IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 14, 2019, accepted April 11, 2019, date of publication April 17, 2019, date of current version May 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911723

Discriminant Affinity Matrix for Deterministic Motion Trajectory Segmentation

GUOBAO XIAO[1], KUN ZENG[1], LEYI WEI[2], TAO WANG[1], AND TAOTAO LAI[1]

[1]Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, College of Computer and Control Engineering, Minjiang University, Fuzhou 350108 , China

[2]College of Intelligence and Computing, Tianjin University, Tianjin 300072, China

Corresponding author: Leyi Wei (weileyi@tju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702431, Grant 61791340, and Grant 61703195, 61702101, and in part by the Fuzhou Technology Planning Project under Grant 2018-G-96.

ABSTRACT We present a novel algorithm (DAM) for deterministic motion trajectory segmentation by using epipolar geometry and adaptive kernel-scale voting. DAM is based on geometric models and exploits the information derived from superpixels to deterministically construct a set of initial correlation matrices. Then DAM introduces a novel adaptive kernel-scale voting scheme to measure each initial correlation matrix. After that, based on the voting scores, the set of initial correlation matrices is accumulated to generate a discriminant affinity matrix, which is utilized for final grouping. The key characteristic of the DAM is its deterministic nature, i.e., DAM is able to achieve reliable and consistent performance for motion trajectory segmentation without randomness. Experimental results on both several traditional datasets (i.e., *Hopkins*155, *Hopkins*12, and *MTPV*62 datasets) and a more realistic and challenging dataset (i.e., *KT*3*DMoSeg*) show the significant superiority of the proposed DAM over several state-of-the-art motion trajectory segmentation algorithms with respect to segmentation accuracy.

INDEX TERMS Motion trajectory segmentation, deterministic fitting, computer vision.

I. INTRODUCTION

Motion trajectory segmentation is a fundamental and critical task in computer vision. Given a video sequence with feature point trajectories, the task of motion trajectory segmentation is to segment the feature point trajectories that belong to different moving objects. Thus, motion trajectory segmentation is usually formulated as the problem of clustering feature point trajectories for a video sequence with respect to their motions in, e.g., [7], [11], [12], [14], [28], [30], [31].

Motion trajectory segmentation has attracted much attention within academic and industrial communities, and a number of motion trajectory segmentation methods [6], [7], [9]–[11], [14]–[16] have been proposed in recent years. Conventional motion trajectory segmentation methods can be classified into two categories according to the number of frames considered during motion trajectory segmentation, i.e., two-frame based methods and multi-frame based methods. Two-frame based methods (e.g., [9], [16]) and multi-frame based methods (e.g., [7], [10], [15]) are usually based on the epipolar geometry and the feature point trajectories, respectively. The former usually shows better computational efficiency than the latter due to less frames used. However, the latter usually achieves more accurate segmentation results than the former since more information between frames is considered. Some other methods (e.g., [6], [11], [14]) take into account both the epipolar geometry and the feature point trajectories, and these methods can achieve better segmentation results within reasonable time.

Note that, the recent homography-based method [11] is able to achieve excellent performance with a mean error of 0.83% in the popular benchmark, i.e., the *Hopkins*155 dataset [19]. The success of [11] is based on the condition that the scene only contains compact objects or piecewise smooth structures, which can be fitted with a valid homography. However, this condition is not always satisfied in real-world scenes. For that, [28] proposed to combine multiple geometric models (i.e., homography and fundamental matrices) together to deal with real-world effects.

Although [28] is able to improve the performance of motion segmentation in that way, it requires to generate a large number of model hypotheses for multiple models,

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Zhao.

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

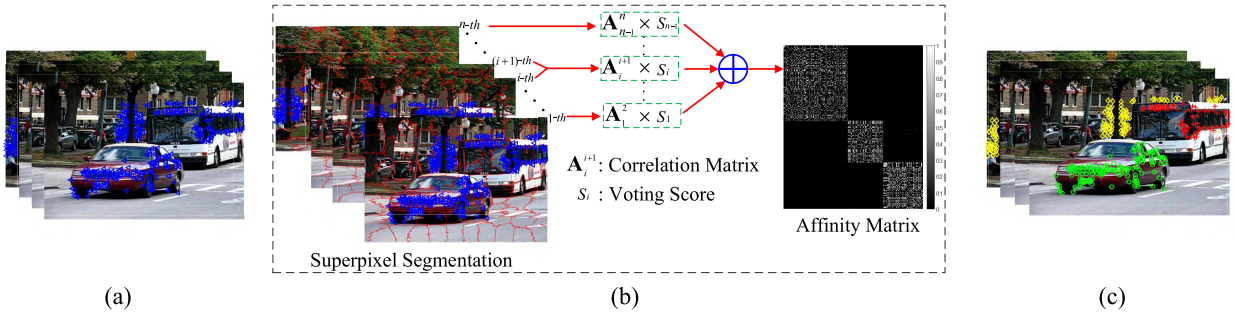


FIGURE 1. Overview of the proposed method for motion trajectory segmentation. (a) A video sequence with feature point trajectories belonging to three different motions. (b) The procedure of the proposed method. (c) The obtained segmentation result (the feature point trajectories with the same color belong to the same motions).

which is very time-consuming. We also find that [28] does not improve the quality of affinity matrices (that are used to the spectral clustering framework) derived from different geometric models. Thus, the subset constraint used by [28] may not be adequate in the sequence where the scene cannot be fitted with both homography and fundamental matrices.

In this paper, we propose a novel Discriminant Affinity Matrix based motion trajectory segmentation method (called DAM) by using epipolar geometry and adaptive kernel-scale voting. Note that the final affinity matrix (used for final grouping) is accumulated by a set of initial correlation matrices between two frames in [28]. Instead of combing different models in [28], we propose to directly improve the quality of an affinity matrix by two strategies, i.e., generating high-quality model hypotheses for each initial correlation matrix, and introducing a voting scheme for the final affinity matrix. More specifically, we firstly introduce superpixels to generate model hypotheses on every pair of consecutive frames in a video sequence. Then, we construct an initial correlation matrix based on residual values (derived from the generated model hypotheses and feature point matches) for every pair of consecutive frames in the video sequence. After that, we vote each initial correlation matrix according to the generated model hypotheses. Then, based on the voting scores, we accumulate all the initial correlation matrices to exploit the motion information from all frames of the video sequence. The main steps of the proposed method are shown in Fig. 1.

Key contributions of this work are described as follows:

- We propose an adaptive kernel-scale voting scheme to construct a discriminant affinity matrix for motion trajectory segmentation. The voting scheme is able to emphasize the initial correlation matrices where the scene can be generated high-quality model hypotheses while weakening the others where the scene cannot be fitted with a geometric model.
- We exploit the information derived from superpixels to deterministically construct a set of initial correlation matrices. Benefiting from the deterministic nature, the proposed DAM method is able to achieve reliable and consistent performance for motion trajectory segmentation without randomness.
- Experimental results demonstrate that the proposed DAM method is able to achieve highly accurate results on both several traditional datasets (i.e., *Hopkins*155 [19], *Hopkins*12 [17] and *MTPV*62 [14] datasets) and a more realistic and challenging dataset (i.e., *KT*3*DMoSeg* [28]) within reasonable time. Compared with several other state-of-the-art motion trajectory segmentation methods, DAM shows significant superiority on the accuracy of motion trajectory segmentation.

The rest of the paper is organized as follows: In Sec. II, we firstly review some related work. Then, we present the details of the proposed motion trajectory segmentation method in Sec. III. After that, we present the experimental results to verify the effectiveness of the proposed method in Sec. IV. Finally, we draw conclusions in Sec. V.

II. RELATED WORK

In this subsection, we further review the existing multi-frame based motion trajectory segmentation methods (since these methods have attracted more attention than the two-frame based methods), which can be classified into subspace-based and affinity-based methods.

The subspace-based methods assign the feature points belonging to different motions to different subspaces of a measurement matrix. Usually, these methods (e.g., [7], [17], [20]) are mathematically elegant and can achieve promising results on popular benchmarks. This is because that feature points belonging to different motions lie in different subspace of the measurement matrix. SSC [7] is a popular and effective subspace-based method. It can provide a subspace-preserving solution, i.e., there are no connections between points from different subspaces, if the input data satisfy some conditions, e.g., the subspaces are independent to each other [29]. These subspace-based methods are usually effective but they cannot well handle an input video sequence containing missing data (e.g., due to object occlusions).

The affinity-based methods analyze the pairwise or higher-order relationship between feature point trajectories

to alleviate the above problem. Therefore, these methods (e.g., [6], [11], [14], [28]) are more robust and can achieve more stable results. These affinity-based methods are robust to object occlusions, however, they also have some other problems: [6] requires the scales of the inlier noise to be known in advance; [11] assumes that all scenes can be fitted with a valid homography, which is not adequate in real-world scenes; [14], [28] have expensive computational cost.

The proposed motion trajectory segmentation method in this paper is an affinity-based method, but it does not require prior information about the scales of inlier noises, and it can achieve good results for real-world scenes within reasonable time. It is worth pointing out the differences between the proposed DAM method and MSSC [11] and Subset [28]. Although they use the robust model fitting theory to motion trajectory segmentation, they are significantly different: 1) MSSC and Subset use random sampling to generate model hypotheses, while DAM introduces superpixels to deterministically generate model hypotheses. Thus, DAM can achieve more stable results than MSSC and Subset due to the deterministic nature of DAM. 2) DAM can construct more accurate affinity matrices than MSSC and Subset. This is due to the fact that DAM can generate more consistent and reliable model hypotheses based on superpixels, which are used to compute the correlation values between feature point trajectories. Moreover, DAM is able to adaptively select the initial correlation matrices for the final affinity matrix by the proposed voting scheme. 3) DAM is much more efficient than Subset since DAM does not require estimate multiple geometric models. In contrast, Subset generates a much larger number of model hypotheses to estimate multiple geometric models, which is very time-consuming.

As the experimental results shown, the proposed DAM method can achieve stable and accurate results on popular and challenging benchmarks.

III. THE METHODOLOGY

In this section, we describe the details of the proposed discriminant affinity matrix based motion segmentation method. Specifically, we firstly introduce superpixels to deterministically construct a set of initial correlation matrices in Sec. III-A. Then we propose an adaptive kernel-scale voting scheme to construct a discriminant affinity matrix for motion trajectory segmentation in Sec. III-B. Finally, we summarize the complete method in Sec. III-C.

A. INTRODUCING SUPERPIXELS TO DETERMINISTIC MOTION SEGMENTATION

Superpixels can capture powerful prior information of feature appearances, and some works have proposed employing superpixels in motion trajectory segmentation (e.g., [2]–[4]). However, few deterministic motion segmentation methods fully take advantage of the prior information. Inspired by [26], [27], which take advantage of prior information derived from superpixels for deterministic model fitting, we introduce

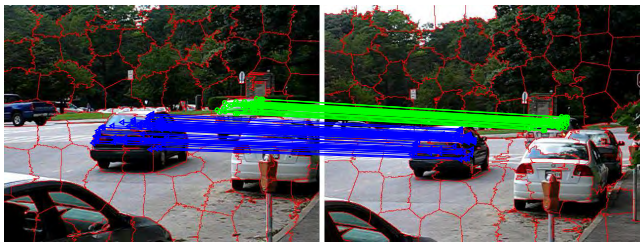


FIGURE 2. An example of the obtained superpixels and feature point matches based on two consecutive frames in the "cars2B" video sequence.

epipolar geometry constraint derived from superpixels for deterministic motion trajectory segmentation.

Specifically, for two consecutive frames in a video sequence, feature points from the two frames have a high possibility of belonging to the same motion if the corresponding feature points in each frame come from the same superpixel. We show an example of superpixels (obtained by superpixel segmentation [1]) and feature point matches (derived from two consecutive frames in the "cars2B" video sequence of the *Hopkins*155 dataset) in Fig. 2. We can see that, all the feature point matches belong to the same motion if the corresponding feature points in each frame come from the same superpixel. Based on this observation, we use the feature point matches, whose feature points belong to the same superpixel in each frame, as sampling subsets to generate model hypotheses for estimating a homography. Note that, a homography is a 3×3 matrix, which is defined as:

$$\theta = \begin{bmatrix} \theta 00 & \theta 01 & \theta 02 \\ \theta 10 & \theta 11 & \theta 12 \\ \theta 20 & \theta 21 & \theta 22 \end{bmatrix}. \tag{1}$$

Consider one feature point match $x_i = \{x_i^z, x_i^{z+1}\}$ of the sampling subsets, a homography θ maps x_i in the following way:

$$\begin{bmatrix} x_{i_x}^z \\ x_{i_y}^z \\ 1 \end{bmatrix} = \theta \begin{bmatrix} x_{i_x}^{z+1} \\ x_{i_y}^{z+1} \\ 1 \end{bmatrix} = \begin{bmatrix} \theta 00 & \theta 01 & \theta 02 \\ \theta 10 & \theta 11 & \theta 12 \\ \theta 20 & \theta 21 & \theta 22 \end{bmatrix} \begin{bmatrix} x_{i_x}^{z+1} \\ x_{i_y}^{z+1} \\ 1 \end{bmatrix}, \tag{2}$$

where $\{x_{i_x}^z, x_{i_y}^z\}$ and $\{x_{i_x}^{z+1}, x_{i_y}^{z+1}\}$ are the coordinate values of two feature points $\{x_i^z, x_i^{z+1}\}$ respectively.

We introduce superpixels to deterministic motion trajectory segmentation in an effective manner, which can provide important information for constructing the affinity matrix in Sec. III-B. However, we cannot avoid degeneracies (it may occur when a model hypothesis is estimated from a sampling subset only containing local feature point matches), during the subset sampling step, due to the over-segmentation caused by superpixels. Therefore, we propose to combine the feature point matches derived from any two superpixels as sampling subsets to generate more model hypotheses, which will effectively enlarge the sampling spans to alleviate degeneracies.

As mentioned above, [26], [27] also use superpixels to deterministically generate model hypotheses, but the proposed DAM method are significantly different with them: 1) DAM selects all feature point matches, whose feature points in each frame come from the same superpixel, as a sampling subset, while [26], [27] only select a small number of matches as a sampling subset. More matches belonging to the same structure will provide more reliable information for estimating the motion parameters. 2) DAM combines the feature point matches derived from any two superpixels in each frame as sampling subsets, while [26], [27] only combine the feature point matches derived from two neighbouring superpixels. Generally, the larger sampling spans will help to better alleviate degeneracies. 3) DAM generates model hypotheses to provide residual information in each frame, and the final affinity matrix is constructed by accumulating the residual information from all frames of a video sequence (see Sec. III-B). Therefore, the performance of DAM does not totally depend on the quality of superpixels in a frame. In contrast, [26], [27] select model instances from the model hypotheses, which are directly derived from superpixels. In other words, the performance of [26], [27] is sensitive to the quality of superpixels.

B. CONSTRUCTING DISCRIMINANT AFFINITY MATRICES

As aforementioned, multi-frame based motion trajectory segmentation methods can usually achieve more accurate results than two-frame based methods, as has been shown in some works [10], [11], [28]. Therefore, we first construct initial correlation matrices based on each pair of consecutive frames, and then accumulate all the initial correlation matrices by a novel voting scheme to yield the final affinity matrix, which includes motion information of all frames in a video sequence.

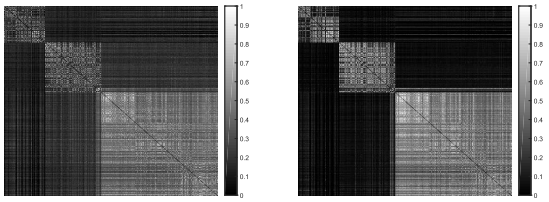
We firstly describe the details of constructing initial correlation matrices as follows: For each pair of consecutive frames (there are $(t-1)$ pairs of consecutive frames in a video sequence with t frames), we

1) Introduce superpixels to generate a set of model hypotheses $\Theta = \{\theta_1, \theta_2, \cdots, \theta_m\}$, where m is the total number of the generated model hypotheses, as described above.

2) Capture the preference of each feature point match x_i for the generated model hypotheses by sorting its absolute residuals $[r_i^1, r_i^2, \ldots, r_i^m]$, such that $r_i^{\lambda_i^1} \leq \cdots \leq r_i^{\lambda_i^m}$. The feature point match preference is encapsulated in the permutation $\mathbf{b}_i = \{\lambda_i^1, \lambda_i^2, \ldots, \lambda_i^m\}$, i.e., x_i has a higher probability of being considered as an inlier of a model hypothesis when the index of the model hypothesis is closer to the top in the permutation.

3) Compute the ''intersection'' between two feature point matches x_i and x_j as their correlation score [5]:

$$f(x_i, x_j) := \frac{1}{h}\left|\mathbf{b}_i^{1:h} \cap \mathbf{b}_j^{1:h}\right|, \tag{3}$$



(a) (b)

FIGURE 3. An example showing the affinity matrices by using different methods on the "1R2RC" video sequence of the *Hopkins*155 dataset. (a) and (b) show the affinity matrices accumulated by [11] and the proposed method, respectively.

where $|\mathbf{b}_i^{1:h} \cap \mathbf{b}_j^{1:h}|$ denotes the number of the common elements shared by $\mathbf{b}_i^{1:h}$ and $\mathbf{b}_j^{1:h}$, and h $(=\lceil 0.1m \rceil)$ is the number of the generated model hypotheses to be taken into account.

These three steps are repeated $(t-1)$ times for a video sequence with t frames. In each time, for two consecutive frames (labelled the z-th and $(z+1)$-th frames), we construct an initial correlation matrix $\mathbf{A}_z^{z+1}$, i.e., $\mathbf{A}_z^{z+1} = \{f(x_i, x_j)\}_{x_i \in \mathbf{X}_z^{z+1}, x_j \in \mathbf{X}_z^{z+1}}$, where $\mathbf{X}_z^{z+1}$ is the feature point matches for the two consecutive frames.

Then we measure the quality of a generated model hypothesis θ_j based on adaptive kernel scales as following [24], [25]

$$v(\theta_j) = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{EK}\left(r_i^j / d(\theta_j)\right)}{\tilde{s}(\theta_i) d(\theta_j)}, \tag{4}$$

where n, $\tilde{s}(\theta_j)$ and $d(\theta_i)$ are the number of feature point matches, the inlier noise scale adaptively estimated by IKOSE [23] and the bandwidth [22], respectively. $\mathbf{EK}(\cdot)$ represents the popular Epanechnikov kernel [23].

According to Eq. (4), if θ_j includes much larger number of inliers and smaller residuals, it will be assigned to a larger value; and vice versa. Thus, we use Eq. (4) to vote the corresponding initial correlation matrix $\mathbf{A}_z^{z+1}$:

$$S_z = \sum_{j=1}^{m_z} v(\theta_j), \tag{5}$$

where m_z is the number of the generated model hypotheses for two consecutive frames.

We can see that, the initial correlation matrix will obtain a higher voting score if it is based on more high-quality model hypotheses. This will help emphasize the matrices where the scene can be generated high-quality model hypotheses while weakening the others where the scene cannot be fitted with a model.

After that, according to the property that feature points on the same moving rigid object have the same matrix [10], we accumulate all the initial correlation matrices from all of the $(t-1)$ pairs of consecutive frames in the video sequence:

$$\hat{\mathbf{A}} := \sum_{z=1}^{t-1} S_z \mathbf{A}_z^{z+1}, \tag{6}$$

We show an example of the affinity matrices by using different methods on the "1R2RC" video sequence of the *Hopkins*155 dataset in Fig. 3. We can see that, we can obtain a more accurate affinity matrix by the proposed method. In an accurate correlation matrix, the entries derived from the same motions will be assigned nonzero values, while the ones derived from the different motions will be assigned zero values. The affinity matrix $\hat{\mathbf{A}}$ will be used to segment motions in the video sequence.

Algorithm 1 The discriminant affinity matrix based motion trajectory segmentation method

Input: A video sequence with feature point trajectories.

1: Perform the superpixel segmentation algorithm [1] on every frame of the video sequence.
2: Generate model hypotheses based on the superpixels for every pair of consecutive frames (see Sec. III-A).
3: Construct an initial correlation matrix for every pair of consecutive frames in the video sequence.
4: Vote all the initial correlation matrices by Eq. (5).
5: Accumulate all the initial correlation matrices to generate the final affinity matrix by Eq. (6).
6: Use sparsity constraint on the affinity matrix.
7: Perform spectral clustering based on the affinity matrix to label all feature point trajectories.

Output: The labels of feature point trajectories in the video sequence.

C. THE COMPLETE METHOD

With all the ingredients developed in the previous sections, we summarize the proposed Discriminant Affinity Matrix based motion trajectory segmentation (DAM) method in Algorithm 1. DAM contains two key elements, i.e., constructing the initial correlation matrix and the final affinity matrix. For constructing the initial correlation matrix, DAM introduces superpixels to deterministically generate model hypotheses on every pair of consecutive frames in a video sequence, and then computes the residual values (from the feature point matches to the generated model hypotheses) to construct initial correlation matrices. For constructing the final affinity matrix, DAM involves a simple and effective voting scheme to emphasize the matrices where the scene can be generated high-quality model hypotheses while weakening the others where the scene cannot be fitted with a model. For Step 6 in Algorithm 1, we follow [11], [28] to use ϵ-neighborhood scheme for improving the final performance (see the details on [11]). For Step 7 in Algorithm 1, we use the common spectral clustering method to obtain the final results as the popular segmentation method, i.e., SSC [7]. It is worth pointing that, compared with most other motion trajectory segmentation methods, DAM is able to achieve reliable and consistent motion trajectory segmentation results due to its deterministic nature.

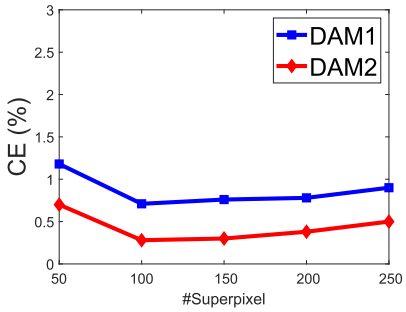


FIGURE 4. The clustering errors obtained by DAM1/DAM2 with different numbers of superpixels on the *Hopkins*155 dataset.

The computational complexity of DAM is mainly governed by Step 3 in Algorithm 1 for constructing the correlation matrices for every pair of consecutive frames. The other steps in Algorithm 1 take much less time than Step 3. Therefore, the total complexity approximately amounts to $O(tn^2)$, where t and n are the number of frames and feature point trajectories in a video sequence, respectively. The computational efficiency can be further improved by employing parallel computations, since constructing each of the initial correlation matrices is an independent step.

IV. EXPERIMENTS

In this section, we compare the proposed DAM method with several state-of-the-art motion trajectory segmentation methods, including: SSC [7], LRR [15], GPCA [21], ALC [17], MTPV [14], RSIM [8], S^3C [13], MSSC [11] and Subset [28]. We adopt three popular datasets, i.e., *Hopkins*155 [19], *Hopkins*12 [17] and *MTPV*62 [14], and a more realistic and challenging dataset, i.e., *KT*3*DMoSeg* [28], for performance evaluation. The *Hopkins*155 dataset is one of the most popular benchmarks for evaluating different motion trajectory segmentation methods. The *Hopkins*12 dataset includes some video sequences with missing data. The *MTPV*62 dataset includes some video sequences with stronger perspective effects. The *KT*3*DMoSeg* dataset includes some video sequences with both strong perspective effects and forward translations. Moreover, to show the effectiveness of the proposed voting scheme, we test two versions of DAM, i.e, DAM1 (without the voting scheme) and DAM2 (with the voting scheme). Here we manually specify the number of motions in each video sequence of all datasets for all competing methods and all experiments are run on MS Windows 7 with Intel Xeon CPU 2.80GHz and 8GB RAM.

We use the clustering error (CE) to measure the segmentation accuracy as [7], [18]:

$$CE = \frac{number\ of\ misclassified\ points}{total\ number\ of\ points} * 100\%. \quad (7)$$

A. PARAMETER ANALYSIS

In this subsection, we first analyze the influence of the number of superpixels on the performance of DAM1/DAM2. We test different numbers of superpixels for DAM1/DAM2 on

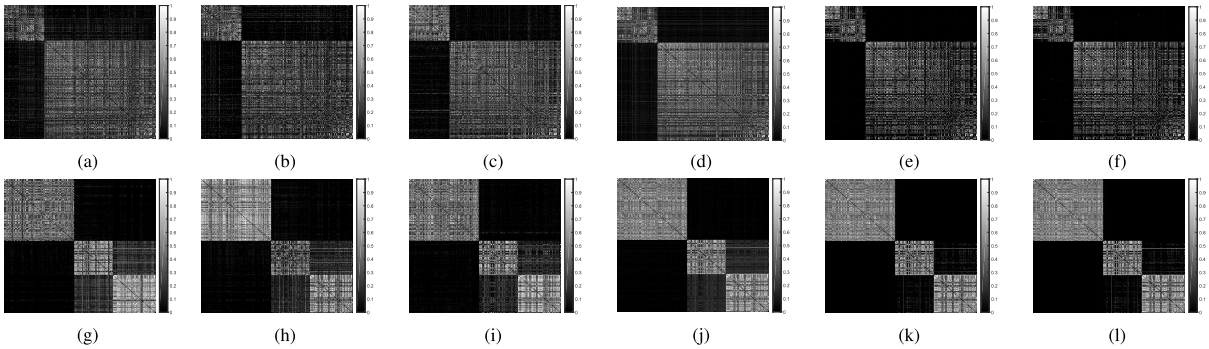


FIGURE 5. Two examples showing different affinity matrices for the "1RT2RCRT_g23" ((a)-(f)) and "cars10" ((g)-(l)) video sequence of the *Hopkins*155 dataset. (a)-(c) and (g)-(i) show the initial correlation matrix (with voting scores). (d) and (h) show the accumulate matrix. (e) and (k) show the final affinity matrix with sparsity constraint. (f) and (l) show the final affinity matrix without the voting scheme.

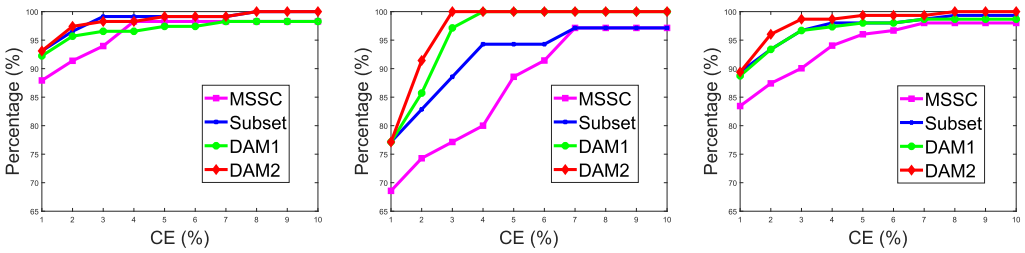


FIGURE 6. Distributions of the clustering errors obtained by four methods (i.e., MSSC, Subset, DAM1 and DAM2) on the *Hopkins*155 dataset. The vertical axis shows the percentage of the cases whose errors are smaller than the value in horizontal axis. (a) Two motions. (b) Three motions. (c) All.

the *Hopkins*155 dataset, and show the mean clustering errors in Fig. 4.

We can see that, DAM1/DAM2 are able to achieve low clustering errors when the number of superpixels is 100−200. This is consistent with the discussion in [26] and [27], which also sets the number of superpixles as 100 − 200.

We also analyze the influence of the proposed voting scheme and the sparsity constraint on the effectiveness of the affinity matrix. We show two examples of different affinity matrices for the ''1RT2RCRT_g23'' and ''cars10'' video sequence of the *Hopkins*155 dataset in Fig. 5. From Fig. 5(a)-(c) and (g)-(i), we can see that the more effective matrix will be assigned to a higher voting score. From Fig. 5(d)-(e) and (j)-(k), we can see that some wrong information can be removed by the sparsity constraint. From Fig. 5(e)-(f) and (k)-(l), we also can see that the affinity matrix becomes more effective by the proposed voting scheme.

B. DISTRIBUTION ANALYSIS

In this subsection, we analyze the distribution of the clustering errors obtained by two most relevant methods (i.e., MSSC and Subset) and two versions of the proposed methods on the popular *Hopkins*155 dataset. We show the distribution for different motions in Fig. 6.

We can see that, for two motions, all four methods are able to achieve over 85% cases whose clustering errors are smaller than 1%. However, for all cases of two motions, only DAM2 obtains all the errors that are smaller than 8%. For three motions, DAM1/DAM2 are able to achieve stable results for all cases, where the corresponding errors are smaller than 4%, due to the deterministic nature. For all cases of the *Hopkins*155 dataset, DAM2 obtains the best performance, i.e., it is reliable and consistent with very small errors. Note that, the *Hopkins*155 dataset does not include missing data, but the voting scheme, which is used by DAM2, is still important for the final grouping since it can help construct a more effective affinity matrix.

C. RESULTS ON POPULAR DATASETS

In this subsection, we evaluate the performance of the proposed DAM method on three popular datasets, i.e., *Hopkins*155, *Hopkins*12 and *MTPV*62. Firstly, we carry out the experiments of motion trajectory segmentation. Then we show some examples obtained by DAM2 on the three popular datasets in Fig. 7, from which we can see that, DAM2 can successfully segment the feature points belonging to different motions and label the feature points with a high accuracy.

To provide the quantitative comparisons, we compare the proposed DAM1/DAM2 method with nine state-of-the-art motion trajectory segmentation methods: SSC [7], LRR [15], GPCA [21], ALC [17], MTPV [14], RSIM [8], S^3C [13], MSSC [11] and Subset [28]. We report the clustering errors (in percentage) obtained by all the eleven competing

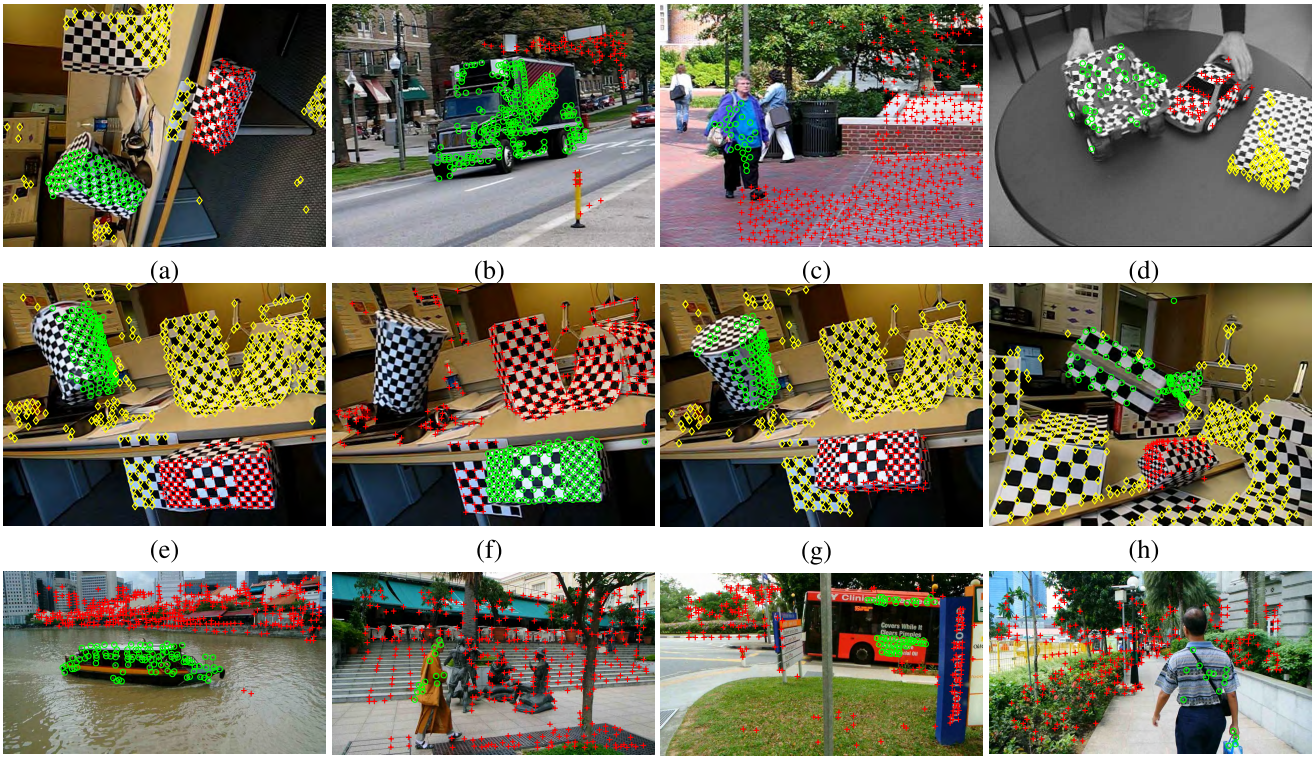


FIGURE 7. Some motion trajectory segmentation results obtained by the proposed DAM2 method on three popular datasets (namely *Hopkins*155, *Hopkins*12 and *MTPV*62 from 1st to 3rd rows, respectively) are shown. (a) 1RT2TC. (b) Truck2. (c) People2. (d) Three-cars. (e) oc1R2RC. (f) oc1R2RC_g23. (g) oc1R2RCT. (h) oc2R3RCRT. (i) Boat. (j) Monk. (k) Bus. (l) Man.

TABLE 1. The clustering errors (in percentage) obtained by the eleven competing methods on the *Hopkins*155, *Hopkins*12, *MTPV*62 and *KT3DMoSeg* dataset. '-' denotes that the corresponding value is not reported or no public code is available.

Method		SSC [7]	LRR [15]	GPCA [21]	ALC [17]	MTPV [14]	RSIM [8]	S^{3}C [13]	MSSC [11]	Subset [28]	DAM1	DAM2
*Hopkins*155 [19]	2 motions: 120 sequences											
	Mean	1.52	1.33	4.59	2.40	1.57	0.78	1.94	0.54	0.23	0.61	**0.21**
	Median	0.00	0.00	0.38	0.43	-	0.00	0.00	0.00	0.00	0.00	0.00
	3 motions: 35 sequences											
	Mean	4.40	2.51	28.66	6.69	4.98	1.77	4.92	1.84	0.58	0.60	**0.51**
	Median	0.56	0.00	28.26	0.67	-	0.28	0.89	0.30	0.00	0.00	0.00
	All: 155 sequences											
	Mean	2.18	1.59	10.02	3.36	2.34	1.01	2.61	0.83	0.31	0.62	**0.28**
	Median	0.00	0.00	-	0.50	-	0.00	0.00	0.00	0.00	0.00	0.00
*Hopkins*12 [17]	Mean	-	-	-	0.89	-	0.68	-	-	**0.06**	0.18	**0.06**
	Median	-	-	-	0.44	-	0.70	-	-	0.00	0.09	0.00
*MTPV*62 [14]	Clips with missing data: 12 clips											
	Mean	17.22	-	28.77	0.43	0.91	-	-	0.65	**0.30**	0.36	0.33
	Clips without missing data: 50 clips											
	Mean	2.01	-	16.20	18.28	2.78	-	-	0.65	0.77	0.46	**0.39**
	All: 62 clips											
	Mean	5.17	-	16.58	14.88	2.37	-	-	0.65	0.65	0.44	**0.37**
KT3DMoSeg [28]	Mean	33.88	33.67	34.60	24.31	-	-	-	-	8.08	6.47	**4.61**
	Median	33.54	36.01	33.95	19.04	-	-	-	-	**0.71**	1.80	1.30

methods in Table 1. We can see that, four competing methods (i.e., MSSC, Subset and DAM1/DAM2) that consider both the epipolar geometry and the feature point trajectories are able to achieve quite lower errors on the *Hopkins*155 dataset than the other seven competing methods. Subset improves the performance over MSSC by combing different models while DAM1 does it by introducing superpixels. However, DAM2 further improves the performance of Subset and DAM1 by using a simple voting scheme. For the *Hopkins*12 dataset, which includes missing data, although DAM1 is able

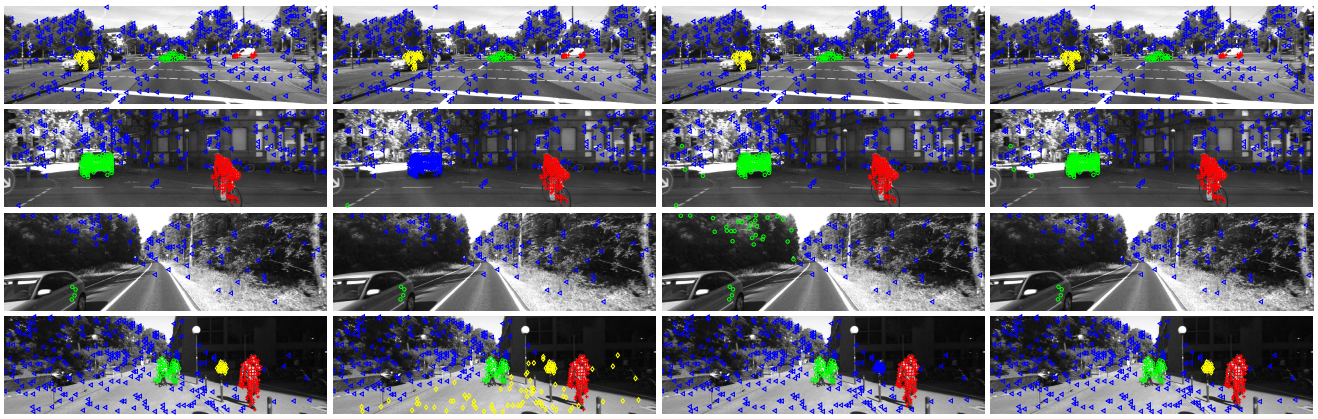


FIGURE 8. Some motion trajectory segmentation results on four video sequences (namely *Seq*011_*clip*01, *Seq*005_*clip*01, *Seq*028_*clip*03 and *Seq*038_*clip*02 from 1^{st} to4^{th} rows, respectively) of the *KT3DMoSeg* dataset are shown. (a) GroundTruth. (b) Subset. (c) DAM1. (d) DAM2.

to obtain lower errors than ALC and RSIM, Subset and DAM2 achieve better performance than DAM1 since they respectively use multiple models and a voting scheme to replace a single model that DAM1 used. On the *MTPV*62 dataset, DAM2 achieves similar errors as Subset for the clips with missing data, but it significantly improves the performance over Subset for the clips without missing data. DAM2 also achieves lower errors than DAM1 due to the effectiveness of the proposed voting scheme.

Overall, after fusing an adaptive kernel-scale voting scheme and superpixels, we can observe a boost in segmentation accuracy compared to several state-of-the-art motion trajectory segmentation methods on all three popular datasets, i.e., 0.28% on *Hopkins*155, 0.06% on *Hopkins*12 and 0.37% on *MTPV*62.

D. RESULTS ON THE KT3DMoSeg DATASET

In this subsection, we evaluate the performance of the proposed DAM2 method on the *KT*3*DMoSeg* dataset, which shows more challenges (including strong perspective effects and forward translations). The *KT*3*DMoSeg* dataset is collected by the authors in [28] and contains 22 video sequences.

We show some motion trajectory segmentation results obtained by three methods (i.e., Subset and DAM1/DAM2) that obtained the top three best performance in Fig. 8. The four video sequences in Fig. 8 are very challenging for the motion trajectory segmentation task. This is because that they not only contain strong perspective effects and forward translations, but also include other challenges, e.g., the unbalanced number of feature points and mutual interference among different motions. From Fig. 8, we can see that all three segmentation methods are able to successfully segment all moving objects on the *Seq*011_*clip*01 video sequence. However, on the *Seq*005_*clip*01 video sequence, Subset fails to segment the car in same times due to the instability of random sampling while DAM1/DAM2 have not this problem. On the *Seq*028_*clip*03 video sequence, only Subset and

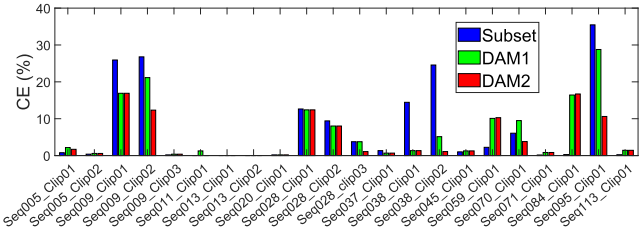


FIGURE 9. The clustering errors obtained by Subset and DAM1/DAM2 on all video sequences of the *KT3DMoSeg*dataset.

DAM2 successfully segment the background since they use multiple models and the voting scheme respectively. In contrast, DAM1 only uses a single model, which is hard to deal with unbalanced data. The *Seq*038_*clip*02 video sequence almost includes all challenges. Both Subset and DAM1 fails to segment all objects since they accumulate all information of all frames, which may mislead the final grouping. The adaptive kernel-scale voting scheme that DAM2 used plays an important role on such video sequences, which also shows its effectiveness.

We also provide some quantitative comparisons in Table 1 and show individual cluster errors obtained by Subset and DAM1/DAM2 in Fig. 9. From Table 1 and Fig. 9, we also saw the consistent boost in segmentation accuracy compared to several state-of-the-art methods as well. Note that the superpixels (which DAM1 used) are able to help yield very competitive performance even not using the adaptive kernel-scale voting scheme. Of course, the voting scheme can help further improve the performance of motion trajectory segmentation.

E. COMPUTATIONAL TIME ANALYSIS

In this subsection, we further analyze the computational time used by the three segmentation methods (i.e., Subset and DAM1/DAM2) that obtained the top three best performance on segmentation accuracy. We show the total processing time

TABLE 2. The CPU time (in seconds) obtained by the three competing methods on the *Hopkins*155, *Hopkins*12, *MTPV*62 and *KT3DMoSeg* dataset.

Method	Subset [28]	DAM1	DAM2
$Hopkins155$ [19]	2621.80	433.37	478.74
$Hopkins12$ [17]	447.40	318.67	340.50
$MTPV62$ [14]	1063.74	241.35	245.32
$KT3DMoSeg$ [28]	696.48	140.69	146.37

of different datasets obtained by Subset and DAM1/DAM2 in Table 2.

From Table 2, we can see that, the proposed DAM1/DAM2 methods significantly reduce the total processing time of Subset for all datasets. This is because Subset combines multiple models (that requires to generate a large number of model hypotheses), which is very time-consuming. Although DAM1 is a little fast than DAM2 on all four datasets since it does not use the voting scheme, DAM2 improves the performance of segmentation accuracy over DAM1.

V. CONCLUSIONS

This paper proposes a discriminant affinity matrix based deterministic motion trajectory segmentation (DAM) method, which can provide reliable and consistent results on popular datasets, i.e., *Hopkins*155, *Hopkins*12 and *MTPV*62, and a more realistic and challenging dataset, i.e., *KT3DMoSeg*. DAM firstly considers the epipolar constraint between each pair of consecutive frames in a video sequence and utilizes superpixels to construct a set of initial correlation matrices for the first time. And then DAM is also the first to introduce a simple and effective voting scheme to construct a discriminant affinity matrix for final grouping. The voting scheme is able to help emphasize the matrices where the scene can be generated high-quality model hypotheses while weakening the others where the scene cannot be fitted with a model.

Compared with several state-of-the-art motion trajectory segmentation methods, the most significant superiority of DAM is its deterministic nature, by which it will yield the same results for the same input data. A deterministic motion trajectory segmentation method is much more trackable than a method with randomized nature, which is important for practical tasks in the real world. Furthermore, DAM can achieve the best performance among all the competing methods on most of the test video sequences with respect to segmentation accuracy.

REFERENCES

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
[2] J. M. Álvarez, A. M. López, T. Gevers, and F. Lumbreras, "Combining priors, appearance, and context for road detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 3, pp. 1168–1178, Jun. 2014.
[3] A. Ayvaci and S. Soatto, "Motion segmentation with occlusions on the superpixel graph," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2009, pp. 727–773.
[4] A. Bódis-Szomorú, H. Riemenschneider, and L. V. Gool, "Fast, approximate piecewise-planar modeling based on sparse structure-from-motion and superpixels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 469–476.
[5] T.-J. Chin, J. Yu, and D. Suter, "Accelerated hypothesis generation for multistructure data via preference analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 625–638, Apr. 2012.
[6] R. Dragon, B. Rosenhahn, and J. Ostermann, "Multi-scale clustering of frame-to-frame correspondences for motion segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 445–458.
[7] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.
[8] P. Ji, M. Salzmann, and H. Li, "Shape interaction matrix revisited and robustified: Efficient subspace clustering with corrupted and incomplete data," in *Proc. IEEE Int. Conf. Comput. Vis.*, Aug. 2015, pp. 4687–4695.
[9] Y.-D. Jian and C.-S. Chen, "Two-view motion segmentation with model selection and outlier removal by ransac-enhanced Dirichlet process mixture models," *Int. J. Comput. Vis.*, vol. 88, no. 3, pp. 489–501, 2010.
[10] H. Jung, J. Ju, and J. Kim, "Rigid motion segmentation using randomized voting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Aug. 2014, pp. 1210–1217.
[11] T. Lai, H. Wang, Y. Yan, T.-J. Chin, and W.-L. Zhao, "Motion segmentation via a sparsity constraint," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 973–983, Apr. 2017.
[12] C.-M. Lee and L.-F. Cheong, "Minimal basis subspace representation: A unified framework for rigid and non-rigid motion segmentation," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 209–233, 2017.
[13] C.-G. Li and R. Vidal, "Structured sparse subspace clustering: A unified optimization framework," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Aug. 2015, pp. 277–285.
[14] Z. Li, J. Guo, L.-F. Cheong, and S. Z. Zhou, "Perspective motion segmentation via collaborative clustering," in *Proc. IEEE Int. Conf. Comput. Vis.*, May 2013, pp. 1369–1376.
[15] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
[16] B. Poling and G. Lerman, "A new approach to two-view motion segmentation using global dimension minimization," *Int. J. Comput. Vis.*, vol. 108, no. 3, pp. 165–185, 2013.
[17] S. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1832–1845, Oct. 2010.
[18] R. Tennakoon, A. Bab-Hadiashar, Z. Cao, R. Hoseinnezhad, and D. Suter, "Robust model fitting using higher than minimal subset sampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 350–362, Feb. 2016.
[19] R. Tron and R. Vidal, "A benchmark for the comparison of 3-D motion segmentation algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2007, pp. 1–8.
[20] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1945–1959, Dec. 2005.
[21] R. Vidal, R. Tron, and R. Hartley, "Multiframe motion segmentation with missing data using powerfactorization and GPCA," *Int. J. Comput. Vis.*, vol. 79, no. 1, pp. 85–105, 2008.
[22] M. Wand and M. Jones, *Kernel Smoothing*. Boston, MA, USA: Chapman Hall, 1994.
[23] H. Wang, T.-J. Chin, and D. Suter, "Simultaneously fitting and segmenting multiple-structure data with outliers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1177–1192, Jun. 2012.
[24] H. Wang, G. Xiao, Y. Yan, and D. Suter, "Searching for representative modes on hypergraphs for robust geometric model fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 697–711, Mar. 2018.
[25] G. Xiao, H. Wang, T. Lai, and D. Suter, "Hypergraph modelling for geometric model fitting," *Pattern Recogn.*, vol. 60, no. 1, pp. 748–760, Mar. 2016.
[26] G. Xiao, H. Wang, Y. Yan, and D. Suter, "Superpixel-based two-view deterministic fitting for multiple-structure data," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 517–533.
[27] G. Xiao, H. Wang, Y. Yan, and D. Suter, "Superpixel-guided two-view deterministic geometric model fitting," *Int. J. Comput. Vis.*, vol. 1, no. 1, pp. 1–17, Mar. 2018.
[28] X. Xu, L. Fah Cheong, and Z. Li, "Motion segmentation by exploiting complementary geometric models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Aug. 2018, pp. 1–9.
[29] C. You, D. P. Robinson, and R. Vidal, "Scalable sparse subspace clustering by orthogonal matching pursuit," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 3918–3927, 2016.

[30] C. Zhang, Z. Liu, C. Bi, and S. Chang, "Dependent motion segmentation in moving camera videos: A survey," *IEEE Access*, vol. 6, pp. 55963–55975, 2018.
[31] J. Zhang and Y. Shen, "High-order affinity extension of normalized cut and its applications," *IEEE Access*, vol. 6, pp. 866–870, 2018.

GUOBAO XIAO received the Ph.D. degree in computer science and technology from Xiamen University, China, in 2016, where he was a Postdoctoral Fellow with the School of Aerospace Engineering, from 2016 to 2018. He is currently a Professor with Minjiang University, China. He has published more than 30 papers in the international journals and conferences, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the *International Journal of Computer Vision*, *Pattern Recognition*, *Pattern Recognition Letters*, *Computer Vision and Image Understanding*, ICCV, ECCV, ACCV, AAAI, ICIP, and ICARCV. His research interests include machine learning, computer vision, pattern recognition, and bioinformatics. He received the Best Ph.D. Thesis in Fujian Province and the Best Ph.D. Thesis Award from the China Society of Image and Graphics. He serves on the reviewer panel for some international journals and top conferences.

KUN ZENG received the B.Eng., M.Sc., and Ph.D. degrees from the Department of Computer Science, Xiamen University, Xiamen, China, in 2005, 2008, and 2015, respectively, where he holds a postdoctoral position with the Department of Electronic Science. He is currently a Lecturer with the College of Computer and Control Engineering, Minjiang University, Fuzhou, China. His current research interests include image processing, machine learning, and medical image reconstruction.

LEYI WEI received the Ph.D. degree from Xiamen University, China, in 2016. He was a Project Researcher with the Institute of Medical Sciences, The University of Tokyo, Japan He is currently an Assistant Professor and an Independent Principle Investigator with the School of Computer Science and Technology, Tianjin University, China. He has more than 45 peer-reviewed papers (more than 30 papers as first or corresponding author) published on top-tier journals, such as bioinformatics, briefings in bioinformatics, and bmc genomics. He has got around 800 citations in Google Scholar, and his h-index is 15. His research interests include bioinformatics and machine learning. He served as the Editorial Board Member of one well-known journal named *PLoS ONE*, and also the Guest Editors of three high-impact journals, *Current Bioinformatics*, *Current Protein and Peptide Science*, and *Computational and Structural Biotechnology Journal*.

TAO WANG received the B.E. degree in information engineering from the South China University of Technology, Guangzhou, China, in 2009, and the Ph.D. degree in computer science from The Australian National University, Canberra, ACT, Australia, in 2016. He was also a member of the Computer Vision Research Group, National ICT Australia, Canberra. He is currently a Lecturer with College of Computer and Control Engineering, Minjiang University, Fuzhou, China. His research interests include scene understanding, object detection, and semantic instance segmentation.

TAOTAO LAI received the Ph.D. degree in computer science and technology from Xiamen University, China, in 2016. He was a Postdoctoral Fellow with the College of Computer and Information Sciences, Fujian Agriculture and Forestry University, China. He is currently a Lecturer with the College of Computer and Control Engineering, Minjiang University, Fuzhou, China. He has published several papers in the international journals, including the IEEE TCYB, IEEE TIE, IEEE TITS, PR, and CVIU. His research interests include robust model fitting, structure from motion, scene understanding, object detection, and semantic instance segmentation.