Research Article

Motion Segmentation by New Three-View Constraint from a Moving Camera

Fuyuan Xu, Guohua Gu, Kan Ren, and Weixian Qian

Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing 210094, China

Correspondence should be addressed to Fuyuan Xu; xu_fuyuan@hotmail.com

We propose a new method for the motion segmentation using a moving camera. The proposed method classifies each image pixel in the image sequence as the background or the motion regions by applying a novel three-view constraint called the "parallax-based multiplanar constraint." This new three-view constraint, being the main contribution of this paper, is derived from the relative projective structure of two points in three different views and implemented within the "Plane + Parallax" framework. The parallax-based multiplanar constraint overcomes the problem of the previous geometry constraint and does not require the reference plane to be constant across multiple views. Unlike the epipolar constraint, the parallax-based multiplanar constraint modifies the surface degradation to the line degradation to detect the motion objects followed by a moving camera in the same direction. We evaluate the proposed method with several video sequences to demonstrate the effectiveness and robustness of the parallax-based multiplanar constraint.

1. Introduction

The ground motion detection is an essential challenge in many computer vision and video processing tasks, such as vision-based motion analysis, intelligent surveillance, and regional defense. When the prior knowledge of the motion object appearance and shape is not available, the change detection or optical flow can still provide powerful motion-based cues for segmenting and localizing the objects, even when the objects move in a cluttered environment or are partially occluded. The aim of the ground motion object detection is to segment the motion objects out according to the motions in the image sequence whether the platform is moving or not.

For extensive study in detecting the motion objects in the image sequence captured by a moving camera, the scene contains multiple objects moving in the background and the background may also contain a strong parallax produced by the 3D structures. The motion segmentation in dynamic image background is inherently difficult, for the moving camera induces 2D motion for each pixel. The motion of pixels in moving objects is generated by both the independent object motion and the camera motion. In contrast, the motion of pixels in the static background is strictly due to the camera motion. Our goal is to utilize multiview geometric constraints to segment the motion objects from the video sequence.

The first geometric constraint used in detecting motion objects is the homography constraint in 2D plane mode [1, 2]. The homography matrix is a global motion model which can compensate for the camera motion between consecutive images. The pixels which are consistent with homography constraint are considered to belong to the static background [3]. However, those inconsistent pixels may correspond to the motion objects or parallax regions [4, 5]. Because the homography constraint cannot distinguish the parallax regions and the moving objects, the epipolar constraint as the supplement of the homography constraint is used in the motion segmentation [6, 7].

The epipolar constraint is a commonly used constraint for motion segmentation between two views [1, 8, 9]. There are two corresponding feature points in two images from the different views. If a feature point in one image does not lie on the epipolar line induced by its matched feature point in another image, then the corresponding 3D point will be determined to be moving [10]. However, the epipolar constraint is not sufficient to detect all kinds of 3D motion.

Indeed, when the motion objects move on a special plane in 3D, then the epipolar constraint cannot detect them [1]. This phenomenon is called "surface degradation." The 3D point moves on the epipolar plane which is formed by two camera centers of the different views and the point itself, so its 2D projections move along the epipolar lines. In this case, the motion objects cannot be detected by the epipolar constraint. The surface degradation often happens when the moving camera follows the objects moving along the same line.

In order to overcome the surface degradation of the epipolar constraint, the geometric constraints over more than two views need to be imposed. The trilinear constraint can be applied to segment the motion objects across three views [1, 11]. However, estimating the parameters of the trifocal tensor is a nontrivial task, which requires accurate correspondence of the points and large camera motion.

In this paper, inspired by [12], we propose a novel three-view constraint which is named "parallax-based multiplanar constraint." The parallax-based multiplanar constraint as the supplement of the epipolar constraint modifies the surface degradation to the line degradation. Compared to the previous method based on the "Plane + Parallax" framework [12–15], the parallax-based multiplanar constraint can segment the motion objects without a fixed reference plane. The main contributions can be summarized as follows.

(1) The parallax-based multiplanar constraint can segment the motion objects without a fixed reference plane. The traditional methods [13–15] assume that the reference plane is consistent across three views. However, this assumption is not valid sometimes. The parallax-based multiplanar constraint is inspired by [12] and segmenting the motion object without a fixed reference plane based on "Plane + Parallax" framework. This is the main contribution in this paper.

(2) A reference point is introduced to replace the epipole. The calculation of the epipole is inaccurate when the motion vectors of the feature points are followed by a moving camera in the same direction [13]. The reference point can improve the accuracy of the parameters to get better results for the motion segmentation.

(3) A motion segmentation framework based on the parallax-based multiplanar constraint is proposed. In the motion segmentation framework, the parallax-based multiplanar constraint and the homography constraint are applied in the "Plane + Parallax" framework. This motion segmentation framework can reduce the run time and apply the parallax-based multiplanar constraint into real-time system.

The paper is organized as follows. In Section 2, we briefly review the existing approaches related to our work. Section 3 formally describes the epipolar constraint and the surface degradation that the epipolar constraint is unable to handle. In Section 4, we briefly review the definition of the parallax-based rigidity constraint in [12]. We then introduce the parallax-based multiplanar constraint and its degenerate cases in Section 5. The application of the parallax-based multiplanar constraint is explained in Section 6. The experimental results are shown and discussed in Section 7. Section 8 concludes our paper and presents possible directions for future research.

2. Related Work

From what is described in Section 1, the methods of the motion objects detection from a moving camera are vast. The parallax-based multiplanar constraint is based on "Plane + Parallax" framework to segment the motion objects. It considers that the image sequence can be decomposed into the reference plane, the parallax, and the motion objects. Thus, the motion segmentation methods based on the background subtraction and the motion segmentation with the strong parallax are the most related topics to this paper.

The background subtraction method has a wide range of applications in the static camera [14]. A novel framework to segment the motion objects by detecting contiguous outliers in the low-rank representation is proposed in [4, 15]. It avoids the complicated motion computation by formulating the problem as outlier detection and makes use of the low-rank modeling to deal with complex background. A new method is based on Dirichlet process Gaussian mixture models, which are used to estimate per-pixel background distributions. It is followed by the probabilistic regularization. Using a nonparametric bayesian method allows per-pixel mode counts to be automatically inferred, avoiding over-/underfitting [16]. These methods have get the better effect in motion segmentation without the strong parallax.

For motion segmentation with the strong parallax, sparse motion field estimation is a common approach [17, 18]. The sparse motion field of the corners is recovered and the corners that belong to the same motion pattern are classified according to their motion consistency [17, 19, 20]. The constraint equations are applied in the optical flow to decompose the background and foreground [21]. An effective approach is to do background subtraction for complex videos by decomposing the motion trajectory matrix into a low rank one and a group sparsity one. Then the information from these trajectories is used to further label foreground at the pixel level [22]. The motion segmentation approaches [23] segment the point trajectories based on subspace analysis. These algorithms provide interesting analysis on sparse trajectories, though they do not output a binary mask as many background subtraction methods do. However, most of the methods based on sparse motion field estimation assume that the motion object can be represented by the feature points (example as Harris corners). This assumption is invalid in many cases, so the detect rate is poor for these methods.

3. Epipolar Constraint and Surface Degradation

The epipolar geometry is the intrinsic projective geometry between two views. It is independent of the scene structure and only depends on the cameras' internal parameters and
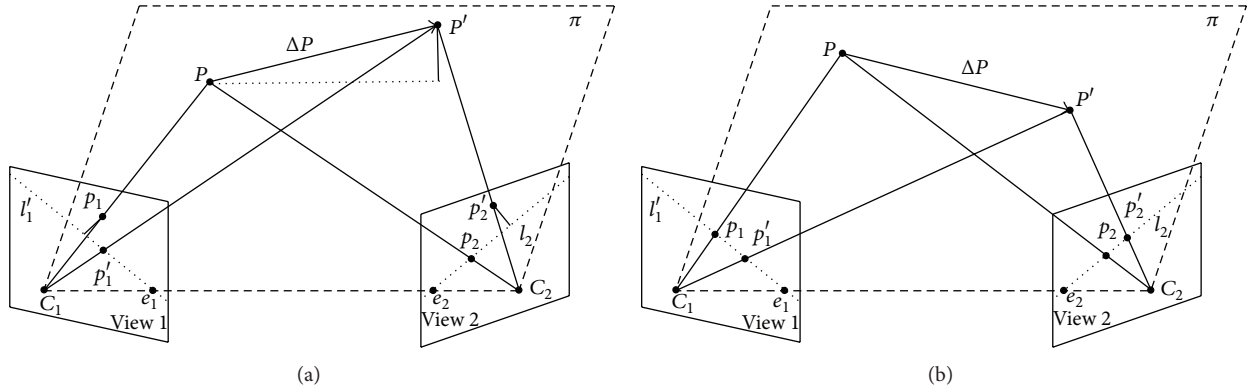
FIGURE 1: Application of the epipolar constraint. (a) Motion object detected by the epipolar constraint. (b) Surface degradation: motion object moving on the plane.

relative pose of the camera. The epipolar constraint is usually used in the motion segmentation in two views. The fundamental matrix is the algebraic representation of epipolar geometry [1, 8].

Suppose that there are two images acquired by cameras with noncoincident centers; then the fundamental matrix F_{21} is the homogeneous matrix between view 1 and view 2 which satisfies

$$p_2^T F_{21} p_1 = 0, \tag{1}$$

for all corresponding points p_1 and p_2. If P is a 3D static point, p_1 and p_2 are the projection of P in image 1 and image 2 which are from view 1 and view 2.

If point P moves between view 1 and view 2, the position of the 3D point P in view 2 is denoted as P'; p_2' is therefore the projection of point P' in image 2 and the epipolar line is $l_2' = F_{21}^T p_2'$. In this case, p_1 does not lie on l_1' and p_2' does not lie on $l_2 = F_{21} p_1$. The pixel-to-line distance d_{epi} is used to measure how the point pair deviates from the epipolar lines

$$d_{\text{epi}} = \frac{\left(\left| l_1' \cdot p_1 \right| + \left| l_2 \cdot p_2' \right| \right)}{2}, \tag{2}$$

where $|l_1' \cdot p_1|$ and $|l_2 \cdot p_2'|$ are the perpendicular distances from p_1 to l_1' and from p_2' to l_2, respectively, as shown in Figure 1(a).

Furthermore, d_{epi} is used to detect whether the 3D point P is moving or not. If $d_{\text{epi}} > 0$, P is moving. However, there is a special kind of case, called "surface degradation," where the moving points cannot be detected by the epipolar constraint. The surface degradation happens when the objects are moving in a special plane, as illustrated in Figure 1(b). If point P and the camera centers C_1 and C_2 constitute a plane π in 3D Euclidean space and point P moves to P' and $P' \in \pi$, point p_2' lies on l_2 and point p_1 lies on l_1' in 2D images. In this situation, $d_{\text{epi}} = 0$ and the surface degradation happens.

Unfortunately, the camera follows the motion objects moving in the same direction in many practical cases [12, 24]. If the camera follows the objects moving in the same direction, the surface degradation may happen. In order to solve the surface degradation, multiview constraints need to
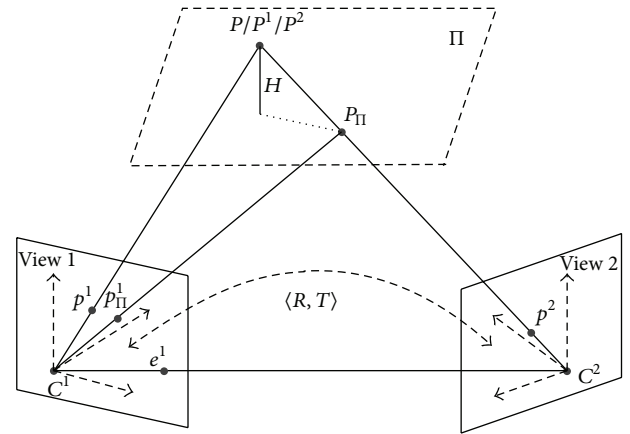
be introduced. Therefore, in Sections 4 and 5, the novel three-view constraints are proposed to segment the motion object.

4. Parallax-Based Rigidity Constraint

The "Plane + Parallax" framework [12, 15, 25] extends the 2D parametric registration approach to general 3D scenes. The plane registration process (using the dominant 2D parametric transformation) removes all effects of camera rotation, zoom, and calibration, without explicitly computing them. The residual image motion after the plane registration is only due to the translational motion of the camera and to the deviations of the scene structure from the planar surface.

4.1. "Plane + Parallax" Framework. Figure 2 provides the geometric interpretation of the "Planar + Parallax" framework. Let $P = (X, Y, Z)$ denote a 3D static point and let $P^1 = (X^1, Y^1, Z^1)^T$ and $P^2 = (X^2, Y^2, Z^2)^T$ denote the coordinates of P in different camera views. Let the 3×3 rotation matrix R and the 3×1 translation vector $T = (T_X, T_Y, T_Z)$ denote the rotation and translation between the camera systems.

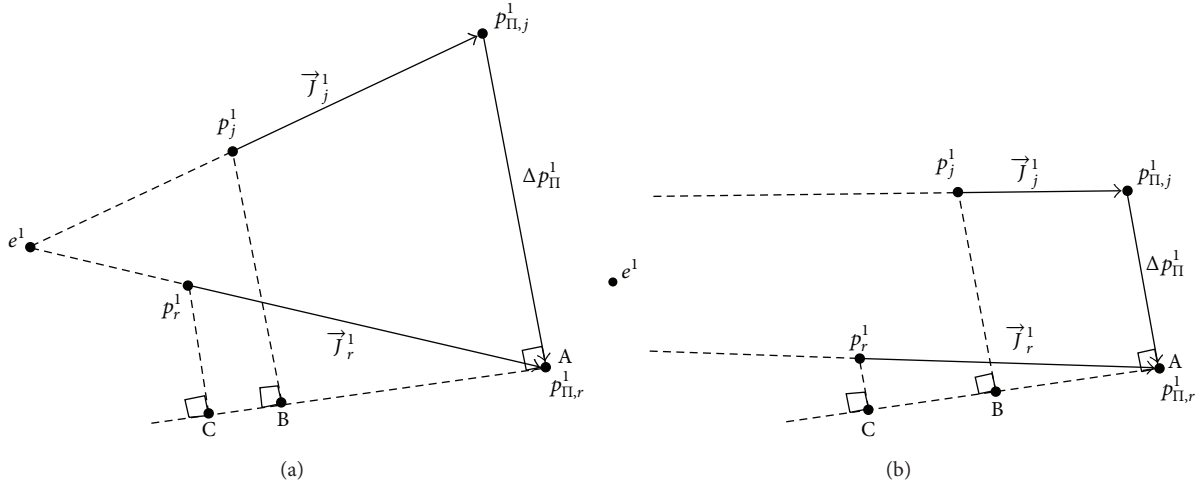FIGURE 2: Geometric interpretation of the "planar + parallax" framework.

FIGURE 3: Pairwise parallax-based shape constraint. (a) Interpretation of the relative structure constraint. (b) When the parallax vectors are nearly parallel, the epipole estimation is unreliable.

Let (x^1, y^1) and (x^2, y^2) denote the image coordinates of the 3D point P projected onto two different views. The homogeneous expression can be denoted as $p^1 = (x^1, y^1, 1)^T$ and $p^2 = (x^2, y^2, 1)^T$. Let Π be an arbitrary planar surface and $P \in \Pi$. A denote the homography matrix that aligns the planar surface Π between two different views. We can describe it as $p^1 = Ap^2$ [1].

Define $= p^2 - p^1 = (u, v, 0)^T$, where (u, v) is the 2D image displacement vector of the 3D point P between two different views. It can be shown, as well, that

$$J = J_\Pi + \vec{J}^1, \tag{3}$$

where J_Π denote the planar part of the 2D image motion and $\vec{J}$ denote the residual planar parallax 2D motion. When $T_Z \neq 0$,

$$J_\Pi = \left(p^2 - p_\Pi^1\right); \qquad \vec{J}^1 = \gamma^1 \frac{T_z}{d_\Pi^2} \left(e^1 - p_\Pi^1\right), \tag{4}$$

where p_Π^1 denotes the image point in view 1 which results from warping the corresponding p^2 in view 2, by the 2D parametric transformation of the reference plane Π. The first view is referred to as the reference view. Also, d_Π^2 is the perpendicular distance from the second camera center to the reference plane Π, and noted e^1 denotes the epipole. γ^1 is a measure of the 3D shape of the 3D point P. In particular, $\gamma^1 = H/Z^1$, where H is the perpendicular distance from P to the reference plane Π, and Z^1 is Z-distance of point P in the first camera coordinate systems. We refer γ^1 to the projective 3D structure of point P. The use of the "Plane + Parallax" framework for ego-motion estimation is described in [12], and for 3D shape recovery it is described in [26]. The "Plane + Parallax" framework is more general than the traditional decomposition in terms of rotational and translational motion.

4.2. Parallax-Based Rigidity Constraint

Theorem 1. *Given the planar-parallax displacement vectors $\vec{J}_j^1$ and $\vec{J}_r^1$ of two points that belong to the static background scene, their relative 3D projective structure γ_j^1/γ_r^1 is given by*

$$\frac{\gamma_j^1}{\gamma_r^1} = \frac{\left(\vec{J}_j^1\right)^T \left(\Delta p_\Pi^1\right)_\perp}{\left(\vec{J}_r^1\right)^T \left(\Delta p_\Pi^1\right)_\perp^j}, \tag{5}$$

where, as shown in Figure 3(a), p_j^1 and p_r^1 are the image locations of two points that are part of the static scene, $\Delta p_\Pi^1 = p_{\Pi,j}^1 - p_{\Pi,r}^1$, the vector connects the "warped" locations of the corresponding points in view 2, and $v_\perp$ signifies a vector perpendicular to v [12].

Form Figure 3(a), $\gamma_j^1/\gamma_r^1 = (\vec{J}_j^1)^T(\Delta p_\Pi^1)_\perp/(\vec{J}_r^1)^T(\Delta p_\Pi^1)_\perp^j = AB/AC$ when the epipole is stable. However, when the parallax vectors are nearly parallel, the epipole estimation is unreliable. However, the relative structure AB/AC can be reliably computed in this case (see Figure 3(b)).

Theorem 1 is called "parallax-based shape constraint," proved in [12], and it is noted that this constraint directly relates the relative projective structure of two points to their parallax displacements alone: no camera parameters, in particular the epipole (FOE), which is difficult to calculate accurately [27–29]. This is different from the traditional methods that use the two parallax vectors to recover the epipole and then use the magnitudes and distances of the points from the computed epipole to estimate their relative projective structure. The benefit of the constraint (5) is that it provides this information directly from the positions and parallax vectors of the two points, without the need to go through the computation of the epipole [12].

Theorem 2. *Given the planar-parallax displacement vectors of two points that belong to the background static scene over view 1, view 2, and view 3, the following constraint must be satisfied:*

$$\frac{\left(\vec{J}_j^1\right)^T \left(\Delta p_\Pi^1\right)_\perp}{\left(\vec{J}_r^1\right)^T \left(\Delta p_\Pi^1\right)_\perp^j} - \frac{\left(\vec{J}_j^2\right)^T \left(\Delta p_\Pi^2\right)_\perp}{\left(\vec{J}_r^2\right)^T \left(\Delta p_\Pi^2\right)_\perp} = 0, \qquad (6)$$

where $\vec{J}_j^1$, $\vec{J}_r^1$ are the parallax displacement vectors of the two points between the reference plane and view 1. $\vec{J}_j^2$, $\vec{J}_r^2$ are the parallax vectors between the reference plane and view 2, and Δp_Π^1, Δp_Π^2 are the corresponding distances between the warped points [12].

In the case of the parallax-based shape constraint, the parallax-based rigidity constraint (Theorem 2) relates the parallax vectors of the pairs of points over three views without referring to any camera parameters. However, the parallax-based rigidity constraint assumes that the reference plane is consistent across three views. This assumption is not valid sometimes, since the interframe homographies are automatically estimated and the reference planes may correspond to different parts of the scene.

5. Parallax-Based Multiplanar Constraint

In this work, we propose a novel three-view constraint, which is called the "parallax-based multiplanar constraint." This constraint is capable of detecting the motion object that the epipolar constraint cannot detect without a fixed reference plane across three views.

5.1. Description of the Parallax-Based Multiplanar Constraint

Theorem 3. *The image points p_j^1 and p_j^2 given in view 1 and view 2, which are projected by the 3D point P which belongs to the background, must all satisfy the following constraint:*

$$\begin{bmatrix} p_j^2 & \dfrac{\gamma_j^2}{\gamma_r^2} \end{bmatrix} \mathcal{N}_{4\times 4} \begin{bmatrix} p_j^1 \\ \dfrac{\gamma_j^1}{\gamma_r^1} \end{bmatrix} = 0, \qquad (7)$$

where γ_j^1/γ_r^1 is the relative projective structure for view 1 to view 2 and γ_j^2/γ_r^2 is the relative projective structure for view 2 to view 3. $\mathcal{N}$ is a 4×4 matrix (proof: see Appendix A).

Theorem 3 is called "parallax-based multiplanar constraint." The parallax-based multiplanar constraint represents a constraint for the same point in the background by their relative 3D projective structure. This constraint can detect the moving objects from the moving camera without a fixed reference plane across three views. Its degenerate case is modified from the surface degradation to the line degradation (it is discussed in Section 5.2).
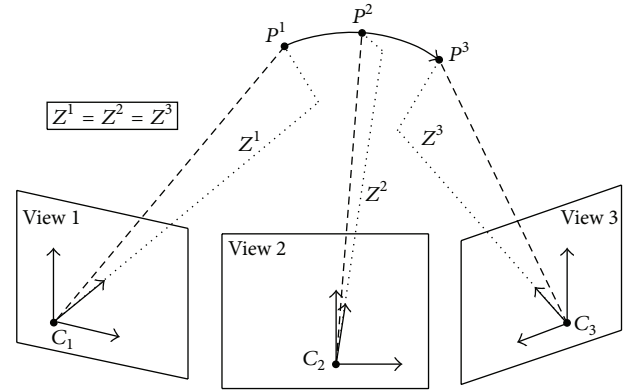

FIGURE 4: Degenerate case for parallax-based multiplanar constraint.

5.2. Degradation of the Parallax-Based Multiplanar Constraint. The parallax-based multiplanar constraint uses the relative 3D projective structure from three views to detect the moving objects. This constraint is capable of detecting most of the degenerate cases mentioned in this paper. However, there still exists a degenerate case that cannot be detected.

Result 4. There is a 3D moving point *P* and its *Z* axis coordinates in the view 1, view 2 and view 3 are equal. The parallax-based multiplanar constraint cannot detect this moving point (proof: see Appendix B).

Figure 4 shows the degenerate case for parallax-based multiplanar constraint. Fortunately, these cases happen much less frequently in reality, because the proportional relationship is not easily satisfied.

6. Application of the Parallax-Based Multiplanar Constraint

In this section, we present some implementation details of a detecting and tracking moving objects system based on the parallax-based multiplanar constraint. As shown in Figure 5, the system is built as a pipeline of five stages: feature point matching, plane segmentation, dense optical flow, object extraction, and spatiotemporal tracking.

This system starts with the feature point matching. Then, the homography parameters and the parallax-based multiplanar constraint parameters are estimated by the feature points matching. We can get the plane residual image which is composed of the pixels which are not satisfied for the homography constraint. The motion field of the binary of the plane residual image can be obtained by the dense optical flow. The parallax-based multiplanar constraint can distinguish the parallax pixels and motion pixels from the plane residual image. Finally, the 2D motion pixels obtained from each frame are linked into motion trajectories by a spatiotemporal tracking algorithm.

The Kanade-Lucas-Tomasi (KLT) feature tracking [30–32] is applied to extract and track feature points in the image sequence $\{I_{t+i} \mid i = -\Delta, \ldots, 0, \ldots, \Delta\}$. Δ is the temporal window size.
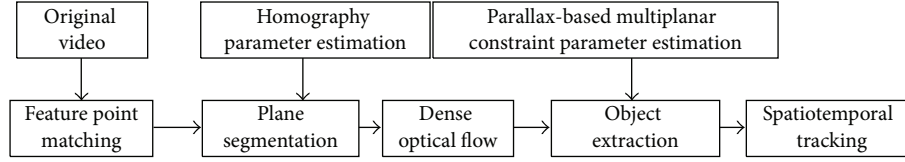
Figure 5: Pipeline of detecting the motion object.

The homography parameters can be estimated by the method described in [1]. I_{t+i} can be warped to I_t by the homography matrix. Then, after estimating the background model [33, 34] (we use the single gauss algorithm in this work), we obtain the binary image of the plane residual image which is composed by the pixels with intensity differences larger than threshold Th_{hom}.

We chose three images ($I_{t+\tau}$, I_t, and $I_{t-\tau}$, (τ is the time interval)) from the image sequence and estimate parallax-based multiplanar constraint parameters by the corresponding feature points. The parallax-based multiplanar constraint parameters are estimated by the similar method of estimating the fundamental matrix [1]. $\mathcal{N}$ is obtained by singular value decomposition. The random sample consensus (RANSAC) scheme is a common choice, which finds a solution with the largest inlier support [1].

The motion field of the pixel in the binary image of the plane residual image can be acquired by the dense optical flow in [31, 32]. We define an algebraic error function through (7):

$$ d_{\text{parallax}} = \left| \left[p_j^2 \quad \dfrac{\gamma_j^2}{\gamma_r^2} \right] \mathcal{N}_{4\times 4} \begin{bmatrix} p_j^1 \\ \gamma_j^1 \\ \dfrac{\gamma_j^1}{\gamma_r^1} \end{bmatrix} \right|. \tag{8} $$

When $d_{\text{parallax}} > \text{th}_{\text{para}}$, this pixel is in the motion region. On the contrary ($d_{\text{parallax}} \leq \text{th}_{\text{para}}$), the pixel is in parallax region. th_{para} is a parallax threshold. So we can extract the motion object from the plane residual image and get the motion binary image.

The motion binary images are further refined by standard morphological operations such as erosion and dilation. The connected pixels are grouped into compact motion regions, whereas scattered pixels are removed. The tracking step takes image appearance, 2D motion vectors, and motion likelihood of these regions as its observation and link similar regions into object trajectories. Since the object tracking is not the focus of this paper, interested readers can refer to [35, 36].

7. Experiment and Analysis

In this section, we present the experimental results obtained by a number of video sequences. In all the video sequences, the camera undergoes general rotation and translation. Both the qualitative and quantitative results demonstrate the effectiveness and robustness of our method.

7.1. Qualitative Evaluation. There are five video sequences that have been adopted to qualitatively demonstrate the effectiveness and robustness of the parallax-based multiplanar constraint.

In Figure 6, we show the segmenting results of a video sequence which is captured in the laboratory. We use checkerboard pattern that only has black and white checks to compose the background of the video and a cylindrical object as a motion object. So, we can call this video "chessboard." This background can ensure that there are enough feature points (Harris corners are accepted in this paper) in it. The video is captured by a moving gray camera; the resolution of the video is 315×235; the frame frequency is 25 fps. In this paper, the parameters are $\Delta = 40$, $\tau = 5$, $\text{Th}_{\text{hom}} = 0.2$, and $\text{th}_{\text{para}} = 0.75$. We show the three-frame images (#148, #153, and #158) in the video sequence which are shown in Figures 6(a), 6(b), and 6(c) and the red points are defined as the reference points. The camera translates from the left to right. In this video sequence the reference plane is the checkerboard. There are two static objects as the parallax regions. After computing by 2D registration [1], the parallax and motion regions are obtained from the plane residual image which is shown in Figure 6(d). As can be seen in Figure 6(d), the two parallax regions are clear. Figure 6(e) is the residual image of the parallax-based multiplanar constraint. The intensity of the motion region is greater than the other region whether including or not the parallax region. In Figure 6(f), it is the binary result of the residual image of the parallax-based multiplanar constraint and shown that the parallax regions (the two static objects) are eliminated finally.

The second video sequence is the experiment video of [12] and is named "car 1." Its resolution is 320×240; the frame frequency is 25 fps. In this paper, the parameters are $\Delta = 12$, $\tau = 3$, $\text{Th}_{\text{hom}} = 0.31$, and $\text{th}_{\text{para}} = 0.68$. Figure 7(a) is the original image (#17) and the red points are defined as the reference points. In this sequence the camera is in motion (translating from the left to right), inducing parallax motion of different magnitudes on the house, road, and road sign. The car moves independently from the left to right. Figure 7(b) is the plane residual image and Figure 7(c) is the binary result of the plane residual image. Because the car is followed by a moving camera in the same direction, the surface degradation of the epipolar constraint happened and is shown in Figure 7(d). Figure 7(e) is the residual of the parallax-based multiplanar constraint computed over three frames. The final binary result is shown in Figure 7(f). From Figure 7, the parallax-based multiplanar constraint modifies the surface degradation to the line degradation to segment
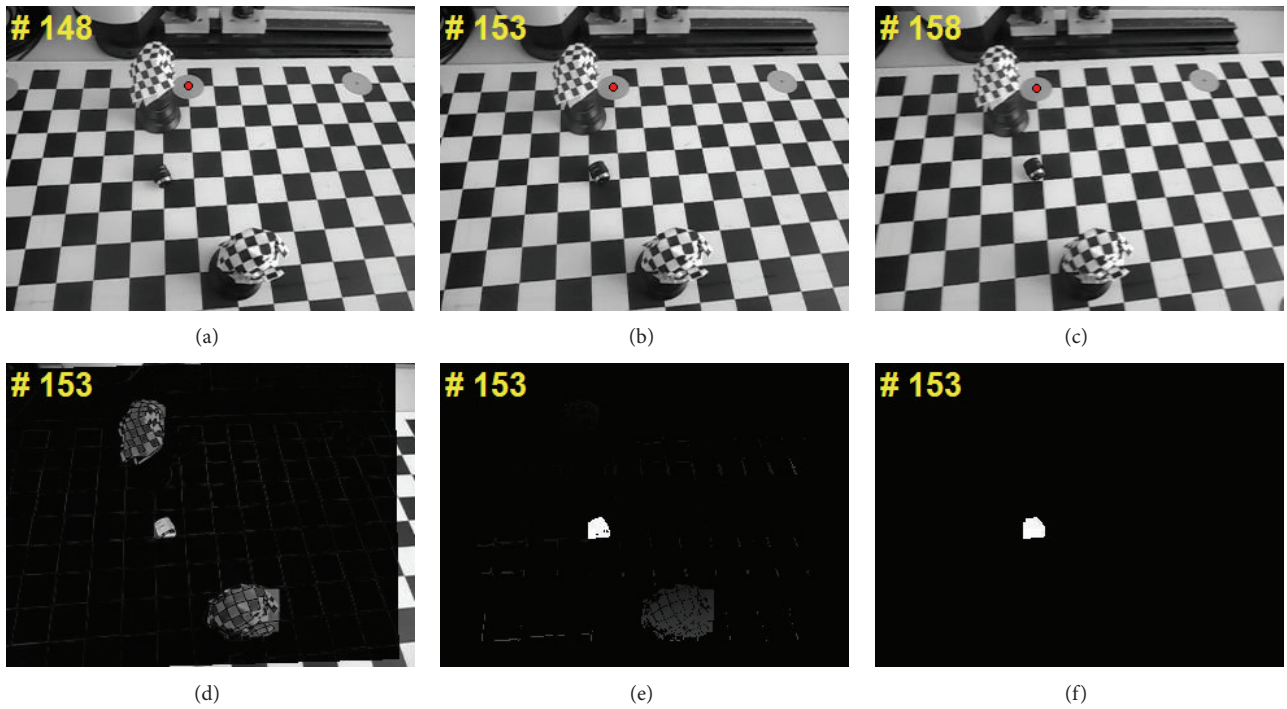
FIGURE 6: Motion segmentation result of the "chessboard." (a) Original image of frame 148. (b) Original image of frame 153. (c) Original image of frame 158. (d) Plane residual image. (e) Residual image of the parallax-based multiplanar constraint. (f) Binary result of the parallax-based multiplanar constraint.
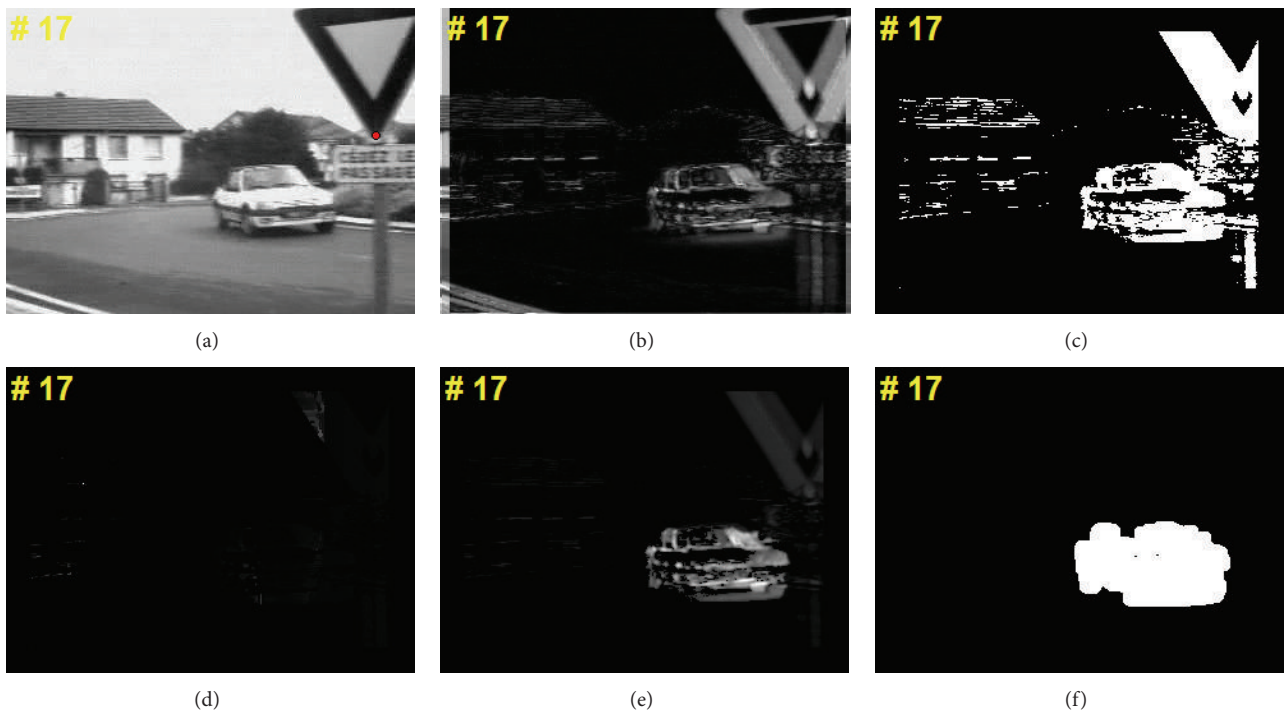

FIGURE 7: Motion segmentation result of "car 1." (a) Original image of frame 17. (b) Plane residual image. (c) Binary result of the plane residual image. (d) Residual image of the epipolar constraint. (e) Residual image of the parallax-based multiplanar constraint. (f) Binary result of the parallax-based multiplanar constraint.
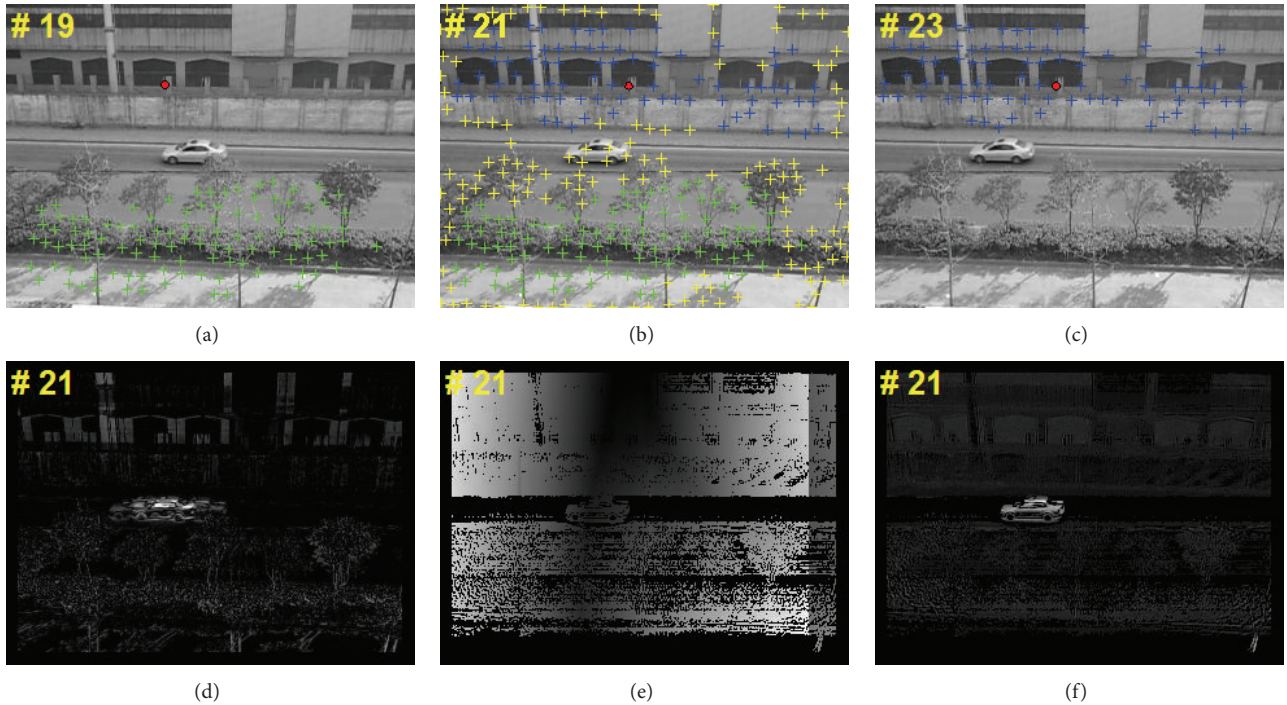
(a)   (b)   (c)

(d)   (e)   (f)

FIGURE 8: Motion segmentation result of "car 2." (a) Original image of frame 19. (b) Original image of frame 21. (c) Original image of frame 23. (d) Plane residual image. (e) Residual image of the parallax-based rigidity constraint. (f) Residual image of the parallax-based multiplanar constraint.

motion objects followed by a moving camera in the same direction.

In Figure 8, it is shown that a camera moves from the left to right and there is a car that moves from the right to left. We can call it "car 2." This video is captured by a gray camera. Its resolution is 315×235; the frame frequency is 25 fps. In this paper, the parameters are $\Delta = 40$, $\tau = 2$, $Th_{hom} = 0.2$, and $th_{para} = 0.8$. In this video sequence, three-frame images (#19, #21, and #23) are shown in Figures 8(a), 8(b), and 8(c) and the red points are defined as the reference points. In Figures 8(a) and 8(b), the green points are the corner points which are the inner points for reference plane between frames 19 and 21. In Figures 8(b) and 8(c), the blue points are the corner points which are the inner points for reference plane between frames 21 and 23. So, the reference plane is changed from frame 19 to frame 23. Figure 8(d) is the plane residual image. The motion regions cannot be segmented from the residual image of the parallax-based rigidity constraint, which is shown in Figure 8(e), because of the change of the reference plane. Figure 8(f) is the residual image of the parallax-based multiplanar constraint. From Figure 8, the parallax-based multiplanar constraint can obtain a better effect compared with the parallax-based rigidity constraint, because it does not need a fixed reference plane over three frames.

Figure 9 is an infrared video acquired from the VIVID dataset. Its resolution is 310×246; the frame frequency is 30 fps. In this paper, the parameters are $\Delta = 30$, $\tau = 3$, $Th_{hom} = 0.2$, and $th_{para} = 0.82$. The camera is in unmanned aerial vehicle. There are three cars moving on the road. So, we can call it "cars 1." The building is considered as the parallax

region. The first row is the original images from 71 to 77 and the red points are defined as the reference points. The second row is the plane residual images. The third row is the residual of the parallax-based multiplanar constraint. The final binary results are shown in the fourth row. We demonstrate the potential of the parallax-based multiplanar constraint applied to the motion segmentation problems using the Berkeley motion segmentation dataset in Figure 10. In this video, there is a car moving on the road and the camera moves from the right to left. It is called "car 4." Its resolution is 320×240; the frame frequency is 30 fps. In this paper, the parameters are $\Delta = 30$, $\tau = 2$, $Th_{hom} = 0.32$, and $th_{para} = 0.74$. The first row is the original images from 13 to 17 and the red points are defined as the reference points. The second row is the plane residual images. The third row is the residual of the parallax-based multiplanar constraint. The final binary results are shown in the fourth row.

From all of the above experiments, we can know that the parallax-based multiplanar constraint can segment the motion regions from the "moving" background. First of all, compared with the homography constraint, the parallax-based multiplanar constraint can segment the parallax regions and the motion regions. Secondly, the parallax-based multiplanar constraint modifies the surface degradation of the epipolar constraint to the line degradation and can detect the motion object which follows the direction of the camera move. Thirdly, in the process of motion segmentation, the parallax-based multiplanar constraint does not need a fixed reference plane across three views. Therefore, this method can
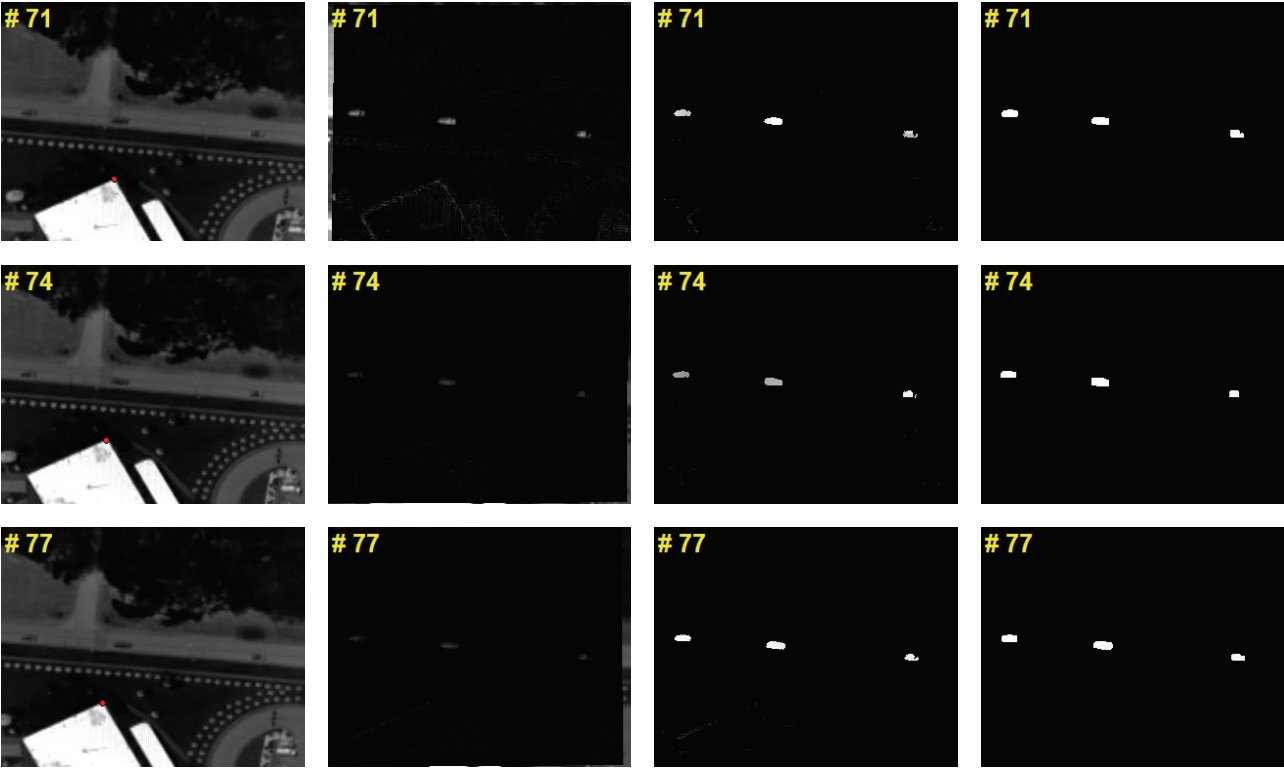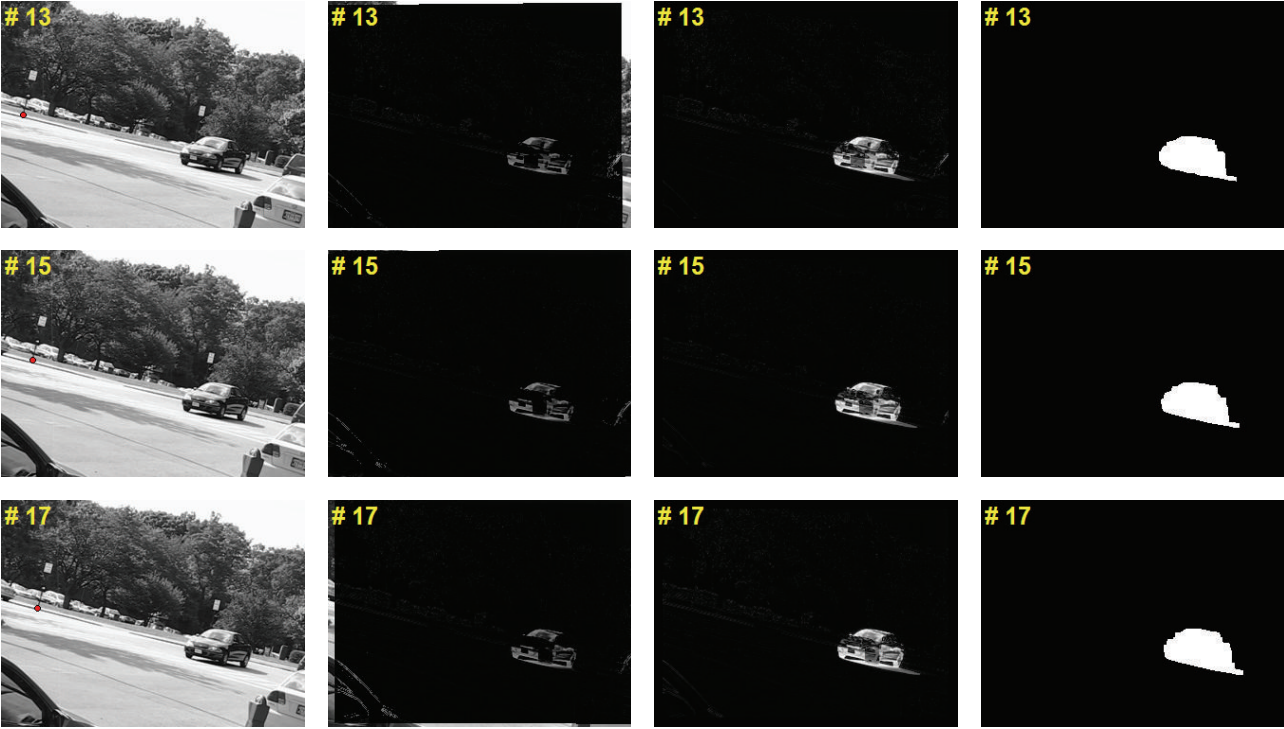
FIGURE 9: Motion segmentation result of "cars 1."
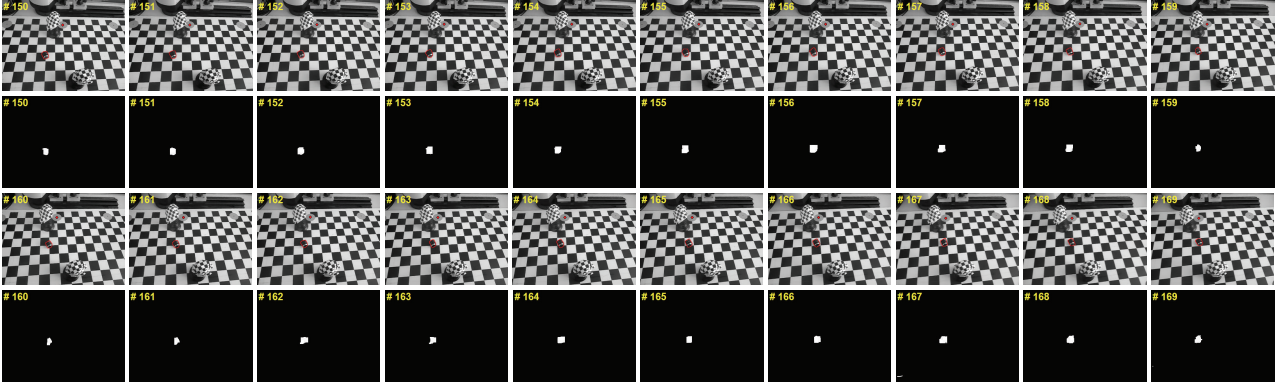

FIGURE 10: Motion segmentation result of "car 3."

FIGURE 11: The ground-truth data and the motion segmentation results of the parallax-based multiplanar constraint for the "chessboard" video.
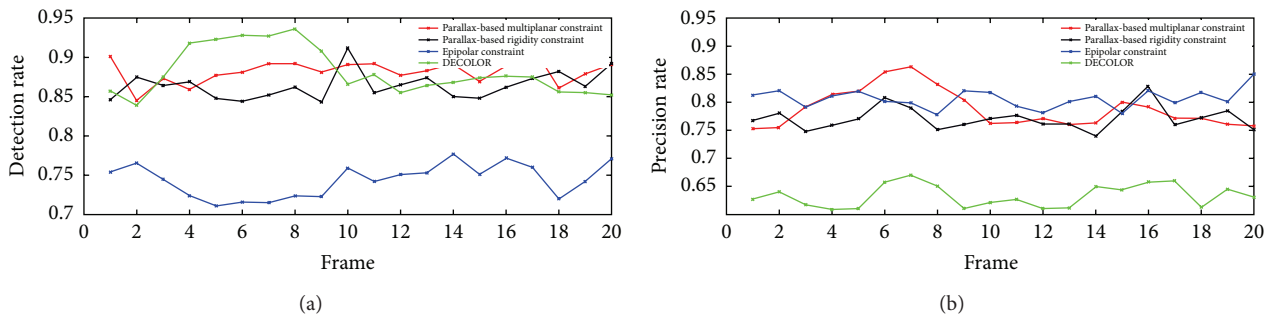

(a)

(b)

FIGURE 12: Quantitative evaluation results for "chessboard" video. (a) Curve of the detection rate. (b) Curve of the precision rate.

effectively extract the motion object from a moving camera and this camera is uncalibrated.

7.2. Quantitative Evaluation. In order to quantitatively evaluate the performance of our system, we have manually labeled the ground-truth data on the above video sequences. The ground-truth data refer to a number of 2D polygons in each video frame, which approximate the contour of motion regions. For the "chessboard" and "car 2" videos, there are 20 frames labeled in different parts.

Based on the ground-truth and detected motion mask images, we define two area-based metrics to evaluate our method [37]. Let Φ_g^t denote the set of pixels that belong to the ground-truth motion regions in frame t and let Φ_d^t denote the set of the actually detected pixels in frame t. We define a detection rate to evaluate how many detected pixels lie in the ground-truth motion regions as

$$R(t) = \frac{N\left(\Phi_d^t \cap \Phi_g^t\right)}{N\left(\Phi_g^t\right)} \tag{9}$$

and a precision rate to evaluate how many detected pixels are indeed motion pixels as

$$P(t) = 1 - \frac{N\left(\Phi_d^t \cap \overline{\Phi}_g^t\right)}{N\left(\Phi_d^t\right)}, \tag{10}$$

where $\overline{\Phi}$ is the complement set of Φ and $N(\Phi)$ is the number of pixels within Φ. In this, $R(t) \in [0,1]$ and $P(t) \in [0,1]$. The higher both measures are, the better the performance of motion segmentation becomes.

The detection rate and the precision rate measures are computed over the labeled video frames to evaluate the performance of our motion segmentation method. For the "chessboard" and "car 1" videos, we evaluate four moving segmentation methods: epipolar constraint [1], parallax-based rigidity constraint [12], detecting contiguous outliers in the low-rank representation (DECOLOR) [4], and our method.

The first line and the third line of Figure 11 are the ground-truth data; the second line and fourth line are the motion segmentation results of the parallax-based multiplanar constraint for the "chessboard" video. The red points are defined as the reference points.

Let us quantitatively compare the performance of the methods based on the curves of the detection rate and precision rate. The detection rate of the epipolar constraint is low compared with the other methods in Figure 12(a) for the surface degradation. The DECOLOR method is based on the homography constraint so that the parallax regions and the motion regions are considered as "motion objects." Its precision rate is lower than the other methods in Figure 12(b). In Figure 12, the parallax-based multiplanar overcomes the surface degradation of the epipolar constraint. it can get the better effect in the detection rate and the precision rate.
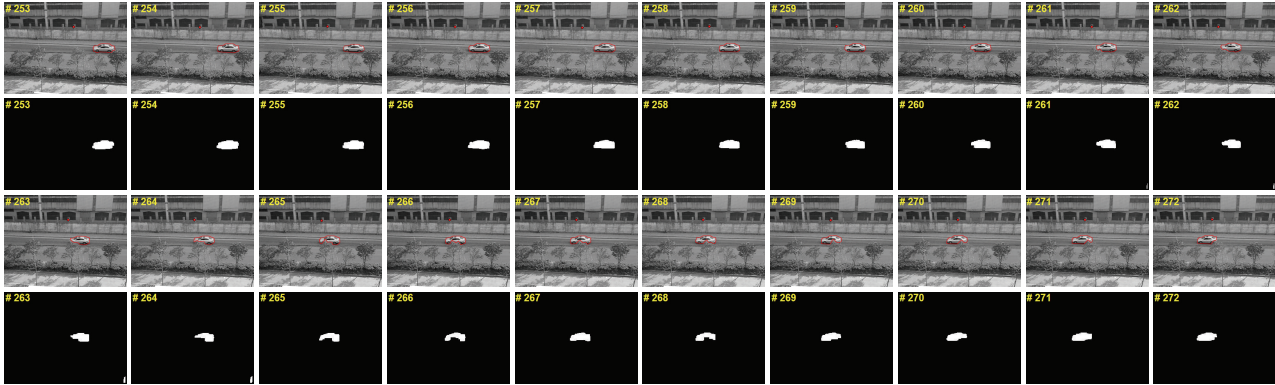
FIGURE 13: The ground-truth data and the motion segmentation results of the parallax-based multiplanar constraint for "car 2" video.
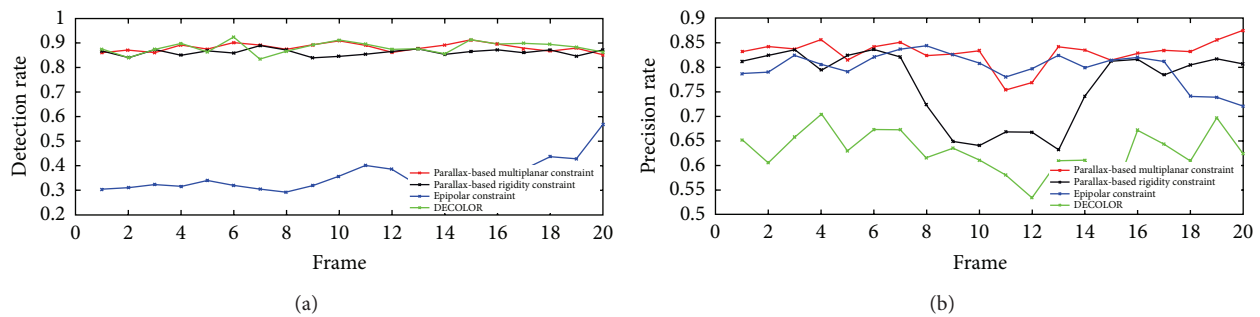

(a)

(b)

FIGURE 14: Quantitative evaluation results for "car 2" video. (a) Curve of the detection rate. (b) Curve of the precision rate.

Figure 13 is the ground-truth data and the motion segmentation results of the parallax-based multiplanar constraint for "car 2" video which is similar to Figure 11.

In Figure 13, when the reference planar is changed from frame 262 to frame 264, there are a lot of false alarms detected by the parallax-based rigidity constraint and DECOLOR which are shown in Figure 14(b). In contrast, the parallax-based multiplanar constraint segments the motion objects without a fixed reference plane, so it performs better in precision rate.

7.3. Parameter Selection. There are a few parameters that are found to be critical for system performance.

The first important measure is the temporal window size Δ. This parameter is used by the homograph image registration to get the plane background. Δ relates to the frame frequency and the size of camera motion. If Δ is set too small, the detect rate may decline. On the contrary, if it is set too big, the overlap region is too small to not detect the motion object and the false alarm probability may increase for the accumulated errors. Δ is proportional to the frame frequency and inversely proportional to the size of camera motion.

The second one, that is, the time interval τ, is used for the estimation of the parallax-based multiplanar constraint parameters. τ also relates to the frame frequency and the size of camera motion. If the difference of continuous image is rather small, τ needs to be increased for a stable estimation of the parallax-based multiplanar constraint parameters.

The third parameter is the homograph threshold Th_{hom}. Th_{hom} is set at low value to make sure that there are enough pixels to compute the d_{parallax}. This threshold needs to be adjusted to different scene configurations in order to include all the possible motion pixels and enough parallax pixels as well. However, if Th_{hom} is set too small, the run time may increase.

The fourth parameter is the parallax threshold th_{para}. This parameter is used to segment the parallax distance d_{parallax} to detect the motion objects. th_{para} relates to the time interval τ and is proportional to it.

8. Conclusion

We have presented a novel method for detecting moving objects in video sequences captured from moving cameras. It uses the multiview geometric constraints for motion detection in three or multiple views. Moreover, the parallax-based multiplanar constraint this paper proposed overcomes the problem of the previous geometry constraint and does not require the reference plane is constant across the multiple views and modifies the surface degradation of the epipolar constraint to the line degradation. It can detect the motion objects followed by a moving camera in the same direction. The experimental results demonstrate the effectiveness and robustness of our approach.

There are several doable directions for future work to be carried out. An appropriate reference point can be found for computing the parallax. If the camera projection matrices

are known or obtained by the self-calibration techniques [1], then both the static background and the moving objects can be reconstructed and aligned together in the 3D Euclidean space.

Appendices

A. Derivation of the Parallax-Based Multiplanar Constraint

In this appendix, we prove Theorem 3; we derive (7).

Let $P_j = (X_j, Y_j, Z_j)$ be a 3D static point and its 3D coordinates in view 1, view 2, and view 3 are expressed as $P_j^1 = (X_j^1, Y_j^1, Z_j^1)$, $P_j^2 = (X_j^2, Y_j^2, Z_j^2)$, and $P_j^3 = (X_j^3, Y_j^3, Z_j^3)$. There is another 3D static point $P_r = (X_r, Y_r, Z_r)$ as the reference point and its 3D coordinates are expressed in three views $P_r^1 = (X_r^1, Y_r^1, Z_r^1)$, $P_r^2 = (X_r^2, Y_r^2, Z_r^2)$, and $P_r^3 = (X_r^3, Y_r^3, Z_r^3)$. p_j^1, p_j^2, and p_j^3 of the 3D point P_j are the homogeneous image coordinates in image 1, image 2, and image 3 and p_r^1, p_r^2, and p_r^3 of the 3D point P_r are the homogeneous image coordinates in image 1, image 2, and image 3, respectively.

From Section 4, for view 1 and view 2, we know that the 3D projective structure of point P_j and the 3D projective structure of point P_r are

$$\gamma_j^1 = \frac{H_j^1}{Z_j^1}; \qquad \gamma_r^1 = \frac{H_r^1}{Z_r^1}. \tag{A.1}$$

Because point P_r is reference point, the 3D projective structure of point γ_r^1 is an invariant for all image points. We can define $\gamma_r^1 = \alpha_1$ and α_1 is a constant factor for the other points.

From (A.1), we can know that

$$\frac{\gamma_j^1}{\gamma_r^1} = \frac{H_j^1}{Z_j^1} \frac{Z_r^1}{H_r^1} = \frac{H_j^1}{\alpha_1 Z_j^1}. \tag{A.2}$$

For the 3D point P_j^1,

$$H_j^1 = v_1^T P_j^1 - 1, \tag{A.3}$$

where v_1 is the normal vector of plane Π scaled by $1/d_\Pi$. d_Π is the perpendicular distance from the camera center of view 1 to the reference plane Π.

Substituting (A.3) in (A.2) obtains

$$\frac{\gamma_j^1}{\gamma_r^1} = \frac{v_1^T P_j^1 - 1}{\alpha_1 Z_j^1}. \tag{A.4}$$

The camera model can be represented as

$$Z_j^1 K^{-1} p_j^1 = P_j^1. \tag{A.5}$$

Substituting (A.5) in (A.4) obtains

$$\frac{1}{Z_j^1} = v_1^T K^{-1} p_j^1 - \alpha_1 \frac{\gamma_j^1}{\gamma_r^1}. \tag{A.6}$$

Similarly, we can get

$$\frac{1}{Z_j^2} = v_2^T K^{-1} p_j^2 - \alpha_2 \frac{\gamma_j^2}{\gamma_r^2} \tag{A.7}$$

for view 2 and view 3.

Let $r_{2,1}$ denote the third row of rotation matrix $R_{2,1}$ and let $t_{2,1}$ denote third component of translation vector $T_{2,1}$. The 3D depth of point P_j^1 could be related to that of P_j^2 by extracting the third row in $P_j^2 = R_{2,1} P_j^1 + T_{2,1}$ as

$$Z_j^2 = r_{2,1} P_j^1 + t_{2,1}. \tag{A.8}$$

Substituting (A.5) into (A.8), we have

$$\frac{1}{Z_j^1} = \frac{r_{2,1} K^{-1} p_j^1}{Z_j^2} + \frac{t_{2,1}}{Z_j^1 Z_j^2}. \tag{A.9}$$

Substituting (A.6) and (A.7) into (A.9), we can obtain

$$v_1^T K^{-1} p_j^1 - \alpha_1 \frac{\gamma_j^1}{\gamma_r^1} = \left(v_2^T K^{-1} p_j^2 - \alpha_2 \frac{\gamma_j^2}{\gamma_r^2} \right)$$
$$\cdot \left(\left(r_{2,1} + t_{2,1} v_1^T \right) K^{-1} p_j^1 - t_{2,1} \alpha_1 \frac{\gamma_j^1}{\gamma_r^1} \right). \tag{A.10}$$

By rewriting (A.10), we have

$$\begin{bmatrix} p_j^2 & \dfrac{\gamma_j^2}{\gamma_r^2} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} v_1^T K^{-1} & -\alpha_1 \end{bmatrix} \begin{bmatrix} p_j^1 \\ \dfrac{\gamma_j^1}{\gamma_r^1} \end{bmatrix}$$
$$= \begin{bmatrix} p_j^2 & \dfrac{\gamma_j^2}{\gamma_r^2} \end{bmatrix} \begin{bmatrix} v_2^T K^{-1} \\ \alpha_2 \end{bmatrix} \begin{bmatrix} r_{2,1} + t_{2,1} v_1^T & -t_{2,1} \alpha_1 \end{bmatrix} \begin{bmatrix} p_j^1 \\ \dfrac{\gamma_j^1}{\gamma_r^1} \end{bmatrix}. \tag{A.11}$$

So we can get the parallax-based multiplanar constraint

$$\begin{bmatrix} p_j^2 & \dfrac{\gamma_j^2}{\gamma_r^2} \end{bmatrix} \mathcal{N}_{4 \times 4} \begin{bmatrix} p_j^1 \\ \dfrac{\gamma_j^1}{\gamma_r^1} \end{bmatrix} = 0. \tag{A.12}$$

B. Degradation of the Parallax-Based Multiplanar Constraint

In this appendix, we prove Result 4 by the algebraic approach; we describe the degradation of the parallax-based multiplanar constraint.

Let $P_j^1 = (X_j^1, Y_j^1, Z_j^1)$, $P_j^2 = (X_j^2, Y_j^2, Z_j^2)$, and $P_j^3 = (X_j^3, Y_j^3, Z_j^3)$ denote the 3D corresponding points in the three

views. Assume $Z_j^1 = Z_j^2 = Z_j^3$ and, according to (A.6) and (A.7), we can get

$$v_2^T K^{-1} p_j^2 - \alpha_2 \frac{\gamma_j^2}{\gamma_r^2} = v_1^T K^{-1} p_j^1 - \alpha_1 \frac{\gamma_j^1}{\gamma_r^1}. \tag{B.1}$$

Substituting (B.1) into (A.10), we can eliminate the left polynomial

$$1 = \left(\left(r_{2,1} + t_{2,1} v_1^T \right) K^{-1} p_j^1 - t_{2,1} \alpha_1 \frac{\gamma_j^1}{\gamma_r^1} \right). \tag{B.2}$$

Decomposing (B.2), we can get

$$Z_j^1 = r_{2,1} P_j^1 + t_{2,1}. \tag{B.3}$$

Because $Z_j^1 = Z_j^2 = Z_j^3$, (B.3) is valid in any case. We can derive that the degradation of the parallax-based multiplanar constraint is that the parallax-based multiplanar constraint cannot detect the motion object when the Z-distance of the 3D point in camera coordinate systems at time $\{i \mid i = 1, 2, 3\}$ is equal to ($Z_j^1 = Z_j^2 = Z_j^3$).

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK, 2nd edition, 2003.

[2] S. Ayer and H. S. Sawhney, "Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding," in *Proceedings of the 5th International Conference on Computer Vision*, pp. 777–784, IEEE, June 1995.

[3] S. W. Kim, K. Yun, K. M. Yi, S. J. Kim, and J. Y. Choi, "Detection of moving objects with a moving camera using non-panoramic background model," *Machine Vision and Applications*, vol. 24, no. 5, pp. 1015–1028, 2013.

[4] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 597–610, 2013.

[5] J. Kang, I. Cohen, and G. Medioni, "Continuous tracking within and across camera streams," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. I-267–I-272, IEEE, June 2003.

[6] J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg, "A three-frame algorithm for estimating two-component image motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 886–896, 1992.

[7] T. Darrell and A. Pentland, "Robust estimation of a multi-layered motion representation," in *Proceedings of the IEEE Workshop on Visual Motion*, pp. 173–178, October 1991.

[8] B. Micusik and T. Pajdla, "Estimation of omnidirectional camera model from epipolar geometry," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 485–490, 2003.

[9] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78, no. 1-2, pp. 87–119, 1995.

[10] W. B. Thompson and T. C. Pong, "Detecting moving objects," *International Journal of Computer Vision*, vol. 4, no. 1, pp. 39–57, 1990.

[11] R. Hartley and R. Vidal, "The multibody trifocal tensor: motion segmentation from 3 perspective views," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 1, pp. 769–775, June-July 2004.

[12] M. Irani and P. Anandan, "A unified approach to moving object detection in 2d and 3d scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6, pp. 577–589, 1998.

[13] S. Dey, V. Reilly, I. Saleemi, and M. Shah, "Detection of independently moving objects in non-planar scenes via multi-frame monocular epipolar constraint," in *Computer Vision—ECCV 2012*, Lecture Notes in Computer Science, pp. 860–873, Springer, Berlin, Germany, 2012.

[14] H. Sajid and S.-C. S. Cheung, "Background subtraction under sudden illumination change," in *Proceedings of the 16th IEEE International Workshop on Multimedia Signal Processing (MMSP '14)*, pp. 1–6, Jakarta, Indonesia, September 2014.

[15] D. Szolgay, J. Benois-Pineau, R. Megret, Y. Gaestel, and J.-F. Dartigues, "Detection of moving foreground objects in videos with strong camera motion," *PAA: Pattern Analysis and Applications*, vol. 14, no. 3, pp. 311–328, 2011.

[16] T. S. F. Haines and T. Xiang, "Background subtraction with Dirichlet process mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pp. 670–683, 2014.

[17] H. Zhang, H. Yuan, and J. Li, "Moving object detection in complex background for a moving camera," in *Fifth International Conference on Machine Vision (ICMV 2012): Computer Vision, Image Analysis and Processing*, vol. 8783 of *Proceedings of SPIE*, 2013.

[18] Y. Wan, X. Wang, and H. Hu, "Automatic moving object segmentation for freely moving cameras," *Mathematical Problems in Engineering*, vol. 2014, Article ID 574041, 11 pages, 2014.

[19] S. W. Sun, Y. C. F. Wang, F. Huang, and H. Y. M. Liao, "Moving foreground object detection via robust SIFT trajectories," *Journal of Visual Communication and Image Representation*, vol. 24, no. 3, pp. 232–243, 2013.

[20] Z. Ren, L.-T. Chia, D. Rajan, and S. Gao, "Background subtraction via coherent trajectory decomposition," in *Proceedings of the 21st ACM International Conference on Multimedia (MM '13)*, pp. 545–548, ACM, October 2013.

[21] H. S. Sawhney, Y. Guo, J. Asmuth, and R. Numar, "Independent motion detection in 3D scenes," in *Proceedings of the 7th IEEE*

International Conference on Computer Vision (ICCV '99), pp. 612–619, September 1999.

[22] X. Cui, J. Huang, S. Zhang et al., "Background subtraction using low rank and group sparsity constraints," in *Computer Vision— ECCV 2012*, vol. 7572 of *Lecture Notes in Computer Science*, pp. 612–625, Springer, Berlin, Germany, 2012.

[23] S. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1832–1845, 2010.

[24] P. H. S. Torr and D. W. Murray, "Outlier detection and motion segmentation," in *6th Sensor Fusion*, vol. 2059 of *Proceedings of SPIE*, pp. 432–443, International Society for Optics and Photonics, Boston, Mass, USA, September 1993.

[25] M. I. A. Lourakis, A. A. Argyrost, and S. C. Orphanoudakis, "Independent 3D motion detection using residual parallax normal flow fields," in *Proceedings of the IEEE 6th International Conference on Computer Vision*, pp. 1012–1017, January 1998.

[26] M. Irani and P. Anandan, "Parallax geometry of pairs of points for 3d scene analysis," in *Computer Vision—ECCV'96*, vol. 1064 of *Lecture Notes in Computer Science*, pp. 17–30, Springer, Berlin, Germany, 1996.

[27] R. Kumar, P. Anandan, and K. Hanna, "Direct recovery of shape from multiple views: a parallax based approach," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Conference A: Computer Vision & Image Processing*, vol. 1, pp. 685–688, Jerusalem, Israel, October 1994.

[28] Z. Chen, C. Wu, P. Shen, Y. Liu, and L. Quan, "A robust algorithm to estimate the fundamental matrix," *Pattern Recognition Letters*, vol. 21, no. 9, pp. 851–861, 2000.

[29] T. Migita and T. Shakunaga, "One-dimensional search for reliable epipole estimation," in *Advances in Image and Video Technology*, vol. 4319 of *Lecture Notes in Computer Science*, pp. 1215–1224, Springer, Berlin, Germany, 2006.

[30] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th international joint conference on Artificial intelligence (IJCAI '81)*, vol. 2, pp. 674–679, 1981.

[31] C. Tomasi and T. Kanade, *Detection and Tracking of Point Features*, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pa, USA, 1991.

[32] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '94)*, pp. 593–600, June 1994.

[33] D.-S. Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 827–832, 2005.

[34] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, 2006.

[35] J. Huang, D. Abendschein, V. G. Dâvila-Român, and A. A. Amini, "Spatio-temporal tracking of myocardial deformations with a 4-D B-spline model from tagged MRI," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 957–972, 1999.

[36] C. Zhang, S. C. Chen, M. L. Shyu et al., "Adaptive background learning for vehicle detection and spatio-temporal tracking," in *Proceedings of the Joint 4th International Conference on Information, Communications and Signal Processing, and the 4th Pacific Rim Conference on Multimedia*, vol. 2, pp. 797–801, 2003.

[37] J. C. Nascimento and J. S. Marques, "Performance evaluation of object detection algorithms for video surveillance," *IEEE Transactions on Multimedia*, vol. 8, no. 4, pp. 761–773, 2006.