3D shape reconstruction with a multiple-constraint estimation approach

Xia Chen[1,2,3], Zhan-Li Sun[4,5]* and Ying Zhang[4]

[1]School of Information and Computer, Anhui Agricultural University, Hefei, China, [2]Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Institute of Physical Science and Information Technology, Anhui University, Hefei, China, [3]Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, Anhui University, Hefei, China, [4]School of Electrical Engineering and Automation, Anhui University, Hefei, China, [5]Information Materials and Intelligent Sensing Laboratory of Anhui Province, Anhui University, Hefei, China

In this study, a multiple-constraint estimation algorithm is presented to estimate the 3D shape of a 2D image sequence. Given the training data, a sparse representation model with an elastic net, i.e., l_1−norm and l_2−norm constraints, is devised to extract the shape bases. In the sparse model, the l_1−norm and l_2−norm constraints are enforced to regulate the sparsity and scale of coefficients, respectively. After obtaining the shape bases, a penalized least-square model is formulated to estimate 3D shape and motion, by considering the orthogonal constraint of the transformation matrix, and the similarity constraint between the 2D observations and the shape bases. Moreover, an Augmented Lagrange Multipliers (ALM) iterative algorithm is adopted to solve the optimization of the proposed approach. Experimental results on the well-known CMU image sequences demonstrate the effectiveness and feasibility of the proposed model.

1. Introduction

As an important component of computer vision, 3D shape reconstruction has been widely used in many applications (Li et al., 2016, 2018; Adamkiewicz et al., 2022; Chiang et al., 2022; Fombona-Pascual et al., 2022; Jang et al., 2022; Lu et al., 2022; Nian et al., 2022a,b; Wang et al., 2022; Wen et al., 2022). Among the various 3D shape reconstruction methods, non-rigid structure from motion (NRSFM) offers a technique to simultaneously recover the 3D structures and motions of an object, by using the 2D landmarks in a series of images (Graßhof and Brandt, 2022; Kumar and Van Gool, 2022; Song et al., 2022). Nevertheless, NRSFM is still an underconstrained and challenging issue because of lacking any prior knowledge of 3D structure deformation.

To alleviate the uncertainty, the various constraints are exploited constantly. Bregler et al. (2000), proposed a low-rank constraint-based approach to decompose the observation matrix into a motion factor and a shape basis. In order to reduce the number of the unknown variables proposed by Bregler et al. (2000), a point trajectory approach was presented by Akhter et al. (2010) by using the predefined bases of discrete cosine transform (DCT). However, the high-frequency deformation cannot be reconstructed well via this trajectory representation because of the low-rank constraint. Gotardo and Martinez (2011) modeled a smoothly deforming 3D shape as a single point moving along a smooth time trajectory within a linear shape space. In addition to the low-rank constraint, the higher frequency DCT was adopted to capture the high-frequency deformation.

For the low-rank constraint methods, it is difficult to determine the optimal number of shape bases or trajectory bases. To solve this problem, a Procrustean normal distribution (PND) model was presented by Lee et al. (2013) to separate the motion and deformation components strictly, without any additional constraints or prior knowledge. The experimental results demonstrate the performance of PND. Subsequently, the Procrustean Markov Process (PMP) algorithm was proposed by Lee et al. (2014), by combing in a first-order Markov model representing the smoothness between two adjacent frames with PND. Lee et al. (2016) reported a consensus of non-rigid reconstruction (CNR) approach to estimate 3D shapes based on local patches. However, the reconstruction performance of these methods may degrade significantly when the number of images becomes small, especially for a single image.

Referring to the active shape model (Cootes et al., 1995), a limb length constraint-based approach was presented by Wang et al. (2014) to estimate the 3D shape of an object from a single 2D image, by solving a l_1-norm minimization problem. Zhou et al. (2013) proposed a sparse representation-based convex relaxation approach (CRA) to guarantee global optimality. The shape bases were extracted from a given training data by using a sparse representation model. The corresponding coefficients were obtained by adopting a convex relaxation assumption. A prominent advantage of CRA is that the algorithm can deal with a single image.

To further enhance the performance of the CRA algorithm, a multiple-constraint-based estimation approach is proposed to estimate the 3D shape of a 2D image sequence. Inspired by Zhang and Xing (2017), a dictionary learning model with l_1-norm and l_2-norm, i.e., elastic net, is constructed to extract more effective shape bases from a given training set. Referring to (Cheng et al., 2015), a penalized least-square model is constructed to estimate 3D shape and motion, by considering the orthogonal constraint of the transformation matrix and the similarity constraint between the

One frame of those eight categories.

2D observations and the shape bases. In addition, an augmented Lagrange multipliers (ALM) iterative algorithm is developed to optimize the reconstruction model. The effectiveness and feasibility of the proposed algorithm are verified on the well-known CMU image sequences.

The rest of this article is organized as follows. A detailed description of the designed MCM-RR approach is introduced in Section 2. In Section 3, we report the experimental results. Finally, the article is concluded in Section 4.

2. Methods

According to the shape-space model by Zhou et al. (2013), the unknown 3D shape $\mathbf{S} \in \mathbb{R}^{3 \times p}$ is constructed as a linear combination of a few shape bases $\mathbf{B}_i \in \mathbb{R}^{3 \times p}$, i.e.,

$$\mathbf{S} = \sum_{i=1}^{K} c_i \mathbf{R}_i \mathbf{B}_i, \qquad (1)$$

where p and K are the numbers of feature points and shape bases, respectively. The parameter c_i and $\mathbf{R}_i \in \mathbb{R}^{3 \times 3}$ denote the coefficient and rotation matrix, respectively. In terms of the weak-perspective projection model, the corresponding 2D observations are modeled as a matrix $\mathbf{W} \in \mathbb{R}^{2 \times p}$,

$$\mathbf{W} = \sum_{i=1}^{K} \mathbf{M}_i \mathbf{B}_i. \qquad (2)$$

The matrix $\mathbf{M}_i \in \mathbb{R}^{2 \times 3}$ can be represented as

$$\mathbf{M}_i = c_i \tilde{R}_i, \qquad (3)$$

where $\tilde{R}_i \in \mathbb{R}^{2 \times 3}$ is the first two rows of $\mathbf{R}_i$. Combining the orthogonal constraint, the matrix $\mathbf{M}_i$ satisfies

$$\mathbf{M}_i \mathbf{M}_i^T = c_i^2 \mathbf{I}_2, \qquad (4)$$

where $\mathbf{I}_2 \in \mathbb{R}^{2 \times 2}$ is an identity matrix. The 3D shape, i.e., $z-$coordinates, and the motion parameters c_i and $\mathbf{R}_i$, are estimated by utilizing the observations $\mathbf{W}$, i.e., the (x, y) coordinates of feature points.

In the proposed method, the shape bases $\mathbf{B} \in \mathbb{R}^{3K \times p}$ are extracted via a sparse model with the elastic net constraint. The $\mathbf{B}$ is the stacking of $\mathbf{B}_i(i = 1, ..., K)$. The matrix $\mathbf{M}$ are solved by a penalized least-square model. Given $\mathbf{M}$, the parameters c_i and $\mathbf{R}_i$ are derived via refinement decompose (Zhou et al., 2013). After obtaining c_i, $\mathbf{R}_i$ and $\mathbf{B}_i$, the unknown 3D shape can be computed via (1). The pseudocode of the proposed algorithm is summarized in 1. The pseudocode of the proposed algorithm is summarized in Algorithm 1.

2.1. Extraction of shape bases via a sparse model with elastic net constraint

For a given 3D training set $\mathbf{A} \in \mathbb{R}^{3p \times F}$, i.e., the (x, y, z) coordinates of feature points of training images, the shape bases
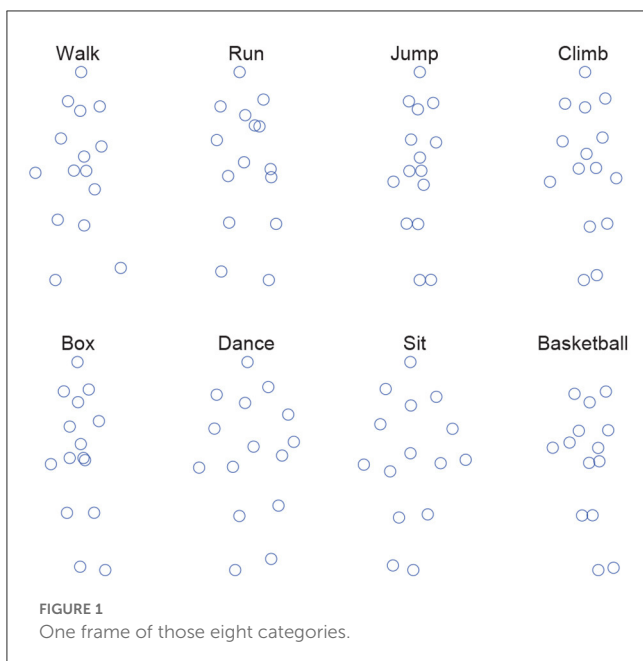
TABLE 1 Mean and standard deviation ($\mu \pm \sigma$) of the 3D reconstruction errors ξ of eight motion categories for five methods.

Sequence	PMP	CNR	PND2	CRA	MCM-RR
Walk	97.06 ± 17.35	78.28 ± 15.70	104.20 ± 26.13	38.98 ± 19.64	35.37 ± 18.49
Run	119.37 ± 31.37	65.92 ± 23.69	124.54 ± 28.82	55.69 ± 18.13	52.64 ± 17.05
Jump	102.22 ± 30.74	61.66 ± 40.35	84.64 ± 41.80	57.08 ± 41.56	44.56 ± 27.30
Climb	119.08 ± 39.39	69.36 ± 30.21	87.72 ± 56.04	58.87 ± 24.73	50.25 ± 25.88
Box	252.61 ± 41.28	82.83 ± 33.65	146.91 ± 45.17	72.90 ± 30.64	65.28 ± 26.82
Dance	118.24 ± 35.34	105.73 ± 38.81	118.52 ± 62.07	102.36 ± 44.93	83.59 ± 34.88
Sit	96.31 ± 32.77	69.58 ± 42.18	73.20 ± 32.47	75.68 ± 36.29	62.72 ± 26.79
Basketball	121.26 ± 44.83	67.63 ± 38.97	105.38 ± 72.17	63.66 ± 27.92	57.57 ± 22.96

TABLE 2 Corresponding 3D reconstruction error decreasing percentage $\xi_p(\%)$ of MCM-RR compared to CRA for eight motion categories.

Sequence	ξ_p
Walk	9.26
Run	5.48
Jump	21.93
Climb	14.64
Box	10.43
Dance	18.34
Sit	17.12
Basketball	9.57

$\mathbf{N} \in \mathbb{R}^{3p \times K}$ and the coefficient matrix $\mathbf{X} \in \mathbb{R}^{K \times F}$ can be obtained from the following sparse model:

$$\min_{\mathbf{N}_1, \cdots, \mathbf{N}_K} \frac{1}{2} \|\mathbf{A} - \mathbf{NX}\|_F^2 + \lambda \left(\tau \|\mathbf{X}\|_1 + (1 - \tau) \|\mathbf{X}\|_2^2 \right) \quad (5)$$
$$\text{s.t.} \|\mathbf{N}_i\|_F \leq 1, \qquad X_{ij} \geq 0, \forall i \in [1, K], j \in [1, F],$$

where F and τ are the number of frames and a weight coefficient, respectively. The $\mathbf{N}_i \in \mathbb{R}^{3p \times 1}$ is the i-th column of $\mathbf{N}$. The linear combination of l_1-norm and l_2-norm, called elastic net constraint, are enforced to constraint the sparsity of coefficients $\mathbf{X}$ as well as scale. The parameter λ is a trade-off parameter between the reconstruction error and the elastic net constraint.

For (5), we first compute the partial differentials of $\mathbf{X}$ and $\mathbf{N}$, i.e.,

$$\partial \mathbf{X} = (\mathbf{N}^t)^T \left(\mathbf{A} - \mathbf{N}^t \mathbf{X} \right) + \lambda \left(\tau \mathbf{I}_{KF} + 2(1 - \tau)\mathbf{X} \right), \quad (6)$$

$$\partial \mathbf{N} = \left(\mathbf{A} - \mathbf{N} \left(\mathbf{X}^{t+1} \right)^T \right) \left(\mathbf{X}^{t+1} \right)^T, \quad (7)$$

where $\mathbf{I}_{KF}$ is a $K \times F$ identity matrix. Thereafter, $\mathbf{X}$ and $\mathbf{N}$ can be updated alternately as

$$\mathbf{X}^{t+1} = \mathbf{X}^t - \phi_1 \partial \mathbf{X}, \quad (8)$$

$$\mathbf{N}^{t+1} = \mathbf{N}^t - \phi_2 \partial \mathbf{N}, \quad (9)$$

```
1: Compute the shape bases B via the elastic net
   based sparse model (5).
2: Initialize α, β, γ.
3: Initialize M⁰, Z⁰, Y⁰, μ⁰, t = 0.
4: while t <= 1000 do
5:    Compute the optimized Mᵗ⁺¹ according to (15) by
      fixing Zᵗ, Yᵗ, and μᵗ,
6:    Update Zᵗ⁺¹ via (17) by fixing Mᵗ⁺¹, Yᵗ, and μᵗ,
7:    Update Yᵗ⁺¹ via (18) by fixing Mᵗ⁺¹, Zᵗ⁺¹,
8:    if δ₁ < ε  &  δ₂ < ε then
9:       break,
10:   else
11:      if δ₁ > 10δ₂ then
12:         μᵗ⁺¹ = 2μᵗ,
13:      else {δ₂ > 10δ₁}
14:         μᵗ⁺¹ = μᵗ/2.
15:      end if
16:   end if
17:   Update t ← t + 1.
18: end while
19: if refinement reconstruction then
20:   Compute R and c according to (22) via the
      alternating minimization (Zhou et al., 2013).
21: end if
22: Estimate S by using (1)
```

Algorithm 1. Pseudocode of the MCM-RR algorithm.

where ϕ_1 and ϕ_2 are the step size of $\partial \mathbf{X}$ and $\partial \mathbf{N}$, respectively. After convergence, the shape bases $\mathbf{B}$ can be obtained by a re-arrangement of $\mathbf{N}$.

2.2. 3D shape estimation via a penalized least-square model with similarity constraint

In terms of (2), the proposed penalized least-square model, including a relaxed orthogonality constraint (Zhou et al., 2013) and

a similarity constraint (Cheng et al., 2015) can be formulated as

$$\min_{\widetilde{\mathbf{M}},\mathbf{Z}} \frac{1}{2} \left\| \mathbf{W} - \mathbf{Z}\widetilde{\mathbf{B}} \right\|_F^2 + \alpha \sum_{i=1}^{K} \|\mathbf{M}_i\|_2 + \frac{\beta}{2} \|\mathbf{Z}\mathbf{D}\|_2^2 \qquad (10)$$
$$\text{s.t.} \ \widetilde{\mathbf{M}} = \mathbf{Z},$$

where $\mathbf{Z} \in \mathbb{R}^{2 \times 3K}$ is an auxiliary variable and $\widetilde{\mathbf{M}} = [\mathbf{M}_1, \cdots, \mathbf{M}_K]$, $\widetilde{\mathbf{B}} = \left[\mathbf{B}_1^T, \cdots, \mathbf{B}_K^T \right]^T$. The parameters α and β are used to weight the two regularization terms. The diagonal matrix $\mathbf{D} \in \mathbb{R}^{3K \times 3K}$ is represented as

$$\mathbf{D} = \left(\bar{\mathbf{D}} \otimes \mathbf{I}_3 \right). \qquad (11)$$

For the diagonal similarity matrix $\bar{\mathbf{D}} \in \mathbb{R}^{K \times K}$, the diagonal element d_i is computed as

$$d_i = \exp \left(\frac{\|\mathbf{W} - \Pi \mathbf{B}_i\|}{2\gamma^2} \right), \qquad (12)$$

where $\Pi = [1, 0, 0; 0, 1, 0]$, γ^2 is the parameter of an exponential function.

With the ALM iterative algorithm, the penalized least-square model (10) can be reformulated as

$$L = \frac{1}{2} \left\| \mathbf{W} - \mathbf{Z}\widetilde{\mathbf{B}} \right\|_F^2 + \alpha \sum_{i=1}^{K} \|\mathbf{M}_i\|_2 + \langle \mathbf{Y}, \widetilde{\mathbf{M}} - \mathbf{Z} \rangle$$
$$+ \frac{\beta}{2} \|\mathbf{Z}\mathbf{D}\|_2^2 + \frac{\mu}{2} \left\| \widetilde{\mathbf{M}} - \mathbf{Z} \right\|_F^2, \qquad (13)$$

where $\mathbf{Y}$ and μ are a dual variable and a weight of penalty term, respectively. In (13), there are four unknown variables $\widetilde{\mathbf{M}}$, $\mathbf{Z}$, $\mathbf{Y}$, and μ. The solutions can be solved by the alternating direction method of multipliers (ADMM).

First, the optimal $\widetilde{\mathbf{M}}$ at the $(t + 1)^{th}$ iteration can be formulated as

$$\widetilde{\mathbf{M}}^{t+1} = \arg \min_{\widetilde{\mathbf{M}}} \sum_{i=1}^{K} \frac{1}{2} \left\| \mathbf{M}_i - \mathbf{P}_i^t \right\|_F^2 + \frac{\alpha}{\mu} \|\mathbf{M}_i\|_2, \qquad (14)$$

where $\mathbf{P}_i^t$ is the ith column-triple of $\mathbf{Z}^t - \frac{1}{\mu}\mathbf{Y}^t$. According to the proximal problem (Zhou et al., 2013), $\mathbf{M}_i^{t+1}$ can be computed as

$$\mathbf{M}_i^{t+1} = U \mathrm{diag} \left(\Sigma - \frac{\alpha}{\mu} \mathcal{P}_{l_1} \left(\frac{\Sigma \mu}{\alpha} \right) \right) V^T, i \in [1, K], \qquad (15)$$

where $U \Sigma V^T = svd(\mathbf{P}_i^t)$. The operation $\mathcal{P}_{l_1}(\cdot)$ denotes the projection of a vector to the unit l_1-norm ball (Zhou et al., 2013).

Similarity, the optimal $\mathbf{Z}$ at the $(t + 1)^{th}$ iteration can be formulated as

$$\mathbf{Z}^{t+1} = \arg \min_{\mathbf{Z}} \frac{1}{2} \left\| \mathbf{W} - \mathbf{Z}\widetilde{\mathbf{B}} \right\|_F^2 + \langle \mathbf{Y}^t, \widetilde{\mathbf{M}}^{t+1} - \mathbf{Z} \rangle$$
$$+ \frac{\beta}{2} \|\mathbf{Z}\mathbf{D}\|_2^2 + \frac{\mu}{2} \left\| \widetilde{\mathbf{M}}^{t+1} - \mathbf{Z} \right\|_F^2. \qquad (16)$$

We compute the one-order partial derivative of (16) with respect to $\mathbf{Z}$ and set it as zero. Thereafter, $\mathbf{Z}^{t+1}$ can be given by

$$\mathbf{Z}^{t+1} = \left(\mathbf{W}\widetilde{\mathbf{B}}^T + \mu \widetilde{\mathbf{M}}^{t+1} + \mathbf{Y}^t \right) \left(\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}^T + \mu \mathbf{I} + \beta \mathbf{D}\mathbf{D}^T \right)^{-1}. \qquad (17)$$

Afterward, the optimal $\mathbf{Y}$ at the $(t + 1)^{th}$ iteration can be computed as

$$\mathbf{Y}^{t+1} = \mathbf{Y}^t + \mu \left(\widetilde{\mathbf{M}}^{t+1} - \mathbf{Z}^{t+1} \right). \qquad (18)$$

Given a weight τ, the coefficient μ at the $(t + 1)^{th}$ iteration can be given by

$$\mu^{t+1} = \begin{cases} 2\mu^t, & if \quad \delta_1 > \tau \delta_2, \\ \mu^t/2, & if \quad \delta_2 > \tau \delta_1, \end{cases} \qquad (19)$$

where

$$\delta_1 = \frac{\left\| \widetilde{\mathbf{M}}^{t+1} - \mathbf{Z}^{t+1} \right\|_F}{\left\| \mathbf{Z}^t \right\|_F}, \delta_2 = \frac{\left\| \mathbf{Z}^{t+1} - \mathbf{Z}^t \right\|_F}{\left\| \mathbf{Z}^t \right\|_F}. \qquad (20)$$

The iterations are repeated until

$$\delta_1 < \varepsilon \quad \& \quad \delta_2 < \varepsilon, \qquad (21)$$

where ε is a small threshold value. After obtaining $\mathbf{M}_i$, the unknown 3D shape can be reconstructed by refinement reconstruction (Zhou et al., 2013).

In the refinement reconstruction, we assume that the rotation matrices of each shape base are equal, denoted as $\bar{\mathbf{R}}$. Thereafter, c_i and $\bar{\mathbf{R}}$ can be estimated by the following rotation synchronization model

$$\min_{\mathbf{c},\bar{\mathbf{R}}} \sum_{i=1}^{k} \left\| \mathbf{M}_i - c_i \bar{\mathbf{R}} \right\|_F^2, \qquad (22)$$
$$\text{s.t.} \quad \bar{\mathbf{R}}\bar{\mathbf{R}}^T = \mathbf{I}_2,$$

which can be solved via the alternating minimization (Zhou et al., 2013). Finally, the 3D shape $\mathbf{S}$ can be estimated after $\mathbf{M}_i$ is obtained.

3. Experimental results

3.1. Experimental comparison of different algorithms

The performance evaluation of the proposed 3D shape reconstruction model (denoted as MCM-RR) is carried out on eight motion categories (walk, run, jump, climb, box, dance, sit, and basketball) from the CMU motion capture dataset (Zhou et al., 2013). Figure 1 shows one frame of those eight categories.
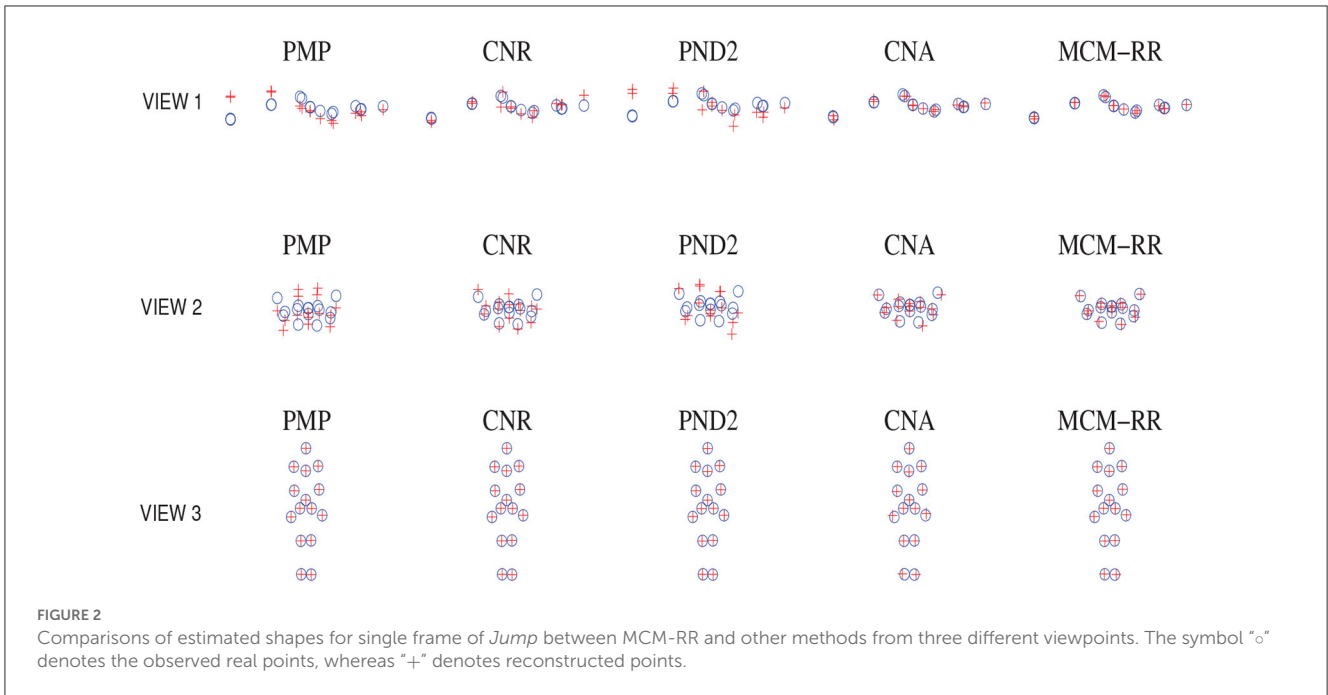
In the experiments, the performance of several state-of-the-art 3D shape estimation methods are used to compare with the presented approach, including PND2 (Lee et al., 2013), CNR (Lee et al., 2016), PMP (Lee et al., 2014), and CRA (Zhou et al., 2013).

Mean error ξ of 3D shapes is calculated as the performance indicator to measure the estimation results:

$$\xi = \frac{1}{F} \sum_{t=1}^{F} \|\tilde{\mathbf{S}}_t - \mathbf{S}_t\|_F^2, \qquad (23)$$

where $\tilde{\mathbf{S}}_t \in \mathbb{R}^{3 \times p}$ and $\mathbf{S}_t \in \mathbb{R}^{3 \times p}$ are the reconstructed 3D structure and real 3D structure of tth frame, respectively.

Table 1 displays the mean and standard deviation ($\mu \pm \sigma$) of reconstruction errors ξ of eight motion categories for the five

FIGURE 2
Comparisons of estimated shapes for single frame of *Jump* between MCM-RR and other methods from three different viewpoints. The symbol "○" denotes the observed real points, whereas "+" denotes reconstructed points.

TABLE 3 Mean and standard deviation ($\mu \pm \sigma$) of the 3D reconstruction errors ξ of eight motion categories for four methods.

Sequence	CRA	CRA-EN	CRA-SC	MCM-RR
Walk	38.98 ± 19.64	36.56 ± 19.18	38.64 ± 19.03	35.37 ± 18.49
Run	55.69 ± 18.13	52.60 ± 16.70	56.06 ± 18.03	52.64 ± 17.05
Jump	57.08 ± 41.56	46.61 ± 33.79	56.42 ± 39.52	44.56 ± 27.30
Climb	58.87 ± 24.73	49.99 ± 25.53	58.99 ± 24.88	50.25 ± 25.88
Box	72.90 ± 30.64	65.32 ± 27.64	73.02 ± 30.10	65.28 ± 26.82
Dance	102.36 ± 44.93	85.23 ± 35.63	101.49 ± 44.01	83.59 ± 34.88
Sit	75.68 ± 36.29	63.12 ± 26.79	74.92 ± 34.80	62.72 ± 26.79
Basketball	63.66 ± 27.92	57.81 ± 22.58	63.28 ± 28.29	57.57 ± 22.96

methods, respectively. The best results are highlighted in red, whereas the second best is in blue.

Table 1 shows the estimation errors of the last two methods are clearly less than that of the first triple algorithms. Among eight categories, the mean reconstruction errors of MCM-RR are the lowest compared to CRA. Moreover, the standard deviations of MCM-RR are less than that of CRA among most categories. Therefore, compared to CRA, both accuracy and robustness are effectively improved for the proposed method.

Compared to CRA, the 3D reconstruction error decreased the percentage $\xi_p(\%)$ of MCM-RR can be computed as

$$\xi_p = \frac{\xi_{\text{CRA}} - \xi_{\text{MCM-RR}}}{\xi_{\text{CRA}}} \times 100\%. \qquad (24)$$

From Table 2, we can see that the mean reconstruction errors of MCM-RR decreased about 5.48% ⌣ 21.93% compared to CRA. Thus, MCM-RR has a better 3D reconstruction performance than CRA for the eight motion categories.

Take one frame of *Jump* as an example. Figure 2 displays a comparison of reconstructed shapes between MCM-RR and the other methods from three different viewpoints. From Figure 2, we can see that compared to other methods, most estimated shapes of MCM-RR are closer to real points than that of the other methods.

3.2. Ablation experiment

In order to verify the feasibility of the proposed two strategies, the elastic net (denoted as CRA-EN) and similarity constraint (denoted as CRA-SC) are separately applied to the original algorithm CRA. Table 3 displays the mean and standard deviation ($\mu \pm \sigma$) of 3D reconstruction errors ξ of eight motion categories for the four methods, respectively. Compared to CRA, both the elastic net and similarity constraint can decrease the 3D reconstruction errors. Therefore, the 3D reconstruction performance can be effectively improved once the two methods are simultaneously designed into CRA.

4. Conclusion

In this study, a multiple-constraint algorithm is devised to estimate the 3D shape of a 2D image sequence. Experimental results on the well-known CMU datasets demonstrated that the proposed methods have higher accuracies and more robustness. Compared with CRA, the 3D reconstruction error is decreased by at least 5.48%.

Data availability statement

The datasets used in this article is from a public datesets, and it can be found in the CMU Graphics Lab Motion Capture Database.

Author contributions

XC proposed the initial research idea, conducted the experiments, and wrote the manuscript. Z-LS supervised the work and advised the entire research process. YZ collected the dataset, analyzed the formal, and revised the manuscript. All authors reviewed and approved the final manuscript.

Funding

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adamkiewicz, M., Chen, T., Caccavale, A., Gardner, R., Culbertson, P., Bohg, J., et al. (2022). Vision-only robot navigation in a neural radiance world. *IEEE Robot. Automat. Lett.* 7, 4606–4613. doi: 10.1109/LRA.2022.3150497

Akhter, I., Sheikh, Y., Khan, S., and Kanade, T. (2010). Trajectory space: a dual representation for nonrigid structure from motion. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 1442–1456. doi: 10.1109/TPAMI.2010.201

Bregler, C., Hertzmann, A., and Biermann, H. (2000). "Recovering non-rigid 3d shape from image streams," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition* (Hilton Head, SC: IEEE), 690–696.

Cheng, J., Yin, F., Wong, D. W. K., Tao, D., and Liu, J. (2015). Sparse dissimilarity-constrained coding for glaucoma screening. *IEEE Trans. Biomed. Eng.* 62, 1395–1403. doi: 10.1109/TBME.2015.2389234

Chiang, F.-K., Shang, X., and Qiao, L. (2022). Augmented reality in vocational training: a systematic review of research and applications. *Comput. Hum. Behav.* 129, 107125. doi: 10.1016/j.chb.2021.107125

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models-their training and application. *Comput. Vis. Image Understand.* 61, 38–59.

Fombona-Pascual, A., Fombona, J., and Vicente, R. (2022). Augmented reality, a review of a way to represent and manipulate 3d chemical structures. *J. Chem. Inform. Model.* 62, 1863–1872. doi: 10.1021/acs.jcim.1c01255

Gotardo, P. F., and Martinez, A. M. (2011). Computing smooth time trajectories for camera and deformable shape in structure from motion with occlusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2051–2065. doi: 10.1109/TPAMI.2011.50

Graßhof, S., and Brandt, S. S. (2022). "Tensor-based non-rigid structure from motion," in *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition* (Waikoloa, HI: IEEE), 3011–3020.

Jang, H., Sedaghat, S., Athertya, J. S., Moazamian, D., Carl, M., Ma, Y., et al. (2022). Feasibility of ultrashort echo time quantitative susceptibility mapping with a 3d cones trajectory in the human brain. *Front. Neurosci.* 16, 1033801. doi: 10.3389/fnins.2022.1033801

Kumar, S., and Van Gool, L. (2022). "Organic priors in non-rigid structure from motion," in *Proceedings of the European Conference on Computer Vision* (Springer), 71–88.

Lee, M., Cho, J., Choi, C.-H., and Oh, S. (2013). "Procrustean normal distribution for non-rigid structure from motion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Portland, OR: IEEE), 1280–1287.

Lee, M., Cho, J., and Oh, S. (2016). "Consensus of non-rigid reconstructions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV: IEEE), 4670–4678.

Lee, M., Choi, C.-H., and Oh, S. (2014). "A procrustean markov process for non-rigid structure recovery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Columbus, OH: IEEE), 1550–1557.

Li, T., Cheng, B., Ni, B., Liu, G., and Yan, S. (2016). Multitask low-rank affinity graph for image segmentation and image annotation. *ACM Trans. Intell. Syst. Technol.* 7, 1–18. doi: 10.1145/2856058

Li, T., Wang, Y., Hong, R., Wang, M., and Wu, X. (2018). PDisVPL: probabilistic discriminative visual part learning for image classification. *IEEE MultiMedia* 25, 34–45. doi: 10.1109/MMUL.2018.2873499

Lu, W., Li, Z., Li, Y., Li, J., Chen, Z., Feng, Y., et al. (2022). A deep learning model for three-dimensional nystagmus detection and its preliminary application. *Front. Neurosci.* 16, 930028. doi: 10.3389/fnins.2022.930028

Nian, F., Li, T., Bao, B.-K., and Xu, C. (2022a). Relative coordinates constraint for face alignment. *Neurocomputing* 395, 119–127. doi: 10.1016/j.neucom.2017.12.071

Nian, F., Sun, J., Jiang, D., Zhang, J., Li, T., and Lu, W. (2022b). Predicting dose-volume histogram of organ-at-risk using spatial geometric-encoding network for esophageal treatment planning. *J. Ambient Intell. Smart Environ.* 14, 25–37. doi: 10.3233/AIS-210084

Song, J., Patel, M., Jasour, A., and Ghaffari, M. (2022). A closed-form uncertainty propagation in non-rigid structure from motion. *IEEE Robot. Automat. Lett.* 7, 6479–6486. doi: 10.1109/LRA.2022.3173733

Wang, C., Wang, Y., Lin, Z., Yuille, A. L., and Gao, W. (2014). "Robust estimation of 3d human poses from a single image," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (Columbus, OH: IEEE), 2361–2368.

Wang, T., Chen, B., Zhang, Z., Li, H., and Zhang, M. (2022). Applications of machine vision in agricultural robot navigation: a review. *Comput. Electron. Agric.* 198, 107085. doi: 10.1016/j.compag.2022.107085

Wen, X., Zhou, J., Liu, Y.-S., Su, H., Dong, Z., and Han, Z. (2022). "3d shape reconstruction from 2d images with disentangled attribute flow," in *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition* (New Orleans, LA: IEEE), 3803–3813.

Zhang, S., and Xing, W. (2017). "Object tracking with adaptive elastic net regression," in *Proceedings of the IEEE International Conference on Image Processing* (Honolulu, HI: IEEE), 2597–2601.

Zhou, X., Zhu, M., Leonardos, S., and Daniilidis, K. (2013). Sparse representation for 3d shape estimation: a convex relaxation approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1648–1661. doi: 10.1109/TPAMI.2016.2605097