show that Na-doping has a great influence on the number of electrons near the Fermi level. When x = 0.01 mol (shown in **Figure 3B**), the PDOS peak is about 148 eV that is higher than that of the pristine (shown in **Figure 3A**). When x = 0.03–0.06 mol, the PDOS peak near Fermi level increases to 157 eV and its conductivity becomes better; when x = 0.06–0.09 mol, the PDOS peak begins to decline, its conductivity becomes worse; when x = 0.10 mol, the PDOS peak is lower than that of the pristine. Thus, the right amount of Na-doping should be x < 0.06 mol.

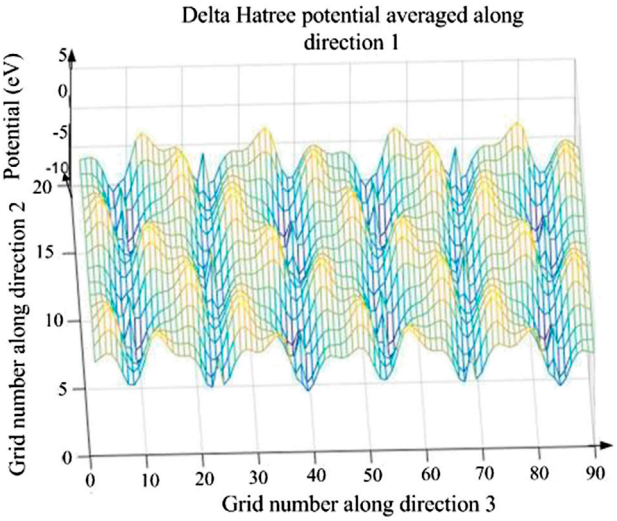

FIGURE 6 | The 3D potential energy map of $Li_{0.95}Na_{0.05}Ni_{1/3}Co_{1/3}Mn_{1/3}O_2$. The potential barrier and potential well take turns, which is consistent with the layered structure of $Li_{0.95}Na_{0.05}Ni_{1/3}Co_{1/3}Mn_{1/3}O_2$.

Formation Energy of Intercalation of Li^+ and Cell Volume

The greater the formation energy of metal oxide is, the more stable the matter is, the more difficult it is for atoms to separate from molecules. The formation energy of the intercalation of Li^+ (E_f) is shown in **Equation (4)**.

$$E_f = E_t - E_{dl} - E_{dil}, \quad (4)$$

where E_t is the total energy of the supercell, E_{dl} is the deintercalation energy of Li^+, and E_{dil} is the energy of the supercell after the deintercalation of Li^+. The 4 × 3 × 2 supercell models are built before $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ is intercalated/deintercalated of Li^+ and when it is intercalated/deintercalated of Li^+ and deintercalated of Li^+. The calculation of the cell's volume shows that the volume of Na-doped $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ is not extremely getting bigger though Na^+ is bigger than Li^+ and the structure keeps stable very well. The results of the energies in **Equation (4)** show that E_f has changed with the Na-doping amount (shown in **Figure 4**). It can be seen

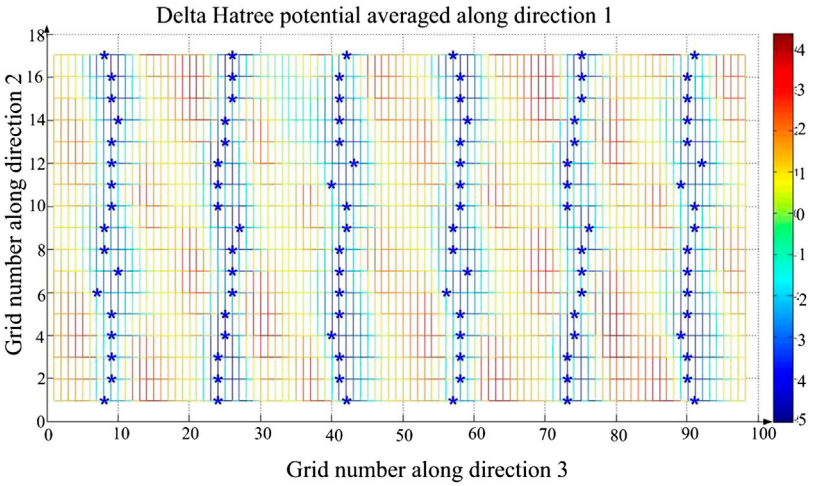

FIGURE 7 | The 2D potential energy map of $Li_{0.95}Na_{0.05}Ni_{1/3}Co_{1/3}Mn_{1/3}O_2$. The blue mark "*" represents the minimum potential energy, so the electrons will move along the path marked blue "*".

that Na-doping can effectively enhance the cycling performance of $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$. E_f decreases when $x < 0.05$ mol, and it rises gradually when $x > 0.06$ mol, so the right amount x is 0.05–0.06 mol.

Electron Density Difference

The change of electrons' distribution near the local atoms with Na-doping amount x is analyzed by the electron density difference of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$. The results are shown in **Figure 5**. Compared with the pristine (**Figure 5A**), there is no change in the color of the electron cloud around atoms when $x = 0.01$ mol (**Figure 5B**), but the electron cloud has more coverage, which exhibits that its conductivity is becoming better; when $x = 0.05$ mol, their color starts to get darker, meaning that the free electrons increase, but the electron cloud's coverage shrinks a little (**Figure 5C**). The electron cloud's color gets darker than before when $x > 0.05$ mol, and the electron cloud's coverage expands again (**Figure 5D**). So, Na-doping can improve the conductivity of $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$, and the best amount of Na-doping is $x =$ 0.05–0.10 mol.

Potential Energy of Electrons

For a system, its internal energy cannot change when there is no external force. In a potential well, the particles have the greatest kinetic energy and smallest potential, and particles can move freely in this potential well. If there is no external energy, it is difficult for them to cross over the potential well to other positions in the molecule. Therefore, the potential energy of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ can be analyzed to judge how difficult it is for electrons to escape from the potential well after doping. The 3D potential energy map of $Li_{0.95}Na_{0.05}Ni_{1/3}Co_{1/3}Mn_{1/3}O_2$ is shown in **Figure 6**. The blue represents the smaller negative potential energy and the red represents the larger positive potential energy. To analyze the influence of Na-doping on the potential energy well, the 2D potential energy diagram of $Li_{0.95}Na_{0.05}Ni_{1/3}Co_{1/3}Mn_{1/3}O_2$ is also obtained (shown in **Figure 7**). If electrons in a potential well (marked blue "*") get enough energy to escape from the potential well, they can move to other places in the molecule. At the same time, the barriers will prevent electrons from moving along the x-axis, so electrons must move along the minimum potential energy path marked blue "*" (shown in **Figure 7**). Finally, the electrons are deintercalated from the cathode to the anode.

In **Figure 7**, the blue mark "*" corresponds to the position and the potential energy. All the potential energy decreases from 17 to 1, and the change in each set is not exactly the same. As the amount of Na-doping increases, the average of the minimum potential energy decreases, which indicates that Na-doping will make the $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ potential well lower and it is easier for the electrons to escape from the potential well, and so it is for the charging reaction. Thus, its rate performance of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ can be effectively improved.

CONCLUSION

The physical and electrochemical properties of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ are calculated and analyzed by DFT. The band structure of $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ can keep stable after Na-doping, and the electronic potential energy becomes lower when the amount x of Na-doping rises; based on the analysis of the band gap and partial density of state, $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ has the best conductivity and rate performance when $x = 0.03$–0.06 mol; when $x = 0.05$ mol, the formation energy of intercalation of Li^+ is lowest, the electrons are easiest to remove; according to the analysis of the electron density

difference of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$, it has the best conductivity when x = 0.05–0.10 mol. In conclusion, the Na-doped $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ has better electrochemical performance when x = 0.05 mol, which is coherent with experimental results. The sample of $Li_{1-x}Na_xNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ can be synthesized by coprecipitation reaction combined with solid-phase sintering steps. Though the electrochemical performance of $Li_{0.95}Na_{0.05}Ni_{1/3}Co_{1/3}Mn_{1/3}O_2$ is proved theoretically and experimentally to be better than that of $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$, energy densities and specific capacity should be improved further with other modification methods. Our method and analysis based on the first-principles can provide some theoretical proposals for the electrochemical performance study of doping, and this method can be applied to the performance study of new optoelectronic devices.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YG designed models, analyzed some data, and wrote the manuscript. KS, PL, and WY carried out calculations. LL and FC gave some proposals. XH analyzed some results.

FUNDING

This research was funded by the Science and Technology Project Foundation of Zhongshan City of Guangdong Province of China (no. 2018B1127), the Guangdong Basic and Applied Basic Research Foundation (no. 2020A1515010420), the Key Research Platforms and Research Projects in Universities and Colleges of Guangdong Provincial Department of Education (no. 2018KQNCX334), the Zhongshan Innovative Research Team Program (no. 180809162197886), Guangdong Provincial Key Laboratory of Optical Information Materials and Technology (no. 2017B030301007), the National Natural Science Foundation of China (no. 11775047), the union project of National Natural Science Foundation of China and Guangdong Province (no. U1601214), Science and Technology Program of Guangzhou (no. 2019050001), and the Scientific and Technological Plan of Guangdong Province (no. 2018B050502010).

REFERENCES

1. Li M, Hou XH, Sha YJ, Wang J, Hu SJ, Liu X, et al. Facile spray-drying/pyrolysis synthesis of core-shell structure graphite/silicon-porous carbon composite as a superior anode for Li-ion batteries. *J Power Sources* (2014) 248:721–8. doi:10.1016/j.jpowsour.2013.10.012
2. Jung SK, Gwon H, Hong J, Park K. Understanding the degradation mechanisms of Li $Ni_{0.5}Co_{0.2}Mn_{0.3}O_2$ cathode material in lithium ion batteries. *Adv Energy Mater* (2014) 4(1):94–8. doi:10.1002/aenm.201300787
3. Zou BK, Ding CX, Chen CH. Research progress in ternary cathode materials Li (Ni, Co, Mn) O_2 for lithium ion batteries. *Sci Sin Chim* (2014) 44(7):1104–15. doi:10.1360/N032014-00019
4. Xu QJ, Zhou LZ, Liu MS, Pan HT, Deng XQ. Research progress in cathode material of Li-Ni-Co-Mn-O for lithium ion battery. *J Shanghai Univ Electr Power* (2012) 28(2):143–8. doi:10.3969/j.issn.1006-4729.2012.02.010
5. Zhang HL, Zhao HB, Xu JJ, Zhang JJ. Optimizing Li_2O-$2B_2O_3$ coating layer on $LiNi_{0.8}Co_{0.1}Mn_{0.1}O_2$ (NCM811) cathode material for high-performance lithium-ion batteries. *Int J Green Energy* (2020) 17(7):447–55. doi:10.1080/15435075.2020.1763362
6. Li JH, Liu ZQ, Wang YF, Wang RG. Investigation of facial B_2O_3 surface modification effect on the cycling stability and high-rate capacity of $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ cathode. *J Alloys Compd* (2020) 834:155150. doi:10.1016/j.jallcom.2020.155150
7. Ren XY, Du JL, Pu ZH, Wang RB, Gan L, Wu Z. Facile synthesis of Li_2MoO_4 coated $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ composite as a novel cathode for high-temperature lithium batteries. *Ionics* (2020) 26:1617–27. doi:10.1007/s11581-020-03474-z
8. Li L, Liu Q, Huang JJ, Luo SY, Feng CQ. Synthesis and electrochemical properties of Zn-doping $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ cathode material for lithium-ion battery application. *J Mater Sci Mater Electron* (2020) 31:12409–16. doi:10.1007/s10854-020-03787-9
9. Shao ZC, Guo J, Zhao Z, Xia JL, Zhang Y. Preparation and properties of Al_2O_3-doping $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ cathode materials. *Adv Manuf Process* (2016) 31(8):1004–8. doi:10.1080/10426914.2015.1117618
10. Chen JW, Xing LD, Yang XR, Liu X, Li TJ, Li WS. Outstanding electrochemical performance of high-voltage $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ cathode achieved by application of $LiPO_2F_2$ electrolyte additive. *Electrochim Acta* (2018) 290:568–76. doi:10.1016/j.electacta.2018.09.077
11. Teng R, Yu HT, Guo CF, Wang XD, Yi TF. Effect of cation doping on the electrochemical properties of li_2MoO_3 as a promising cathode material for lithium-ion battery. *Ionics* (2020) 26(17):4413–22. doi:10.1007/s11581-020-03607-4
12. Xing L, Peng H, Wang MS, Zhao X, Xu J, Wang ZQ, et al. Enhanced electrochemical performance of Zr^- modified layered $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ cathode material for lithium-ion batteries. *Chemelectrochem* (2016) 3(1):130–7. doi:10.1002/celc.201500360
13. He R, Zhang LH, Yan MF, Gao YH, Liu ZF. Effects of Cr_2O_3-modified $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ cathode material on the electrochemical performance of lithium-ion batteries. *J Mater Sci* (2016) 52:4599–607. doi:10.1007/s10853-016-0704-z
14. Yang ZG, Guo XD, Xiang W, Hua WB, Zhang J, He FR, et al. K-doped layered $LiNi_{0.5}Co_{0.2}Mn_{0.3}O_2$ cathode material: towards the superior rate capability and cycling performance. *J Alloys Compd* (2017) 699:358–65. doi:10.1016/j.jallcom.2016.11.245
15. Li YH, Liu JY, Lei YK, Lai CY, Xu QJ. Enhanced electrochemical performance of Na-doped cathode material $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ for lithium-ion batteries. *J Mater Sci* (2017) 52:13596–605. doi:10.1007/s10853-017-1449-z
16. Kabiraj A, Mahapatra S. High-throughput first-principles-calculations based estimation of lithium ion storage in monolayer rhenium disulfide. *Commun Chem* (2018) 1:81. doi:10.1038/s42004-018-0082-3
17. Ng MF, Sullivan MB. First-Principles Characterization of Lithium Cobalt Pyrophosphate as a cathode material for solid-State li-ion batteries. *J Phys Chem C* (2019) 123(49):29623–9. doi:10.1021/acs.jpcc.9b09946
18. Wang WW, Zhong Y, Zheng DH, Liu HD, Kong YF, Zhang LX, et al. P-type conductivity mechanism and defect structure of nitrogen-doped $LiNbO_3$ from first-principles calculations. *Phys Chem Chem Phys* (2020) 22(1):20–7. doi:10.1039/c9cp05019a
19. Thomas LH. The calculation of atomic fields. *Math Proc Camb Phil Soc* (1927) 23:542. doi:10.1017/S0305004100011683
20. Fermi EZ. Eine statistische methode zur bestimmung einiger eigenschaften des atoms und ihre anwendung auf die theorie des periodischen systems der elemente. *Ztschrift für Physik A Hadrons and Nucl* (1928) 48(1):73–9. doi:10.1007/BF01351576
21. Hohenberg P, Kohn W. Inhomogeneous electron gas. *Phys Rev* (1964) 136(3B):B864–B871. doi:10.100/CUL-ID:1483532

22. Jones RO, Gunnarsson O. The density functional formalism, its applications and prospects. *Rev Mod Phys* (1989) 61(3):689–746. doi:10.1103/RevModPhys.61.689
23. Kohn W, Sham L. Self-consistent equations including exchange and correlation effects. *Phys Rev A* (1965) 140:1133–8. doi:10.1103/physrev.140.a1133
24. Perdew JP, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Singh DJ et al. Erratum: atoms, molecules, solids, and surface: applications of the generalized gradient approximation for exchange and correlation. *Phys Rev B* (1992) 46(11):6671–87. doi:10.1103/PhysRevB.46.6671
25. Kresse G, Joubert D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys Rev B* (1999) 59:1758–75. doi:10.1103/PhysRevB.59.1758
26. Perdew JP, Burke K, Ernzerhof M. Generalized gradient approximation made simple. *Phys Rev Lett* (1996) 77:3865–8. doi:10.1103/PhysRevLett.77.3865
27. Vanderbilt D. Soft self-consistent pseudopotentials in generalized eigenvalue formalism. *Phys Rev B* (1990) 41(11):7892–5. doi:10.1103/PhysRevB.41.7892
28. Monkhorst HJ, Pack JD. Special points for brillouin-zone integrations. *Phys Rev B* (1976) 13:5188–92. doi:10.1103/PhysRevB.16.1748

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 16 March 2021
doi: 10.3389/fphy.2020.610510

The Parameters of the Field Emission Model and the Fabrication of Zinc Oxide Nanorod Arrays/Graphene Film

JianJun Yang[1], ShuangShuang Shao[1,2], YaXin Li[1,2], YueHui Wang[1] and Chi Feng[1]*

[1]College of Electron and Information, University of Electronic Science and Technology of China, Zhongshan Institute, Zhongshan, China, [2]School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

A large-scale growth of zinc oxide (ZnO) nanorod arrays on graphene sheets was fabricated by a hydrothermal technique, and the Fowler–Nordheim theory was used to build a model to describe the properties of the arrays' field emission. The results indicated that the morphological characteristics of the ZnO nanorods grown on the graphene sheets can be easily tuned by varying the reaction time and concentrations of the reaction solution. The regular ordered ZnO nanorods arrays on the graphene sheets were obtained under the appropriate experimental conditions. Further, this composite cathode was demonstrated to possess excellent field emission properties due to the outstanding mechanical and electrical properties of graphene. The field emission current density of the composite cathode reached 1,448 μA cm^{-2} at the electric field of 16.5 V μm^{-1}. The key parameter, field enhancement factor, reached 6,366, while the pure graphene cathode field is about 1,660. These specific nanorod arrays with enhancement of the field emission properties would be useful to sensor or modulator units for accessing networks.

Keywords: zinc oxide nanorod, graphene, field emission, sensor, modulator

OPEN ACCESS

Edited by:
Guofu Zhou,
South China Normal University, China

Reviewed by:
Li Lu,
Chongqing University of Arts and Sciences, China
Lei Xi,
Department of Biomedical Engineering, Southern University of Science and Technology, China

****Correspondence:***
JianJun Yang
sdyman@uestc.edu.cn

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *26 September 2020*
Accepted: *29 October 2020*
Published: *16 March 2021*

Citation:
Yang J, Shao S, Li Y, Wang Y and Feng C (2021) The Parameters of the Field Emission Model and the Fabrication of Zinc Oxide Nanorod Arrays/Graphene Film. Front. Phys. 8:610510. doi: 10.3389/fphy.2020.610510

INTRODUCTION

Graphene is usually used as a single cathode field emission material due to its excellent properties, such as the perfect quantum tunneling effect, excellent electrical conductivity, and mechanical properties [1]. It can also form sharp edges and concaves which can obtain a larger field enhancement factor. All of these made it possible to have the excellent field emission properties [2, 3]. It is advantageous for electron tunneling, and its two-dimensional conductive surface has the advantage in larger current [4–6]. Therefore, graphene can be used as a kind of field emission cathode material [4–8]. However, the emission performance of the graphene cathode field and the stability of the emission are still need to be solved urgently. The sharpness of the ground of graphene prepared by the chemical vapor deposition (CVD) method was limited due to the choice of flat copper or nickel as the substrate, which also limited the field emission properties [9–12].

ZnO nanorods with a wide bandgap (3.2eV) are good candidates for use in next-generation electronics and optoelectronics [13–17]. Zinc oxide/graphene hybrids have excellent emitting behavior characteristic, which is suitable for application in the field emission technology [18, 19]. Raman spectroscopy has figured out many excellent properties of graphene layers [20–23]. Therefore, the compound of graphene sheets with ZnO nanorods had recently been reported for multifunctional applications, such as gas sensors, light-emitting diodes, and solar cells [24–30].

So far, there was seldom report on the transfer of graphene sheets onto the silicon substrate with the silica film by wet transfer and the growth of ZnO nanorods on the graphene sheets. In this study, the two-step method was used to prepare ZnO nanorods on the surface of graphene, and the effect of the experimental conditions on the morphology of the ZnO nanostructure had been studied. The field emission properties of such ZnO nanorod/graphene cathodes have been described by a model and been tested.

PRINCIPLE

As we know, the expression of the field emission current density is as follows, based on Fowler–Nordheim theory (F–N equation):

$$J = \frac{AF^2}{\phi t^2}\exp\left[-B\frac{\phi^{3/2}}{F}v(y)\right] \quad (1)$$

where J is the current density and F is the local field strength of the cathode surface; A and B are constants with the value of 1.54×10^{-6} and 6.87×10^{7}, respectively; ϕ is the work function; and y represents $3.79 \times 10^{-4} \times E^{1/2}/\phi$, $v(y) = (0.95 - y^2)$, where E is the applied field strength. And, t^2 has a value close to 1.1.

Considering that the field emission cathode is usually with a special shape, the field enhancement factor (β) is introduced into the expression **Eq. 1**, and F can be expressed by $\beta \times E$. And, **Eq. 1** can be transformed as follows:

$$\ln\left(\frac{J}{E^2}\right) = -\frac{B\phi^{3/2}}{\beta E} - \ln\left(\frac{\phi}{A\beta^2}\right) \quad (2)$$

It can be used as a basis to illustrate the field emission mechanism with a synthesized cathode. The field emission properties can be described by $\ln(J/E^2) - (1/E)$ curve. And, the field enhancement factor can be obtained according to the curve slope.

MATERIALS AND METHODS

Materials

Copper foil (Cu) was purchased from Xiamen Xicheng Graphene Technology Co., Ltd., acetone (CH_3COCH_3, ≥99.5%) and isopropyl alcohol (C_3H_8O, 99.7%) were purchased from Zhongshan Jingke Chemical Instrument Co., Ltd., silica wafer (SiO_2/Si, 300 nm) was purchased from Harbin Tebo Technology Co., Ltd., and poly (methyl methacrylate) (PMMA, 60 mg/ml) was purchased from Sigma-Aldrich (Shanghai) Trading Co., Ltd. Methane (CH_4), hydrogen (H_2), and argon (Ar) were purchased from Zhongshan Fu'an Gas Co., Ltd. The ZnO sol–gels consist mainly of three kinds of materials, zinc acetate, ethanolamine, and ethanol. Zinc acetate ($Zn(CH_3COO)_2$, 99.99%) and ethanolamine (C_2H_7NO, ≥99.0%) were purchased from Shanghai Aladdin Biochemical Technology Co., Ltd. Ethanol absolute (CH_3CH_2OH, ≥99.7%) was purchased from Zhongshan Jingke Chemical Instrument Co., Ltd. Zinc nitrate hexahydrate ($Zn(NO_3)_2 \bullet 6H_2O$, 99.998%) and hexamethylenetetramine ($C_6H_{12}N_4$, ≥99.5%) were purchased from Shanghai Aladdin Biochemical Technology Co., Ltd.

METHOD

Synthesis of Graphene Sheets

Graphene sheets were synthesized on the Cu substrate by chemical vapor deposition (CVD). Cu foils ($40 \times 25 \times 0.1\ mm^3$) were washed with a mixed solvent of acetone, isopropyl alcohol, and deionized water by ultrasonic cleaning. Copper foil was loaded in the CVD quartz tube with the pressure of below 1.1 Pa and heated at 1,030°C for 65 min in hydrogen without changing the flowing rate. Then, the mixture of gases (H_2: CH_4) with the flowing rate ratio of 50:1 was introduced into the flow as the carbon source for 20 min. After reaction, the quartz tube was dropped to room temperature. The sample was removed from a tube furnace.

Transfer of Graphene Sheets

The transfer process was based on the utilization of the poly (methyl methacrylate) (PMMA) film as a mediator. PMMA was spin-coated on the surface of graphene sheets with the Cu substrate. After baking at 140°C for 10 min, the organic solvent was removed from PMMA. The sample was placed in a 30-mg/ml ammonium persulfate solution for etching to strip the copper foil. After etching, PMMA/graphene sheets were moved into a beaker containing deionized water and transferred onto the silicon substrate and dried at room temperature. Then, the sample was baked at 120°C for an hour. Finally, the sample was soaked in an acetone solution at 50°C to remove PMMA, and then washed with isopropyl alcohol.

Preparation of ZnO Nanorods/Graphene Composite Film

A certain amount of zinc nitrate hexahydrate and hexamethylenetetramine was mixed in water as the reactant solution. ZnO sol–gel as a ZnO seed layer was spin-coated onto the graphene sheets and transferred onto the silicon substrate. Then, the samples were placed in a reaction kettle with a reactant solution at 95°C for the suitable time for the growth of ZnO nanorods through a hydrothermal method, followed by natural cooling to room temperature and washed with deionized water for several times and dried at room temperature. The morphological characteristics of the ZnO nanorods grown on the graphene sheets were investigated at different reaction solution concentrations and temperatures.

A schematic diagram of the fabrication process is illustrated in **Figure 1**.

Field Emission Measurement

The field emission properties were measured by using a parallel plate electrode configuration. The sample was fixed on a stainless stage with a conductive tape as the cathode, and a copper electrode with the bottom diameter of 2 mm was used as an anode. The field emission measurement needed a vacuum below

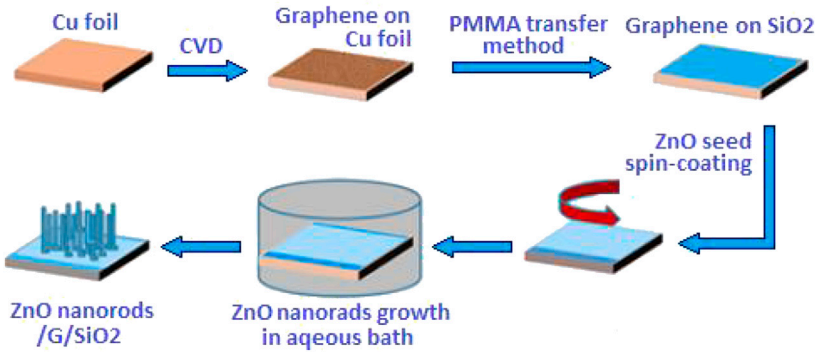

FIGURE 1 | Schematic of the fabrication process of the graphene/ZnO composite film.

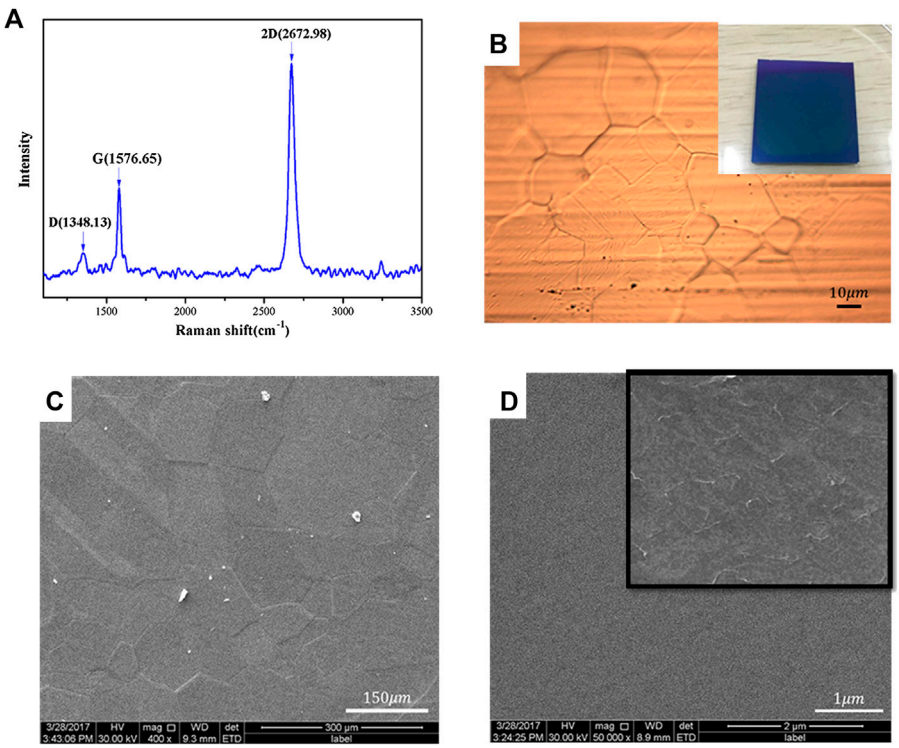

FIGURE 2 | Raman spectrum of graphene transferred onto the silicon substrate **(A)** and the optical microscope image of copper-based graphene **(B)** and SEM images of copper-based graphene **(C)** and the flattened graphene transferred onto the SiO2/Si substrate **(D)**.

2×10^{-4} Pa. The emission current was measured at voltages ranging from 100 to 5000 V. The spacing between the anode and the cathode was maintained at 300 μm.

RESULTS AND DISCUSSIONS

Figure 2 shows the Raman spectrum of the graphene sheets transferred onto the silicon substrate (**Figure 2A**), and the optical image (**Figure 2B**) and SEM images of the graphene film onto the copper substrate (**Figure 2C**) and silicon substrate (**Figure 2D**), respectively. **Figures 2B** and **2D** are photos of the graphene sheets on the silicon substrate and local magnification, respectively.

Raman spectroscopy is a promising method for inspecting the ordered/disordered crystal structures of carbonaceous materials and the different layer characteristics of graphene. From **Figure 2A**, it is observed that the Raman spectrum of graphene exhibited the D peak at 1,348.31°cm^{-1}, the G peak at 1,576.65°cm^{-1}, and the 2D peak at 2,672.98°cm^{-1}. The intensity of D peak was very small, indicating that the local defects and disorder of graphene transferred onto the silicon substrate were little. The number of layers of graphene can be determined by the

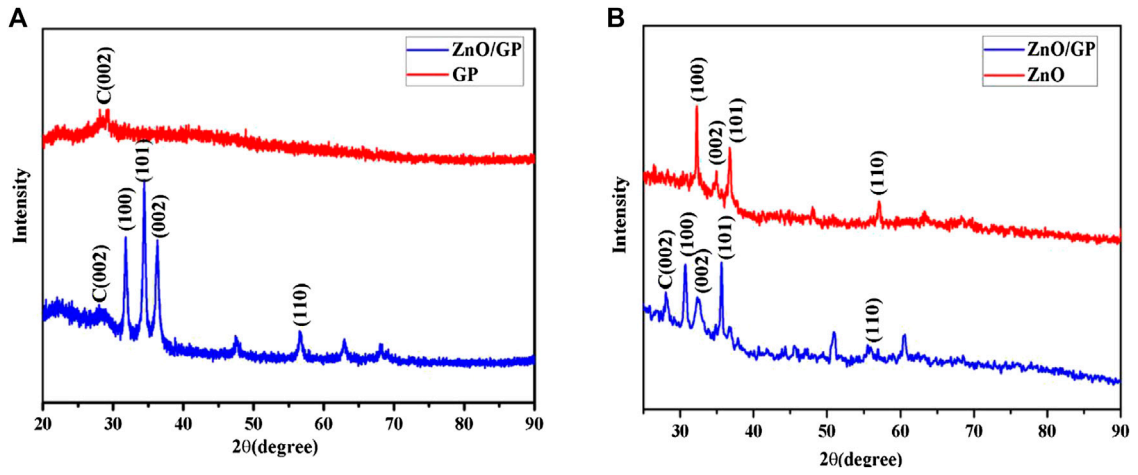

FIGURE 3 | XRD spectra of graphene and graphene/ZnO nanorods **(A)**. ZnO and graphene/ZnO nanorods **(B)**.

peak intensity ratio of the G peak to that of the 2D peak of the Raman spectrum [20–23]. For our sample, the peak intensity ratio of the G peak to that of the 2D peak was approximately 2, indicating that graphene is a monolayer [20]. The distribution of grain and grain boundary of graphene on the copper substrate was observed in the optical image (**Figures 2B,C**). The color of the graphene film transferred onto the silicon substrate was dark blue (inserted in **Figure 2B**). The graphene sheets on the copper substrate were flat. However, the wrinkles of graphene sheets on the silicon substrate were observed (**Figure 2D**). The edge of the graphene film would be rolled up when transferring. And there would be "fracture" resulting in an incomplete film due to the softness of the copper foil. It was inevitable that the transfer process causes the graphene sheets to be wrinkled because of the soft monolayer structure. In addition, the surface of graphene will absorb impurities during the transfer.

ZnO nanorod arrays on graphene sheets were obtained through hydrothermal routes. **Figure 3** shows the XRD patterns of the ZnO nanorod arrays on graphene sheets and the pure graphene, respectively. A diffraction peak of the pure graphene near 2θ value of 26.4° was observed in **Figure 3**. There was a strong diffraction peak in the vicinity of 2θ value of 31.8°, corresponding to the diffraction peak of the wurtzite-type zinc oxide (100) crystal plane. There were also some diffraction peaks near 34.4°, 36.3°, and 56.6°, corresponding to zinc oxide with crystal planes of (002), (101), and (110). The comparison between ZnO and the composite films in **Figure 3B** shows that there were some other diffraction peaks except the diffraction peaks in about the same position as those in **Figure 3A**, but the intensity was relatively small, indicating the quality of the composite film was better.

To observe the growth of the ZnO nanorods on the graphene sheets, SEM images were taken, as shown in **Figures 4** and **5**. **Figure 4** shows SEM images of the ZnO nanorods on the graphene sheets with zinc nitrate hexahydrate of 10 mM (**Figure 4A**), 15 mM (**Figure 4B**), and 25 mM (**Figure 4C**), respectively, for 5 h. **Figure 4D** is the local magnification of **Figure 4C**. The density and diameter and size of ZnO nanorods increased with increasing of the reaction solution concentration. When the reaction solution concentration was 25 mM, the uniform ZnO nanorods were successfully grown on the graphene surface, and the order of ZnO nanorod arrays could be obtained. The average length and diameter of ZnO nanorods is 2 μm and 200 nm, respectively. The hexagonal structure of ZnO naorods can be clearly seen from **Figure 4D**.

Figure 5 shows the SEM images of the ZnO nanorods on the graphene sheets with zinc nitrate hexahydrate of 25 mM and the reaction time of 3 h (**Figure 5A**), 4 h (**Figure 5B**), and 5 h (**Figure 5C**), respectively. The inset in **Figure 5C** is the local amplification and the sectional drawing of the sample (**Figure 5D**). It can be seen that the density and diameter and size and order of ZnO nanorods increased with increasing of the reaction time. The ZnO seed layer was preferably the growth of the c-axis–oriented ZnO nanostructures [26]. The formation of ZnO was a process of increasing the entropy from a chemical point of view [30]. The adhesion of hexamethylenetetramine (HMTA) to the nonpolar surface accelerated the growth of the crystal in the [0001] direction and slowly hydrolyzed HCHO and NH3 in the solution. All were beneficial to the orientation growth of ZnO nanorods to obtain longer arrays.

To analyze the field emission performance, the curves of field emission current density on the applied electric field (J–E) from graphene/ZnO nanorods based on CVD were plotted. Emission current density vs. electric field curves and F–N plot at different emission cathodes are shown in **Figures 6A and 6B**, respectively. From **Figure 6A**, when the emission current density is 1 $\mu A^{\circ}cm^{-2}$, the turn-on field of graphene/ZnO nanorod composite films ($4\ V^{\circ}\mu m^{-1}$) was obviously smaller than that of the pure graphene films ($7\ V^{\circ}\mu m^{-1}$). The reason to the graphene film with larger turn-on field could be that there were fewer sharp edges and concaves in graphene by CVD, which led to the linear change of the current and the slow growth at the beginning of the applied electric filed. The field emission current density of the composite cathode could reach to 1,448 $\mu A\ cm^{-2}$ when the electric field reaches 16.5 $V\ \mu m^{-1}$ (**Figure 6A**). Further, from **Figures 4B–D**, with the increase of reactant concentration, the current density decreased and the emission performance became

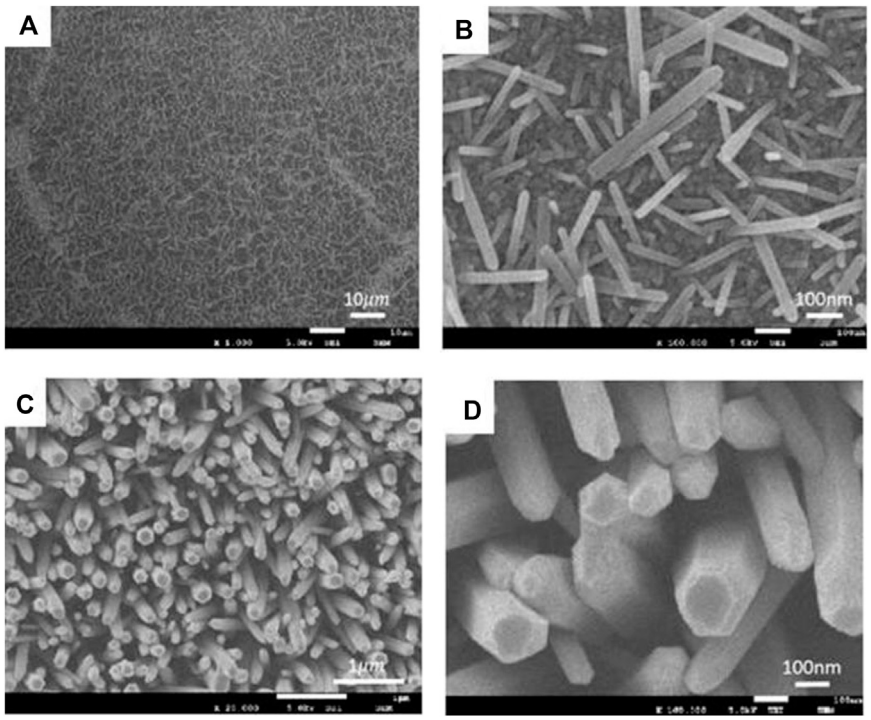

FIGURE 4 | SEM images of the ZnO nanorods on the graphene sheets with zinc nitrate hexahydrate of 10 mM **(A)**, 15 mM **(B)**, and 25 mM **(C)**, respectively, for 5 h.

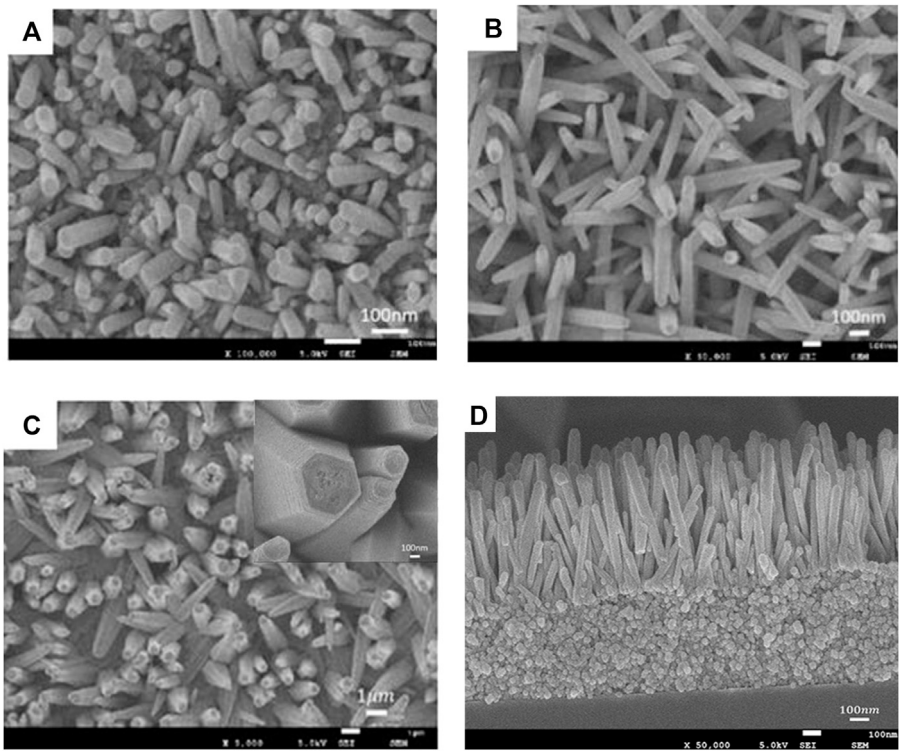

FIGURE 5 | SEM images of the ZnO nanorods on the graphene sheets with zinc nitrate hexahydrate of 25 mM and the reaction time of 3 h **(A)**, 4 h **(B)**, and 5 h **(C)**, respectively.

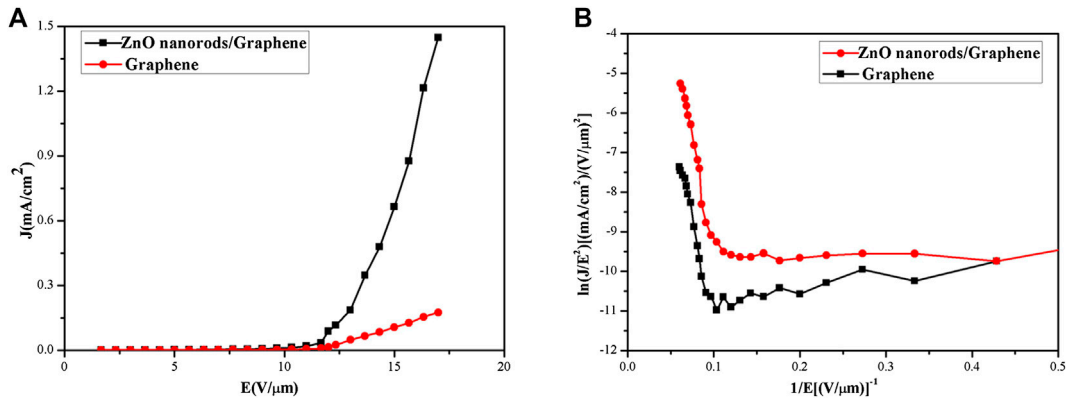

FIGURE 6 | (A) Emission current density vs. electric field curves and **(B)** F–N plot at different emission cathodes.

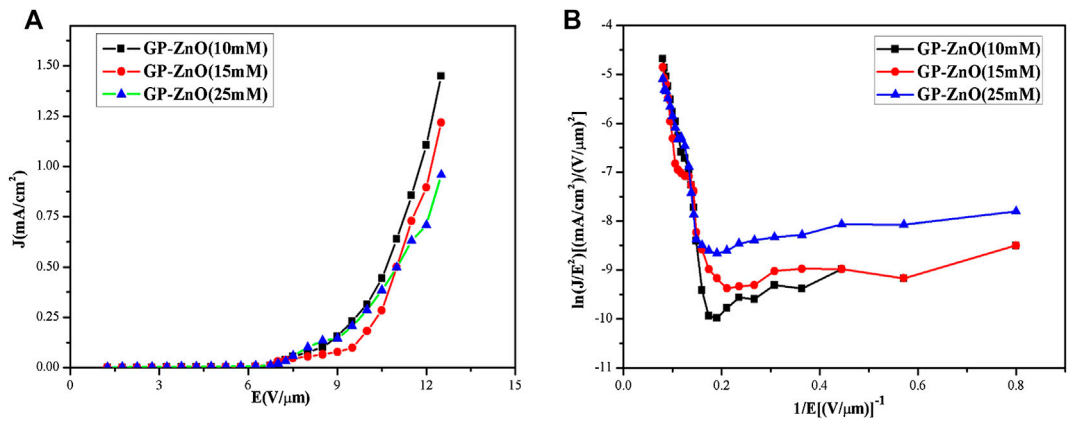

FIGURE 7 | (A) Emission current density vs. electric field curves and **(B)** F–N plot of graphene/ZnO nanorods at different concentrations of reactant.

worse. Emission current density vs. electric field curves and F–N plot of graphene/ZnO nanorods at different concentrations of reactants were shown in **Figures 7A and 7B**, respectively. From **Figure 7**, it is observed that current density decreased and the emission performance became worse with the increase of the reactant concentration.

The field enhancement factor can be used to described the field emission effect of this synthesized cathode [31]. From the F–N equation model, the field enhancement factor of the graphene cathode field is about 1,660, and the field enhancement factor of the graphene/ZnO composite cathode field could reach 6,366. It can be indicated from the F–N equation model that the field enhancement factor was determined by the slope of the ln (J/E^2) and (1/E) curves and the work function of the material. The work functions of graphene and zinc oxide are 5eV and 5.3eV, respectively [32]. Graphene/ZnO composites had a smaller work function than graphene or ZnO [33]. Therefore, the field enhancement factor of the composites was larger and the emission current density was higher, so the field emission performance was better.

Graphene had excellent electrical conductivity and electronic transport capability, which can form good ohmic contact with ZnO, and also improved the efficiency of electron transport. Therefore, the introduction of the graphene layer not only solved the contact problem between ZnO and the substrate but also promoted the transmission of electrons. In addition, the graphene films prepared by the CVD method mostly appeared as concaves. If the graphene film is deposited in a short time, there will be more defects and raised edges. And, they would be sharp at the edge after the wet transfer. Furthermore, the introduction of ZnO nanorods could make up for the defect of insufficient emission in the local area and obtained excellent field emission properties.

CONCULSIONS

Graphene sheets were synthesized on the Cu substrate by chemical vapor deposition and transferred onto the silicon substrate with poly (methyl methacrylate) (PMMA) as a mediator. Zinc oxide nanosructures grown on graphene sheets were prepared by the hydrothermal method. The results indicated that the morphological characteristics of the ZnO nanorods grown on the graphene sheets could be easily tuned by

varying the reaction time and concentration of the reaction solution. With increasing of the reaction time and concentration of the reaction solution, the density and size of ZnO nanorods increased. The regular ordered ZnO nanorod arrays on the graphene sheets were obtained at the conditions of zinc nitrate hexahydrate of 10 mM for 5 h. The field emission current density of ZnO nanorods arrays/graphene composite cathode reached 1,448 $\mu A\,cm^{-2}$ at the electric field of 16.5 $V\,\mu m^{-1}$. The field enhancement factor of ZnO nanorod arrays/graphene composite cathode field reached 6,366, while the pure graphene cathode field is about 1,660. The enhancement of the field emission properties indicated that ZnO nanorods arrays/graphene had lower turn-on fields and threshold fields as well as higher field enhancement factors than pure graphene.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JY, YW, and JY and YW designed this project. SS, YL, and FC carried out most of the experiments and data analysis.

FUNDING

This work was sponsored by National Key R&D Program of China (Grant No. 2018YFB0407100) and supported by Zhongshan Innovative Research Team Program (180809162197886).

ACKNOWLEDGMENTS

This work was supported by the State Key Laboratory of Electronic Thin Films and Integrated Devices and Zhongshan division of the State Key Laboratory of Electronic Thin Films and Integrated Devices. And, we thank Jian Zhong for this work constructive comments.

REFERENCES

1. Berger C, Song Z, Li X, Wu X, Brown N, Naud C, et al. Electronic confinement and coherence in patterned epitaxial graphene. *Science.* (2006). 312(5777): 1191–6. doi:10.1126/science.1125925
2. Deng JH, Yu B, Li GZ, Hou XG, Zhao ML, Li DJ, et al. Self-assembled growth of multi-layer graphene on planar and nano-structured substrates and its field emission properties. *Nanoscale.* (2013). 5:12388–93. doi:10.1039/c3nr04145j
3. Wang BB, Zheng K, Cheng QJ, Wang L, Zheng MP, Ostrikov K. Formation and electron field emission of graphene films grown by hot filament chemical vapor deposition. *Mater Chem Phys.* (2014). 144:66–74. doi:10.1016/j.matchemphys.2013.12.013
4. Zhang Y, Du J, Tang S, Liu P, Deng S, Chen J, et al. Optimize the field emission character of a vertical few-layer graphene sheet by manipulating the morphology. *Nanotechnology.* (2012). 23:015202. doi:10.1088/0957-4484/23/1/015202
5. Peng Y, Huang D. Fabrication of patterned reduced graphene oxide nanosheet field-emission cathodic film at room-temperature. *Appl Surf Sci.* (2013). 283: 81–6. doi:10.1016/j.apsusc.2013.06.029
6. Chen LF, He H, Lei D, Menggen Q, Hu LQ, Yang DR. Field emission performance enhancement of Au nanoparticles doped graphene emitters. *Appl Phys Lett.* (2013). 103:233105. doi:10.1063/1.4837895
7. Liu J, Zeng B, Wang W, Li N, Guo J, Fang Y, et al. Graphene electron cannon: high-current edge emission from aligned graphene sheets. *Appl Phys Lett.* (2014). 104:023101. doi:10.1063/1.4861611
8. Spindt CA, Brodie I, Humphrey L, Westberg ER. Physical properties of thinfilm field emission cathodes with molybdenum cones. *J Appl Phys.* (1976). 47: 5248–63. doi:10.1063/1.322600
9. Xiao ZM, She JC, Deng SZ, Tang ZK, Li ZB, Lu JM, et al. Field electron emission characteristics and physical mechanism of individual single-layer graphene. *ACS Nano.* (2010). 4(11):6332–6. doi:10.1021/nn101719r
10. Fowler RH, Nordheim L. Electron emission in intense electric fields. *Proc Roy Soc Lond A.* (1928). 119:173–81. doi:10.1098/rspa.1928.0091
11. Chang HC, Li CC, Jen SF, Lu CC, Bu IYY, Chiu PW, et al. All-carbon field emission device by direct synthesis of graphene and carbon nanotube. *Diam Relat Mater.* (2013). 31:42–6. doi:10.1016/j.diamond.2012.10.011
12. Jiang J, Du J, Wang Q, Zhang X, Zhu W, Li R, et al. Enhanced field emission properties from graphene-TiO2/DLC nanocomposite films prepared by ultraviolet-light assisted electrochemical deposition. *J Alloys Compd.* (2016). 686:588–92. doi:10.1016/j.jallcom.2016.06.084
13. Ng HT, Han J, Yamada T, Nguyen P, Chen YP, Meyyappan M. Single crystal nanowire vertical surround-gate field-effect transistor. *Nano Lett.* (2004). 4(7): 1247–52. doi:10.1021/nl049461z
14. Lim SY, Brahma S, Liu CP, Wang RC, Huang JL. Effect of indium concentration on luminescence and electrical properties of indium doped ZnO nanowires. *Thin Solid Films.* (2013). 549:165–71. doi:10.1016/j.tsf.2013.09.001
15. Li X, Li C, Hou S, et al. Thickness of ITO thin film influences on fabricating ZnO nanorods applying for dye sensitized solar cell. *Compos B.* (2015). 74:147. doi:10.1016/j.compositesb.2015.01.017
16. Wang WB, Gu H, He XL, Xuan WP, Chan JK, Wang XZ, , et al.Luo JK. Thermal annealing effect on ZnO surface acoustic wave-based ultraviolet light sensors on glass substrates. *Appl Phys Lett.* (2014). 104:212107. doi:10.1063/1.4880898
17. Garry S, McCarthy É, Mosnier JP, McGlynn E. Influence of ZnO nanowire array morphology on field emission characteristics. *Nanotechnology.* (2014). 25:135604. doi:10.1088/0957-4484/25/13/135604
18. Ding JJ, Yan XB, Xue QJ. Study on field emission and photoluminescence properties of ZnO/graphene hybrids grown on Si substrates. *Mater Chem Phys.* (2012). 133:405–9. doi:10.1016/j.matchemphys.2012.01.051
19. Liu J, Zhang Z, Lv Y, Yan J, Yun J, Wu Z, et al. Synthesis and characterization of ZnO NWAs/graphene composites for enhanced optical and field emission performances. *Compos B Eng.* (2016). 99:366–72. doi:10.1016/j.compositesb.2016.05.036
20. Hao Y, Wang Y, Wang L, Ni Z, Wang Z, Wang R, et al. Probing layer number and stacking order of few-layer graphene by Raman spectroscopy. *Small.* (2010). 6(2):195–200. doi:10.1002/smll.200901173
21. Ferrari AC, Meyer JC, Scardaci V, Casiraghi C, Lazzeri M, Mauri F, et al. Raman spectrum of graphene and graphene layers. *Phys Rev Lett.* (2006). 97: 187401. doi:10.1103/PhysRevLett.97.187401
22. Ferrari AC. Raman spectroscopy of graphene and graphite: disorder, electron–phonon coupling, doping and nonadiabatic effects. *Solid State Commun.* (2007). 143:47–57. doi:10.1016/j.ssc.2007.03.052
23. Malard LM, Pimenta MA, Dresselhaus G, Dresselhaus MS. Raman spectroscopy in graphene. *Phys Rep.* (2009). 473:51–87. doi:10.1016/j.physrep.2009.02.003
24. Özgür Ü, Alivov YI, Liu C, Teke A, Reshchikov MA, Doğan S, et al. A comprehensive review of ZnO materials and devices. *J Appl Phys.* (2005). 98: 041301. doi:10.1063/1.1992666

25. Greene LE, Law M, Tan DH, Montano M, Goldberger J, Somorjai G, et al. Yang P. General route to vertical ZnO nanowire arrays using textured ZnO seeds. *Nano Lett.* (2005). 5(7):1231–6. doi:10.1021/nl050788p
26. Chen ZW, Hong YY, Lin ZD, Liu LM, Zhang XW. Enhanced formaldehyde gas sensing properties of ZnO nanosheets modified with graphene. *Electron Mater Lett.* (2017). 13:270. doi:10.1007/s13391-017-6245-z
27. Biroiu RK, Giri PK. Strong visible and near infrared photoluminescence from ZnO nanorods/nanowires grown on single layer graphene studied using sub-band gap excitation. *J Appl Phys.* (2017). 122:044302. doi:10.1063/1.4995957
28. Yuan X, Yu LM, Ma HN, Li Y, Li C, Yin ML, et al. Direct synthesis of upstanding graphene/ZnO nanowalls/graphene sandwich heterojunction and its application for NO2 gas sensor. *J Nanosci Nanotechnol.* (2019). 19:7947–52. doi:10.1166/jnn.2019.16856
29. Liu J, Zhang WH, Wang AY, Zhang ZY, Lv YY. Preparation and field emission performance of metal-doped ZnO NRAs/graphene composite. *Material Integr Ferroelectr.* (2020). 209(1):98–109. doi:10.1080/10584587.2020.1728815
30. Xu S, Wang ZL. One-dimensional ZnO nanostructures: solution growth and functional properties. *Nano Res.* (2011). 4(11):1013–98. doi:10.1007/s12274-011-0160-7
31. Wu CX, Li FS, Zhang YG, Guo TL. Improving the field emission of graphene by depositing zinc oxide nanorods on its surface. *Carbon.* (2012). 50:3622–6. doi:10.1016/j.carbon.2012.03.033
32. Sun L, Zhou XT, Zhang YA, Guo TL. Enhanced field emission of graphene–ZnO quantum dots hybrid structure. *J Alloys Compd.* (2015). 632:604–8. doi:10.1016/j.jallcom.2015.01.105
33. Zhang S, Zhang Y, Huang S, Liu H, Wang P, Tian H. First-principles study of field emission properties of graphene-ZnO nanocomposite. *J Phys Chem C.* (2010). 114:19284–8. doi:10.1021/jp107780q

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership