Moving object detection using unstable camera for consumer surveillance systems

Article · January 2013

Namwook Kim
Hanyang University

Monson Hayes
George Mason University

Joonki Paik
Chung-Ang University

Optik 126 (2015) 2436–2441

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Moving object detection using unstable camera for video surveillance systems

Seungwon Lee, Nahyun Kim, Kyungwon Jeong, Kyungju Park, Joonki Paik*

Department of Image, Chung-Ang University, Seoul, South Korea

ARTICLE INFO

Article history:
Received 23 April 2014
Accepted 2 June 2015

Keywords:
Moving object detection
Surveillance system
Camera motion estimation
Unstable camera

ABSTRACT

This paper presents a robust moving object detection method by compensating for motion of an unstable camera. Assuming that global camera motion results in affine transform between two successive frames, local affine motions are separately estimated in multiple pre-specified regions for fast, robust estimation of the global motion. The global camera motion is then estimated by the least squares method using the pre-estimated multiple local affine motions. Given a current frame as the reference, the subsequent frame is registered to the current frame using the estimated global motion. The moving objects are finally detected using difference of Gaussian and non-parametric kernel density estimation from the set of registered three frames. Experimental results show that the proposed method can robustly detect moving objects in unstable imaging environment for intelligent surveillance systems using various types of cameras including pan-tilt-zoom (PTZ) and unmanned aerial vehicle (UAV) cameras.

1. Introduction

Moving object detection is a fundamental problem in computer vision and video processing, and its applications have widely extended to robot vision, human computer interfaces, and intelligent surveillance systems [1–3].

The detection of a moving object can be considered as a low- or intermediate-level processing step for various object recognition and analysis applications. In general, the background subtraction method using background modeling and the motion-based object detection are used to detect moving object [4,5]. Since the background subtraction method assumes that a camera is fixed, it may fail in detecting objects under critical environments containing as camera jitter, noisy sequences, and illumination changes. The multimodal strategies have been presented for overcoming these limitations. To solve the problems of the repetitive motion like rippling water, flickering monitors, and periodic jitter of a camera, the mixture of Gaussians-based background modeling methods were proposed by estimating the Gaussian distribution in the pixel level [6,7]. However, Gaussian assumption for the pixel intensity distribution does not always hold, and as a result the detection of an object fails if the background is updated when the global camera motion occurs. To overcome limitations of the parametric methods, a non-parametric approach to background modeling was proposed [8]. However, this method is memory- and time-consuming because of the computation of the average of all kernels centered at each training sample for every pixel.

The background modeling-based object detection method is not suitable for video sequences acquired by dynamic cameras such as pan-tilt-zoom (PTZ) and unmanned aerial vehicle (UAV) cameras because the background has to be remodeled at each frame. Therefore, the camera motion compensation and a proper object detection methods are needed for such unstable imaging environment. To overcome the problem of the unstable camera, Lee et al. [9] proposed the local motion-based moving object detection method after the global motion compensation. However, the global motion estimation using elastic registration has the high complexity because the affine motion is estimated in the entire image using multiple iterations. In addition, the error of the inaccurately estimated global motion and the corresponding motion compensation result in performance degradation in estimating local motions.

This paper presents a robust moving object detection algorithm in the unstable camera environment as shown in Fig. 1. The proposed method consists of the global motion estimation step for registering frames and the moving object detection step using the proposed frame difference scheme and non-parametric kernel density estimation. In the global motion estimation step, under assumption that a camera's global motion is modeled as an affine transformation, the affine matrix is estimated between successive

* Corresponding author. Tel.: +82 2 8205300.
E-mail addresses: superlsw@gmail.com (S. Lee), skgus1433@gmail.com (N. Kim), fjqmsa@gmail.com (K. Jeong), kjpark@cau.ac.kr (K. Park), paikj@cau.ac.kr (J. Paik).

http://dx.doi.org/10.1016/j.ijleo.2015.06.003

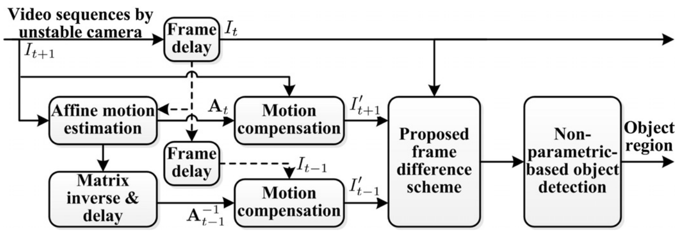

Fig. 1. The proposed moving object segmentation framework.

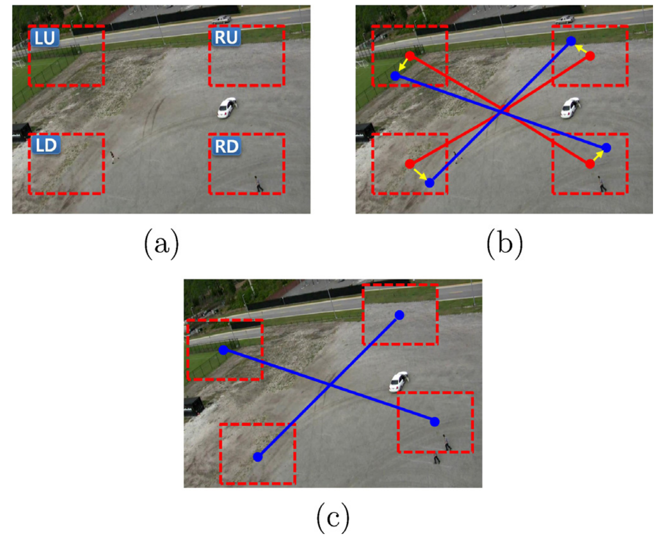

Fig. 2. Concept of the proposed camera motion estimation; (a) local motion estimation at selected four regions (LU: left up, LD: left down, RD: right down, and RU: right up), (b) global motion estimation (red points: center of the each selected region, blue points: moved positions), and (c) refining the global motion estimation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

frames using the estimated global motion. Local affine motions are estimated using the affine model-based optical flow method at the selected four regions for fast global motion estimation [10]. The global camera motion is then estimated by the least squares method using the estimated four local affine motions. The refinement step is finally performed for increasing the accuracy of the global motion estimation.

In the moving object detection step, the subsequent frame is registered to the current frame using the estimated global motion, and the previous frame is also registered to the current frame using the inverse of the estimated global motion of the previous frame. Frame difference between the current and registered adjacent frames may result in the erroneous detection of an object because of the error in estimating affine motions and motion compensated interpolation. To overcome these problems, an average image is generated using the current and two adjacent registered frames. The average frame and two adjacent frames are lowpass filtered by a suitable Gaussian function for robust detection of object by removing spurious patterns.

By subtracting two adjacent lowpass filtered frames from the current frame, two differences of Gaussian (DoG) images are obtained, one of which has an opposite sign to the other in the moving object region [11]. A moving object can be detected using the property of the DoG and the relationship between temporally adjacent frames. Since the DoG has holes in the center of the object, and is still sensitive to noise, the proposed method uses non-parametric kernel density estimation for filling the holes and removing noisy artifacts. The proposed global affine motion estimation algorithm is computationally efficient because of the fusion of multiple local motions by averaging four locally estimated affine motions. Since the existing background modeling methods are not suitable for dynamic imaging system using PTZ and UAV cameras, the proposed method uses only two adjacent frames without background modeling, which can also reduce the memory space and computational cost.

Experimental results show that the proposed method can robustly detect moving objects in unstable imaging environment for intelligent surveillance systems using various types of cameras including PTZ and UAV cameras.

The paper is organized as follows. Section 2 presents the moving object detection in unstable camera environment. Section 3 summarizes experimental results, and Section 4 concludes the paper.

2. Moving object detection in unstable imaging environment

If camera motion occurs, the conventional motion-based object detection method fails to detect objects because of the global camera motion. To overcome this problem, the proposed method uses the global motion estimation and compensation for robust object detection in unstable imaging environment.

2.1. Camera motion estimation

For registering two consecutive frames with irregular camera motions, a motion estimation method used in a digital image stabilization system is adopted. Based on the assumption that the global camera motion is modeled by an affine transformation, the proposed camera motion estimation algorithm estimates the affine matrix between current and subsequent frames.

Fig. 2 illustrates the concept of the proposed camera motion estimation algorithm. After estimating local affine motions in four sub-regions as shown in Fig. 2(a), the global motion is determined using the Gaussian pyramid-based coarse-to-fine approach.

The general affine transform is expressed as follows:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \mathbf{H}_A \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_5 \\ a_3 & a_4 & a_6 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \tag{1}$$

where $\mathbf{H}_A$ represents the affine camera model matrix; a_1, a_2, a_3, and a_4 represent a set of linear affine parameters; and a_5 and a_6 represent the translation parameters. Local affine motions are estimated using the affine model-based optical flow method proposed by Periaswamy [10]. More specifically, if $I(x, y, t)$ is the region in the current image frame, and $I(x, y, t+1)$ is the same region in the subsequent frame, the motion estimation problem assumes that the motion between these two images can be modeled as an affine transform. To estimate the affine matrix, the following quadratic error function is minimized

$$E(\mathbf{a}) = \sum_{x,y \in \Omega} [I(x, y, t) - I(a_1x + a_2y + a_5, a_3x + a_4y + a_6, t+1)]^2, \tag{2}$$

where $\mathbf{a} = (a_1, a_2, a_3, a_4, a_5, a_6)^T$, and Ω represents the support of region.

Since $E(\mathbf{a})$ is a nonlinear function of $\mathbf{a}$, it should be linearized using the first-order Taylor series expansion to explicitly solve the minimization problem for $\mathbf{a}$ as

$$E(\mathbf{a}) \approx \sum_{x,y\in\Omega} [I(x,y,t) - \{I(x,y,t) + (a_1x + a_2y + a_5)I_x(x,y,t) + (a_3x + a_4y + a_6)I_y(x,y,t) - I_t(x,y,t)\}]^2, \quad (3)$$

where I_x and I_y are spatial derivatives of I in the horizontal and vertical directions, respectively, and I_t is the temporal derivatives of I. The approximation error can be simplified by rearranging the terms as

$$E(\mathbf{a}) = \sum_{x,y\in\Omega} \{I_t - \mathbf{c}^T\mathbf{a}\}^2, \quad (4)$$

where $\mathbf{c} = [xI_x, yI_x, xI_y, yI_y, I_x, I_y]^T$. Because $E(\mathbf{a})$ is a quadratic function of $\mathbf{a}$, a closed-form solution for $\mathbf{a}$ that minimizes this error can be found by differentiating with respect to $\mathbf{a}$ and setting the result equal to zero. The result is a set of linear equations that need to be solved for $\mathbf{a}$ as

$$\mathbf{Ca} = I_t\mathbf{c}, \quad (5)$$

where $\mathbf{C} = \mathbf{c}^T\mathbf{c}$ is a 6×6 matrix. Although there is no guarantee that matrix $\mathbf{C}$ is invertible, $\mathbf{C}$ will generally be invertible if the region Ω is sufficiently large and the image has sufficient amount of details. A more accurate estimation of the actual error function can be performed by using the Newton–Raphson type iteration. The estimated transformation is applied to the source image and a new transformation is estimated between the newly translated source and target images at each iteration. Also, since the required spatial and temporal derivatives have finite support, which restricts the length of estimable. For this reason, a coarse-to-fine approach is adopted in order to deal with larger motions. A Gaussian pyramid is built for both source and target images. Six parameters are used to translate the source image in the next level of the pyramid.

The affine transform is performed based on the central axis of the image and the central point is subject to only translational motion regardless of the affine transformation. The central point of each region is defined as $\mathbf{x}_i = [x_i, y_i, 1]^T$ and $\mathbf{X}_{t+1} = [\mathbf{x}_{LU}, \mathbf{x}_{LD}, \mathbf{x}_{RD}, \mathbf{x}_{RU}]^T$. Let $\mathbf{T}_i = [a_5^i, a_6^i, 0]^T$ represent the translation vector, then the translated point $\mathbf{X}'_{t+1}$ is defined as

$$\mathbf{X}'_{t+1} = \begin{bmatrix} \mathbf{x}_{LU}^T \\ \mathbf{x}_{LD}^T \\ \mathbf{x}_{RD}^T \\ \mathbf{x}_{RU}^T \end{bmatrix} + \begin{bmatrix} \mathbf{T}_{LU}^T \\ \mathbf{T}_{LD}^T \\ \mathbf{T}_{RD}^T \\ \mathbf{T}_{RU}^T \end{bmatrix}. \quad (6)$$

When $\mathbf{X}_{t+1}$ moves to $\mathbf{X}'_{t+1}$ by the global affine transformation defined by $\mathbf{H}_A$, which $\mathbf{H}_A$ is calculated using the least squares method as

$$\mathbf{H}_A = [\mathbf{X}'^T_{t+1}\mathbf{X}'_{t+1}]^{-1}\mathbf{X}'^T_{t+1}\mathbf{X}_{t+1}. \quad (7)$$

For more accurate estimation of $\mathbf{H}_A$ the refined motion estimation step is performed using the registered subsequent frame using $\mathbf{H}_A$. $\mathbf{H}'_A$ is estimated using the corresponding four regions in the current and the registered subsequent frames and the solution of (5) without using the Gaussian pyramid. The final affine matrix $\hat{\mathbf{H}}_A$ is computed as

$$\hat{\mathbf{H}}_A = \mathbf{H}_A \cdot \mathbf{H}'_A. \quad (8)$$

Fig. 3 shows the result of the affine transformation of the input image using the estimated camera motion. The difference image between Figs. 3(a) and (b) is shown in Fig. 3(c) with a large camera motion. On the other hand the difference image between Fig. 3(a) and registered subsequent image using the refined affine matrix is shown in Fig. 3(e), which has a smaller difference than Fig. 3(c).

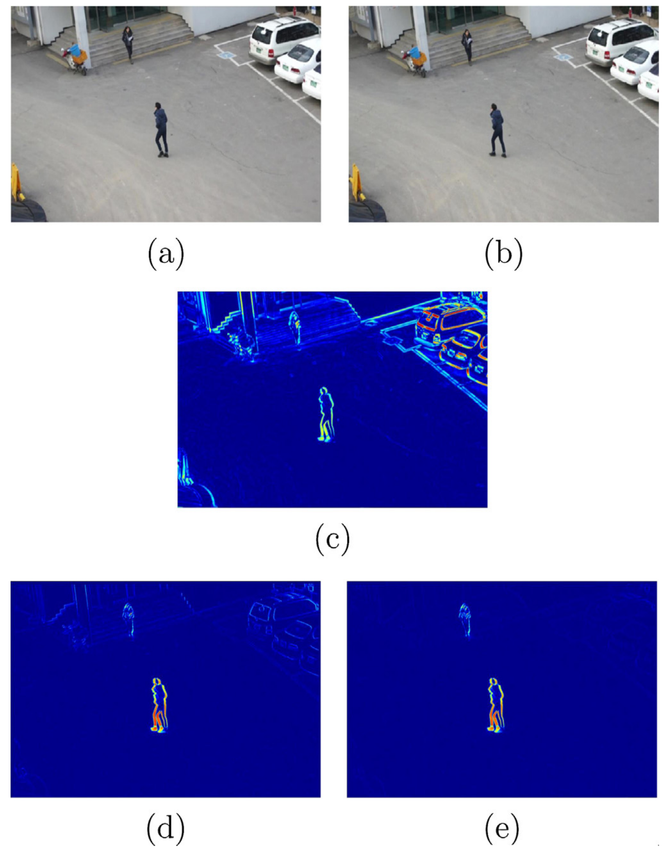

Fig. 3. Results of the camera motion estimation using affine model; (a) current frame, (b) subsequence frame, (c) difference image between (a) and (b), (d) difference image between (a) and the compensated subsequence frame using affine model, and (e) difference image between (a) and the compensated subsequent image using refined affine model.

2.2. Moving object detection using non-parametric estimation

In the moving object detection step, the moving object region is detected using the proposed frame difference scheme based on the DoG and non-parametric kernel density estimation [11]. The subsequent frame is first registered to the current frame using the estimated global motion, and the previous frame is registered to the current frame using the inverse of the estimated global motion of the previous frame.

An average image is generated using the current frame and two registered adjacent (the previous and the subsequent) frames. The average image is defined as

$$M(x,y) = (I'_{t-1}(x,y) + I_t(x,y) + I'_{t+1}(x,y))/3, \quad (9)$$

where I'_{t-1} and I'_{t+1} represent the registered previous and subsequent frames, respectively.

To illustrate the proposed DoG scheme, Fig. 4 shows results of detecting the one-dimensional (1D) moving signal using the proposed method. Let S_{t-1} be the previous signal, and S_{t+1} be the subsequent signal. Fig. 4(a) shows the moving signal region using the signal difference between S_{t-1} and S_{t+1}. If the signals move according to the global motion as well as local motions, the moving region detected from the registered S_{t-1} and S_{t+1} has the large error because of the inaccurate motion estimation and the corresponding motion compensation. If the moving signal region is detected using the blurred signals by Gaussian kernel for reducing the noise, the region is roughly detected as shown in Fig. 4(b). To solve this problem, the moving signal region is detected using the proposed DoG scheme. The moving signal region is finally detected by

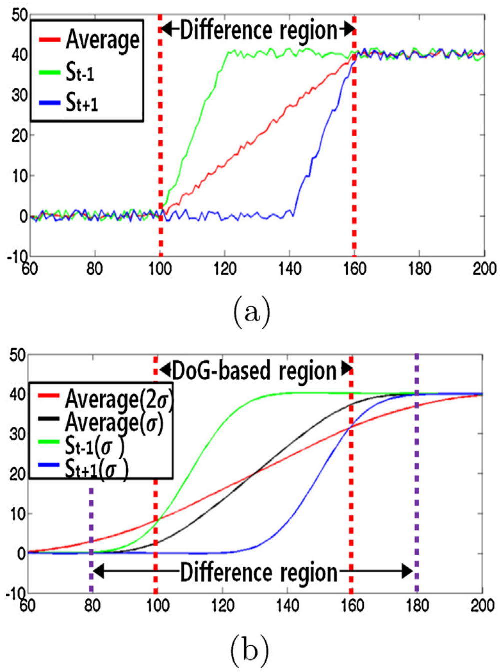

Fig. 4. Result of moving signal region detection using the DoG scheme, (a) using original signal and (b) using Gaussian blurred signal.

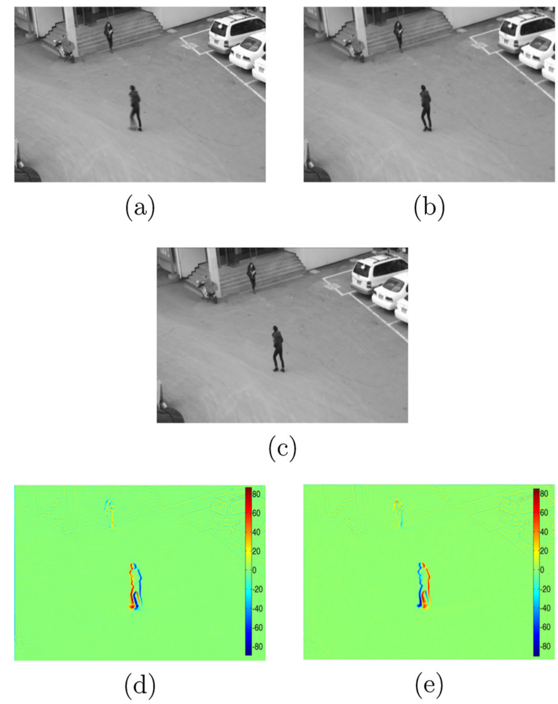

Fig. 5. Result of DoG; (a) average image, (b) compensated previous frame, (c) compensated subsequent frame, (d) DoG between (a) and (b), and (e) DoG between (a) and (c).

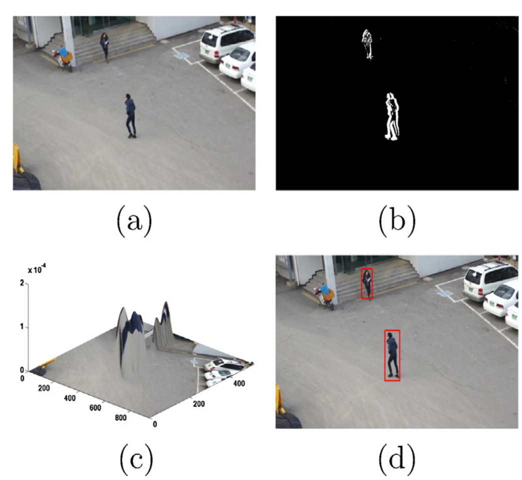

Fig. 6. Result of the moving object detection; (a) current frame, (b) result of the moving region detection, (c) $p(x, y|O)$, and (d) result of the moving object detection (red: object region). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

finding the region of the opposite sign between two DoGs as shown in Fig. 4(b). Note that the standard deviation of the Gaussian kernel for the average signal is twice larger than that of the two registered signals.

The difference between the average image and two registered adjacent frames results in DoGs such as

$$D_1(x, y) = G(x, y, 2\sigma) * M(x, y) - G(x, y, \sigma) * I'_{t-1}(x, y), \tag{10}$$

$$D_2(x, y) = G(x, y, 2\sigma) * M(x, y) - G(x, y, \sigma) * I'_{t+1}(x, y), \tag{11}$$

where $G(x, y, \sigma)$ represents the Gaussian kernel. DoG using the average image can reduce the noise and detect the moving object region at the same time. The resulting DoG is shown in Fig. 5. The region without any moving objects has the similar values between D_1 and D_2. On the other hand, the generated two difference images have a property that signs of the difference values are opposite to each other in the moving object region as shown in Fig. 5(d) and (e).

The moving object region is detected using the property of opposite sign between D_1 and D_2 defined as

$$O(x, y) = \begin{cases} 1, & \begin{array}{l} [sign(D_1(x, y)) \neq sign(D_2(x, y))] \\ \cap D_1(x, y) > T_D \end{array} \\ 0, & \text{otherwise} \end{cases}, \tag{12}$$

where $sign(\cdot)$ represents the sign of the value, T_D denotes a pre-specified threshold for noise reduction. However, the frame difference cannot avoid the hole artifact in the center of the object and noise sensitivity. The moving object region is finally detected using non-parametric kernel density estimation (KDE) to overcome these problems. The probability functions $p(x, y|O)$ is estimated by Gaussian-based kernel density estimation as

$$p(x, y|O) = \frac{1}{N_O} \sum_{i=1}^{N_O} \frac{1}{\sqrt{2\pi h}} e^{\frac{-(x-x_i)^2-(y-y_i)^2}{2h}}, \tag{13}$$

where N_O represents the total number of $O(x, y) = 1$, x_i and y_i denote the coordinates of $O(x, y) = 1$. h represents the kernel bandwidth. The moving object detection is then defined as

$$L(x, y) = \begin{cases} 1, & p(x, y|O) > T_L \\ 0, & \text{otherwise} \end{cases}, \tag{14}$$

where T_L represents a threshold which is adaptively determined using Otsu's threshold selection method [12]. The detected object region is labeled using the eight-directional labeling method and the small object region is removed.

The result of the proposed moving object detection is shown in Fig. 6. The moving object region is detected using (12) as shown in

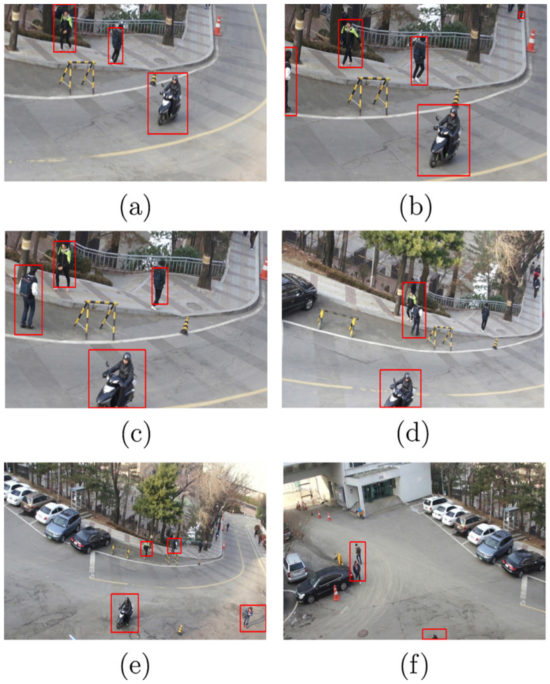

Fig. 7. Experimental results of the proposed moving object detection method in the unstable camera environment.

Fig. 6(b), and the final object region is then detected using KDE and labeling methods as shown in Fig. 6(c) and (d).

3. Experimental results

To evaluate performance of the proposed method for detecting moving objects under the unstable imaging environment, the proposed algorithm was tested under two scenarios including an in-house handheld camera and the UCF aerial action data set [13]. The test sequences have the resolution of 960 × 540 at the rate of

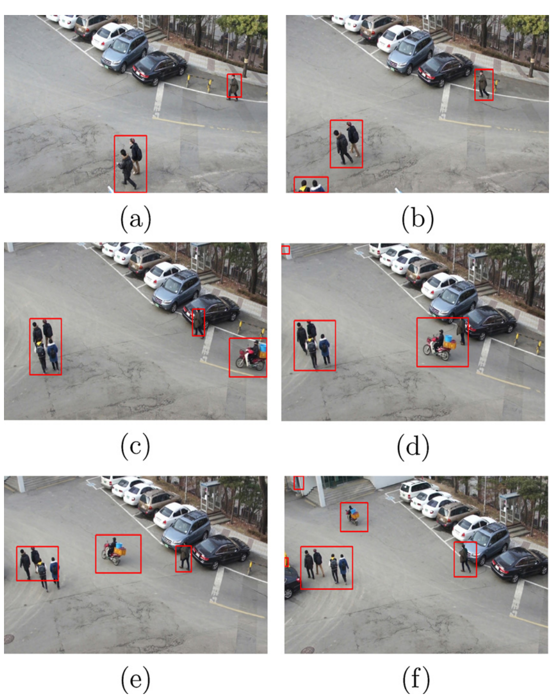

Fig. 8. Experimental results of the proposed moving object detection method in the unstable camera environment.

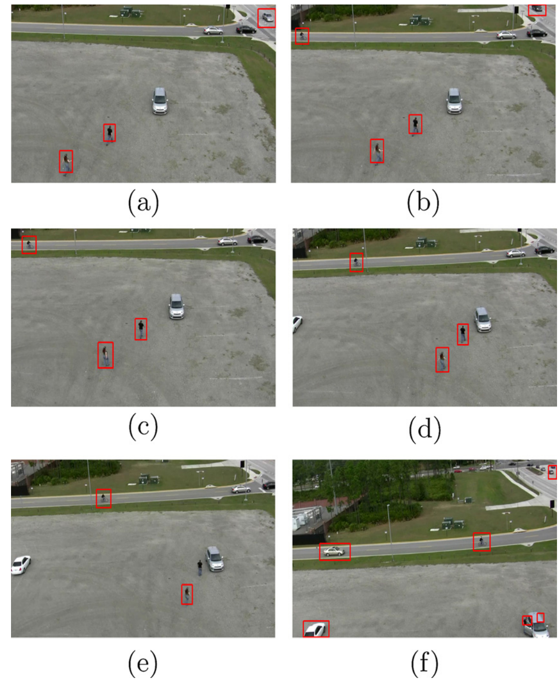

Fig. 9. Experimental results of the proposed moving object detection method in the UAV camera environment.

24 frames per second. A personal computer with 3.07 GHz CPU and 4 GB RAM is used for implementing the proposed method.

In the experiment, the proposed method set the size of four local regions as 61 × 61 for camera motion estimation and the positions of these regions were selected in the set of images except the previously detected object regions. Gaussian kernel of size 15 × 15 was used, $\sigma = 0.5$, and $T_D = 3$.

Fig. 7 shows the experimental result in the unstable camera with zoom and the drastic camera motion. The proposed method accurately detects the object regions even with zooming-in and zout as shown in Fig. 7(a)–(f).

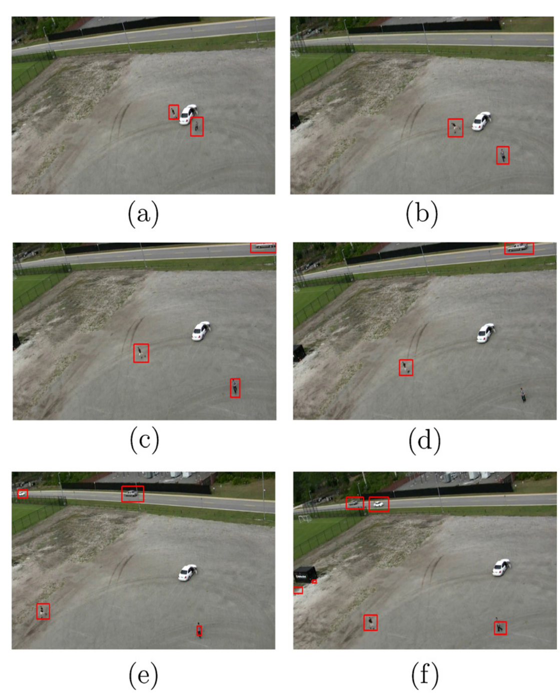

Fig. 10. Experimental results of the proposed moving object detection method in the UAV camera environment.

Fig. 8 shows the experimental result of the proposed moving object detection method in the unstable camera environment. The proposed method can robustly detect the moving object while the camera was moving. However, the erroneous object regions were detected outside of the image as shown in Fig. 8(d) and (f). These errors are caused by the lens distortion with the increased error of the affine transform towards the outside of the image.

The experimental results in the UAV camera environments are shown in Figs. 9 and 10. The conventional background modeling-based method is not suitable to detect objects in the UAV camera environments. On the other hand, the proposed method can successfully detect the object as shown in Figs. 9 and 10.

4. Conclusion

A novel moving object detection method is presented by compensating the camera motion in the unstable camera environment. The proposed method consists of two steps: (i) global motion estimation for motion compensation and (ii) moving object detection using difference of Gaussian (DoG)-based frame difference and non-parametric kernel density estimation.

Background modeling assumes that the camera must be kept static, so it cannot deal with global camera motion environment. Since the proposed method estimates the kernel density instead of background modeling, it is more suitable to detect objects in unstable imaging environment. It can also reduce the memory space and computational cost for background modeling.

Experimental results show that the proposed method can robustly detect moving objects in the unstable camera environment and be able to be embedded in an image signal processing chip for high-level image processing functions in high-definition video surveillance systems using various types of cameras including PTZ and UAV cameras.

Acknowledgments

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (B0101-15-0525, Development of global multi-target tracking and event prediction techniques based on real-time large-scale video analysis), the Technology Innovation Program (Development of Smart Video/Audio Surveillance SoC & Core Component for Onsite Decision Security System) under Grant 10047788, and Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program.

References

[1] S. Lee, J. Paik, Simultaneous object tracking and depth estimation using color shifting property of a multiple color-filter aperture camera, in: Proceedings of IEEE Int. Conf. Acoustics, Speech, Signal Processing, 2011, pp. 1401–1404.
[2] T. Zin, P. Tin, H. Hama, T. Toriu, Unattended object intelligent analyzer for consumer video surveillance, IEEE Trans. Consum. Electron. 57 (2) (2011) 549–557.
[3] J. Kim, D. Yeom, Y. Joo, Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems, IEEE Trans. Consum. Electron. 57 (3) (2011) 1165–1170.
[4] R. Zhang, S. Zhang, S. Yu, Moving objects detection method based on brightness distortion and chromaticity distortion, IEEE Trans. Consum. Electron. 53 (3) (2007) 1177–1185.
[5] S. Lee, J. Lee, E. Chon, M. Hayes, J. Paik, Moving object segmentation using motion orientation histogram in adaptively partitioned blocks for consumer surveillance system, in: Proceedings of Int. Conf. Consumer Electronics, 2012, pp. 202–203.
[6] C. Stauffer, W. Grimson, Adaptive background mixture models for real-time tracking, in: Proceedings of Int. Conf. Computer Vision and Pattern Recognition, 1999, pp. 246–252.
[7] Z. Zivkovic, F. Heijden, Efficient adaptive density estimation per image pixel for the task of background subtraction, Pattern Recognit. Lett. 27 (7) (2006) 773–780.
[8] A. Elgammal, R. Duraiswami, D. Harwood, L. Davis, Background and foreground modeling using nonparametric kernel density estimation for visual surveillance, Proc. IEEE 90 (2002) 1151–1163.
[9] S. Lee, N. Kim, I. Paek, M. Hayes, J. Paik, Moving object detection using unstable camera for consumer surveillance systems, in: Proceedings of IEEE Int. Conf. Consumer Electronics, 2013, pp. 149–150.
[10] S. Periaswamy, H. Farid, Elastic registration in the presence of intensity variations, IEEE. Trans. Med. Imaging 22 (7) (2003) 865–874.
[11] D. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[12] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Systems, Man Cybern. 9 (1) (1979) 62–66.
[13] University of Central Florida (UCF) Aerial Action Data Set, http://crcv.ucf.edu/data/UCF_Aerial_Action.php.