A Method for Moving Target Detection on a Moving Camera in the Presence of a Strong Parallax

Chen Haixin　Gu Guohua　Qian Weixian　Chen Qian　Xu Fuyuan

（*College of Electric & Optic*，*Nanjing University of Science and Technology*，*Nanjing*，*Jiangsu* 210094，*China*）

Abstract　Based on projective geometry，a relationship between the moving cameras in the presence of a strong parallax is analyzed，and unique constraints are proposed based on surface homography model. Generally，previous works focus on Planar ＋ Parallax or simple geometric constraints such as fundamental matrix，but the degradation cannot be solved. A much more strong constraint is proposed to modify the surface degradation based on fundamental matrix to line degradation based on surface homography，and solves the degradation by modeling，learning and detecting. Unlike previous works，main planar is not needed and degradation is not ignored with no reason，through surface homography model and modeling-learning-detecting framework. Experiments are performed based on actual image data，and the results show that this model can learn the motion of cameras efficiently and be practical on moving target detection with a moving camera with a strong parallax.

Key words　machine vision；moving target detection；surface homography constraints；computer vision；monocular vision；image processing

OCIS codes　100.6890；150.1135；330.4150；330.7326

一种运动相机强视差下运动目标检测方法

陈海欣　顾国华　钱惟贤　陈　钱　徐富元

（南京理工大学电光学院，江苏 南京 210094）

摘要　从多视角几何的角度分析了强视差下基于运动相机的运动目标检测问题，提出了曲面单应模型，给出了多视角下的三维静止场景的约束条件。现在一般的方法是平面加视差或直接用基础矩阵等简单的几何约束，但无法解决约束条件退化的问题。该算法框架提出了比基础矩阵更强的约束条件，将基础矩阵面退化情况改进为曲面单应下的线退化，并根据曲面单应引入了建模、学习、检测的过程，从而解决了约束条件退化的问题。与以往算法不同，该算法不需要主平面，也不需要防止摄像机出现退化的情况。根据实际采集的图像序列进行了分析对比，真实的数据测试表明该算法在保证算法准确性与稳定性的前提下，能快速有效地学习摄像机的运动，极大地提高了运动目标的探测率，解决了约束条件退化的问题。

关键词　机器视觉；运动目标检测；曲面单应约束；计算机视觉；单目视觉；图像处理

中图分类号　TP391　　　**文献标识码**　A　　　**doi**：10.3788/CJL201441.0509002

1　Introduction

Moving target detection on an uncalibrated moving camera is a complex and key technology in the research of computer vision. Because of the camera movement，the three-dimensional（3D）scene and independent moving targets are in motion on the image plane simultaneously. It's difficult to detect the moving targets，as the global motion induced by the perspective projection of the moving camera. At this point，the traditional algorithm can only be limited to the parallax negligible aerial video[1]. However，due to the obvious depth of 3D scene on the handheld，automotive and

other mobile platforms, the strong parallax must be considered and estimated effectively[2-3].

The latest mobile platform moving target detection methods, mostly around the fundamental matrix, add artificial priori constraints specific target, but cease to be effective with a moving monocular camera on the complex background. For general cases, especially for a moving monocular camera, the latest and popular methods applied to moving target detection are based on multiple view geometry such as the fundamental constraints[4], shape constraints[5], and structure constraints[3]. To improve the detection ratio, some priori constraints to specific targets on specific scenes are introduced artificially, such as the flow vector bound (FVB) constraint[6], and the Planar + Parallax induced by the homography matrix[3]. Generally, the degenerate case is ignored with no reason, while there are few constraints of nature innovation[3,5,7].

In order to overcome the limitations of the traditional methods, a fresh new modeling, learning and detecting (MLD) model based on the surface homography constraint (SHC) is proposed. The SHC constraint is for camera motion parameter estimation, and the camera motion model is built by MLD. Degenerate sequences of frames are removed and learning library is updated to estimate a no degradation SHC constraints. Moving targets are detected by the non-degraded SHC.

2　Surface homography constraint

In order to compensate the parallax caused by the moving camera, a SHC constraint is proposed, which is induced by a planar surface based on the multiple view geometry[4]. The degenerate case of SHC is analyzed.

2.1　Induction of surface homography matrix

The homography matrix is derived by a common virtual planar as follows[4]

$$H = K'(A - av^{\mathrm{T}})K^{-1}, \qquad (1)$$

where K and K' are internal parameters of the camera to the specific 3D point. As for common cases, not only for the specific 3D point, the homography matrix is needed to be corrected as follows

$$\boldsymbol{K} = \frac{1}{Z_{\mathrm{c}}}\begin{bmatrix} \alpha_x & 0 & c_x & 0 \\ 0 & \alpha_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \frac{1}{Z_{\mathrm{c}}}M. \qquad (2)$$

The corrected homography matrix is showed in formula (3), and named SH for short.

$$SH = \frac{Z'_{\mathrm{c}}}{Z_{\mathrm{c}}}H = M'\left(A - \frac{an^{\mathrm{T}}}{d}\right)M^{-1}, \qquad (3)$$

$$x'_{\mathrm{w}} = SH \cdot x_{\mathrm{w}}, \qquad (4)$$

where x_{w} is of homogeneous coordinates, x_{h} is of normalized homogeneous coordinates, and x is of non-homogeneous coordinates[4].

To be short and simple as follows:

$$SH = R + t\gamma^{\mathrm{T}}, \qquad (5)$$

where,

$$R = M'AM^{-1}, \quad t = -M'a, \quad \gamma^{\mathrm{T}} = v^{\mathrm{T}}M^{-1}. \quad (6)$$

2.2　Point mapping based on surface homography

As shown in Fig. 1, 3D point X lies on the planar $\pi = (v^{\mathrm{T}},1)^{\mathrm{T}}$, it will satisfy formula (7):

$$v^{\mathrm{T}}X = -1, \qquad (7)$$

expand formula (7) as follows:

$$v^{\mathrm{T}}X = Z_{\mathrm{c}}v^{\mathrm{T}}M^{-1}x_{\mathrm{h}} = Z_{\mathrm{c}}\gamma^{\mathrm{T}}x_{\mathrm{h}} = -1, \qquad (8)$$

let $\kappa = 1/Z_{\mathrm{c}}$ for the derivation:

$$x^{\mathrm{T}}\gamma = -\kappa, \qquad (9)$$

where there exists a relationship of matched pixels in two views as follows:

$$x' = SHx = (R + t\gamma^{\mathrm{T}})x = Rx - t\kappa. \qquad (10)$$

SHM is mamed for short of surface homography mapping, and the mapping relationship of pixels in two views is corresponded to the location of pixels, internal /external parameters of the camera and the depth of the 3D point.
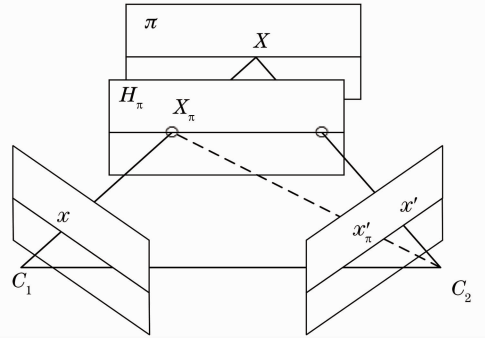

Fig. 1　Plane induced surface homography

2.3　Induction of first constraint of SHC

The SHM representation in formula (10) is indeed a general homography relationship on a surface. If $t = [t_x, t_y, t_z]\mathrm{T}$ is a 3-vector, then one defines a corresponding skew-symmetric matrix $[t]_x$ as follows:

$$[t]_x = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}. \qquad (11)$$

To the matching points $x_1 \leftrightarrow x_2 \leftrightarrow x_3$ in three views, the first constraint of SHC (SHC1) is written in the form as follows:

$$\begin{cases} x_1^T[t_{12}]_x R_{12}x_2 = 0 \\ x_3^T[t_{32}]_x R_{32}x_2 = 0 \end{cases}. \qquad (12)$$

SHC1 is indeed an epipolar constraint equivalent to the fundamental matrix[4], but formula (12) has a much more simple form.

2.4　Induction of second constraint of SHC

In order to overcome the limitations of the epipolar constraint, geometric constraints over more than two views which include the depth information of 3D points, need to be imposed on the matched pixels $x_1 \leftrightarrow x_2 \leftrightarrow x_3$ as follows:

$$\begin{cases} t_{32}^{\mathrm{T}}(x_1 - R_{12}x_2) = -t_{32}^{\mathrm{T}}t_{12}\kappa_2 \\ t_{12}^{\mathrm{T}}(x_3 - R_{32}x_2) = -t_{12}^{\mathrm{T}}t_{32}\kappa_2 \end{cases}. \qquad (13)$$

Therefore, there exists the second constraint of SHC (SHC2) as follows:

$$t_{32}^{\mathrm{T}}x_1 + (t_{12}^{\mathrm{T}}R_{32} - t_{32}^{\mathrm{T}}R_{12})x_2 - t_{12}^{\mathrm{T}}x_3 = 0. \quad (14)$$

2.5 Degenerate cases for SHC
2.5.1 Degenerate cases for SHC1

The SHC1 constraint is a commonly used geometric constraint as the epipolar constraint for moving targets detection in two views， which encapsulates the relative orientation between two cameras[3]. The degenerate case for SHC1 is represented in Fig. 2. As the motion of an uncalibrated camera， any static 3D point P can only move on the planar π， that passes the optical center C_1， C_2 and C_3.

The most serious defect of SHC1 is that the degeneration will always exist no matter what the motion of the camera is in three views.
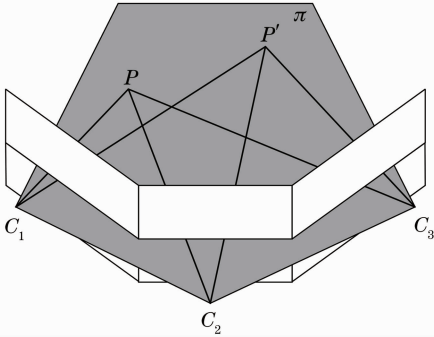

Fig. 2　Degradation of SHC-1

2.5.2 Degenerate cases for SHC2

The SHC2 constraint is a fresh new constraint in three views. The degenerate case for SHC2 is represented in Fig. 3. As the motion of an uncalibrated camera， any static 3D point P can only move on the planar π， that is parallel to the image planar of three views. Only when there exists no motion of yaw or pitch， there exists a degeneration.
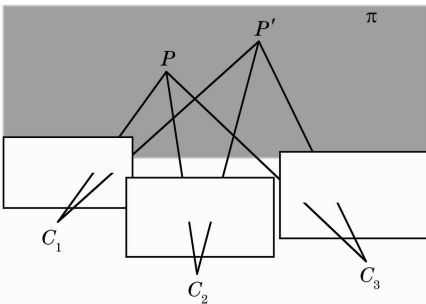

Fig. 3　Degradation of SHC-2

2.5.3 Degenerate cases for SHC

The degenerate case for SHC including SHC1 and SHC2 is represented in Fig. 4. As the motion of an uncalibrated camera， any static 3D point P can only move on the line L on the planar π， when the camera moves without motion of yaw or pitch.

Fortunately， this degeneration happens much less frequently than the whole set of degenerate motion， as only when the camera moves without motion of yaw or pitch. And the MLD process will minimize the risk of degradation.
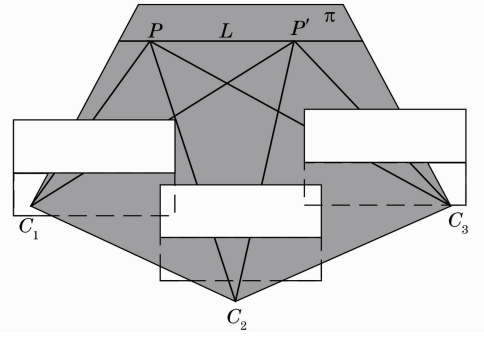

Fig. 4　Degradation of SHC-1 and SHC-2

2.6 Robust estimation of SHC

In order to estimate the surface homography model， the SHC is written as follows：

$$\begin{cases} x_1^{\mathrm{T}}F_{12}x_2 = 0 \\ x_3^{\mathrm{T}}F_{32}x_2 = 0 \\ \gamma_{12}^{\mathrm{T}}x_1 + \gamma_2^{\mathrm{T}}x_2 + \gamma_{32}^{\mathrm{T}}x_3 = 0 \end{cases} \quad (15)$$

Similar to the robust estimation of fundamental matrix in the previous paper[4]， parameters of SHM can be estimated through the normalized algorithm[8-10].

3　Modeling， learning and detecting

In this section， an automatic framework is presented for modeling， learning and detecting in video scenes from uncalibrated moving cameras， inspired by TLD[11]. As shown in Fig. 5， the MLD framework consists of three steps： surface homography modeling， online learning and moving targets detecting. By learning the motion of the camera， the learning library can be updated to refresh the SHC model， and independent moving targets can be filtered out by the learning library.

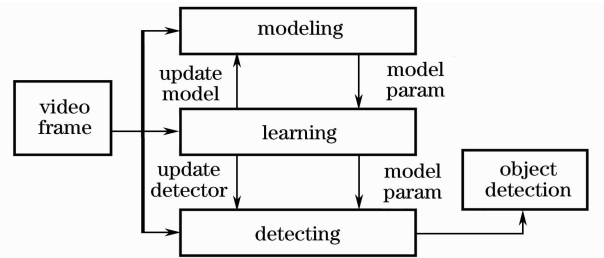The extreme degradation is avoided by the online learning of camera motionin an adaptive strategy[11-12].

Fig. 5　Process of MLD

3.1　Modeling

As shown in formula (15)， matched points in three views are obtained according to kanade lucas tomasi (KLT) algorithm[13-14]， and the history library is updated by robust estimation of SHC.

3.2　Learning

As shown in Fig. 6， an online learning process of camera motion is introduced to avoid the degenerate cases and increase the detecting rate.

According to formulas (12) and (15)， the rotation matrix of cameras is written as follows：

$$R = [t]_x^{-1} F. \tag{16}$$

In order to reduce the nonlinear degrees of freedom, the rotation matrix of cameras is converted to Euler angle representation[4]. It is easy to see that the rotation angle satisfies follow formulas:

Euler angle:

$$\theta = \arccos\left[\frac{1}{2}(R_{11} + R_{22} + R_{33} - 1)\right]. \tag{17}$$

Yaw:

$$e_x = \frac{R_{32} - R_{23}}{2\sin\theta}. \tag{18}$$

Pitching:

$$e_y = \frac{R_{13} - R_{31}}{2\sin\theta}. \tag{19}$$

Roll:

$$e_y = \frac{R_{13} - R_{31}}{2\sin\theta}. \tag{20}$$

As shown in SHC constraints, the necessary condition is that motion of yaw or pitch happens, and it will satisfy $e_x = e_y = 0$. Online learning of history learning library is updated to retain the camera motion of yaw and pitch.
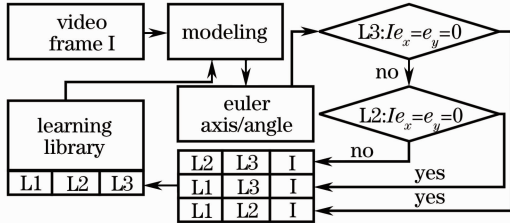

Fig.6　Process of learning

The renovation of the learning library is judged by the using of the motion of yaw and pitch parameters as judged conditions. As shown in Fig.6, the diamond of L_3 is the judged conditions between current frame I and L_3 in the library, and the diamond of L_2 represents the judged conditions between current frame I and L_2 in the
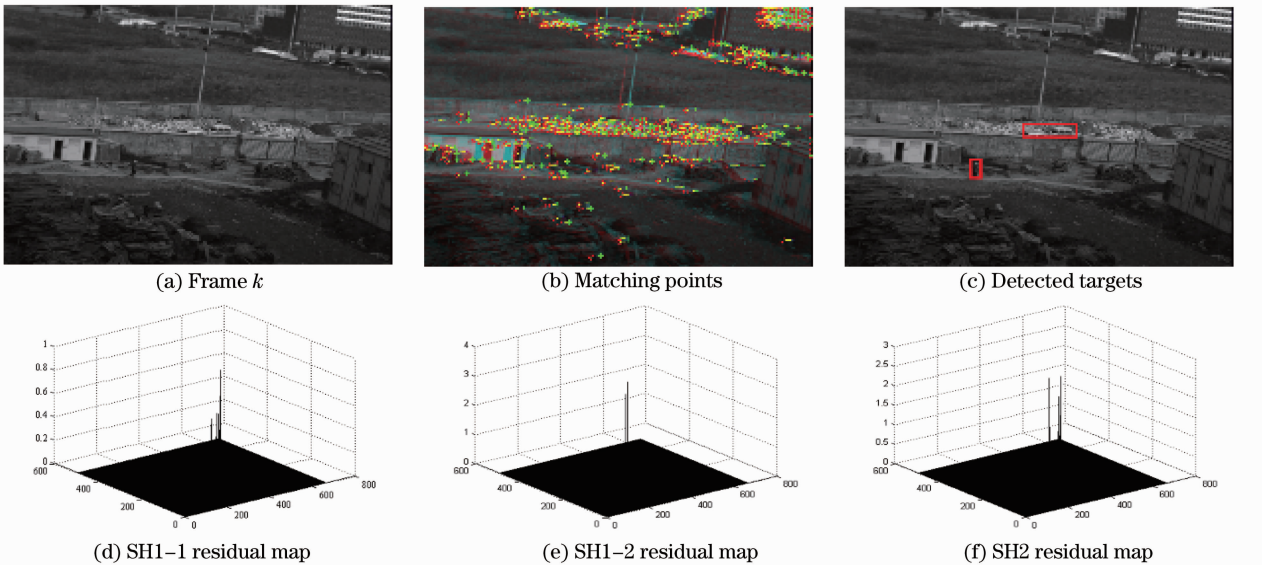
library. Learning library is dynamically updated, and the input image sequences are filtered and with less degeneration.

3.3　Detecting

The moving targets detection on a moving camera is finished by the learning library online updated and the surface homography model, as shown in Fig.7. Rather than the original image sequences, the detecting step bases on the filtered image sequences, and degenerate cases are avoided generally. As object tracking and subsequent processing methods are not the focus of this paper, interested readers are referred to paper[2,11].
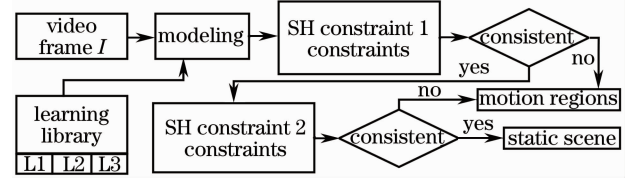

Fig.7　Process of detecting

4　Experimental results

Experimental results obtained by the MLD system on a number of real world video sequences. In all the sequences, the camera undergoes general rotation, translation and zooming. The performance of the MLD of SHC is based on real data analysis. The experiment under the proposed MLD algorithm framework for the actual sequence of image processing, demonstrates the effectiveness and robustness of our approach.

The proposed algorithm has been implemented on a 2 GHz PC with various cameras. Figs.8 – 10 illustrate the moving object segmentation results for the three consecutive frames.

4.1　SHC performance evaluation based on a moving Camera

Figure 8(a)　current frame captured by a moving

(a) Frame k

(b) Matching points

(c) Detected targets

(d) SH1−1 residual map

(e) SH1−2 residual map

(f) SH2 residual map

Fig.8　MLD moving target detection

camera. Figure 8(b) depicts the computed optical flow by KLT. Figures 8(d) − (f) show the computed residual map of SH1-1，SH1-2，SH2. Figure 8(c) represents detected independent moving objects（IMOs）in red rectangle boxes which are a moving vehicle and human. As the motion of the camera is not parallel to the targets，both the SHC1 and SHC2 constraints have no denegation according to the residual maps.

4.2　Evaluation of degenerate cases for SHC based on a moving camera

The MLD process is shown in Fig. 9 when there exists a degenerate case. The history library frame $k-1$ is updated to frame $k+1$ to avoid the latest degenerate case. As shown in Fig. 9(g)，the degeneration of SH1-1 is obvious as the limitations of the fundamental matrix，many independent moving points failed to be detected. The degeneration of SH1-2 fails in Fig. 9(h). In Fig. 9(i)，SH2 has not degenerated and detected moving vehicle and human，without degeneration by updating of MLD.

(a) Frame $k-1$　　　　　　　(b) Frame k　　　　　　　(c) Frame $k+1$

(d) Library $k-1$　　　　　　(e) Library k　　　　　　(f) Library $k+1$

(g) SH1−1 residual map　　　(h) SH1−2 residual map　　　(i) SH2 residual map

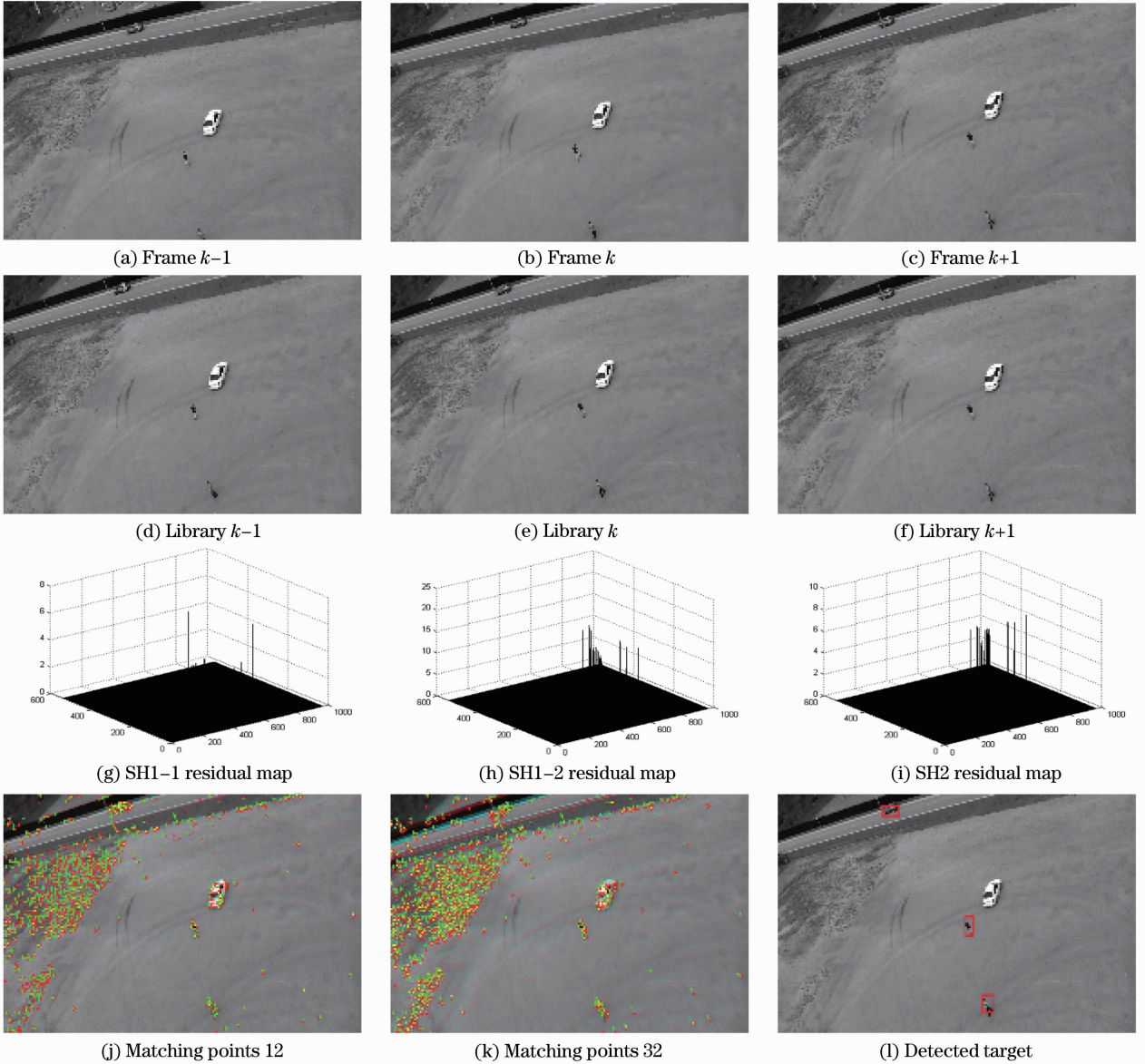(j) Matching points 12　　　　(k) Matching points 32　　　　(l) Detected target

Fig. 9　MLD moving target detection（with a plat ground）

Figure 10 presents a sequence of frames containing very strong parallax due to the varying depth of 3D structures in the scene captured from a handheld camera. A number of vehicles on the road are followed by a handheld camera moving in the same direction，as shown in the three original and matched frames in Figs. 10(a) − (c). This is a typical example of the degenerate motion discussed in Section 2.5.3. In Figs. 10(a) − (c). As one can observe from the residual maps，the insufficiency of the SH1（epipolar constraint）is illustrated in Figs. 10(g) and （h）. The filtered detection result is presented in Figs. 10(k) − (m). The proposed SHC approach successfully filters regions in case of the strong parallax and the degenerate case.
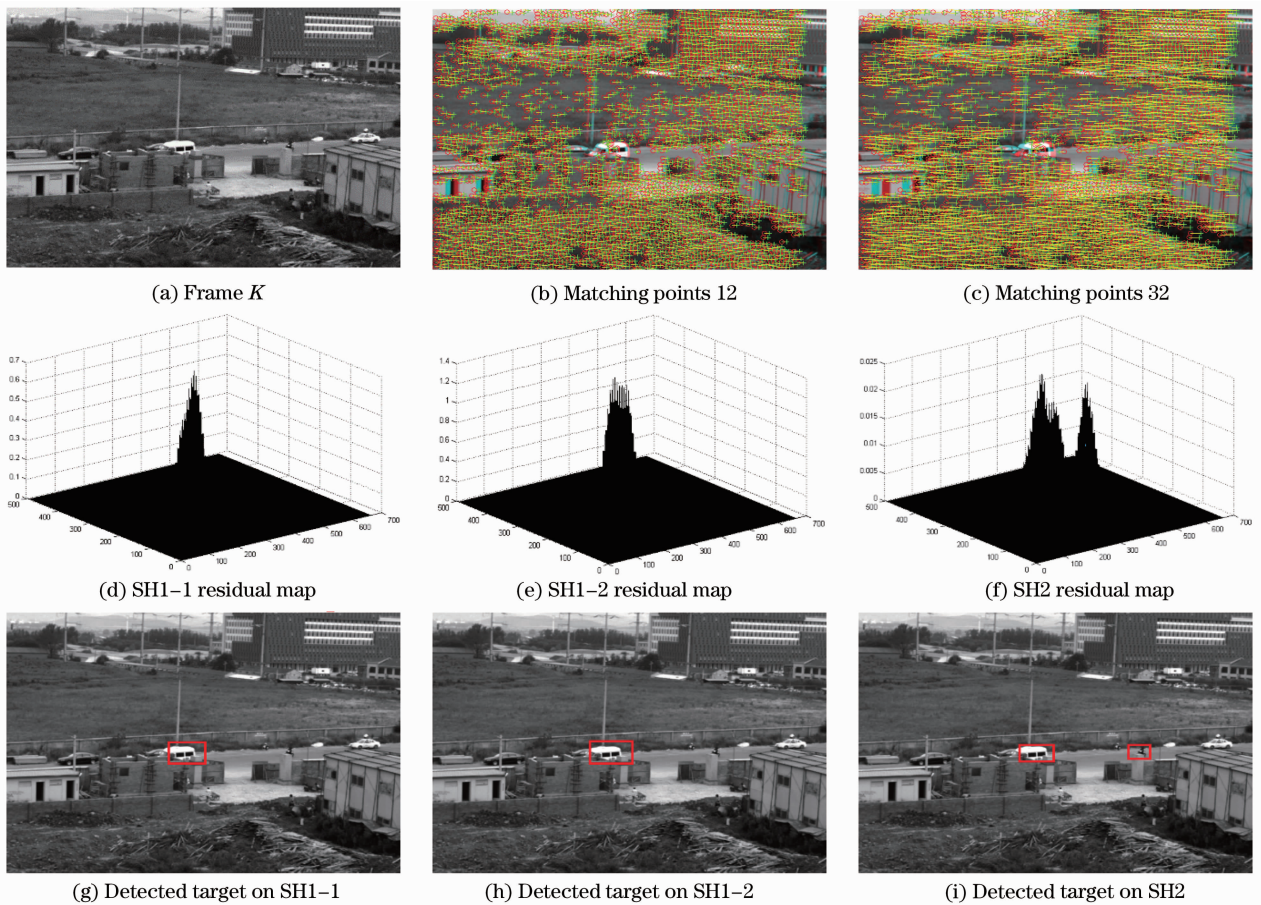
(a) Frame K (b) Matching points 12   (c) Matching points 32

(d) SH1−1 residual map   (e) SH1−2 residual map   (f) SH2 residual map

(g) Detected target on SH1−1   (h) Detected target on SH1−2   (i) Detected target on SH2

Fig. 10　MLD moving target detection（with varying depth of 3D structures）

5　Conclusion

Moving target detection on an uncalibrated moving camera is a complex and key technology in the research of computer vision. A fast method is proposed for moving targets detection on a moving camera in the presence of a strong parallax. A much more strong constraint is proposed to modify the surface degradation based on fundamental matrix to line degradation based on SHC，and solves the degradation by modelling，learning and detecting. Unlike previous works，main planar is not needed and degradation is not ignored with no reason，through SHM and MLD framework. Experiments are performed based on actual image data，and the results show that this model can learn the motion of cameras efficiently and be practical on moving target detection with a moving camera with a strong parallax.

References

1 Eshel R，Moses Y. Homography based multiple camera detection and tracking of people in a dense crowd［C］. IEEE Computer Vision and Pattern Recognition，2008：1 − 8.

2 Yan Zhang，Stephen J Kiselewich，et al.. Robust Moving Object Detection at Distance in the Visible Spectrum and Beyond Using A Moving Camera ［C］. Computer Vision and Pattern Recognition Workshop，2006，1：246 − 252.

3 Yuan C，Medioni G，Kang J，et al.. Detecting motion regions in the presence of a strong parallax from a moving camera by multiview geometric constraints［J］. IEEE Pattern Analysis and Machine Intelligence，2007，29(9)：1627 − 1641.

4 Hartley R，Zisserman A. Multiple View Geometry in Computer Vision［M］. Cambridge：Cambridge University Press，2003.

5 Sawhney H S，Guo Y，Asmuth J，et al.. Independent motion detection in 3D scenes［C］. IEEE Computer Vision，1999，1：612 − 619.

6 Kundu A，Krishna K M，Sivaswamy J. Moving object detection by multi-view geometric techniques from a single camera mounted robot ［C］. IEEE Intelligent Robots and Systems，2009：4306 − 4312.

7 Raguram R，Frahm J M，Pollefeys M. A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus［M］. Computer Vision-ECCV 2008. Berlin Heidelberg：Springer，2008. 500 − 513.

8 Barlow J L. More accuratebidiagonal reduction for computing the singular value decomposition［J］. SIAM Journal on Matrix Analysis and Applications，2002，23(3)：761 − 798.

9 Goldstein A，Fattal R. Video stabilization using epipolar geometry［J］. ACM Transactions on Graphics，2012，31(5)：126.

10 Torr P H S，Murray D W. The development and comparison of robust methods for estimating the fundamental matrix［J］. International J Computer Vision，1997，24(3)：271 − 300.

11 Kalal Z，Mikolajczyk K，Matas J. Forward-backward error：Automatic detection of tracking failures［C］. IEEE Pattern Recognition (ICPR)，2010：2756 − 2759.

12 Kalal Z，Mikolajczyk K，Matas J. Tracking-learning-detection［J］. IEEE Pattern Analysis and Machine Intelligence，2012，34(7)：1409 − 1422.

13 Lee H K，Choi K W，Kong D，et al.. Improved Kanade-Lucas-Tomasi tracker for images with scale changes［C］. IEEE Consumer Electronics (ICCE)，2013：33 − 34.

14 Jiang J，Yilmaz A. Good features to track：A view geometric approach ［ C ］. IEEE Computer Vision Workshops (ICCV Workshops)，2011：72 − 79.