Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-019-01581-y

ORIGINAL RESEARCH

Vision-based vehicle behavior monitoring method using a novel clustering algorithm

Xuan Wang[1] · **Zhe Dai**[2]

Received: 9 March 2019 / Accepted: 12 November 2019

Abstract
Vehicle behavior analysis is an important task in the area of intelligent transportation systems. The purpose of this study is to monitor vehicle real-time risk and behavior using soft computing techniques. It is inspired by the fact that feature points of the same vehicle have the same motion status in the 3D space. This paper proposed an effective vehicle motion segmentation method using rigid motion constraints. Specifically, we first recovered the 3D information of the feature points using the height enumeration and camera affine matrix. Then, the rigid motion constraints of two feature point trajectories in two directions were computed to analyze the characteristic. Thereafter, an affinity matrix was built by using Gaussian kernel, where the larger value in the affinity matrix means that the feature point trajectories are more likely belonging to the same vehicle. Finally, a segment and merge procedure is adopted to cluster the feature point trajectories. Experimental results on the traffic videos and the Hopkins 155 datasets demonstrate that the proposed method achieves a good performance. This paper also gives a new idea to the field of intelligent transportation system and application.

Keywords Motion segmentation · Affinity-based method · Camera calibration · Rigid motion constraint · Spectral clustering

1 Introduction

Vehicle motion segmentation is a necessary process for many applications in intelligent transportation systems, such as traffic parameter detection, vehicle tracking and vehicle behavior analysis. The purpose of vehicle motion segmentation is to detect and separate different moving vehicles based on their different motion patterns, where each moving vehicle is regarded as a coherent rigid body. Recently, numerous methods (Dragon et al. 2012; Jeon et al. 2009; Liu et al. 2010; Jung et al. 2014; Li et al. 2013; Lai et al. 2017; Dragon et al. 2012; Xuan et al. 2018) have been proposed to achieve the vehicle motion segmentation.

The 2D motion segmentation techniques (Wu et al. 2016; Wang et al. 2018; Kanhere and Birchfield 2008; Fang et al. 2012) aimed to divide each frame of the video sequence into different target regions of a number of 2D motions. However, in many application environments, multiple moving targets may be included within a traffic scene, and each moving target needs to be identified as a coherent moving entity. In this case, the 2D-based segmentation method has many limitations, so it is necessary to implement the specific segmentation task by using the motion hypothesis based on the 3D space. This has motivated the research on 3D motion segmentation over the past decade, and its approach can be broadly divided into two-frame-based and multi-frame-based approaches. The multi-frame-based approach has received more attention because it enables the use of motion information from all frames of the video sequence to achieve accurate motion segmentation.

Various approaches for solving motion segmentation have been proposed such as multibody factorization (Vidal et al. 2005), subspace clustering (Jeon et al. 2009; Liu et al. 2010; Shi et al. 2017), epipolar geometry method (Jung et al. 2014; Li et al. 2013), agglomerative lossy compression (Rao et al. 2010) and other approaches. Those algorithms can be classified into two groups; two and multi frame-based methods.

Two-frame-based methods are usually based on epipolar geometry. One of recently proposed approaches tried to

✉ Xuan Wang
xuanwang_91@126.com

[1] School of Computer and Control Engineering, Yantai University, Yantai 264005, China

[2] School of Information Engineering, Chang'an University, Xi'an 710064, China

deal with outliers by using RANSAC and enhanced Dirichlet process mixture models (Jian and Chen 2010). Another approach (Lee and Bang 2019) defined global dimension minimization in order to reveal the clusters corresponding to the underlying subspace.These two approaches use only two frames for motion segmentation. In other words, even though multiple frames are available, these approaches do not have an ability to utilize motion information in multiple frames.

Multi-frame-based methods utilize the trajectory of feature points unlike two-view-based approaches. PAC (Trobo et al. 2019)and SSC (Jeon et al. 2009) methods have quite accurate results in both two and three motion cases. Furthermore, those algorithms are also robust to noise. However, those algorithms are extremely slow. Latent low-rank representation-based method (LatLRR) (Liu and Yan 2011) is faster and more accurate, but this method is degraded in extremely noisy environments. The ICLM-based approach (Flores-Mangas and Jepson 2013) which has been recently proposed, is very fast, but has lower accuracy than other state-of-the-art approaches.

The multi-frame based method is mainly studied under the affine hypothesis, because under this assumption, the trajectory of rigid motion vehicles across multiple frames is located in an affine subspace with a size of no more than 3, or at most four-dimensional subspace. Then the motion segmentation problem can be solved using either a factorization or a subspace separation framework (Jeon et al. 2016; Shi et al. 2017; Liu et al. 2013; Rao et al. 2010; Schindler and Suter 2006; Sugaya and Kanatani 2004; Rao et al. 2008; Yan and Pollefeys 2006). Two-frame-based methods are usually on the basis of epipolar geometry, and thus these methods are able to handle perspective effects. Their motion model are achieved by either statistical methods (Jian and Chen 2010; Li 2007) or algebraic methods (Rao et al. 2009; Vidal et al. 2005). The multi-frame-based methods have been better developed recently, mainly due to the release of the Hopkins 155 database (Rao et al. 2008). Such multi-frame-based methods are more accurate than two-view-based approaches, but they do not have a good performance when there are only a few frames. However, we considered that the current multi-frame-based methods do not encounter several real-world problems, although the classification rate of Hopkins 155 is declining and near to perfect. Compared to the two-frame-based methods, the multi-frame-based methods have the following three major drawbacks.

1.1 Perspective problem

The multi-frame-based methods are affected by the inability to handle the perspective effect, and there is no such problem in the two-frame-based methods; this becomes an important consideration when shooting outdoor image sequences with shorter lenses.

1.2 Full-length trajectories

The multi-frame-based methods typically require that the target trajectory obtained should be full length. When the tracking process fails, it will cause the trajectory to break. Therefore, it is necessary to filter out the trajectories that do not exist in some frames, and the density of the overall trajectory data set is significantly reduced. This in turn results in a lack of trajectory coverage for many parts of the target in the sequence. The full length of the trajectory requirements also makes it difficult to handle the situation of objects that enter or leave the scene and suffer temporary occlusion.

1.3 The number of motion groups

The number of motion groups is usually assumed to be known for multi-frame affine methods. It is indeed a strong indication that model selection is actually difficult for motion segmentation. Related to this issue, it is the fact that the number of motion groups in each group of the Hopkins 155 dataset remains unchanged throughout the entire video sequence tracking process, which makes it easy to indulge in the aforementioned assumption. However, the number of motion groups may change throughout a sequence as moving objects enter or leave the scene in real-time traffic video. If this fundamental problem is not solved, the application of these jobs in real life will be seriously hindered.

In our approach, we use correspondences from feature points (Rublee et al. 2011; Rosten and Drummond 2006; Calonder et al. 2010; Rosin 1999; Lowe 2004; Bay et al. 2006; El-Alfy and Binsaadoon 2019; Leutenegger et al. 2011; Alcantarilla et al. 2012) as entries for motion segmentation. In contrast to tracked features, feature points are detected independently in different frames. On one hand, comparing feature point descriptors allows establishing correspondences over missing trajectory elements. On the other hand, we use the KCF method to track the feature points in order to ensure the trajectory points are continuous and stable. In this paper, we show that by building a bottom-up frame-to-frame motion segmentation framework on multiple time scales, we can obtain stable trajectory point entries. As such, we are able to accurately segment motions of vehicles which are temporarily or spatially not visible, for example as they are occluded or leave the field of view. Furthermore, our approach is robust to outliers which are explicitly taken into account during the frame-to-frame motion segmentation. The main contribution of our method are summarized as follows:

- We propose a simple but effective method to convert the 2D feature points to 3D feature points under an affinity camera. The affinity matrix of the camera should be obtained in the processing step.
- Based on the rigid motion constraint, we propose to use the constraint model to construct an affinity for motion segmentation.
- To solve the problem of unknown number of motions, we adopt an over-segment and merging strategy to accurately segment the vehicles.

The rest of this paper is organized as follows. Section 2 introduces the processing steps. The rigid motion constraints are analyzed in Sect. 3. Section 4 displays the vehicle motion segmentation method. Experimental results are shown in Sect. 5 and finally Sect. 6 draws our conclusions.

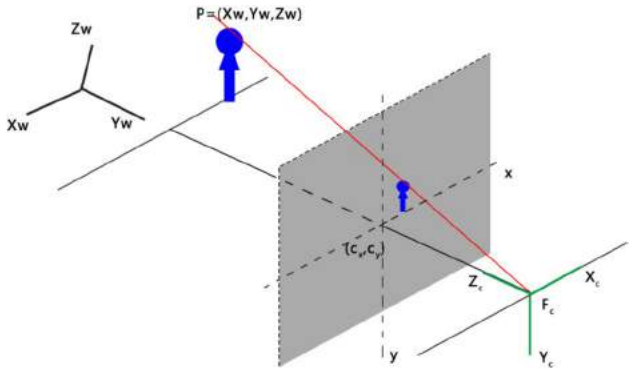

Fig. 1 Positional relationship between coordinate systems

2 Camera calibration

Camera calibration is the key to obtain the target 3D parameters based on video/image. However, the clustering and behavior analysis of the vehicle trajectory under the monocular camera is based on the feature of the vehicle feature point trajectory in 3D space, so the camera is used. It is an important precondition for the subsequent algorithm to calibrate the correspondence between the acquisition target in the 3D space and the image plane. The internal parameters and the external parameters determine the geometric mapping relationship between the 3D world coordinate system and the 2D image coordinate system. Therefore, the camera calibration is the important processing step of this paper, the camera model is shown as Fig. 1.

Without considering the distortion, the camera model is

$$\lambda p = K \begin{bmatrix} R & T \end{bmatrix} P_W \tag{1}$$

where, λ is the scale factor, $p = (u, v, 1)^T$, $P_W = (X_W, Y_W, Z_W, 1)^T$, $K = \begin{bmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix}$.

In the traffic scenes, the vanishing point along the road direction is the first direction vanishing point or the vertical vanishing point. The vanishing point perpendicular to the road direction is the second direction vanishing point or the horizontal vanishing point, the vanishing point perpendicular to the direction of the road plane is the vanishing point in the third direction. In the camera calibration principle of vanishing point, two orthogonal vanishing points can be used to calibrate the camera focal length and rotation parameters. As shown in Fig. 2, this paper selects two vanishing points along the road direction and perpendicular to the road direction in the road plane for calibration, because the

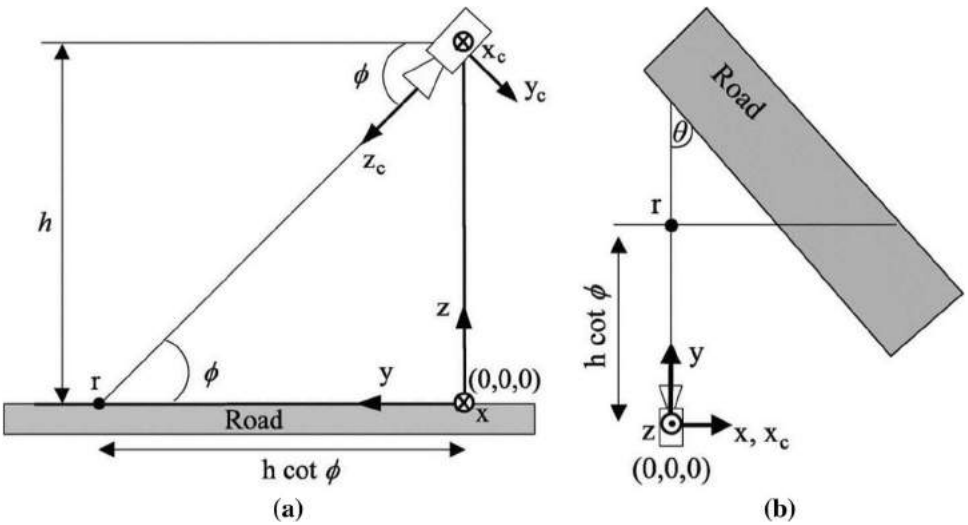

Fig. 2 Camera model under traffic scene

marking lines in the traffic scene are almost all here. In both directions, it is convenient to obtain the vanishing point, which is the direction of the world coordinate X and Y axes.

The vanishing point $vp_0(u_0, v_0)$ of the Y-axis direction is located at infinity in the world coordinate system, and can be expressed as (0, 1, 0, 0). Similarly, the vanishing point $vp_1(u_1, v_1)$ in the X-axis direction can be expressed as (1, 0, 0, 0). According to the literature (Kanhere and Birchfield 2010), the formula for calculating the camera focal length and rotation angle can be obtained as follows:

$$\begin{aligned} f &= \sqrt{-({v_0}^2 + u_0 u_1)} \\ \phi &= \arctan\left(\frac{-v_0}{f}\right) \\ \theta &= \arctan\left(\frac{-u_0 \cos\phi}{f}\right) \end{aligned} \tag{2}$$

3 Rigid motion constraint

Motion segmentation refers to the problem of separating a video sequence into multiple spatiotemporal regions corresponding to different rigid-body motions in the scene. Under the affine projection model, all the trajectories associated with a single rigid motion live in a 3D affine subspace, as we show below. Therefore, the motion segmentation problem reduces to clustering a collection of point trajectories according to multiple affine subspaces.

More specifically, let $\{x_{fi} \in \mathbb{R}^2\}_{i=1,\ldots,N}^{f=1,\ldots,F}$ denote the tracked feature points trajectories in F 2D image frames of N 3D points $\{X_{fi} \in \mathbb{R}^3\}_{i=1,\ldots,N}^{f=1,\ldots,F}$ on a rigidly moving object. The relation between the tracked feature points and the corresponding 3D coordinates of the points on the object under the affine camera model is given by

$$x_{fi} = A_f \begin{bmatrix} X_{fi} \\ 1 \end{bmatrix} \tag{3}$$

where $A_f = K_f \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_f & t_f \\ 0^T & 1 \end{bmatrix} \in \mathbb{R}^{2\times 4}$ is the affine camera matrix at frame f, which depends on the camera intrinsic parameters K_f and the object pose relative to the camera (R_f, t_f).

In this paper, we use the video sequence which are under the same affine camera model to segment vehicles. Therefore, the N points have the same affine camera matrix. The camera matrix can be obtained by the recovery method of vanishing points (Kanhere and Birchfield 2010) in the stage of pre-processing steps.

In this section, we utilize two properties of the 3D trajectories from the same vehicles for motion segmentation as follows:

Property 1. Points on the same moving rigid object have the same displacement.

Property 2. Each pair of points has a constant distance.

For instance, if there are two feature points on the same moving rigid object, they should satisfy the two properties. Then, the rigid motion constraints can be formulated as

$$E(i,j) = \sum_{f=1}^{F-1} |E_i^{(f,f+1)} - E_j^{(f,f+1)}|, \quad i \neq j, \tag{4}$$

$$D(i,j) = \sum_{f=1}^{F-1} |D_f^{(i,j)} - D_{f+1}^{(i,j)}|, \quad i \neq j, \tag{5}$$

where $E_i^{(f,f+1)} \in \mathbb{R}^3$ denotes the displacement of feature point i from fth frame to $(f+1)$th frame, $D_f^{(i,j)}$ denotes the distance between feature points i and j in fth frame. Thus, $E(i,j)$ and $D(i,j)$ indicate the two rigid motion constraints of two different 3D trajectories.

However, only the 2D trajectory points are available in the video sequence. Moreover, 2D coordinates can not be directly converted to 3D coordinates based on (3). Fortunately, the research object of this paper is the vehicle and the 3D height of the vehicle is a bounded value. For this prior knowledge, we can construct a set of tests using the method of 3D height enumeration (the range is 0–4 m and the interval is 0.01 m) to estimate the 3D height. Furthermore, the 3D trajectory points can be estimated by the 2D trajectory points with the enumerated height h_i.

For each two 2D trajectory of $\{x_{fi} \in \mathbb{R}^2\}_{i=1,\ldots,N}^{f=1,\ldots,F}$, we can compute the rigid motion constraint $E(i,j)$ and $D(i,j)$ using the estimated 3D trajectories. Then, we can get two $S \times S$ correlation matrix $\mathbf{A}$ and $\mathbf{B}$, where the size of S is decided by the number of enumerated height h_i.

$$A_{i,j} = E|_{h_i,h_j}, B_{i,j} = D|_{h_i,h_j} \tag{6}$$

As observed in Fig. 3, the two 2D trajectories both have 401 estimated 3D trajectories with h_i. If they are from the same motion, the two correlation matrices have the same region of zero value. Moreover, the corresponding height difference is a constant which indicate the height difference between two feature points in 3D space.

In theory, the values of the correlations between two feature point trajectories from the same motion in A and B are both close to zero. Based on this correlation, we construct the new measurement relation, i.e., motion constraint matrix D. $D_{i,j}$ is decided by $\min(A_{i,j})$ and $\min(B_{i,j})$. As stated in Rosin (1999), there are two ways to construct the affinity matrix: The t-nearest neighbour way and the ϵ

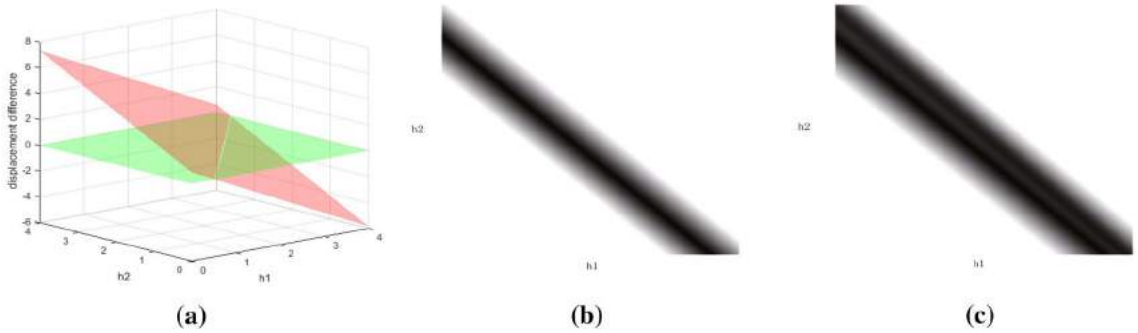

Fig. 3 Correlation matrices by rigid motion constraints

-neighbourhood way. In the t-nearest neighbour way, $D_{i,j}$ is set as $\min(A_{i,j}) + \min(B_{i,j})$ if the $\min(A_{i,j})$ (or $\min(B_{i,j})$) of the two 2D trajectories is among the t-nearest neighbours. In the ϵ-neighbourhood way, $D_{i,j}$ is set to be zero if the value of $\min(A_{i,j})$ (or $\min(B_{i,j})$) is smaller than a threshold ϵ. In this paper, we use the ϵ-neighbourhood way to construct the motion constraint matrix D. The details are given as follows.

For kth column $A_k = [A_{1,k}, A_{2,k}, \ldots, A_{401,k}]$ of the correlation matrix A, there is a minimum ϕ. Then, $\boldsymbol{\Phi}_k$ is given as

$$\boldsymbol{\Phi}_k = \begin{cases} 0 & \phi < \epsilon \\ \phi & \text{otherwise.} \end{cases} \tag{7}$$

Similarly, the $\boldsymbol{\Psi}_k$ of the correlation B is given as

$$\boldsymbol{\Psi}_k = \begin{cases} 0 & \psi < \epsilon \\ \psi & \text{otherwise.} \end{cases} \tag{8}$$

We use the median values of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, respectively, to represent the measure whether two trajectories are from the same motion. If the two trajectories from the same motion, the value is less than threshold ϵ; otherwise, the value is large. Thus, $\mathbf{D}$ is a dissimilarity matrix. $D_{i,j}$ is given as

$$D_{i,j} = \alpha \text{median}\{\boldsymbol{\Phi}\} + (1-\alpha)\text{median}\{\boldsymbol{\Psi}\}, \tag{9}$$

after which a spectral clustering method can be applied to obtain the desired segmentation into the respective subspace.

4 Motion segmentation

In this section, there are two steps involved to group the trajectories into respective motions. Firstly, by utilizing the motion constraint matrix D to construct the affinity matrix W which is applied to spectral clustering, the proposed method is able to segment or over-segment the motions. Afterwards, a merging step is implemented on the over-segment groups to obtain the final motion groups.

4.1 Affinity matrix construction

Let $G = (V, E)$ is a weighted undirected graph, where V denotes the set of trajectory, E is the set of edges connecting the trajectory nodes, $W = [w_{i,j}]_{N\times N}$ is the affinity matrix with weights describing how likely two trajectories belong to the same group, N is the number of trajectory to be grouped. Based on the new measurement relation of motion constraint matrix D, we construct the affinity matrix W, which is also the adjacency matrix, using the Gaussian kernel:

$$W_{i,j} = \begin{cases} \exp\left(-\frac{|D_{i,j}|}{2\sigma^2}\right) & i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

Let M be a diagonal matrix derived from the affinity matrix W, and $M_{i,i} = \sum_{j=1}^{N} W_{i,j}$. The normalized Laplacian matrix L is constructed as $L = M - W$. Then, the eigenvalues and eigenvectors of L are arranged in ascending order. The K eigenvectors of L are corresponding to the smallest K eigenvalues for L, where K is the possible motion number. In this paper, the Gaussian kernel parameter σ is 0.1, which may obtain the accuracy segmentation or over-segmentation. The K eigenvectors are organized to form the matrix $Q = [Q_1, Q_2, \ldots, Q_K]$. Afterwards, K-means algorithm is employed to cluster the matrix Q in order to obtain the category labels of the original trajectory data.

4.2 Merging via coefficient analysis

Spectral clustering is able to segment the trajectory data into different motions. However, sometimes it may be found that the trajectories of the same vehicle is segmented into different motions. That is because there is a tracking error at the trajectory points. Therefore, it is necessary to add a merging step after the spectral clustering.

In this section, epipolar geometry is applied to determine whether the two motions come from the same rigid object. It is represented by a fundamental matrix F, obtained from the moving object, indicating motion because it encapsulates

the intrinsic geometry. As stated in Jung et al. (2014), there are two properties of the fundamental matrix for rigid motion segmentation as follows:

Property 1. Points on the same moving rigid object have the same matrix F.

Property 2. If a set of points lie on the same moving object in an image, the corresponding points in the other frame lie on the corresponding epipolar lines. Even if the points are not used for estimating the matrix F, the same moving object's corresponding points lie on the lines.

Therefore, if there are several motions in which feature points have been clustered by spectral clustering, the feature points in each motion group can be used to estimate the fundamental matrix F. For instance, if the two motion groups are from the same rigid object, the feature points in one group can be used to the estimate fundamental matrix F. Then, the distance between the feature points in the other group and the corresponding epipolar line, computed from the matrix F, is zero. On the other hand, if most of the feature points in the other group does not lie on the corresponding epipolar line, the two motion groups are not from the same moving rigid object and vice-versa. Figure 3 demonstrates these properties. We use the properties to solve the rigid motion segmentation problem.

Various methods, such as algebraic distance, and geometric distance and its approximation, are used to measure the distance between the point and the epipolar line (Dragon et al. 2012). In this paper, we utilize the Sampson distance to determine whether the two motion groups need to be merged, as follows:

$$SD(x, y, F) = \frac{y^T F x}{(Fx)_1^2 + (Fx)_2^2 + (Fy)_1^2 + (Fy)_2^2}, \tag{11}$$

where $(Fx)_j^2$ is the square of the jth entry of the vector Fx. If we know the matrix F for one motion group, the sum of the distances over the feature points in the other group can be obtained as

$$S(X_k, Y_k, F_g) = \sum_{k=1}^{N} SD(x_k, y_k, F_g), \tag{12}$$

where F_g is a set of fundamental matrices estimated from motion group g and X_k and Y_k are the kth feature point and its corresponding point in other frame, respectively. Consequently, if $S(X_k, Y_k, F_g)$ is approximately zero for two motion groups, the two groups should be merged.

5 Experiments

In this section, we evaluate the performance of the proposed method compared with several state-of-the-art motion segmentation methods including SSC, LRR, MTPV and MSMC. In Sect. 5.1, the influence of the parameter α on the performance of the proposed method is evaluated. The motion segmentation results on the different traffic videos are shown in Sect. 5.2. Section 5.3 performs the accuracy on *Hopkins 155* data sets.

5.1 Influence of the parameter α

We compared the influence of the parameter α in (9) on the performance of the proposed method. The value of α will affect the affinity matrix obtained by (10). We set α in the range of [0 1], where $\alpha = 1$ and $\alpha = 0$ means that we only use the first property or second property of rigid motion constraint to construct the affinity matrix, respectively. Figure 4 shows the results obtained by the proposed method on the real trajectory data. The experiment was repeated 100 times on different motions. The median results are regarded as the finial performance.

5.2 Results on different traffic video

We employed our method on different real video sequences in order to demonstrate its effectiveness and robustness. These traffic scenes are tested on a Windows 10 platform and the parameters of video sequences are displayed in Table 1.

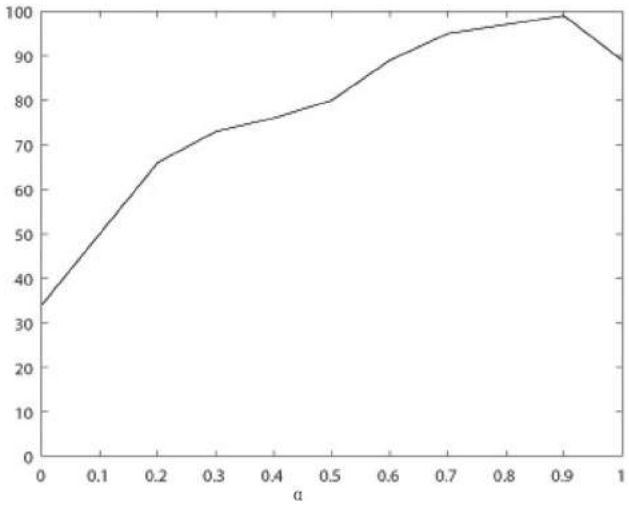

Fig. 4 Influence of the parameter α

Table 1 Parameters of video sequences

Sequences	Number of videos	Size	Frequency (FPS)
Outer Ring Road of Shanghai	22	720 × 288	25
Second South Ring road of Xi'an	23	1280 × 720	29
Jinqu Highway of Hangzhou	31	1280 × 720	25

Table 2 Running time

Numbers of vehicles	2	3	4	5
Running time (s)	2.01	2.89	3.11	3.67

The proposed algorithm is to provide data reference for real-time detection of traffic video. The running time of the algorithm needs to meet the real-time requirements. The time complexity of the algorithm is $O^2(n)$. For different numbers of vehicles, the average running time is shown in Table 2.

In each scene, we compared the results of two motions and three motions. Moreover, the feature points are extracted by SIFT descriptor and some of the traffic videos have poor quality.

As shown in Fig. 5, there are different traffic scenes in the experiments and each scene has different image quality. In

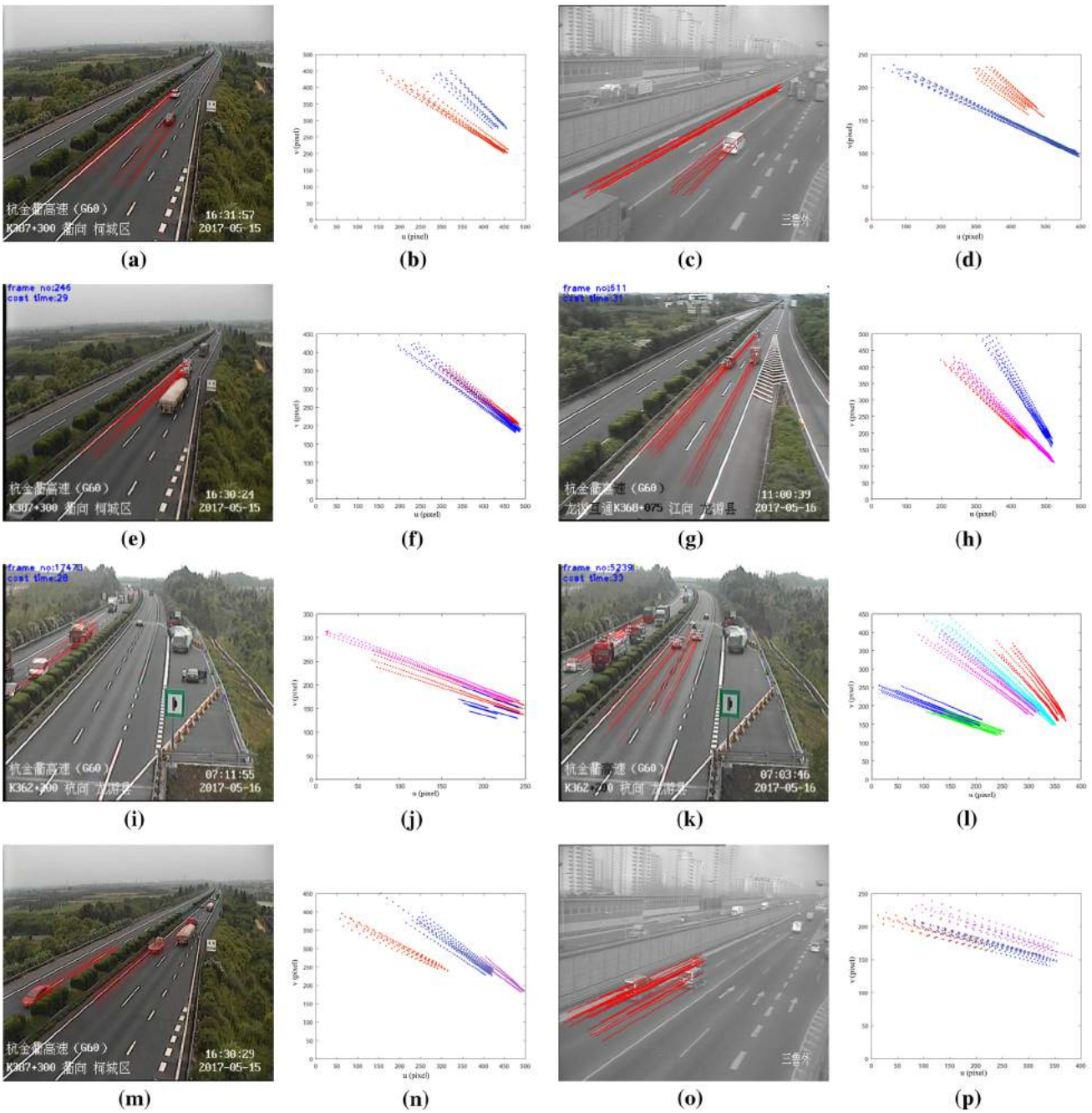

Fig. 5 Results of different traffic video sequences

Table 3 Several parameters of the sequences and trajectory sets used in the experiments

Traffic scenes	Sequences of 2 motion/3 motion	Frames	ϵ	α
Outer Ring Load of Shanghai	15/18	25	0.02	0.6
Second South Ring Road of Xi'an	18/22	20	0.02	0.7
Highway of Hangzhou	21/26	24	0.01	0.4

Table 4 Clustering error (%) obtained by several state-of-the-art methods

Method	RANSAC	SCC	LCV	SSC	PAC	TPV	RV	Proposed method
Two motion	9.48	2.7	1.86	1.24 2	1.24	1.57	0.77	0.87
Three motion	5.56	1.77	1.25	0.82	0.96	4.98	0.44	0.54
Total	22.94	5.89	3.97	2.45	2.22	2.34	1.88	2.11

terms of 2 motions and 3 motions, we compared the clustering result on multi-group experiments. Table 3 shows the relative parameters. It can be seen that ϵ and α are different in different scenes. Moreover, the correct rate of the each scene can reach 95%.

5.3 Comparison with other methods on Hopkins 155 datasets

The proposed method was compared with several state-of-the-art methods: RANSAC (Jian and Chen 2010), SCC (Chen and Lerman 2009), LCV (Zografos and Nordberg 2011), SSC (Jeon et al. 2009), PAC (Trobo et al. 2019), TPV (Li et al. 2013) and RV (Jung et al. 2014). The results obtained by the competing methods and the proposed method are given in Table 4.

Table 4 shows that the average segmentation errors obtained by the proposed method are less than 2% for video sequences with both two and three motions. In contrast to RV, the proposed method outperforms RV on the three-motion video sequences while RV outperforms the proposed method on the two-motion video sequences. On the whole dataset, RV achieves slightly better performance than the proposed method. This is not surprising because RV uses the predefined number of motions. However, the prior knowledge is not available in practice. In contrast, the proposed method is able to obtain the number of motions automatically.

6 Conclusion

In this paper, a novel motion segmentation method has been proposed, which is able to achieve vehicle motion segmentation in the real traffic scene. It is built based on the rigid motion constraints among the entire movement process. More specifically, it first calibrates the camera to obtain the affine matrix of the traffic scene. Then, 3D information of the feature point trajectories can be calculated by the enumeration height and the affine matrix. After that, rigid motion constraints is analysed between each two feature point trajectories. Then we can use the rigid motion constraints of the same vehicle to construct the affinity matrix. Finally, spectral clustering is applied on the affinity matrix to segment vehicle objects. Moreover, we add a merging step to check the clustering results. In the experiments, we evaluate the proposed method on the traffic sequences and Hopkins 155 datasets and we compare it with several state-of-the-art motion segmentation methods. The results demonstrate the proposed method has good performance in different traffic scenes for vehicle motion segmentation.

References

Alcantarilla PF, Bartoli A, Davison AJ (2012) Kaze features. In: Proceedings of European conference on computer vision. https://www.semanticscholar.org/paper/KAZE-Features-Alcantarilla-Bartoli/dd89b0332ae4f980d8200ec2c1230c50cd66259f

Bay H, Tuytelaars T, Gool LV (2006) Surf: speeded up robust features. In: Proceedings of European conference on computer vision. https://www.semanticscholar.org/paper/Speeded-Up-Robust-Features-(SURF)-Bay-Ess/cdbb606ae47c64049262dfbd3bb147d3f4ba8420

Calonder M, Lepetit V, Strecha C, Fua P (2010) Brief: binary robust independent elementary features. In: Proceedings of European conference on computer vision. https://www.semanticscholar.org/paper/BRIEF%3A-Binary-Robust-Independent-Elementary-Calonder-Lepetit/2145f5cbac48df8ed9f7695606c34e98b26cf5a9

Chen G, Lerman G (2009) Motion segmentation by SCC on the Hopkins 155 database. CoRR abs/0909.1608

Dragon R, Rosenhahn B, Ostermann J (2012) Multi-scale clustering of frame-to-frame correspondences for motion segmentation. In:

ECCV. http://citeseerx.ist.psu.edu/viewdoc/download?doi=%5C%5C10.1.1.725.7144%5C&rep=rep1%5C&type=pdf
El-Alfy ES, Binsaadoon A (2019) Automated gait-based gender identification using fuzzy local binary patterns with tuned parameters. J Ambient Intell Hum Comput 10:2495–2504
Fang Y, Wu J, Huang B (2012) 2D sparse signal recovery via 2D orthogonal matching pursuit. Sci China Inf Sci 55(4):889–897
Flores-Mangas F, Jepson AD (2013) Fast rigid motion segmentation via incrementally-complex local models. In: 2013 IEEE conference on computer vision and pattern recognition, pp 2259–2266
Jeon G, Anisetti M, Lee J, Bellandi V, Damiani E, Jeong J (2009) Concept of linguistic variable-based fuzzy ensemble approach: application to interlaced HDTV sequences. IEEE Trans Fuzzy Syst 17(6):1245–1258
Jeon G, Anisetti M, Wang L, Damiani E (2016) Locally estimated heterogeneity property and its fuzzy filter application for deinterlacing. Inf Sci 354:112–130
Jian YD, Chen CS (2010) Two-view motion segmentation with model selection and outlier removal by RANSAC-enhanced dirichlet process mixture models. Int J Comput Vis 88:489–501. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.415.3493%5C&rep=rep1%5C&type=pdf
Jung H, Ju J, Kim J (2014) Rigid motion segmentation using randomized voting. In: 2014 IEEE conference on computer vision and pattern recognition, pp 1210–1217
Kanhere NK, Birchfield ST (2008) Real-time incremental segmentation and tracking of vehicles at low camera angles using stable features. IEEE Trans Intell Transp Syst 9(1):148–160
Kanhere NK, Birchfield ST (2010) A taxonomy and analysis of camera calibration methods for traffic monitoring applications. IEEE Trans Intell Transp Syst 11(2):441–452
Lai T, Wang H, Yan Y, Chin TJ, Zhao WL (2017) Motion segmentation via a sparsity constraint. IEEE Trans Intell Transp Syst 18(4):973–983
Lee YH, Bang SI (2019) Improved image retrieval and classification with combined invariant features and color descriptor. J Ambient Intell Hum Comput 10(6):2255–2264
Leutenegger S, Chli M, Siegwart R (2011) BRISK: binary robust invariant scalable keypoints. In: Proceedings of IEEE conference on computer vision. https://www.semanticscholar.org/paper/BRISK%3A-Binary-Robust-invariant-scalable-keypoints-Leutenegger-Chli/86d27422aac2398cfe132ae8e312a6f2d190f754
Li H (2007) Two-view motion segmentation from linear programming relaxation. In: 2007 IEEE conference on computer vision and pattern recognition, pp 1–8
Li Z, Guo J, Cheong LF, Zhou SZ (2013) Perspective motion segmentation via collaborative clustering. In: 2013 IEEE international conference on computer vision, pp 1369–1376. http://openaccess.thecvf.com/content%5C_iccv%5C_2013/papers/Li%5C_Perspective%5C_Motion%5C_Segmentation%5C_2013%5C_ICCV%5C_paper.pdf
Liu G, Yan S (2011) Latent low-rank representation for subspace segmentation and feature extraction. In: 2011 International conference on computer vision, pp 1615–1622
Liu G, Lin Z, Yu Y (2010) Robust subspace segmentation by low-rank representation. IEEE Trans Cybern 44:1432–1445. https://www.semanticscholar.org/paper/Robust-Subspace-Segmentation-by-Low-Rank-Liu-Lin/047175fb23f6f152d86e81100ba7140dd2847636
Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y (2013) Robust recovery of subspace structures by low-rank representation. IEEE Trans Pattern Anal Mach Intell 35:171–184
Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
Rao SR, Tron R, Vidal R, Ma Y (2008) Motion segmentation via robust subspace separation in the presence of outlying, incomplete, or corrupted trajectories. In: 2008 IEEE conference on computer vision and pattern recognition, pp 1–8
Rao SR, Yang AY, Sastry SS, Ma Y (2009) Robust algebraic segmentation of mixed rigid-body and planar motions from two views. Int J Comput Vis 88:425–446
Rao SR, Tron R, Vidal R, Ma Y (2010) Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories. IEEE Trans Pattern Anal Mach Intell 32:1832–1845
Rosin PL (1999) Measuring corner properties. Comput Vis Image Underst 73(2):291–307
Rosten E, Drummond T (2006) Machine learning for high speed corner detection. In: Proceedings of European conference on computer vision, pp 2091–2096. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.60.3991%5C&rep=rep1%5C&type=pdf
Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: an efficient alternative to SIFT or SURF. In: Proceedings of IEEE conference on computer vision, pp 2564–2571. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.370.4395%5C&rep=rep1%5C&type=pdf
Schindler K, Suter D (2006) Two-view multibody structure-and-motion with outliers through model selection. IEEE Trans Pattern Anal Mach Intell 28:983–995
Shi J, Wu J, Anisetti M, Damiani E, Jeon G (2017) An interval type-2 fuzzy active contour model for auroral oval segmentation. Soft Comput 21(9):2325–2345
Sugaya Y, Kanatani K (2004) Geometric structure of degeneracy for multi-body motion segmentation. In: ECCV workshop SMVP. https://www.semanticscholar.org/paper/Geometric-Structure-of-Degeneracy-for-Multi-body-Sugaya-Kanatani/ab33459c40411e20d2d06c0a9ba4cfa5a07b1d02
Trobo IP, Díaz VG, Espada JP, Crespo RG, Moreno-Ger P (2019) Rapid modeling of human-defined AI behavior patterns in games. J Ambient Intell Hum Comput 10:2683–2692
Vidal R, Ma Y, Soatto S, Sastry SS (2005) Two-view multibody structure from motion. Int J Comput Vis 68:7–25
Wang J, Wu J, Wu Z, Anisetti M, Jeon G (2018) Bayesian method application for color demosaicking. Opt Eng 57(5):053102
Wu J, Anisetti M, Wu W, Damiani E, Jeon G (2016) Bayer demosaicking with polynomial interpolation. IEEE Trans Image Process 25(11):5369–5382
Xuan W, Huansheng S, Qi G, Hua C, Zhaoyang Z, Haiying L (2018) Vehicle motion segmentation using rigid motion constraints in traffic video. Sustain Cities Soc 42:547–557
Yan J, Pollefeys M (2006) A general framework for motion segmentation: independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In: ECCV. https://www.semanticscholar.org/paper/A-General-Framework-for-Motion-Segmentation%3A-Rigid%2C-Yan-Pollefeys/7785c92d471ef3d2ec132f3facf2c1f7ec7e56a5
Zografos V, Nordberg K (2011) Fast and accurate motion segmentation using linear combination of views. In: BMVC. https://www.semanticscholar.org/paper/Fast-and-accurate-motion-segmentation-using-Linear-Zografos-Nordberg/981f32bc4a5d8550025384c9b131a1f721bf9e95