Progress In Electromagnetics Research M, Vol. 65, 101–109, 2018

Visual Image Sequential Motion Detection via Half Quadratic Minimization Method

Ran Zhu*, Yunli Long, and Wei An

Abstract—In this paper, we present a straightforward numerical algorithm for visual image sequential motion detection based on half quadratic minimization method. As for the standard visual image sequences with RGB color representation, an intuitive way is to convert it to grayscale image to achieve an approximate motion detection with relatively low computational load. Instead, we propose a sequential processing scheme for more accurate detection by utilizing the motion detection algorithm separately and then performing fusion on a higher level. In this way, we extend the motion detection technique for grayscale images to be capable of dealing with visual videos. Experiment results show that the proposed algorithm can provide more robust motion detection performance and be successfully utilized in practical visual surveillance applications.

1. INTRODUCTION

Image sequence-based motion detection technologies are widely applied in a variety of practical surveillance applications [1]. Traditional target detection on single frame image generally makes use of imaging characteristics of the targets of interest, such as in synthetic aperture radar (SAR) images [2, 3] and infrared images [4, 5]. However, for moving target detection from image sequences, the target motion is the most distinguishable characteristic to effectively extract targets.

Conventional methods for motion detection include temporal difference [6], spatio-temporal filtering [7] and optical flow estimation [8, 9]. Optical flow is one of the most classical motion detection algorithms which approximately calculate a two-dimensional motion field according to spatio-temporal patterns of the image intensity [10, 11]. A complete overview of the developments and typical processing framework in visual surveillance is presented in [12]. Visual surveillance using the absolute difference technique in practical systems is evaluated in [13]. For automatic target tracking from image sequences, the authors in [14] proposed new image metrics that can effectively differentiate moving targets from backgrounds.

In addition, background subtraction is another effective way to directly detect the moving objects [15, 16]. By assuming that the intensity value of background pixel has the maximum probability along the image sequence, a background subtraction method based on pixel intensity classification is proposed in [17]. A Gaussian mixture model based background subtraction algorithm is proposed in [18], which can constantly update the parameters and select the appropriate number of components for each pixel. [19] presents a real-time foreground-background segmentation algorithm by quantizing each pixel into codebooks to represent the background model over a image sequence. Furthermore, a universal sample-based motion detection technique is proposed in [20], which is a new milestone for the group of background subtraction algorithms.

Received 28 November 2017, Accepted 26 February 2018, Scheduled 5 March 2018
* Corresponding author: Ran Zhu (zhuran@nudt.edu.cn).
The authors are with the College of Electronic Science and Engineering, National University of Defense Technology, Changsha, Hunan 410073, China.

A thorough theoretical analysis on the variational approach for motion segmentation and background restoration of grayscale images has been presented in [21]. By modeling the task as an optimization problem, moving objects are detected by an iterative algorithm via the half quadratic (HQ) minimization method in [22]. As for segmentation in color videos, the authors presented a practical system for coupled background modeling and moving object segmentation in [23]. Variational approach based on the conditional probability of the spatio-temporal image gradient and prior parametric motion model is proposed in [24, 25]. Note that such a minimization processing scheme can also be applied to multiplicative noise models as discussed in [26]. In particular, a variational approach for simultaneous motion detection and background restoration focusing on motion-blurred videos has been well studied in [27].

In this paper, we focus on the problem of motion detection in visual image sequences and propose a variational algorithm based on fusion of the half quadratic minimization results of different channels. The proposed algorithm is more efficient since the batch processing scheme in [22] is extended to a sequential processing scheme after initial stage, and thus the iterative computation operates mainly on the image data of current frame rather than the whole batch of images.

The remainder of the paper is organized as follows. In Section 2, an iterative algorithm for image sequence motion detection based on half quadratic minimization is presented. A motion detection fusion scheme for visual image sequences is described in Section 3. Experiment results of the algorithm tested on real visual images are presented in Section 4. The paper is concluded in Section 5.

2. HALF QUADRATIC MINIMIZATION BASED MOTION DETECTION

We assume that the image data acquired at frame h are denoted as $F_h(x) \in L^\infty(\Omega)$ ($x \in \Omega$, $\Omega \subset \mathbb{R}^2$ is bounded). In the initial stage, we can get the background estimation B_T and motion detection results from 1 to T (i.e., $C_1, \ldots, C_T$) based on the first T frames image data $F_1, \ldots, F_T$ by the batch processing method proposed in [22]. The objective functions for optimization in batch processing scheme are

$$\inf_{B, C_1, \ldots, C_T} \sum_{h=1}^{T} \int_\Omega C_h^2 (B - F_h)^2 dx + \alpha_c \sum_{h=1}^{T} \int_\Omega (C_h - 1)^2 dx + \alpha_b^r \int_\Omega |\nabla B| \, dx \tag{1}$$

Motion detection result C_h ($h = 1, \ldots, T$) is an identification function which takes value between 0 and 1, with 0 indicating the moving object region and otherwise the static background. After the initial stage, we get initial estimate of the background B_T and motion detection results of frame 1 to frame T. Then we are ready to start the sequential processing for the rest incoming images, i.e., F_{T+1} and so on. At time step $T+1$, we fix the detection results $C_2, \ldots, C_T$, then the optimization problem in Eq. (1) is simplified as follows.

$$\inf_{B, C_{T+1}} \sum_{h=2}^{T+1} \int_\Omega \left[C_h^2 (B - F_h)^2 + \alpha_c (C_h - 1)^2 \right] dx + \alpha_b^r \int_\Omega |\nabla B| \, dx \tag{2}$$

In order to overcome the difficulty in solving the Euler Lagrange equation of the above minimization problem, an extra defined functional d_B is introduced, which is essentially important in the half quadratic minimization process. Details on the mathematical proof and analysis please refer to [21]. For brevity, in this section, we present our modified algorithm which is designed specifically for visual image sequence analysis directly as follows.

When the current frame of image F_{T+1} is acquired, we initialize $B^0 \equiv B_T$, $C_{T+1}^0 \equiv 1$ and calculate the summations of existing detection results as follows.

$$S_{01} = \sum_{h=2}^{T} (C_h)^2 \tag{3}$$

$$S_{02} = \sum_{h=2}^{T} (C_h)^2 F_h \tag{4}$$

Then we iteratively perform the following operations:

Step 1 Background convolution (H is a smoothing template)

$$\bar{B} = B^n \otimes H \tag{5}$$

Step 2 Background estimation coefficient update

$$S_1 = S_{01} + (C_{T+1}^n)^2 \tag{6}$$

$$S_2 = S_{02} + (C_{T+1}^n)^2 F_{T+1} \tag{7}$$

Step 3 Background estimation

$$B^{n+1} = \frac{\alpha_b^r d_B^n \bar{B} + S_2}{\alpha_b^r d_B^n + S_1} \tag{8}$$

Step 4 Half quadratic function update

$$d_B^{n+1} = \frac{1}{2\left|\nabla B^{n+1}\right|} \tag{9}$$

Step 5 Motion detection

$$C_{T+1}^{n+1} = \frac{\alpha_c}{\alpha_c + (B^{n+1} - F_{T+1})^2} \tag{10}$$

After a predefined number of iterations N, the background estimation at frame $T+1$ is set as $B_{T+1} = B^N$, which is also the initial value for background convolution step at the next time step. Note that total variation regularization is applied on the estimated background B, which is proved to be effective in preserving the edges in images, thus the characteristics of objects in the scenario can be preserved as well [28, 29]. Motion detection step is straightforward since for simplicity, we do not set regularization on the detection results C_h ($h = T+1, T+2, \ldots$).

3. VISUAL IMAGE SEQUENCE PROCESSING

For motion detection in visual image sequences, an intuitive idea is to convert the RGB color image to a grayscale image and then apply the above algorithm. We assume that F_R, F_G, F_B are the R, G, and B components of a visual image F. Equivalent gray value I for each pixel can be calculated as the average of the R, G, and B component images.

$$I = \frac{F_R + F_G + F_B}{3} \tag{11}$$

Taking the frame 1250 as example, Fig. 1 shows the original image and the corresponding converted grayscale image. More generalized form as a weighted sum of the R, G, and B component images is

$$I = \omega_R F_R + \omega_G F_G + \omega_B F_B \tag{12}$$

For example, in applications of analog color television signals encoding, the weight coefficients are set as $\omega_R = 0.299$, $\omega_G = 0.587$ and $\omega_B = 0.114$. When it comes to digital color encoding, combination of $\omega_R = 0.2125$, $\omega_G = 0.7154$ and $\omega_B = 0.072$ is recommended [1].

In this way, we can operate the motion detection algorithm merely on the grayscale image converted from RGB image. A main advantage is that the computational load is reduced since only one channel of the image data needs to perform the iterative motion detection algorithm. However, as we will see in the next section, this irreversible conversion inevitably degrades the detection performance due to the loss of information.

Figure 2 presents the RGB decomposition of a visual image at frame 1275. In this paper, we perform the motion detection algorithm on component image data F_R, F_G, and F_B, and then we get the moving object index matrices of R, G, and B channels, which are denoted as C_R, C_G, and C_B. The final moving object detection result fusion can be calculated as the pointwise minimum of the above moving object index matrices.

$$C = \min(C_R, C_G, C_B) \tag{13}$$

Note that other operations such as taking the maximization or taking the average can be used as well depending on specific needs and the imaging scene of a practical surveillance system. The processing schemes discussed above are summarized as shown in Fig. 3.

(a) (b)

Figure 1. Scenario at frame 1250. (a) Original visual image F. (b) Converted grayscale image I.

(a) (b)

(c) (d)

Figure 2. Scenario at frame 1275. (a) Original visual image F. (b) R component image F_R. (c) G component image F_G. (d) B component image F_B.

4. EXPERIMENT RESULTS

In this section, we verify the proposed visual image sequence motion detection algorithm by testing the real visual image data of a practical surveillance application. We further compare the proposed motion detection approach with representative methods such as optical flow and ViBe algorithm.

Parameters in the energy function are $a_b^r = 100$ and $a_c = 3000$, and iteration times are set to 10. Length of the temporal sliding window is set as 50. The smoothing template in the background convolution step introduced in Section 2 is selected as

$$H = \frac{1}{12}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 0 & 2 \\ 1 & 2 & 1 \end{bmatrix} \tag{14}$$

More specifically, for discrete images of size $M \times N$, the gradient operation $(\nabla u)_{i,j} = ((\nabla u)^1_{i,j}, (\nabla u)^2_{i,j})$

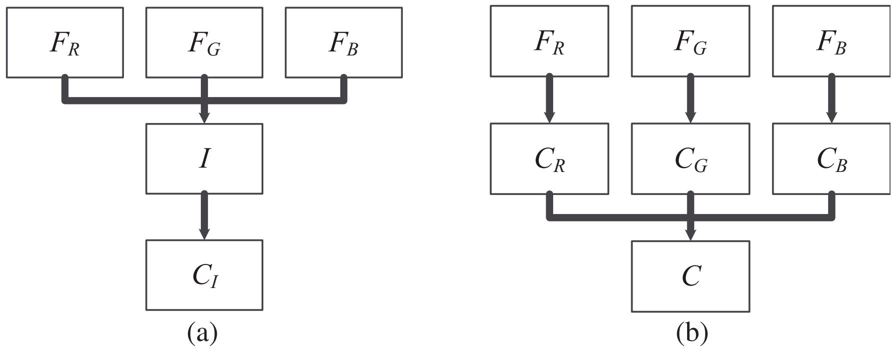


Figure 3. Processing schemes. (a) Grayscale image processing. (b) RGB color image processing.

is calculated according to

$$(\nabla u)^1_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < M \\ 0 & \text{if } i = M \end{cases} \tag{15}$$

$$(\nabla u)^2_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < N \\ 0 & \text{if } j = N \end{cases} \tag{16}$$

Motion detection results under the grayscale image processing scheme are presented in Fig. 4. Set of pixels that are assumed to be the moving target is

$$S_I = \{(i,j) : (C_I)_{i,j} < 0.5\} \tag{17}$$

In Fig. 4(b), we show the extracted objects, i.e., $\{F_{i,j} = (F_R, F_G, F_B)_{i,j} : (i,j) \in S_I\}$. Note that there are some lost part of the objects due to the color of the objects and varying ambient lights, especially when the color of moving object is close to that of the static background.

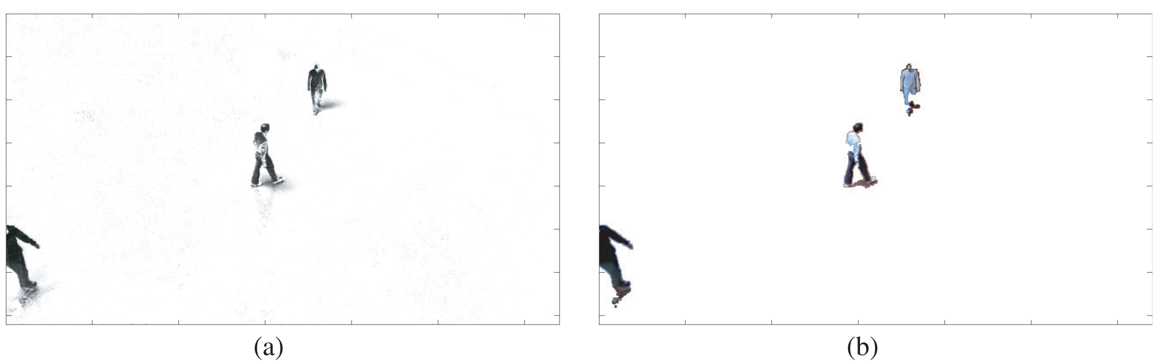


Figure 4. Grayscale image motion detection results. (a) Motion detection result C_I. (b) Moving objects extracted from the scene.

However, under the RGB color image processing scheme, if the pixel indices which are regarded as moving object regions of R, G, and B channels are

$$S_R = \{(i,j) : (C_R)_{i,j} < 0.5\} \tag{18}$$

$$S_G = \{(i,j) : (C_G)_{i,j} < 0.5\} \tag{19}$$

$$S_B = \{(i,j) : (C_B)_{i,j} < 0.5\} \tag{20}$$

then the overall moving region pixel set is simply the union of S_R, S_G, and S_B.

$$S = S_R \cup S_G \cup S_B \tag{21}$$

Note that it is equivalent to performing segmentation by $S = \{(i,j) : (C)_{i,j} < 0.5\}$, where C is the fused moving object index matrix as described in Eq. (13).

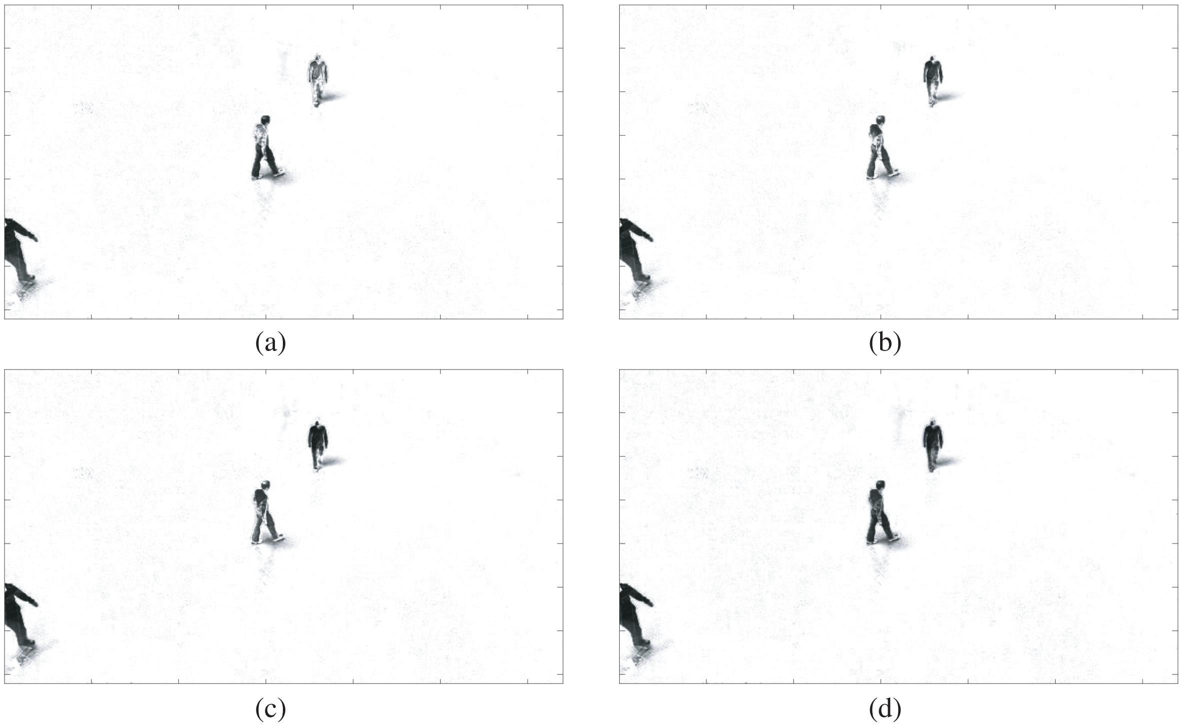
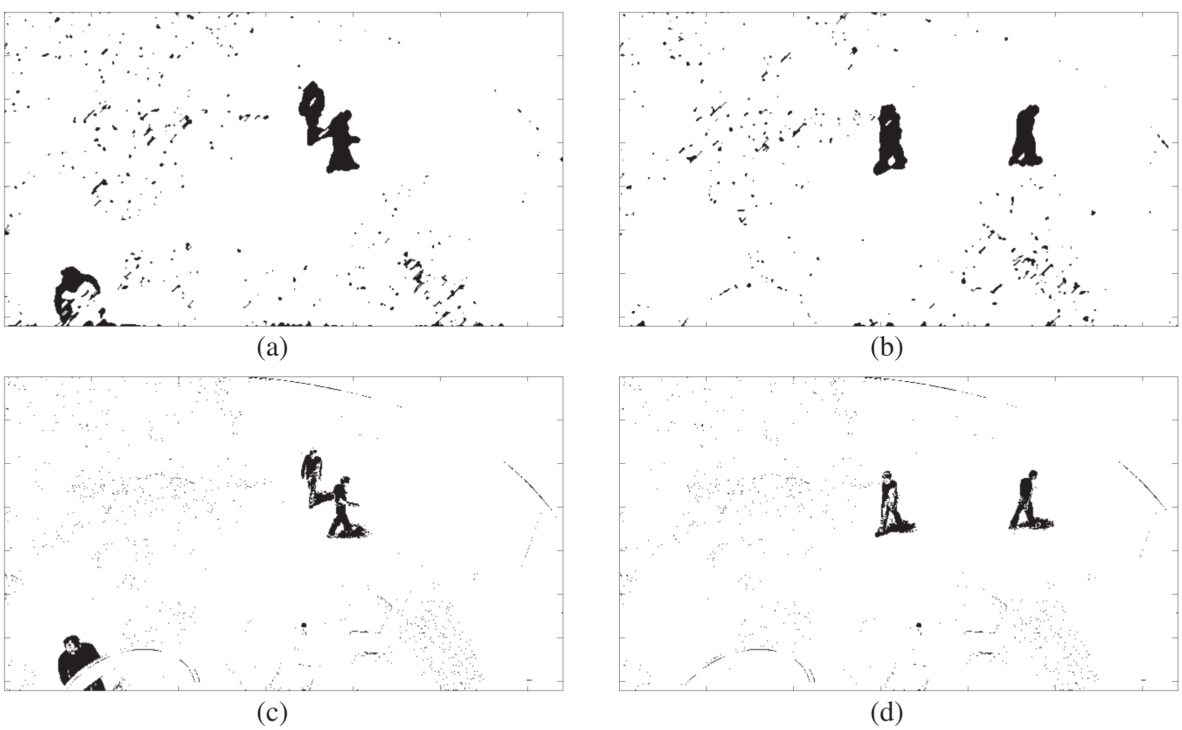
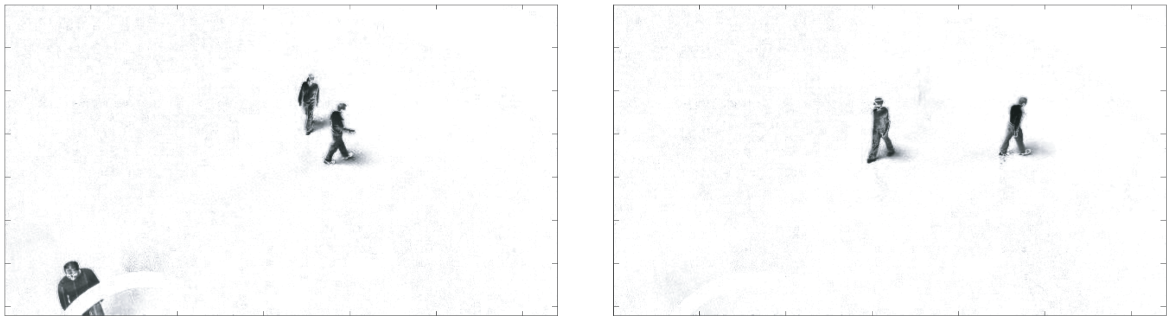


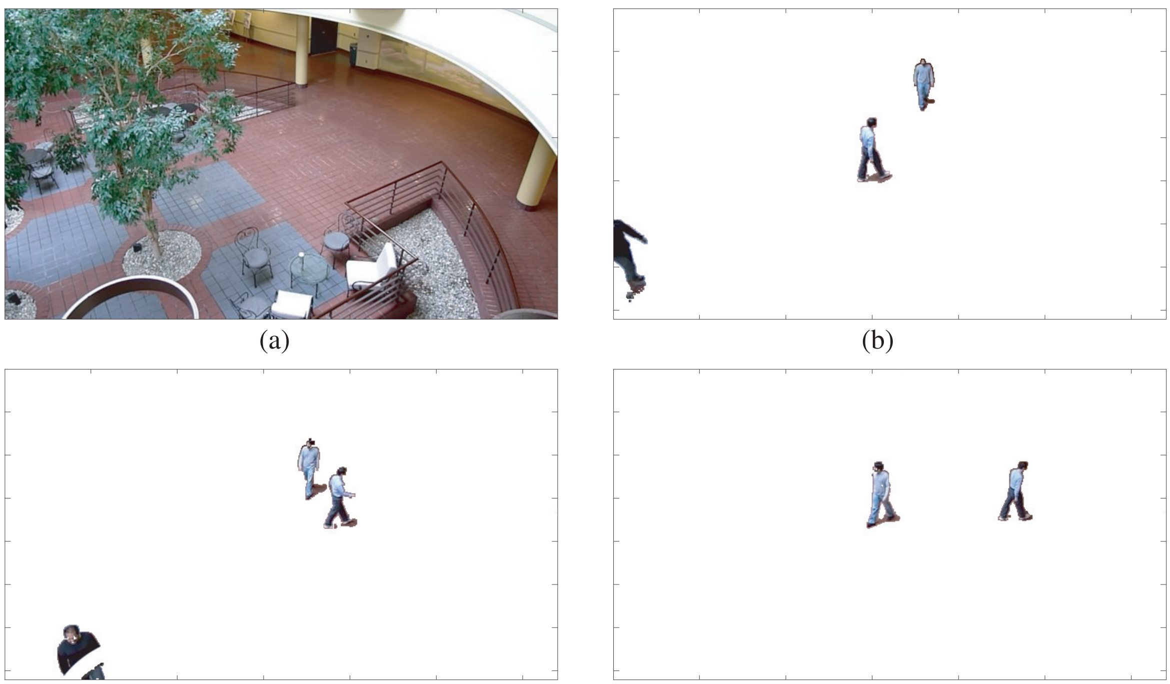
Figure 5. Motion detection results of RGB color image via HQ minimization. (a) Motion detection result of R channel. (b) Motion detection result of G channel. (c) Motion detection result of B channel. (d) Motion detection fusion result.





(e) (f)

Figure 6. Moving object detection results comparison. (a) Optical flow result of frame 1275. (b) Optical flow result of frame 1300. (c) ViBe result of frame 1275. (d) ViBe result of frame 1300. (e) HQ minimization result of frame 1275. (f) HQ minimization result of frame 1300.



(a) (b)

(c) (d)

Figure 7. Moving object extraction results. (a) Background estimation. (b) Frame 1250. (c) Frame 1275. (d) Frame 1300.

Motion detection results of R, G, and B channels and the final fusion result are shown in Fig. 5. We see that the lost part of the moving object in the detection result of R channel (see Fig. 5(a)) can be supplemented by G and B channels. Obviously, compared to the results of grayscale processing scheme, shapes of the moving objects in the fusion results are more complete.

We further compare the proposed algorithm via HQ minimization with typical motion detection algorithms including optical flow and visual background extractor (ViBe). From the result comparison in Fig. 6, we see that the optical flow method is very sensitive to slight variance of light in the scene. The segmentation result is apparently improved by the ViBe method at the cost of losing some character details of moving objects. However, there are still some false detections that are clearly observed in most frames. The proposed method can achieve a better identification result by incorporating a sliding window of image frames and performing background estimation and moving object detection in a coupled way. And fusion of motion detection results of all the channels guarantees the validity and completeness of the extract results.

Extracted objects (i.e., $\{F_{i,j} : (i, j) \in S\}$) at different frames are shown in Fig. 7. Note that in this method, we can also restore the visual background by the combination of the background estimation in R, G, and B channels. Compared to Fig. 4, the visual segmentation of moving objects via proposed method is more accurate and complete. Unlike the conventional motion detection algorithms which usually give hard decision (0 or 1), the proposed HQ minimization method gives soft detection results between 0 and 1. Such a detection result in the probability field can be further applied in the track-before-detect algorithms [30], which has the potential to form a novel detection and tracking processing scheme for visual videos.

5. CONCLUSION

In this paper we propose a motion detection algorithm for visual image sequences, which is a sequential method suitable for video processing in practical surveillance systems. Based on the half quadratic minimization technique, the conventional motion detection algorithm for grayscale images is extended to be applicable to visual videos. Fusion of the motion detection results of multiple channels over a sliding window of frames significantly improves the ability to accommodate variance of the ambient light. Experiment results verify that we can get more accurate and complete moving object detection by the proposed method. Future work includes possible extensions for non-static camera with non-Gaussian imaging noises, which can provide improved robustness of the algorithm and be utilized for more general applications.

REFERENCES

1. Burger, W. and M. Burge, *Principles of Digital Image Processing: Fundamental Techniques*, Springer, Berlin, 2009.
2. Yang, M. and G. Zhang, "Unsupervised target detection in sar images using scattering center model and mean shift clustering algorithm," *Progress In Electromagnetics Research Letters*, Vol. 35, 11–18, 2012.
3. Yang, M. and G. Zhang, "A dictionary-based image fusion for integration of SAR and optical images," *Progress In Electromagnetics Research Letters*, Vol. 49, 87–90, 2014.
4. Diao, W., X. Mao, and V. Gui, "Metrics for performance evaluation of preprocessing algorithms in infrared small target images," *Progress In Electromagnetics Research*, Vol. 115, 35–53, 2011.
5. Zhao, B., S. Xiao, H. Lu, and J. Liu, "Point target detection in space-based infrared imaging system based on multi-direction filtering fusion," *Progress In Electromagnetics Research M*, Vol. 56, 145–156, 2017.
6. Lipton, A. J., H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," *Proceedings of Proc. IEEE Workshop Applications of Computer Vision*, 8–14, Princeton, NJ, USA, 1998.
7. Caspi, Y. and M. Irani, "Spatio-temporal alignment of sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 11, 1409–1424, 2002.
8. Barron, J. L., D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, Vol. 12, No. 1, 43–77, 1994.
9. Barranco, F., J. Diaz, E. Ros, and B. D. Pino, "Visual system based on artificial retina for motion detection," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 39, No. 3, 752–762, 2009.
10. Horn, B. K. P. and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, Vol. 17, No. 1–3, 185–203, 1981.
11. Lucas, B. D. and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proceedings of International Joint Conference on Artificial Intelligence*, 674–679, Vancouver, British Columbia, Canada, 1981.

12. Hu, W., T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Vol. 34, No. 3, 334–352, 2004.
13. Mhatre, S., S. Varma, and R. Nikhare, "Visual surveillance using absolute difference motion detection," *Proceedings of IEEE International Conference on Technologies for Sustainable Development*, 1–5, Mumbai, India, 2015.
14. Diao, W., X. Mao, H. Zheng, Y. Xue, and V. Gui, "Image sequence measures for automatic target tracking," *Progress In Electromagnetics Research*, Vol. 130, 447–472, 2012.
15. Li, C. and Y. Jiang, "An effective background reconstruction method for video objects detection," *Proceedings of IEEE International Conference on Networking and Distributed Computing*, 161–165, Hangzhou, China, 2012.
16. Jiang, S., Z. Wei, S. Wang, Z. Zhou, and J. Zhang, "A new algorithm for background extraction under video surveillance," *Proceedings of IEEE Conference Anthology*, 1–4, China, 2013.
17. Hou, Z. and C. Han, "A background reconstruction algorithm based on pixel intensity classification," *Journal of Software*, Vol. 16, No. 9, 1568–1576, 2005.
18. Zivkovic, Z. and F. V. D. Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, Vol. 27, No. 7, 773–780, 2006.
19. Kim, K., T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imaging*, Vol. 11, No. 3, 172–185, 2005.
20. Barnich, O. and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, Vol. 20, No. 6, 1709–1724, 2011.
21. Aubert, G. and P. Kornprobst, *Mathematical Problems in Image Processing, Partial Differential Equations and the Calculus of Cariations*, Springer, Berlin, 2006.
22. Kornprobst, P., R. Deriche, and G. Aubert, "Image sequence analysis via partial differential equations," *Journal of Mathematical Imaging and Vision*, Vol. 11, No. 1, 5–26, 1999.
23. Francois, A. and G. Medioni, "Adaptive color background modeling for real-time segmentation of video streams," *Proceedings of International Conference on Imaging Science, System and Technology*, 1–6, 1999.
24. Cremers, D. and S. Soatto, "Variational space-time motion segmentation," *Proceedings of IEEE International Conference on Computer Vision*, 886–893, Nice, France, 2003.
25. Cremers, D. and S. Soatto, "Motion competition: A variational approach to piecewise parametric motion segmentation," *International Journal of Computer Vision*, Vol. 62, No. 3, 249–265, 2005.
26. Aubert, G. and J. Aujol, "A variational approach to removing multiplicative noise," *Siam Journal on Applied Mathematics*, Vol. 68, No. 4, 925–946, 2008.
27. Bar, L., B. Berkels, M. Rumpf, and G. Sapiro, "A variational framework for simultaneous motion estimation and restoration of motion-blurred video," *Proceedings of IEEE International Conference on Computer Vision*, 1–8, Rio de Janeiro, Brazil, 2007.
28. Tikhonov, A. N. and V. Y. Arsenin, *Solutions of Ill-posed Problems*, Winston and Sons, Washington, D.C., 1977.
29. Rudin, L., S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, Vol. 60, No. 1-4, 259–268, 1992.
30. Reza, H., B. Ngu, and B. Tuong, "Visual tracking in background subtracted image sequences via multi-bernoulli filtering," *IEEE Transactions on Signal Processing*, Vol. 61, No. 2, 392–397, 2012.