frontiers in Neuroscience

ORIGINAL RESEARCH
published: 11 April 2018
doi: 10.3389/fnins.2018.00135

Event-Based Color Segmentation With a High Dynamic Range Sensor

Alexandre Marcireau, Sio-Hoi Ieng, Camille Simon-Chane† and Ryad B. Benosman*

Institut National de la Santé et de la Recherche Médicale, UMRI S 968, Sorbonne Universites, UPMC Univ Paris 06, UMR S 968, Centre National de la Recherche Scientifique, UMR 7210, Institut de la Vision, Paris, France

OPEN ACCESS

Edited by:
Runchun Mark Wang,
Western Sydney University, Australia

Reviewed by:
Chetan Singh Thakur,
Indian Institute of Science, India
Subhrajit Roy,
IBM Research, Australia
Michael Schmuker,
University of Hertfordshire, United Kingdom

****Correspondence:***
Sio-Hoi Ieng
siohoi.ieng@gmail.com

†Present Address:
Camille Simon-Chane, ETIS UMR 8051, Université Paris Seine, Université Cergy-Pontoise, ENSEA, Centre National de la Recherche Scientifique, Paris, France

Specialty section:
This article was submitted to Neuromorphic Engineering, a section of the journal Frontiers in Neuroscience

Received: *11 October 2017*
Accepted: *20 February 2018*
Published: *11 April 2018*

Citation:
Marcireau A, Ieng S-H, Simon-Chane C and Benosman RB (2018) Event-Based Color Segmentation With a High Dynamic Range Sensor. Front. Neurosci. 12:135. doi: 10.3389/fnins.2018.00135

This paper introduces a color asynchronous neuromorphic event-based camera and a methodology to process color output from the device to perform color segmentation and tracking at the native temporal resolution of the sensor (down to one microsecond). Our color vision sensor prototype is a combination of three Asynchronous Time-based Image Sensors, sensitive to absolute color information. We devise a color processing algorithm leveraging this information. It is designed to be computationally cheap, thus showing how low level processing benefits from asynchronous acquisition and high temporal resolution data. The resulting color segmentation and tracking performance is assessed both with an indoor controlled scene and two outdoor uncontrolled scenes. The tracking's mean error to the ground truth for the objects of the outdoor scenes ranges from two to twenty pixels.

Keywords: event-based signal processing, AER, color segmentation, tracking, silicon retina

1. INTRODUCTION

Primates' ability to discriminate colors is advantageous for survival (Dominy and Lucas, 2001). They use it for long-range detection of edible food in forest environments. This ability is nowadays exploited by humans to efficiently communicate information: road signs, merchandizing and maps are a few examples. Consequently, machine vision systems meant to interface with humans or mimic their vision often rely on colors (Trémeau et al., 2008). Applications include traffic sign recognition (Bahlmann et al., 2005), skin detection (Kakumanu et al., 2007), visual saliency modeling (van de Weijer et al., 2006) or vehicle color classification (Hsieh et al., 2015). However, detecting colored objects in a scene remains a challenge for such systems: even though a considerable amount of research has been carried out on color segmentation, *ad hoc* techniques are still required to solve many problems. The large amount of recently published works (Vantaram and Saber, 2012) tends to demonstrate that automated color segmentation is far from being solved. Current state-of-the-art methods for color video segmentation rely on a model describing tracked objects: superpixels (Pun and Huang, 2016), graphs (Rother et al., 2004; Grundmann et al., 2010; Lezama et al., 2011) or local classifiers (Bai et al., 2010; Lee et al., 2011). These methods yield robust and accurate results, but require heavy computations. Other methods rely on clustering techniques, especially *Mean Shift* derivatives, to segment colors (Fukunaga and Hostetler, 1975; Cheng, 1995).

The results quality and very high computational costs drove research on speed optimization and complexity reduction through structuring of the feature space (Guo et al., 2006; Paris and Durand, 2007; Xiao and Liu, 2010), dynamic bandwidth selection (Yang et al., 2003) or kernel choice improvements (Comaniciu, 2003). Despite these optimizations, state-of-the-art methods still require large computations preventing their use for real-time or low-power applications. This need derives from the dense and exceedingly redundant visual information provided by conventional, frame-based cameras. By contrast, the human eye contains several neuron layers known to reduce redundancy and participate in color information processing (Johansson, 2004).

This work introduces a new direction for color segmentation, with algorithms operating on high temporal resolution data (down to one microsecond) provided by neuromorphic event-based cameras which mimic the human eye. These sensors are based on pixels operating independently. Instead of capturing information at a fixed frame-rate, with no relation to the visual information source (Lichtsteiner et al., 2006), each pixel optimizes its sampling depending on the visual information it receives. If the scene changes quickly, the pixel samples information with a high adaptive rate. Otherwise, the pixel stops acquiring redundant data and goes idle until luminance changes in its field of view, therefore contributing to information processing. The sensor does not require a common frame clock, since event-based cameras' pixels are independent and autonomous. A variety of such sensors were developed in the past few years: temporal contrast vision sensors detecting relative luminance changes (Lichtsteiner et al., 2006; Posch et al., 2008, 2011), gradient-based sensors detecting static edges (Delbruck, 1993), edge-orientation sensitive devices (Etienne-Cummings et al., 1997) and optical flow sensors (Krammer and Koch, 1997).

The *Asynchronous Time-based Image Sensor* (Posch et al., 2011), or ATIS, used in this paper is an asynchronous camera that contains an array of autonomously operating pixels that combine an asynchronous change detection circuit and a separate exposure measurement circuit, the latter being triggered by the former. Each pixel independently and continuously monitors its field of view. The detection of luminance change triggers a local light integration, as illustrated **Figure 1**. The information is output asynchronously with the pixel coordinates, hence providing the new gray level. Consequently, the scene is not acquired frame-wise, but rather continuously and locally, conditionally on visual information changes. In other words, only information that is relevant—because unknown—is acquired, transmitted, stored and eventually processed by machine vision algorithms. Pixel acquisition and readout times range from milliseconds to microseconds, resulting in temporal resolutions equivalent to conventional sensors running at tens to hundreds of thousands frames per second. The sparse nature of the generated visual data benefits subsequent processing in terms of speed and power consumption. Moreover, the data's high temporal resolution allows for simplifying assumptions, with complex behaviors emerging from simple, high-speed algorithms. The event-based formulation of vision problems in the time domain has already produced striking results for many computer vision algorithms, such as stereo-vision (Rogister et al., 2012; Carneiro et al., 2013), optical flow (Benosman et al., 2014) or tracking (Ni et al., 2012).

The benefits of event-based cameras make them well-suited candidates to overcome existing limitations in automated color segmentation. To our knowledge, only two attempts at building a color event-based sensor have been made. The pixel proposed by Berner and Delbruck (2011) is sensitive to both luminance and wavelength changes. Moeys et al. (2017) added a Bayer matrix to an existing DVS sensor. However, both sensors are only sensitive to relative luminance changes, and were not illustrated with concrete applications. Using the ATIS capacity to acquire absolute luminance in an event-based manner, we present in this paper both a functional event-based color sensor, illustrated **Figure 2**, and its application for segmenting colored objects with simple processing techniques requiring little computation power. The absolute luminance information allows for a robust and computationally cheap color segmentation based on clustering, unlike change detectors which need to rely on edges detection. We evaluate the sensor's ability to track colored shapes, using a real-time on-line algorithm. Thanks to the nature of the data generated by event-based cameras, tracking can be implemented with a moving mean algorithm (Drazen et al., 2011), which requires very little computational power. More complex and robust methods have been devised (Lagorce et al., 2015; Reverter Valeiras et al., 2015) for more demanding applications.

2. MATERIALS AND METHODS

2.1. Three-Chip Event-Based Camera

We build an event-based color sensor as an association of three ATIS cameras acquiring red, green and blue light exposures. The sensor captures light through a hot mirror reflecting infra-red light. A beam splitter directs photons with wavelengths larger than 605 nm toward the red sensor. The other photons are reflected toward a second beam splitter, which directs photons with wavelengths smaller than 505 nm toward the blue sensor. The remaining photons are directed toward the green sensor. Before hitting the red, green and blue sensors, photons cross band-pass filters which mimic the filtering functions of conventional Bayer matrices' pixels. Each sensor uses a C-mount objective, as the sensors dimensions prevent using a common objective placed behind the hot-mirror. **Figure 3** illustrates the assembly.

In order to account for the mechanical imperfections of the prototype, a spatial calibration step is required to make sure that the color sensor's cameras share the same field of view. We capture a checker board with the sensor before each recording, and compute the homography linking the green and blue cameras to the red one. The homography is computed by determining the direct linear transformation on normalized points (Hartley, 1995), which were extracted from a reconstructed image of the checker board using corner detection and structure recovery (Geiger et al., 2012). This spatial calibration is valid only for objects within the checker board's

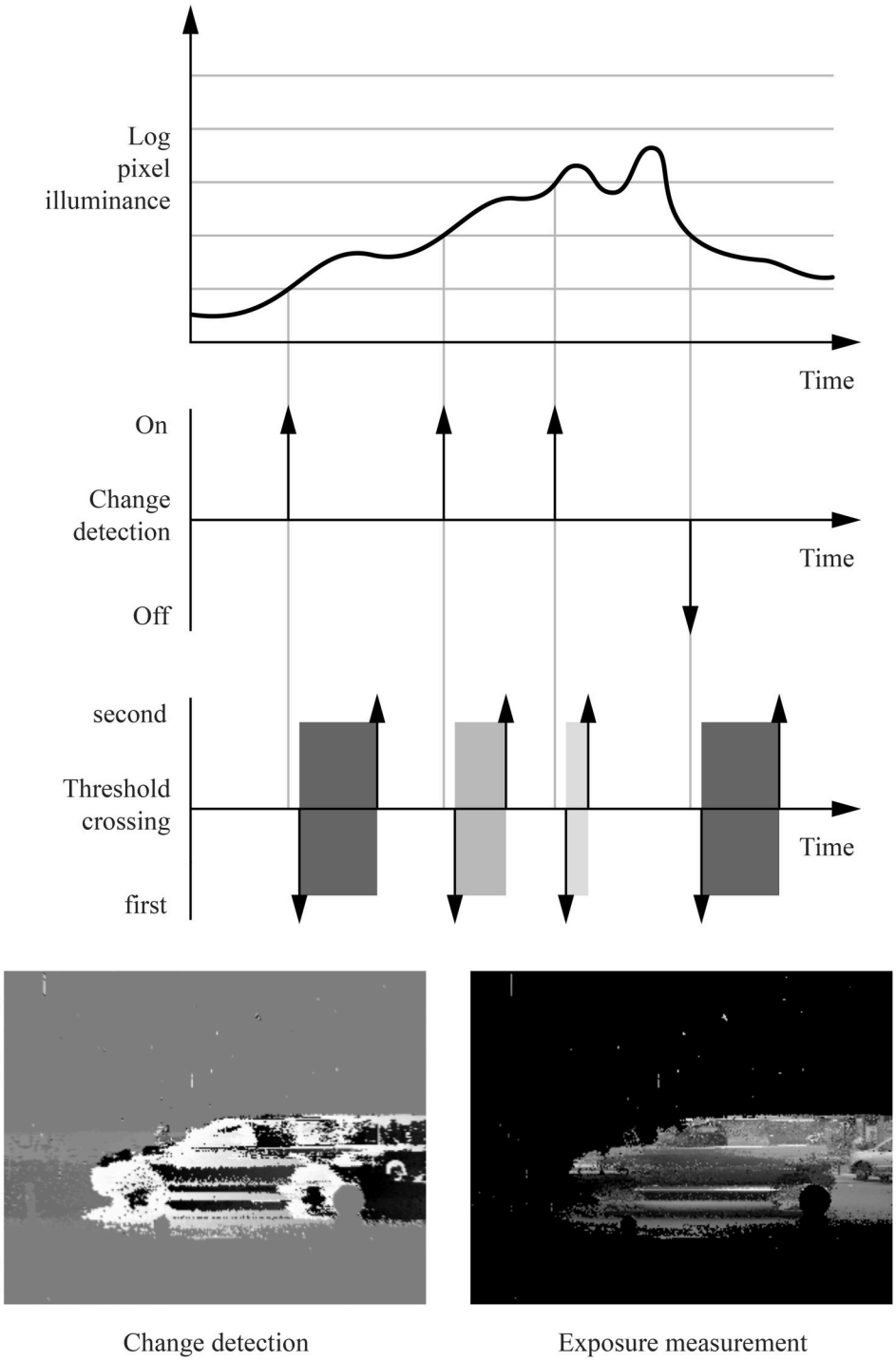


FIGURE 1 | The ATIS is an asynchronous, event-based camera with independent pixels. This figures illustrates the behavior of a single pixel. When the logarithm of the luminance captured by the pixel crosses a threshold, light integration for the pixel starts. The exposure measurement's duration is proportional to the inverse of the luminance, and is notified by two events called threshold crossings: the first one is sent when the integration begins, the second one when it ends.

plan. However, we observe a good pixel matching for objects in other plans as well, as the fields of view differences are small compared to the pixels size.

2.2. Color Events

After applying the spatial calibration step, the red sensor's pixel with index i has the same field of view as the green and blue

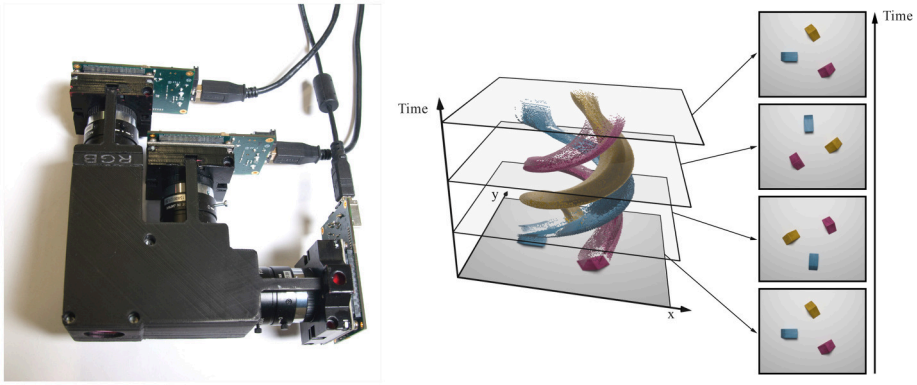


FIGURE 2 | The three-chip event-based camera is an assembly of three ATIS cameras **(Left)**. The cameras share the same field of view. Reconstructed color events can be visualized as a spatio-temporal point-cloud **(Right)**. There are as many color points in the four frames (far right) as in the whole point cloud.

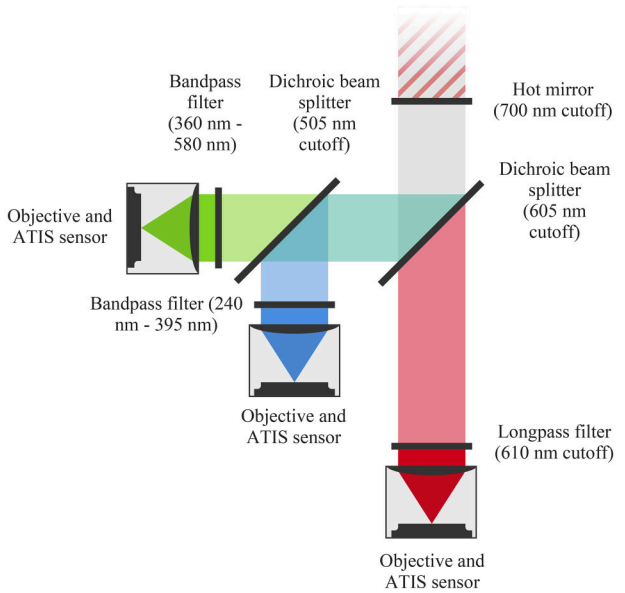


FIGURE 3 | Our three-chip event-based camera uses dichroic filters to split the light beam and ATIS cameras to record the scene as events. We use an objective for each camera instead of a single one in order to reduce the flange distance (constrained by the sensors' size).

sensors' pixels with index i. We call *pixel with index i of the color sensor* the virtual pixel combining the red, green and blue pixels with index i. The signal captured by this pixel can be modeled as a continuous $\mathbb{R}^3$ function s_i of the time t:

$$\begin{aligned} s_i &: \mathbb{R} \to \mathbb{R}^3 \\ t &\mapsto (r, g, b) \end{aligned} \tag{1}$$

where r, g, and b are the red, green and blue components intensities of the signal. i, the pixel's index, is in the range $[1, n]$, where n is the sensor's number of pixels. We want the color sensor to generate events $e_{i,t}$ defined by the tuple of attributes:

$$e_{i,t} = (i, t, r, g, b) \tag{2}$$

Assuming an initial value $s_i(t_0) = (r_0, g_0, b_0)$, the pixel with index i's first event should be generated at the time t_1 such that $s_i(t_1) = (r_1, g_1, b_1)$ and the distance in $\mathbb{R}^3$ between (r_0, g_0, b_0) and (r_1, g_1, b_1) is larger than a tunable threshold. The distance function should not be the euclidean distance in order to mimic human perception, which is highly non-linear in RGB space (Cheng et al., 2001).

The ATIS-based three-chip camera's pixels do not yield the s_i signal directly. Instead, the camera associated with each color component generates an independent stream of events. Since ATIS cameras yield exposure measurements with a delay inversely proportional to the measured exposure, it is not possible to detect temporally coinciding events to generate color events. Therefore, we associate each color sensor's virtual pixel with a memory space storing the three color components. Every time an event is generated by one of the color component cameras, the memory is updated and a color event based on the current memory value is dispatched. This mechanism is illustrated **Figure 4**.

2.3. Color Model

We consider color object tracking as a first practical application of the event-based color sensor. Given a pre-determined set of uniformly colored objects, we want to determine the position of each object in an scene at every moment. We use a two-step approach : first, we build a statistical color signature for each object using a labeled scene. Then, events from an unknown scene are matched against the statistical models and associated with the closest signature.

In order to reduce the required amount of computation for each event, we reduce the problem's dimensionality by converting color events from the RGB space to the CIEL*a*b* space. We use only the a* and b* components of the latter. For each object, we gather events from the labeled scene and project them to the

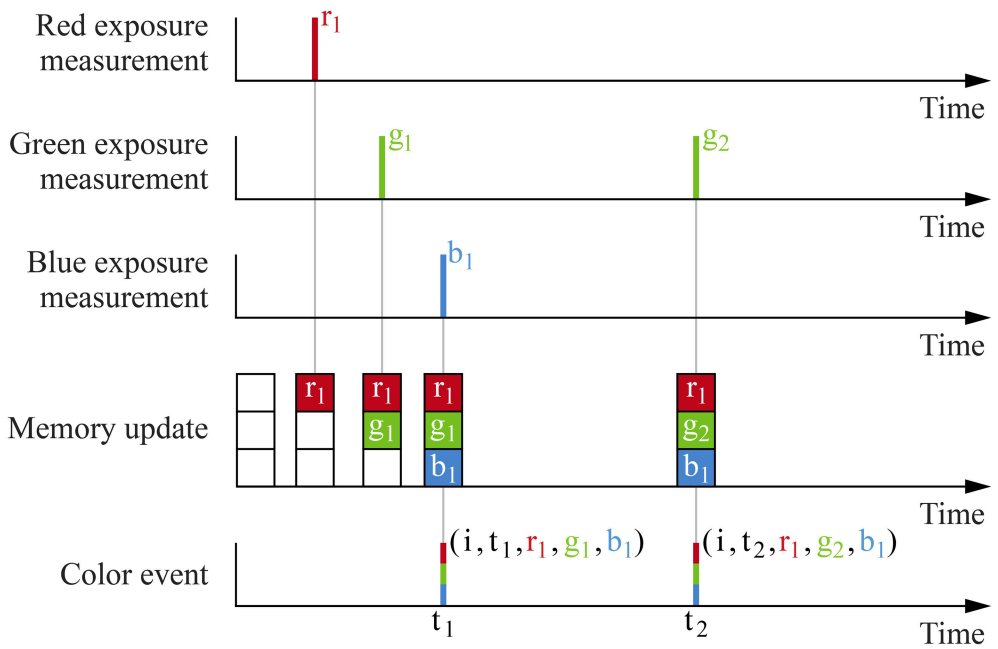


FIGURE 4 | Events generated by the color sensor's three cameras' pixel *i* are merged to generate color events. Every time an exposure measurement is generated by one of the ATIS cameras, it is stored in memory. Then, a color event is dispatched using the current memory value.

a*b* plane of the CIEL*a*b* space. We use a bivariate normal distribution as a statistical model for describing these points.

Converting events from the RGB space to the CIEL*a*b* space requires a color calibration step. ATIS cameras send exposures as a pair of threshold-crossing events to the computer. The actual exposure is—as a first approximation—proportional to the inverse of the time difference between the two thresholds-crossing events. We use a Macbeth ColorChecker to evaluate the required proportionality factor between the time difference inverse and red, green and blue components in RGB space. We observe that the expected red, green and blue values given by the Macbeth ColorChecker as functions of the measured inverse threshold-crossing time differences are well described by affine functions, as shown **Figure 5**. The need for affine functions instead of linear functions can be attributed to the sensor imperfections, including the pixels' dark current. We calculate the affine regression by minimizing the mean squared error for each color component. This method yields good results for displaying the sensor's measurements using an RGB screen. Estimated red, green and blue components can be used to determine the CIEL*a*b* color components using several non-linear relations (Jain, 1989).

However, when using this model to convert the measured colors to the CIEL*a*b* space, we observe a poor fit with the values given by the Macbeth ColorChecker. The difference can be attributed to the uncorrelated regression applied to each component and the ATIS cameras' noise. Therefore, we use the Nelder-Mead simplex algorithm (Lagarias et al., 1998) to optimize the six parameters of the three color components' affine regressions. We minimize the distances between the expected colors given by the Macbeth ColorChecker and the measured points in CIEL*a*b* space. Since the CIEL*a*b* space is perceptually uniform, this method yields the best compromise for converting the measured Macbeth ColorChecker's colors to the CIEL*a*b* space with regards to human perception. **Figure 5** shows the two methods results.

2.4. Signatures

We consider the sequence S of n color events associated with a uniformly colored object:

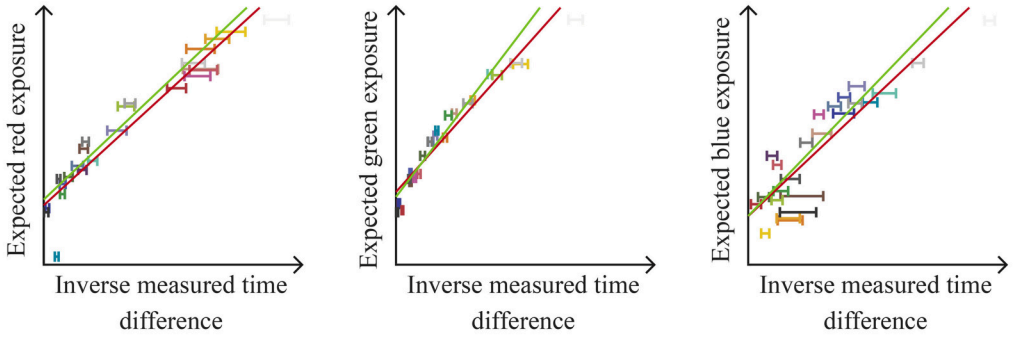
$$S = \left(\left(i_k, t_k, r_k, g_k, b_k \right), k \in [\![0, n-1]\!] \right) \tag{3}$$

We define S_{ab} as the sequence of pairs (a, b) obtained by converting each color event from the sequence S to the CIEL*a*b* space:

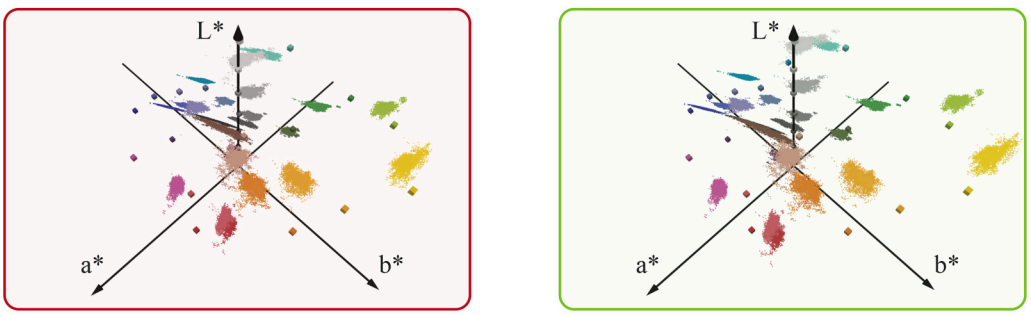
$$S_{ab} = \left(\left(a_k, b_k \right), k \in [\![0, n-1]\!] \right) \tag{4}$$

We call *signature* of the considered object the bivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ estimated from the S_{ab} sequence:

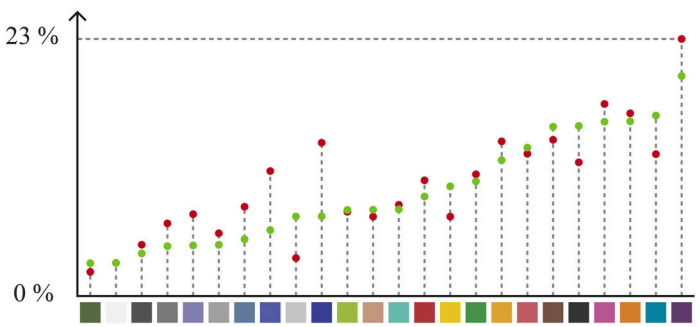
$$\begin{aligned} \boldsymbol{\mu} &= \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} = \frac{1}{n} \sum_{k=0}^{n-1} \begin{pmatrix} a_k \\ b_k \end{pmatrix} \\ \boldsymbol{\Sigma} &= \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix} \\ &= \frac{1}{n-1} \sum_{k=0}^{n-1} \left(\begin{pmatrix} a_k \\ b_k \end{pmatrix} - \boldsymbol{\mu} \right) \left(\begin{pmatrix} a_k \\ b_k \end{pmatrix} - \boldsymbol{\mu} \right)^{\mathsf{T}} \end{aligned} \tag{5}$$



Measurements of a ColorChecker's tiles are plotted against the expected RGB components. Affine functions appear as a good fit to model the transformations. We consider two methods to determine the functions' parameters : mean-squared errors minimization (red) and mean CIEL*a*b* distances minimization (green).



Points represent the sensor's measurements converted to calibrated RGB colors with each model, then converted to the CIEL*a*b* space. Cubes represent the tiles' expected colors' positions given by the ColorChecker.



The mean distance of the measured points to their expected position in the CIEL*a*b* space as a fraction of the largest distance between two colors is plotted for each tile.

FIGURE 5 | Converting the sensor's color events to the CIEL*a*b* space requires a transformation model. We use a Macbeth ColorChecker to compare the expected color values with the measured time differences. We consider two strategies for finding our model's parameters : mean squared error (red) and non-linear optimization to minimize the distance between expected and measured colors in the CIEL*a*b* space (green). The latter yields a better fit for blue and purple colors in the a*b* plane. On the a*b* plane figures, squares represent the expected colors and circles represent the mean measured value. The represented Macbeth ColorCheckers were captured by our sensor.

The experiment presented **Figure 6** illustrates the method to determine the signature of actual objects. Five colored wooden pieces placed on a white background are recorded. Even though the scene is static, the ATIS cameras' noise triggers exposure measurements which are converted to color events. We associate a pixel set—or mask—to each wooden piece. The color events generated by this pixel set make up the S sequence used to fit a signature. The color signature for the background is evaluated as well.

2.5. Tracking

After determining the five wooden pieces' signatures, we consider a scene with the same pieces moving. Let X be the continuous bivariate random variable which associates each color event with

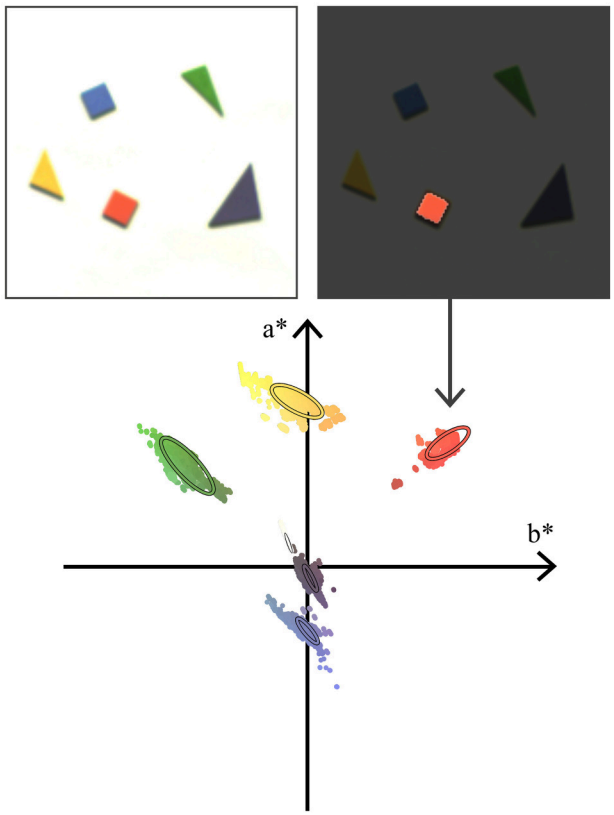


FIGURE 6 | In order to build color signatures for a set of wooden pieces, we accumulate noise-generated events from a static scene. We use the resulting image to build a mask labeling a specific piece (here, the orange one). All the color events associated with the mask's pixels are converted to the CIEL*a*b* space and projected on the a*b* plane. The projected points are used to estimate a bivariate normal distribution, which we call the signature. The bottom diagram shows 95% confidence ellipses of the wooden pieces' signatures in the a*b* plane.

its a* and b* components:

$$\begin{aligned} \mathbf{X}\colon \left(\mathbb{N}, \mathbb{R}^{+}, \mathbb{R}^{3}\right) &\rightarrow \mathbb{R}^{2} \\ e_{i,t} &\mapsto \begin{pmatrix} a \\ b \end{pmatrix} \end{aligned} \tag{6}$$

We note $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$ a color event's a* and b* components.

Writing O_j for the probabilistic event "the color event was generated by the object j," the Bayes' theorem yields the probability that the object j generated the considered color event:

$$P\left(O_j \mid \mathbf{X} = \mathbf{x}\right) = \frac{f_{O_j}\left(\mathbf{x}\right) P\left(O_j\right)}{\sum_{j=0}^{m-1} f_{O_k}\left(\mathbf{x}\right) P\left(O_k\right)} \tag{7}$$

f_{O_j} is the probability density function of the bivariate normal distribution associated with the object j, and m is the number of objects.

The color event is associated with the object with the largest probability to be the source of the event. Assuming an identical probability $P\left(O_j\right) = \frac{1}{m}$ for each object to generate an event (six in the wooden pieces example, background included), we simply need to find the index j maximizing $f_{O_j}\left(\mathbf{x}\right)$, given by:

$$f_{O_j}\left(\mathbf{x}\right) = \frac{1}{2\pi\left|\mathbf{\Sigma}_j\right|} e^{-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_j\right)^{\mathrm{T}} \mathbf{\Sigma}_j^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_j\right)} \tag{8}$$

where $\boldsymbol{\mu}_j$ and $\mathbf{\Sigma}_j$ are the object j' signature' mean and covariance matrix.

In order to track the objects, we use a moving mean algorithm. Each object is given a center $\mathbf{p_j} = \begin{pmatrix} x_j \\ y_j \end{pmatrix}$, where x_j and y_j are the object's mean coordinates in the screen referential. When an event $e_{i,t}$ is generated, the mean associated with the object minimizing expression 8 is updated. The new mean $\mathbf{p_j'}$ is calculated from the previous mean and the event as:

$$\mathbf{p_j'} = \lambda \mathbf{p_j} + (1 - \lambda)\, \mathbf{x_i} \tag{9}$$

where $\mathbf{x_i}$ is the coordinates in the screen referential of the pixel which generated the event. λ is an inertia parameter ranging from zero to one. λ is generally given a value close to $\left(1 - 10^{-3}\right)$. The larger λ, the more robust to noise the tracking. However, large λ values yield more latency and deteriorate the algorithm's ability to account for small variations.

We take into account the camera noise with a spatio-temporal activity filter. Once an event is associated with an object, we count the number of prior events associated with the same object that were generated less than one second before in a six-by-six square window around the event's position. Only events with at least thirty neighbors in this spatio-temporal window are taken into account for updating the object's mean position. Increasing the required count decreases the number of false positive events, while increasing the number of false negative events.

3. RESULTS

We applied the color tracking algorithm to three experiments. Videos illustrating the associated results are provided as Supplementary Materials. The first experiment is recorded under laboratory-controlled conditions. Five colored wooden pieces are placed on a rotating surface captured with a static sensor. **Figure 7** shows the results compared to the ground truth, evaluated with a contour tracing algorithm (Suzuki and be, 1985). The objects and their centers are identified on images reconstructed from the color camera's events. For each event, we calculate the distance between the associated object's estimated mean and its ground truth. The mean distance is given for each object as a fraction of the yellow object's trajectory's radius. We are able to estimate the objects trajectories using only color data. The moving mean algorithm's λ parameter is identical for all the objects, empirically set to $\left(1 - 10^{-3}\right)$. A compromise must be reached between noise robustness and accuracy. Reducing the

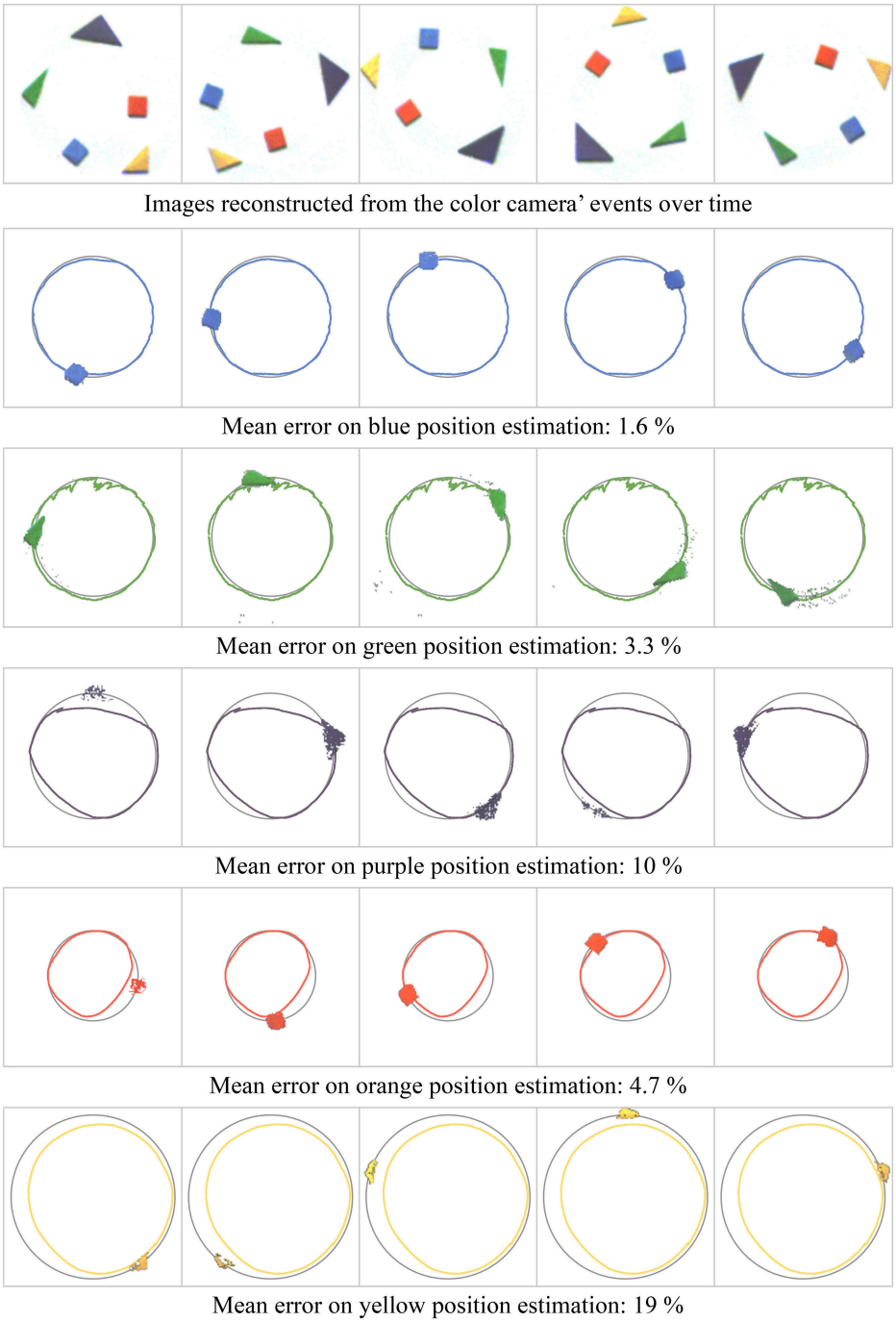


FIGURE 7 | Tracking of five wooden pieces in rotation motion. Figures show the pieces motions estimated with our method (colored trajectories) and the ground truth (gray trajectories) for a whole rotation. The mean error for each object is the average distance between the estimated object's mean position and the ground truth as a ratio of the yellow object's trajectory's radius. The observed errors derive from the compromise between noise robustness and accuracy imposed by the moving mean algorithm.

parameter's value would improve results for the purple object while degrading the results for the green one.

The second experiment consists of a moving camera in an urban scene containing a red road sign and a green pharmacy neon light. The corresponding color signatures are learnt from an initialization step. The latter uses data from a one-second sequence taking place before the experiment's recording. **Figure 8** illustrates the associated color reconstruction process. Due to

FIGURE 8 | The figure's top row shows reconstructed frames from events acquired by each sensor of the three-chip event-based camera. The gray levels are tone-mapped with a logarithmic function in order to be displayed on a regular screen. The bottom-left frames are reconstructed with the linear color model presented in the methodology. The colors are properly reconstructed, but very little detail is left in dark areas. The bottom-right frames use the channels' logarithmic tone-mappings as their color channel, yielding incorrect colors but much more details in dark areas.

the scene's high dynamic range, the linear color model yields little detail in the dark areas of reconstructed frames. Therefore, we generate color frames for display purposes by applying a logarithmic tone-mapping on each channel independently. This operation yields incorrect colors, but shows much more detail in dark areas. The tracking algorithm is not influenced by this operation since it uses colors calculated by the linear model, as presented in our methodology.

The third experiment takes place in an urban scene as well. It consists of two pedestrian wearing colored sweaters walking in front of the sensor. **Figure 9** shows the segmented events for both urban experiments, while **Figure 10** compares the estimated trajectories with the ground truth. The latter is computed with the contour tracing algorithm provided by Suzuki and be (1985) as well. We remind the reader that this technique exploits shape rather than color: spatial constrains yield more robust results, but require a more complex algorithm. The estimated mean position lags behind the ground truth, which is a consequence of the moving mean algorithm. In order to assess the dynamics of our results, we compensate the lag for the road signs experiment and shift the position's reference for the pedestrians one. **Table 1** summarizes the mean errors and standard deviations along the x and y axis for both urban scenes. We observe degraded performances for objects near the sensor's edges, which can be attributed to sensor limits. On the one hand, the ATIS camera used in the assembly lacks sensitivity to blue wavelengths, which is reflected by longer integration times for this component. This leads to timestamp differences between channels, which result in incorrect color reconstructions. On the other hand, our prototype three-chip color camera exhibits optical aberrations which degrade the signal near the edges.

4. DISCUSSION

The event-based three-chip color camera is the first working prototype of an event-based sensor able to acquire absolute color information: the sensor generates packets of data carrying the luminance value integrated over a small time interval. By contrast, the event-based color pixel designed 6 years ago (Berner and Delbruck, 2011) and the DVS camera with a Bayer matrix built in early 2017 (Moeys et al., 2017) can only detect color variations: they send the same message regardless the variation magnitude, and require heavy calculations to retrieve the absolute luminance. Even though our prototype is still at an early stage, we manage to track colored objects in several scenes, using only the color information generated by the sensor. Thanks to the capture of absolute luminance, very little computational power is required to label the events. Therefore, the algorithm is a good candidate for the first stage of a complex chain of computations achieving a higher level task. It proves that color information alone is enough to achieve tracking with event-based cameras. The advantages of the event-based color sensor

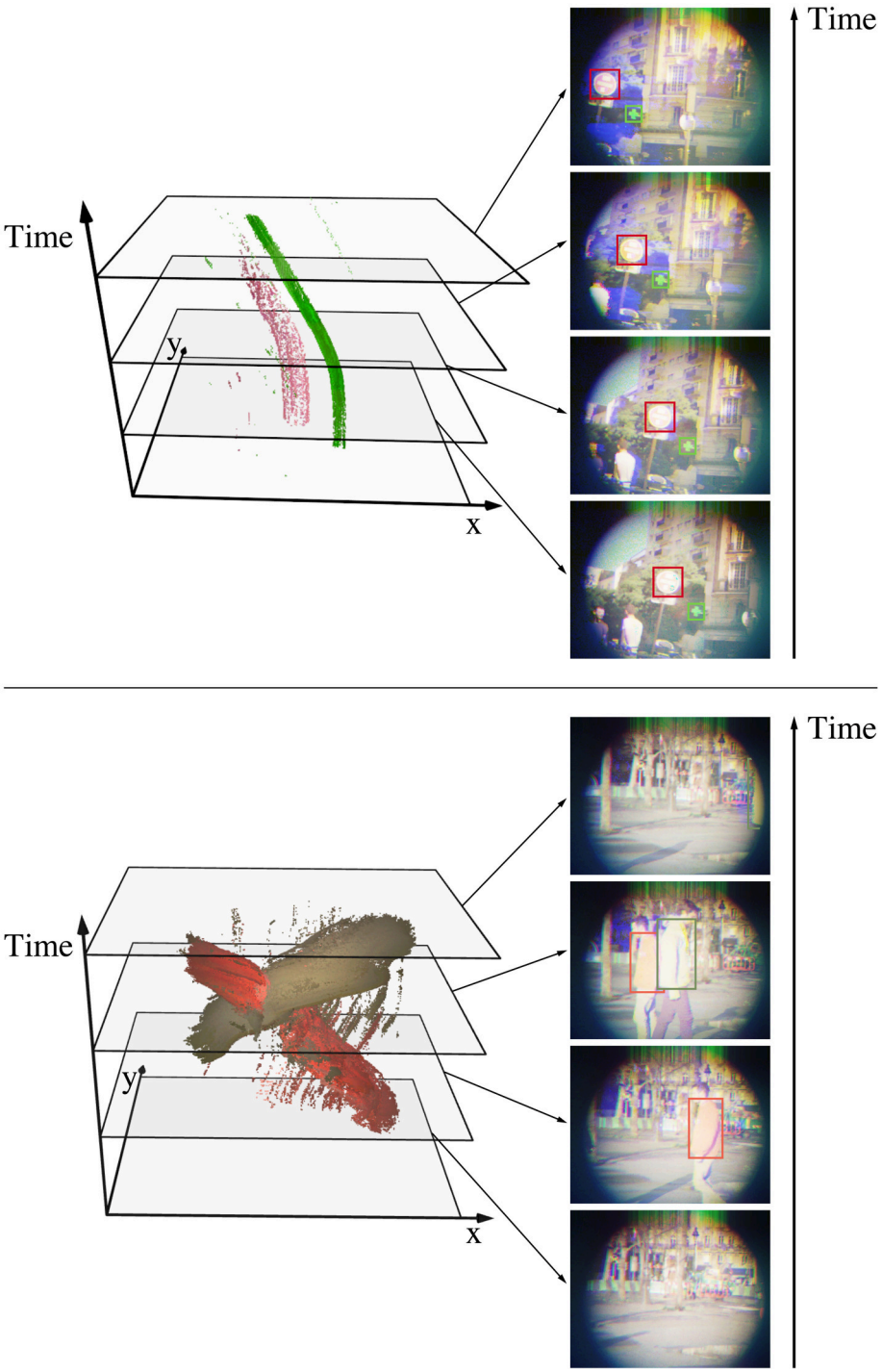


FIGURE 9 | We consider two outdoors scenes to assess the tracking algorithm performance: a moving camera acquiring a red road sign and a green pharmacy sign **(Top)**, and a static camera recording pedestrians wearing colored sweaters **(Bottom)**. The color signatures for the objects are calculated from similar scenes. The point clouds show color events where f_O is larger than 10^{-5} for one of the objects. These events are used to update the estimated center of the associated object with a moving mean algorithm where $\lambda = 1 - 10^{-3}$. The tracked object are framed on the reconstructed frames for better visualization.

presented in this work over frame-based color cameras are similar to the advantages of gray event-based sensors over gray frame-based sensors: carrying out part of the computation on the sensor yields a natural data compression, with an increased temporal resolution. Both greatly reduce the required amount of computation for the processor. However, a proper use of the

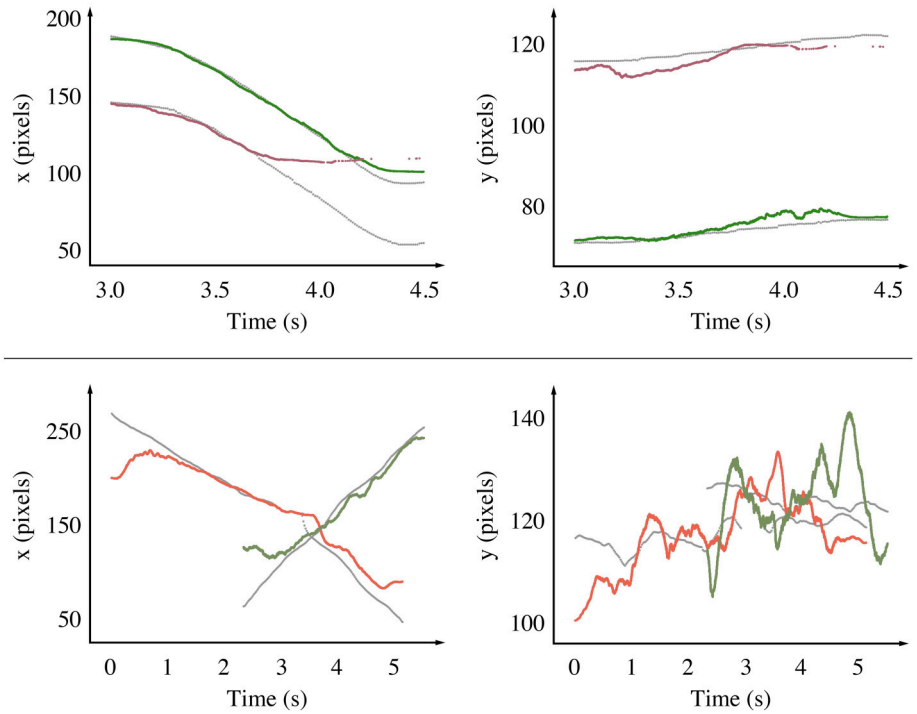


FIGURE 10 | We compare the estimated position of the tracked objects with the ground truth along the x and y axis as functions of time for the road signs experiment **(Top)** and the pedestrians experiment **(Bottom)**. The events' timestamps are corrected to account for the delay induced by the mean-shift algorithm, effectively comparing dynamics rather than absolute values. Tracking is degraded near the edges, especially for the red stop sign (**Top**, x axis, after 3.6 s), which is a consequence of our prototype's optical aberrations. Since the pedestrians move along the x axis, motion along the y axis is relatively small, making the noise appear stronger.

TABLE 1 | The mean error between the estimated position and the ground truth is evaluated for each tracked object in the outdoor scenes.

Object	Mean error (pixels)	
	x	y
Green sign	2.18	1.30
Red sign	18.4	2.09
Orange sweater	14.5	4.58
Brown sweater	11.9	6.20

The larger errors along the x axis for three out of four object can be attributed to the optical aberration near the sensors' edges.

generated data requires a re-design of most computer vision algorithms.

The presented prototype can find applications in embedded systems. When low latency and low power consumption are required—as an example, with drones—conventional vision sensors show limits which can be overcome with event-based cameras. Fast color segmentation on a drone can be useful for several tasks, such as target detection and tracking or environment mapping. Moreover, the high dynamic range of the ATIS camera tackles the luminance adaptation issue, particularly troublesome for self-driving cars. Color makes road sign segmentation and recognition much easier on such systems.

The use of spatial information is out of the scope of this work. However, it should allow for a more robust algorithm thanks to data fusion, and is considered as this work continuation. We also identify several areas of improvement for the sensor that would benefit the algorithm's results. These improvements require hardware development. On the one hand, one of the event-based three-chip color sensor's weaknesses is its lack of sensitivity to short wavelengths. This limitation is shared by most silicon-based photo-detectors, but is aggravated by the ATIS's low sensitivity to low light. Therefore, better results would be achieved with increased sensitivity. On the other hand, the three-chip event-based prototype exhibits optical limits, such as color aberrations near the edges and vignetting. These shortcomings are caused by the large size of the electronics boards used by the ATIS sensors to communicate with the computer: the sensors are 18×18 mm, while the board are 50×50 mm. They impose large distances between the light entry and the sensors, which in turn require the objectives to be placed after the beam splitters. The small defects in the beam splitters' angles and positioning are responsible for the visible aberrations. The circular field of view is a consequence of a compromise reach with the off-the-shelf components used in the assembly: the input hot mirror results in a partially masked field of view, however a larger one would increase the amount of reflections inside the sensor's casing, which degrades capture. Designing a dedicated electronics board

matching the sensor size would allow a casing reduction large enough to place as single objective before the light entry, hence alleviating the optical aberrations. Another solution consists in using a single array of pixels with a Bayer matrix. However, the latter requires designing a chip from the ground up, as Bayer matrix placing is part of the pixel building process.

Assuming the design of a new chip, it would be interesting to consider the following problem. Both the sensor presented in this work and the existing event-based color sensors digitize the analog light signal into events for each channel independently. Processing is then performed on the generated events, including color merging. The parallel drawn with the human eye for such sensors (Posch et al., 2011) ignores part of the eye complexity, including data passed between pixels through the horizontal cells. This data appears to be analog rather than digital. Implementing such a data transfer in the next generation of color neuromorphic vision sensors may be the key to acquiring color information efficiently. It may also help overcoming the following paradox in computer vision: for segmenting natural scenes, color, though helpful, provides a generally small advantage (Hansen and Gegenfurtner, 2017). However, it requires dealing with three times as much data. Assuming an access to the extra power required to deal with this data, one is generally better off with a more complex algorithm working on gray levels. Merging colors on the analog level may help reducing the amount of generated data without tainting its quality.

AUTHOR CONTRIBUTIONS

AM, S-HI, and RB contributed conception and design of the study. AM devised the theoretical model, with substantial contributions from CS-C. AM and CS-C carried out the experiments. AM analyzed the results, and wrote the first draft of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work received the support from the French Medical Research Foundation via the program of Bioinformatics analysis in biology research [DBI20141231328]. This work also received the support from LABEX LIFESENSES [ANR-10-LABX-65], managed by the French state funds (ANR) within the Investissements d'Avenir program [ANR-11-IDEX-0004-02].

ACKNOWLEDGMENTS

The authors would like to thank Guillaume Chenegros for helping in the design of the prototype's optical system, and Vincent Fraillon-Maison and Félix Rutard for participating in the urban experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2018.00135/full#supplementary-material

Supplementary Video 1 | Simple shapes.

Supplementary Video 2 | Road signs.

Supplementary Video 3 | Pedestrians.

REFERENCES

Bahlmann, C., Zhu, Y., Ramesh, V., Pellkofer, M., and Koehler, T. (2005). "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," in *IEEE Proceedings Intelligent Vehicles Symposium, 2005* (Las Vegas, NV), 255–260. doi: 10.1109/IVS.2005.1505111

Bai, X., Wang, J., and Sapiro, G. (2010). "Dynamic color flow: a motion-adaptive color model for object segmentation in video," in *Computer Vision - ECCV 2010 Heraklion, Crete, Greece*, eds K. Daniilidis, P. Maragos, and N. Paragios (Berlin; Heidelberg: Springer), 617–630.

Benosman, R., Clercq, C., Lagorce, X., Ieng, S.-H., and Bartolozzi, C. (2014). Event-based visual flow. *IEEE Trans. Neural Netw. Learn. Syst.* 25, 407–417. doi: 10.1109/TNNLS.2013.2273537

Berner, R., and Delbruck, T. (2011). Event-based pixel sensitive to changes of color and brightness. *IEEE Trans. Circ. Syst. I Reg. Papers* 58, 1581–1590. doi: 10.1109/TCSI.2011.2157770

Carneiro, J., Ieng, S.-H., Posch, C., and Benosman, R. (2013). Event-based 3d reconstruction from neuromorphic retinas. *Neural Netw.* 45, 27–38. doi: 10.1016/j.neunet.2013.03.006

Cheng, H., Jiang, X., Sun, Y., and Wang, J. (2001). Color image segmentation: advances and prospects. *Patt. Recogn.* 34, 2259–2281. doi: 10.1016/S0031-3203(00)00149-7

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Trans. Patt. Anal. Mach. Intell.* 17, 790–799. doi: 10.1109/34.400568

Comaniciu, D. (2003). An algorithm for data-driven bandwidth selection. *IEEE Trans. Patt. Anal. Mach. Intell.* 25, 281–288. doi: 10.1109/TPAMI.2003.1177159

Delbruck, T. (1993). Silicon retina with correlation-based, velocity-tuned pixels. *IEEE Trans. Neural Netw.* 4, 529–541. doi: 10.1109/72.217194

Dominy, N. J., and Lucas, P. W. (2001). Ecological importance of trichromatic vision to primates. *Nature* 410, 363–366. doi: 10.1038/35066567

Drazen, D., Lichtsteiner, P., Hafliger, P., Delbruck, T., and Jensen, A. (2011). Toward real-time particle tracking using an event-based dynamic vision sensor. *Exp. Fluids* 51, 1465–1469. doi: 10.1007/s00348-011-1207-y

Etienne-Cummings, R., Van der Spiegel, J., and Mueller, P. (1997). A focal plane visual motion measurement sensor. *IEEE Trans. Circ. Syst. I Fundament. Theory Appl.* 44, 55–66. doi: 10.1109/81.558442

Fukunaga, K., and Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inform. Theory* 21, 32–40. doi: 10.1109/TIT.1975.1055330

Geiger, A., Moosmann, F., Car, O., and Schuster, B. (2012). "Automatic camera and range sensor calibration using a single shot," in *2012 IEEE International Conference on Robotics and Automation* (St Paul, MN: IEEE). doi: 10.1109/icra.2012.6224570

Grundmann, M., Kwatra, V., Han, M., and Essa, I. (2010). "Efficient hierarchical graph-based video segmentation," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (San Francisco, CA), 2141–2148. doi: 10.1109/CVPR.2010.5539893

Guo, H., Guo, P., and Lu, H. (2006). "A fast mean shift procedure with new iteration strategy and re-sampling," in *2006 IEEE International Conference on Systems, Man and Cybernetics* (Taipei: IEEE). doi: 10.1109/icsmc.2006.385220

Hansen, T., and Gegenfurtner, K. R. (2017). Color contributes to object-contour perception in natural scenes. *J. Vis.* 17:14. doi: 10.1167/17.3.14

Hartley, R. (1995). "In defense of the eight-point algorithm," in *Proceedings of IEEE International Conference on Computer Vision, 1995* (Cambridge, MA). doi: 10.1109/ICCV.1995.466816

Hsieh, J.-W., Chen, L.-C., Chen, S.-Y., Chen, D.-Y., Alghyaline, S., and Chiang, H.-F. (2015). Vehicle color classification under different lighting conditions through color correction. *IEEE Sen. J.* 15, 971–983. doi: 10.1109/JSEN.2014.2358079

Jain, A. K. (1989). *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Johansson, L. (2004). *Human Colour Vision*. Master's thesis, Department of Science and Technology, Linkoping University.

Kakumanu, P., Makrogiannis, S., and Bourbakis, N. (2007). A survey of skin-color modeling and detection methods. *Patt. Recogn.* 40, 1106–1122. doi: 10.1016/j.patcog.2006.06.010

Krammer, J., and Koch, C. (1997). Pulse-based analog vlsi velocity sensors. *IEEE Trans. Circ. Syst. II Analog Digit. Signal Process.* 44, 86–101. doi: 10.1109/82.554431

Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM J. Optim.* 9, 112–147. doi: 10.1137/S1052623496303470

Lagorce, X., Meyer, C., Ieng, S.-H., Filliat, D., and Benosman, R. (2015). Asynchronous event-based multikernel algorithm for high-speed visual features tracking. *IEEE Trans. Neural Netw. Learn. Syst.* 26, 1710–1720. doi: 10.1109/TNNLS.2014.2352401

Lee, Y. J., Kim, J., and Grauman, K. (2011). "Key-segments for video object segmentation," in *2011 International Conference on Computer Vision* (Barcelona: IEEE), 1995–2002. doi: 10.1109/iccv.2011.6126471

Lezama, J., Alahari, K., Sivic, J., and Laptev, I. (2011). "Track to the future: spatio-temporal video segmentation with long-range motion cues," in *CVPR 2011* (Colorado Springs, CO: IEEE), doi: 10.1109/cvpr.2011.6044588

Lichtsteiner, P., Posch, C., and Delbruck, T. (2006). "A 128 x 128 120db 30mw asynchronous vision sensor that responds to relative intensity change," in *2006 IEEE International Solid State Circuits Conference - Digest of Technical Papers* (San Francisco, CA: IEEE). doi: 10.1109/isscc.2006.1696265

Moeys, D. P., Li, C., Martel, J. N., Bamford, S., Longinotti, L., Motsnyi, V., et al. (2017). "Color temporal contrast sensitivity in dynamic vision sensors," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)* (Baltimore, MD: IEEE). doi: 10.1109/iscas.2017.8050412

Ni, Z., Bolopion, A., Agnus, J., Benosman, R., and Regnier, S. (2012). Asynchronous event-based visual shape tracking for stable haptic feedback in microrobotics. *IEEE Trans. Robot.* 28, 1081–1089. doi: 10.1109/TRO.2012.2198930

Paris, S., and Durand, F. (2007). "A topological approach to hierarchical segmentation using mean shift," in *2007 IEEE Conference on Computer Vision and Pattern Recognition* (Minneapolis, MN: IEEE). doi: 10.1109/cvpr.2007.383228

Posch, C., Matolin, D., and Wohlgenannt, R. (2008). "An asynchronous time-based image sensor," in *2008 IEEE International Symposium on Circuits and Systems* (Seattle, WA: IEEE). doi: 10.1109/iscas.2008.4541871

Posch, C., Matolin, D., and Wohlgenannt, R. (2011). A qvga 143 db dynamic range frame-free pwm image sensor with lossless pixel-level video compression and time-domain CDS. *IEEE J. Solid-State Circ.* 46, 259–275. doi: 10.1109/JSSC.2010.2085952

Pun, C.-M., and Huang, G. (2016). On-line video object segmentation using illumination-invariant color-texture feature extraction and marker prediction. *J. Vis. Commun. Image Represent.* 41, 391–405. doi: 10.1016/j.jvcir.2016.10.017

Reverter Valeiras, D., Lagorce, X., Clady, X., Bartolozzi, C., Ieng, S.-H., and Benosman, R. (2015). An asynchronous neuromorphic event-driven visual part-based shape tracking. *IEEE Trans. Neural Netw. Learn. Syst.* 26, 3045–3059. doi: 10.1109/TNNLS.2015.2401834

Rogister, P., Benosman, R., Ieng, S.-H., Lichtsteiner, P., and Delbruck, T. (2012). Asynchronous event-based binocular stereo matching. *IEEE Trans. Neural Netw. Learn. Syst.* 23, 347–353. doi: 10.1109/TNNLS.2011.2180025

Rother, C., Kolmogorov, V., and Blake, A. (2004). ""GrabCut": interactive foreground extraction using iterated graph cuts," in *ACM SIGGRAPH 2004 Papers. Los Angeles, CA, USA*, *SIGGRAPH'04* (New York, NY: ACM), 309–314. doi: 10.1145/1186562.1015720

Suzuki, S., and be, K. (1985). Topological structural analysis of digitized binary images by border following. *Comput. Vis. Graph. Image Process.* 30, 32–46. doi: 10.1016/0734-189X(85)90016-7

Trémeau, A., Tominaga, S., and Plataniotis, K. N. (2008). Color in image and video processing: most recent trends and future research directions. *EURASIP J. Image Video Process.* 2008, 1–26. doi: 10.1155/2008/581371

van de Weijer, J., Gevers, T., and Bagdanov, A. (2006). Boosting color saliency in image feature detection. *IEEE Trans. Patt. Anal. Mach. Intell.* 28, 150–156. doi: 10.1109/TPAMI.2006.3

Vantaram, S. R., and Saber, E. (2012). Survey of contemporary trends in color image segmentation. *J. Electr. Imaging* 21:040901. doi: 10.1117/1.JEI.21.4.040901

Xiao, C., and Liu, M. (2010). Efficient mean-shift clustering using gaussian kd-tree. *Comput. Graph. Forum* 29, 2065–2073. doi: 10.1111/j.1467-8659.2010.01793.x

Yang, C., Duraiswami, R., DeMenthon, D., and Davis, L. (2003). "Mean-shift analysis using quasinewton methods," in *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), Vols. 2, 3* (Barcelona: IEEE), II-447-II-450. doi: 10.1109/icip.2003.1246713

Conflict of Interest Statement: 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.

Copyright © 2018 Marcireau, Ieng, Simon-Chane and Benosman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.