

## Research Article

# Motion Objects Segmentation and Shadow Suppressing without Background Learning

Y.-P. Guan<sup>1,2</sup>

<sup>1</sup> School of Communication and Information Engineering, Shanghai University, 99 Shangda Road, Shanghai 200444, China

<sup>2</sup> Key Laboratory of Advanced Displays and System Application, Ministry of Education, 149 Yanchang Road, Shanghai 200072, China

Correspondence should be addressed to Y.-P. Guan; [ypguan@shu.edu.cn](mailto:ypguan@shu.edu.cn)

Received 1 November 2013; Revised 15 December 2013; Accepted 16 December 2013; Published 23 January 2014

Academic Editor: Haranath Kar

Copyright © 2014 Y.-P. Guan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An approach to segmenting motion objects and suppressing shadows without background learning has been developed. Since wavelet transformation indicates the position of sharper variation, it is adopted to extract the information contents with the most meaningful features based on two successive video frames only. According to the fact that the saturation component is lower in the region of shadow and is independent of the brightness, HSV color space is selected to extract foreground motion region and suppress shadow instead of other color models. A local adaptive thresholding approach is proposed to extract initial binary motion masks based on the results of the wavelet transformation. A foreground reclassification is developed to get an optimal segmentation by fusion of mode filtering, connectivity analysis, and spatial-temporal correlation. Comparative studies with some investigated methods have indicated the superior performance of the proposal in extracting motion objects and suppressing shadows from cluttered contents with dynamic scene variation and crowded environments.

## 1. Introduction

Robust foreground motion objects segmentation plays a crucial role in many computer vision applications such as visual surveillance, intelligent traffic monitoring, athletic performance analysis, perceptual user interface, and others. Many methods have been proposed for motion object segmentation. One of the popular approaches is background subtraction by comparing each new frame with a learned model of the scene taken by a static camera [1–3]. The initial background modeling is critical for this method. Gaussian mixture models (GMM) [4, 5] have been adopted to construct background modeling based on some previous period observations. There are two problems when GMM is applied to model background [1] including the selection of the number of components and the initialization. When the pixels have more Gaussian distribution than a predefined value or the pixels are covered by motion objects, it results in some background pixel errors easily. Nonparametric approaches methods have been developed to tackle this problem [6–8]. However, the size of a temporal window is needed to be

specified. Besides, spatial dependencies are not exploited and the presence of shadows is usually incorrectly classified. Some algorithms based on spatiotemporal characteristics have been developed [9, 10] to overcome above shortcomings. Some fragmentations arise where foreground objects overlap spatially with background ones with similar color in this method. Some attempts have been made to segment motion objects without prelearning [11, 12], which needs some attempts and tests to find a good set of parameter values.

Shadow is another challenging problem in segmenting motion objects. Shadow can cause object merging, object shape distortion and even object losses (due to the shadow casts over another object). It is difficult to remove shadow from object since shadow and object share some similar visual features [13]. A  $c_1c_2c_3$  space [14] is proposed to detect cast shadow through using chrominance color components. Several assumptions are needed regarding the reflecting surfaces and the lightings. A normalized rgb space [15] is proposed to detect shadows. The normalized rgb space suffers from noise at low intensities which would result in unstable chromatic components. Only gray level [16] is adopted for shadow

segmentation. Other approaches were done with the CIE Luv or CIE Lab spaces, respectively. However, it remains open ended how important is the appropriate color space selection and which color space is the most effective regarding shadow detection [17]. Texture analysis can be potentially effective in solving the problem. Lam et al. [18] proposed a method based on texture analysis for outdoor vehicle segmentation. Some assumptions are made such as single and strong light source so that the illumination difference between the shadow and background is reasonable in intensity and homogenous road texture. Heikkilä and Pietikäinen [19] proposed texture-based method for modeling the background and detecting motion objects. Because of a huge amount of different combinations, it needs more or less empirical test to find a good set of parameters. Zhang et al. [20] developed a ratio edge to detected motion cast shadows. An iteration strategy is implemented to estimate the shadow intensity ratio. The reliability of the method improves by increasing the number of neighboring points.

Aiming at some limits mentioned above, a new approach is developed to detect motion objects and suppress shadow without background learning. Although the author in [21] used spatiotemporal motion to segment foreground and suppress shadow, the proposed algorithm in this work has some main differences with that in [21] as follows. Motion objects segmentation developed does not consider background or background learning. Only two successive video frames are used to extract the information contents with the most meaningful features based on a wavelet transformation. Besides a local adaptive thresholding method is proposed to segment initial motion masks instead of a global thresholding. Moreover a foreground reclassification is developed to get an optimal segmentation by connectivity analysis and spatial-temporal correlation. The first contribution of the proposal is that the wavelet transformation is used to get some significant features based on the temporal difference between only two successive frames without background learning. A second contribution is that a local adaptive thresholding is proposed to extract initial motion mask based on the fact that a higher signal-to-noise ratio corresponds to truer boundary with the most meaningful features. A third contribution is that a foreground reclassification is developed to get an optimal segmentation by fusion of multiple cues including mode filtering, connectivity analysis, and spatial-temporal correlation. Another important contribution is efficient in segmenting foreground objects and suppressing shadows from cluttered contents. Comparative studies with some state-of-the-arts have indicated the superior performance of the proposal.

The organization of the rest of the paper is as follows. In the next section, we discuss wavelet transform. Foreground segmentation is discussed in Section 3. Experimental results from both indoor and outdoor scenes are given in Section 4 and followed by some conclusions in Section 5.

## 2. Wavelet Transformation

Since similarity and discontinuity underlie most current popular segmentation algorithms for images, the low-pass

and high-pass filters of the wavelet transformation (WT) naturally break a signal into similar and discontinuous sub-signals, which effectively combines the two basic properties into a single approach [21].

A function  $\psi(x, y)$  is called a wavelet if its average is equal to 0. The WT of  $f(x, y)$  at dyadic scale  $2^j$  and position  $(x, y)$  in orientation  $d$  is defined as

$$W_{2^j}^d f(x, y) = f(x, y) \otimes \psi_{2^j}^d(x, y), \quad d = 1, 2, \quad (1)$$

where  $\otimes$  denotes convolution operator.

The two oriented wavelets  $\psi_{2^j}^d$  can be constructed by taking the partial derivate as

$$\psi^1(x, y) = \frac{\partial \theta(x, y)}{\partial x}, \quad \psi^2(x, y) = \frac{\partial \theta(x, y)}{\partial y}, \quad (2)$$

where  $\theta(x, y)$  is a separable scaling function which plays the role of a smoothing filter.

The 2-D WT defined by (1) gives the gradient of  $f(x, y)$  smoothed by  $\theta(x, y)$  at dyadic scales

$$\begin{aligned} \nabla_{2^j} f(x, y) &\equiv (W_{2^j}^1 f(x, y), W_{2^j}^2 f(x, y)) \\ &= \frac{1}{2^{2j}} \nabla f(x, y) * \theta_{2^j}(x, y). \end{aligned} \quad (3)$$

Consider the local maxima of the gradient magnitude at various scales which is given by

$$\begin{aligned} M_{2^j} f(x, y) &\equiv \|\nabla_{2^j} f(x, y)\| \\ &= \sqrt{(W_{2^j}^1 f(x, y))^2 + (W_{2^j}^2 f(x, y))^2}. \end{aligned} \quad (4)$$

A point  $(x, y)$  is a multiscale edge point at scale  $2^j$  if the magnitude of the gradient  $M_{2^j} f(x, y)$  attains a local maximum along the gradient direction  $A_{2^j} f(x, y)$  defined as

$$A_{2^j} f(x, y) \equiv \arctan \left[ \frac{W_{2^j}^2 f(x, y)}{W_{2^j}^1 f(x, y)} \right]. \quad (5)$$

For each scale, collect the edge points together with the corresponding values of the gradient, namely, the WT values at that scale. The resulting local gradient maxima set at scale  $2^j$  is

$$P_{2^j}(f) = \{p_{2^j,i} = ((x_i, y_i); \nabla_{2^j} f(x_i, y_i))\}, \quad (6)$$

where  $M_{2^j} f(x_i, y_i)$  has local maximum at  $p_{2^j,i} = (x_i, y_i)$  along the direction  $A_{2^j} f(x_i, y_i)$ .

For a  $J$ -level 2-D dyadic WT, the set is called a multi-scale edge representation of the image  $f(x, y)$ .

$$\rho(f) = \{O_{2^j} f(x, y), [P_{2^j}(f)]_{1 \leq j \leq J}\}, \quad (7)$$

where  $O_{2^j} f(x, y)$  is the low-pass approximation of  $f(x, y)$  at the coarsest scale  $2^J$ .

The above algorithm is based on a dyadic scale WT, which can reach large scales quickly. However, when it is used to deal

with noisy images, the noise level is sensitive to the change of scales [21]. A dyadic sequence of scales cannot always optimally adapt to the effect of noise. In order to overcome the drawback, an interesting image  $E(x, y)$  for the image  $f(x, y)$  is given as

$$E(x, y) = \left| |f(x, y) \otimes h| \geq |(f(x, y) \otimes h) \otimes s| \right| + \left| |f(x, y) \otimes v| \geq |(f(x, y) \otimes v) \otimes s| \right|, \quad (8)$$

where  $h, v$ , and  $s$  are three filters in different directions defined as

$$h = \begin{bmatrix} -3 & 0 & 3 \\ -5 & 0 & 5 \\ -3 & 0 & 3 \end{bmatrix}, \quad v = h^T, \quad s = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad (9)$$

where  $T$  is a matrix transposition operator.

### 3. Foreground Segmentation

**3.1. Motion Detection and Shadow Suppressing.** Since the WT indicates the positions of sharper variation, it can be used to extract the information content of signals with the most meaningful features. The solution provided uses HSV color space instead of other complex color models. The main reason for using HSV is that the HSV color space corresponds closely to the human perception of color and it explicitly separates chromaticity and luminosity. Since the saturation is independent of the brightness [22] and it is lower in shadow of point, we use the saturation component (S) and intensity one (V) to segment motion objects and suppress shadow. Let the inputs of a pair of gray level frames be  $I_{t-1}$  and  $I_t$  taken at times  $t-1$  and  $t$ , respectively. The output is the variance regions of an image with significant changes. To detect the region with significant changes, we calculate WT on the current frame and temporal differences between the two successive frames for S and V components, respectively as follows:

$$\begin{aligned} W_S &= W_{S_t} \\ W_V &= W_{V_t} \\ W_{D_S} &= W_{|S_t - S_{t-1}|} \\ W_{D_V} &= W_{|V_t - V_{t-1}|} \end{aligned} \quad (10)$$

where  $W_X$  is a WT operator on  $X$ ,  $S_t$  represents S level taken at  $t$  time,  $V_t$  is V level taken at  $t$  time, and  $||$  is absolute operator.

For a spatial point  $(x, y)$ , the significant changes exist only if the WT on the above components have the same information with the most meaningful ones. So the motion mask  $M_{(x,y)}$  is computed as

$$M_{(x,y)} = \begin{cases} 1, & \text{if } \prod W_i * W_{D_i} \geq T, \quad (i = S, V) \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

where  $T$  is a threshold value obtained by a local adaptive thresholding method (described later).

**3.2. Local Adaptive Thresholding.** The thresholding can be roughly categorized as global and local adaptive methods. Global method such as Otsu [23] tries to find a single threshold value to assign foreground or background based on its grey value. The major problem with global thresholding is that only the intensity is considered, not any relationships between the pixels. There is no guarantee that the pixels identified by the thresholding process are contiguous. Local adaptive thresholding method tries to overcome the above problem by computing thresholds individually for each pixel using information from the local neighborhood of the pixel. The most classical local adaptive method is Niblack method [24] based on the calculation of the local mean and standard deviation of image intensity  $I(x, y)$  as follows:

$$B(x, y) = \begin{cases} 1, & \text{if } I(x, y) > T(x, y) \\ 0, & \text{otherwise,} \end{cases} \quad (12)$$

where  $T(x, y) = m_{w \times w}(x, y) + k \cdot \sigma_{w \times w}(x, y)$ ;  $m_{w \times w}(x, y)$  and  $\sigma_{w \times w}(x, y)$  represent the mean and standard deviation of the pixel intensities in a window with  $w \times w$  size, respectively.

This method does not work well when the background area contains local variations due to uneven illumination. To solve this problem, Sauvola and Pietikäinen [25] presented another modified version of Niblack's method. The thresholds are computed with the dynamic range of standard deviation  $R$  as follows:

$$T(x, y) = m_{w \times w}(x, y) \left[ 1 + k \left( \frac{\sigma_{w \times w}(x, y)}{R} - 1 \right) \right]. \quad (13)$$

One of the flaws involved with the standard deviation is that if the data is spread out over a large range of values, the standard deviation  $\sigma$  will be large.

Signal-to-noise ratio, which is equal to the mean divided by the standard deviation, describes the quality of a measurement and refers to the relative magnitude of the signal compared to the uncertainty in that signal on a per-pixel basis [26]. It means that higher signal-to-noise ratio corresponds to truer boundary with the most meaningful features in an image. A novel local adaptive thresholding approach is developed as follows:

$$T(x, y) = m_{w \times w}(x, y) + k \frac{m_{w \times w}(x, y)}{\sigma_{w \times w}(x, y)}, \quad (14)$$

where  $m_{w \times w}(x, y)$  and  $\sigma_{w \times w}(x, y)$  represent the mean and standard deviation of the pixel intensities in a window with  $w \times w$  size, respectively;  $k$  is a thresholding coefficient.

**3.3. Foreground Reclassification.** Some results are given in Figure 1 for an outdoor scene about *highway 1* from <http://cvrr.ucsd.edu/aton/shadow> where crowded cars move in highway and cast heavy shadows on the ground foreground.

Figure 1 illustrates the moving cast shadows are suppressed by the proposed method (seen from Figure 1(c)). The detected moving object region is not complete in which some points are misclassified. To overcome such open issue [27, 28],

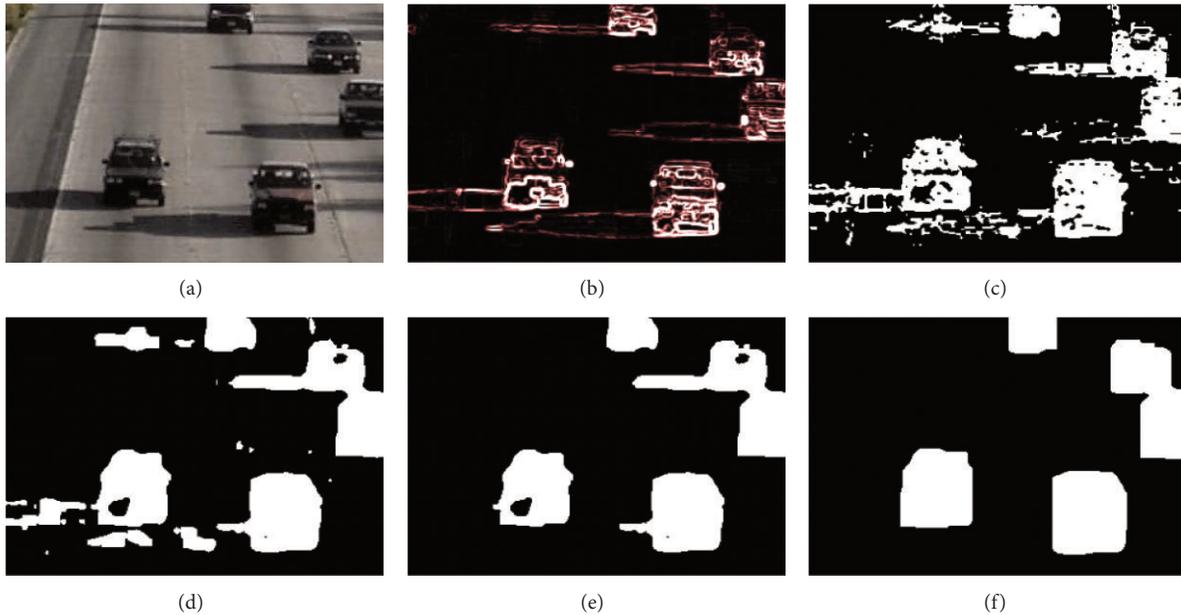


FIGURE 1: WT based adaptive thresholding segmentation for *highway I*. (a) The current frame (number 14); (b) image after WT; (c) adaptive thresholding segmentation results using  $w = 5 \times 5$  and  $k = 0.5$  in (14); (d) binary masks filtered by  $5 \times 5$  mode factor; (e) small-area regions less than 100 removed by connective characteristics; (f) final results based on spatial-temporal correlation characteristics.

a foreground reclassification method is developed as follows. The proposed approach is based on the fact that the detected prominent boundaries are connected with foreground objects to a great extent. Since local mode filtering [29] can preserve edges and detail, it is used to extract details and suppress noises from the initial binary masks (seen from Figure 1(c)) in a  $5 \times 5$  window size factor (seen from Figure 1(d)). Afterwards a four-neighbor connected component analysis is employed to extract some meaningful blobs by rejecting some isolated and disconnected regions with its area less than a certain size (seen from Figure 1(e) where the threshold for area size is 100). Some points are misclassified still after the steps mentioned above. A spatial-temporal correlation is proposed to solve this problem. According to chromatic similarity within a neighborhood of segmented foreground patch, the mutual information between the pixels in the foreground patch and their counterparts at the current frame is selected to compute the region similarity. Besides, interframe continuity is incorporated to get an optimal segmentation as mentioned in [30]. The final binary results are given in Figure 1(f) after the developed spatial-temporal correlation characteristics.

For foreground object with relatively large size such as a heavy truck running in highway (seen from Figure 2(a)), some results as mentioned above are shown in Figure 2.

One can find that the developed approach can be applied to segment foreground objects with different scales from Figures 1 and 2.

## 4. Experimental Results

**4.1. Test Set and Processing Time.** Extensive experiments have been carried out on different video sequences. Among these sequences, both indoor and outdoor scenes with different

contents are selected to test the proposed approach. The selected indoor sequences include *intelligent room* from a ATON project <http://cvrr.ucsd.edu/aton/shadow>, where people move slowly in the scene and shadow project on the ground wall, table, and so forth; a *CAVIAR* from the project <http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/>, where people move slowly, fight in situ, and move fast moreover some fluorescent lights in the scene turn on and off randomly. The selected outdoor sequences are *campus* from the ATON project <http://cvrr.ucsd.edu/aton/shadow>, where car and people move in the scene and *PETS'09* from <http://www.cvg.rdg.ac.uk/PETS2009/a.html> where crowded people move in high-lights environment. These choices of such different contexts are to emphasize the reliability and robustness of the proposed approach. Besides, its performance and comparisons with some state-of-the-arts are presented in this section.

The experiment is performed at Matlab7.10.0 environment, PC with PIV 2.1GHz CPU 4.0G Byte RAM. The processing time is dependent on the image dimension. We have processed all images for each sequence during which people or cars move in the scenes and evaluated the average processing time. The results together with the image dimensions are reported in Table 1.

**4.2. Segmentation Results.** In the experiments, two main parameters affect the performance of the proposal including the window size  $w$  and the thresholding coefficient  $k$  in (14). For the window size  $w$ , a selection of larger  $w$  means more foreground and background pixels are covered to the neighborhood and increase the possibilities to smooth the noise effects, whereas a selection of small window size gives rise to misdetection for the foreground object contours. In all

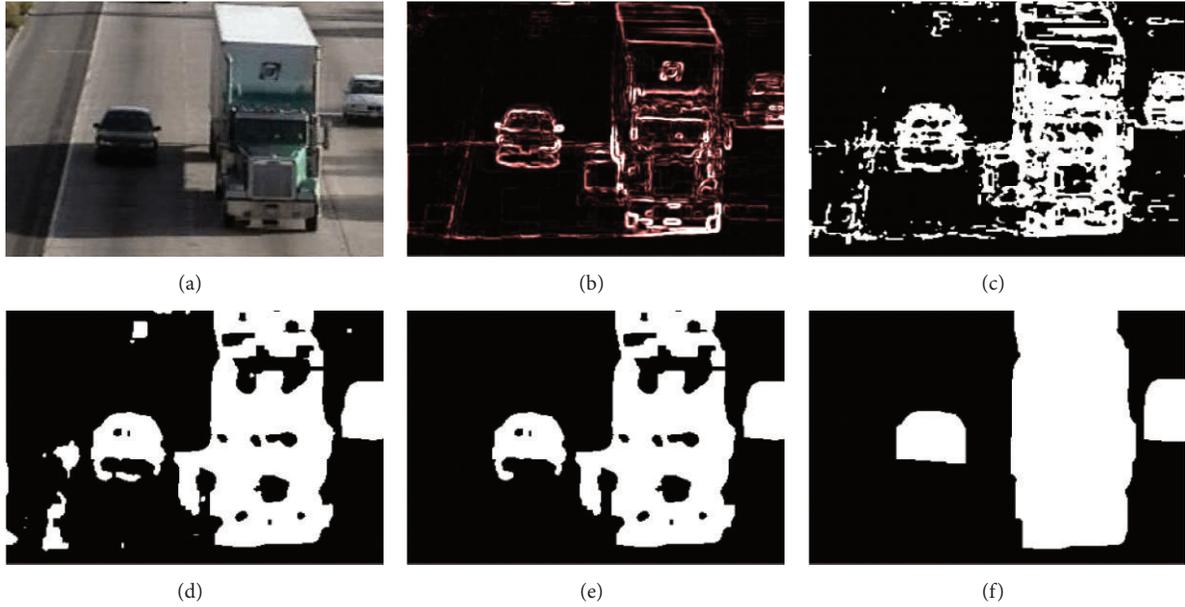


FIGURE 2: Some segmentation results for a heavy truck running in highway with relatively large size. (a) the current frame (number 363); (b) image after WT; (c) adaptive thresholding segmentation results using  $w = 5 \times 5$  and  $k = 0.5$  in (14); (d) binary masks filtered by  $5 \times 5$  mode factor; (e) small-area regions less than 100 removed by connective characteristics; (f) final results based on spatial-temporal correlation characteristics.

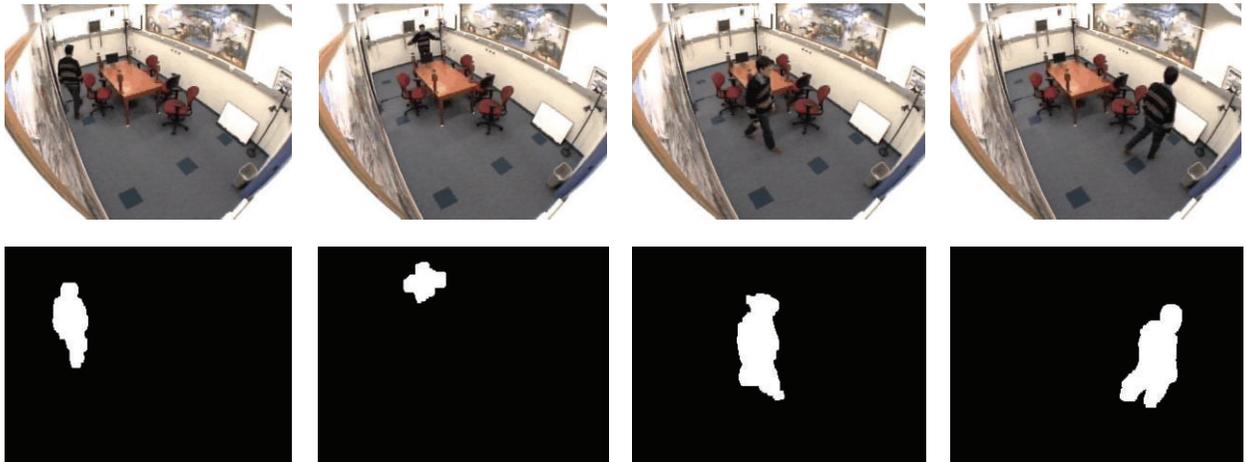


FIGURE 3: Some results for the *intelligent room* sequence. The first row is the current frame; the second row is the segmented masks.

TABLE 1: Average processing time for the tested sequences with different kinds of motion.

Sequence	Image size	Average processing time (s)
<i>intelligent room</i>	$240 \times 320$	0.1763
<i>CAVIAR</i>	$288 \times 384$	0.2530
<i>campus</i>	$288 \times 352$	0.2269
<i>PETS'09</i>	$576 \times 768$	0.9801

contexts considered, the window size  $w$  is selected as  $5 \times 5$  in the experiment. For the thresholding coefficient  $k$ , a larger  $k$  leads to less false segmentation and less true positive one. In

all contexts tested, the thresholding coefficient  $k$  is set as 0.5 and kept the same for the whole testing.

Some experimental results are given in Figures 3–6 based on the above selected parameters for the selected sequences, respectively.

The developed algorithm correctly identifies the people moving slowly, and some shadows casting on the ground wall and table have been suppressed from Figure 3.

People who move in the scene and the lighting conditions are different from that of the *intelligent room*; especially the people moving in the turbulent blossom are extracted efficiently. Besides, the developed algorithm correctly identifies the people moving in different styles from Figure 4 including moving slowly (the first column and the second column),

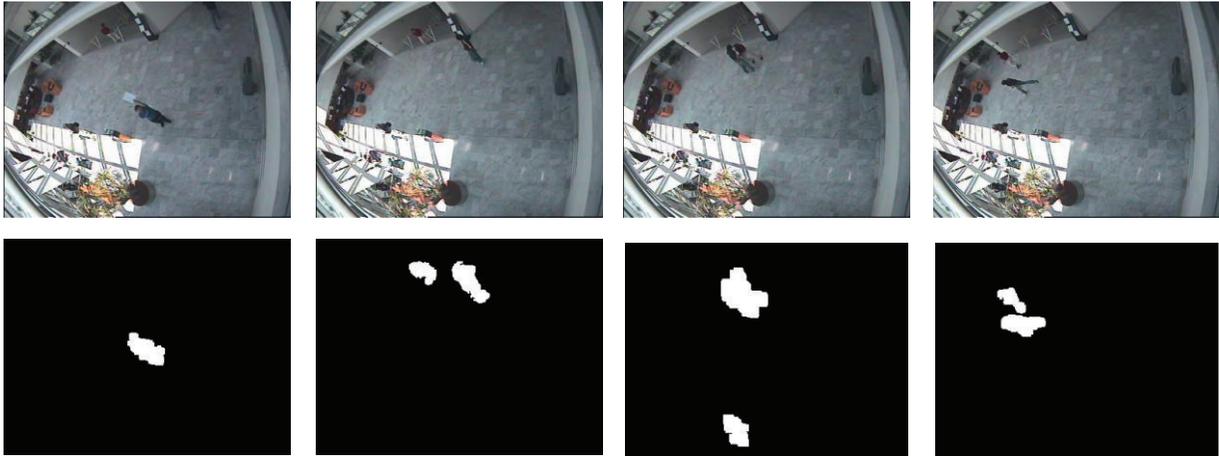


FIGURE 4: Some results for the *CAVIAR* sequence. The first row is the current frame; the second row is the segmented masks.



FIGURE 5: Some results for the *campus* sequence. The first row is the current frame; the second row is the segmented masks.

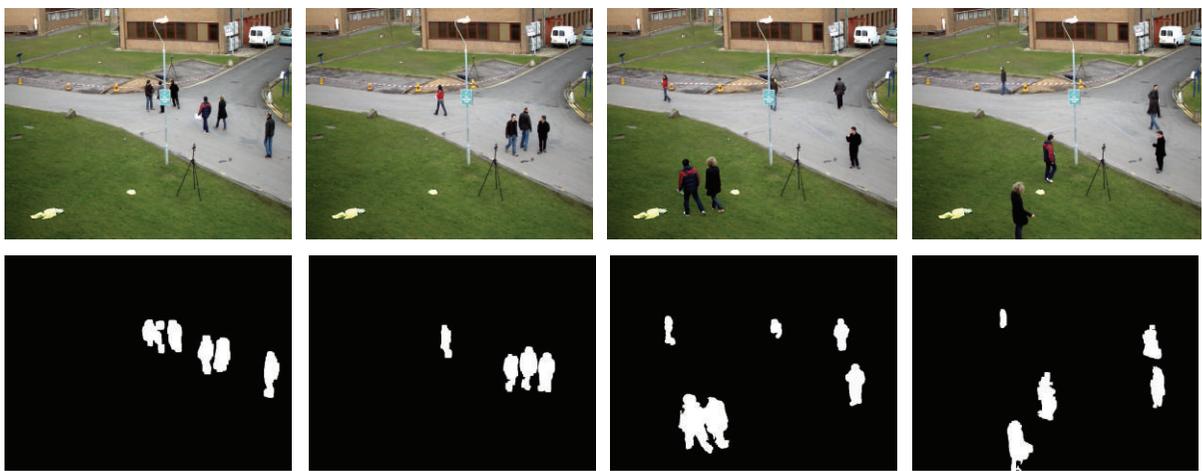


FIGURE 6: Some results for the *PETS'09* sequence with crowded people moving in the high-lights environment. The first row is the current frame; the second row is the segmented masks.

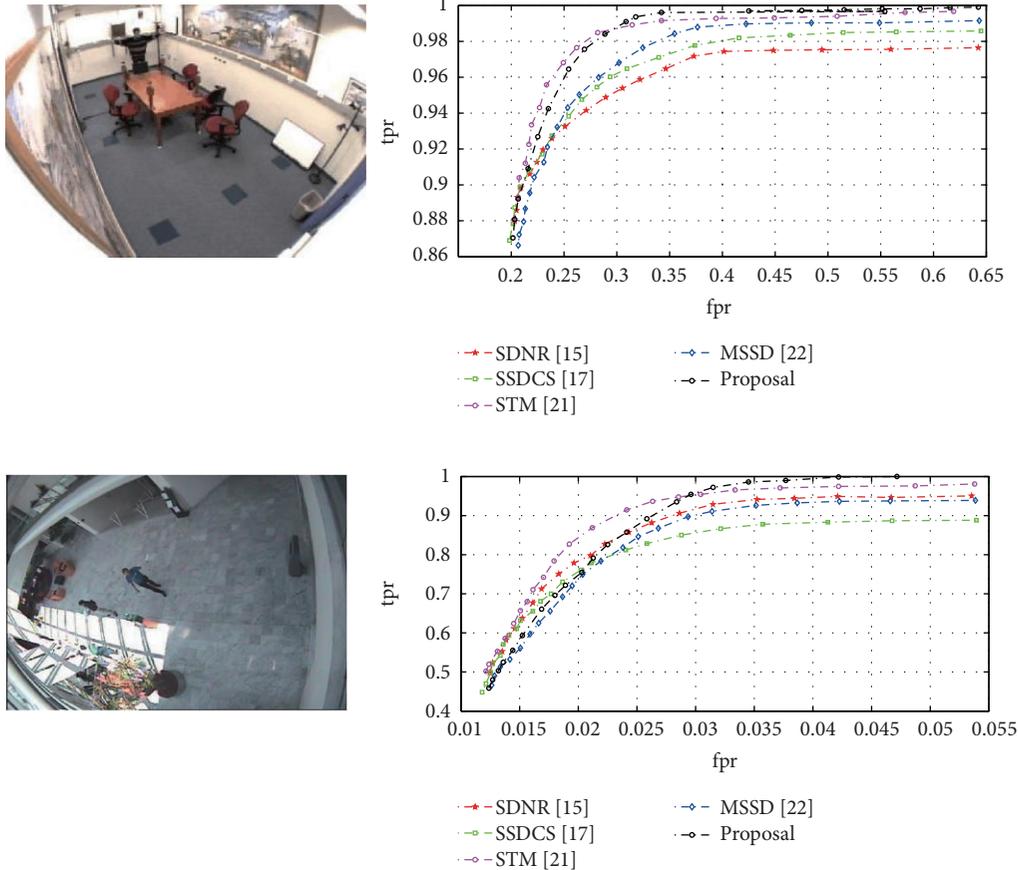


FIGURE 7: Comparisons in different approaches for the tested indoor sequences.

fighting in situ (the third column), and moving quickly (the last column). Some effects caused by the fluorescent lights turning on and off randomly have been suppressed also.

Car and people move in the scene and the lighting conditions are different from that of the indoor contexts mentioned above. One can find that the developed algorithm correctly identifies the car and people moving in different styles, and the shadow caused by outside light has been suppressed from Figure 5.

Some results of another outdoor scene *PETS'09* are shown in Figure 6, in which the crowded people move in high-lights and the lighting condition is different from that mentioned above.

The developed algorithm correctly identifies the crowded people moving in high lights from Figure 6.

**4.3. Objective Performance Evaluation.** The above provided results are quite good, considering different scenes with different lighting conditions and different objects moving in different styles. The above results show qualitative information about the effectiveness of the developed approach. To evaluate quantitatively the performance of the proposal and establish a fair comparison, the public video sequences mentioned above and some published articles are selected to perform comparisons. For tested sequences, ground-truth pixels are

segmented manually except the ground truth of *intelligent room* is available from <http://cvrr.ucsd.edu/aton/shadow>. The investigated approaches include Shadow Detection based on Normalized rgb (SDNR) [15], Space Selection for Detecting Cast Shadows (SSDCS) [17], SpatioTemporal Motion (STM) [21], and Motion Segmentation based on Shadow Detection (MSSD) [22]. The segmented outputs in the above investigated algorithms are compared to the ground-truth data. Some results of ROC [31] for both indoor and outdoor scenes are given in Figures 7 and 8, respectively. The ROCs for the tested indoor scenes are shown in Figure 7.

The performance of the proposed method outperforms that of investigated ones across the tested indoor scenes as a whole by comparisons from Figure 7.

The ROCs for the tested outdoor scenes are given in Figure 8.

The performance of proposed method outperforms that of investigated ones across the tested outdoor scenes by further comparisons from Figure 8.

## 5. Conclusions

The WT based temporal difference between two successive frames is used to extract motion regions and suppress shadow without any background learning. A new local adaptive

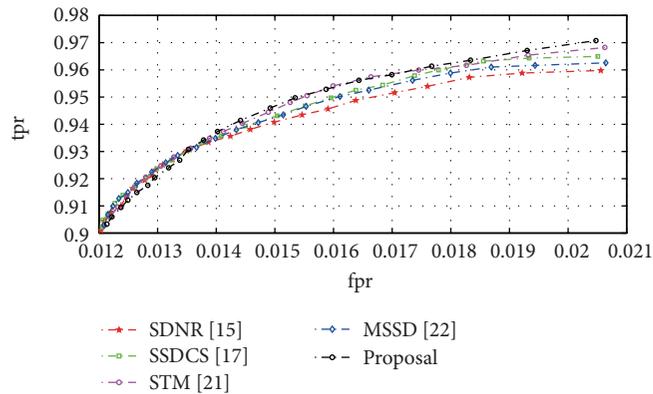
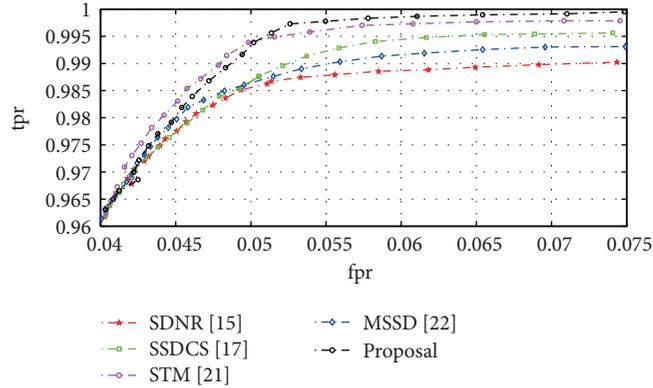


FIGURE 8: Comparison results for the tested outdoor sequences in different approaches.

thresholding approach is proposed based on the fact that a higher signal-to-noise ratio corresponds to truer boundary with the most meaningful features. Since the detected prominent boundaries are connected with foreground objects to a great extent, foreground reclassification is developed to get an optimal segmentation based on connectivity analysis and spatial-temporal correlation.

The experimental results both in different indoor and outdoor contexts show the robustness and reliability of the proposed algorithm. Comparative study with some state-of-the-arts has indicated the superior performance of the proposal. By comparisons, it has been highlighted that the proposal is robust and efficient in detecting motion objects and suppressing shadow for some different challenging contexts.

While the results have been promising so far, the algorithm needs to be tested on a larger number of visual environments with complex backgrounds further.

### Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This work is supported partly by the National Natural Science Foundation of China (Grant no. 11176016) and Specialized

Research Fund for the Doctoral Program of Higher Education (Grant no. 20123108110014).

### References

- [1] J. Cheng, J. Yang, Y. Zhou, and Y. Cui, "Flexible background mixture models for foreground segmentation," *Image and Vision Computing*, vol. 24, no. 5, pp. 473–482, 2006.
- [2] L. Cheng, M. Gong, D. Schuurmans, and T. Caelli, "Real-time discriminative background subtraction," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1401–1414, 2011.
- [3] T. Lim, B. Han, and J. H. Han, "Modeling and segmentation of floating foreground and background in videos," *Pattern Recognition*, vol. 45, no. 4, pp. 1696–1706, 2012.
- [4] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, vol. 2, pp. 246–252, Fort Collins, Colo, USA, June 1999.
- [5] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.
- [6] Y. Liu, H. Yao, W. Gao, X. Chen, and D. Zhao, "Nonparametric background generation," *Journal of Visual Communication and Image Representation*, vol. 18, no. 3, pp. 253–263, 2007.
- [7] A. Tavakkoli, M. Nicolescu, G. Bebis, and M. Nicolescu, "Non-parametric statistical background modeling for efficient foreground region detection," *Machine Vision and Applications*, vol. 20, no. 6, pp. 395–409, 2009.

- [8] N. Joshi and M. Brady, "Non-parametric mixture model based evolution of level sets and application to medical Images," *International Journal of Computer Vision*, vol. 88, no. 1, pp. 52–68, 2010.
- [9] C.-W. Ngo, T.-C. Pong, and H.-J. Zhang, "Motion analysis and segmentation through spatio-temporal slices processing," *IEEE Transactions on Image Processing*, vol. 12, no. 3, pp. 341–355, 2003.
- [10] J. Shao, Z. Jia, Z. Li, F. Liu, J. Zhao, and P. Y. Peng, "Spatiotemporal energy modeling for foreground segmentation in multiple object tracking," in *Proceedings of the International Conference on Robotics and Automation (ICRA '11)*, pp. 3875–3880, Shanghai, China, 2011.
- [11] Z. Kuang, H. Zhou, and K.-Y. K. Wong, "Accurate foreground segmentation without pre-learning," in *Proceedings of the 6th International Conference on Image and Graphics (ICIG '11)*, pp. 331–337, Hefei, China, August 2011.
- [12] W.-C. Hu, "Real-time on-line video object segmentation based on motion detection without background construction," *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 4, pp. 1845–1860, 2011.
- [13] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: algorithms and evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 918–923, 2003.
- [14] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," *Computer Vision and Image Understanding*, vol. 95, no. 2, pp. 238–259, 2004.
- [15] J. Gao, Y. Tian, and Y. Liu, "Ellipsoidal method for shadow detection based on normalized RGB values," in *Proceedings of the International Conference on Networking and Digital Society (ICNDS '09)*, vol. 1, pp. 104–107, Guiyang, China, May 2009.
- [16] A. Bevilacqua, "Effective shadow detection in traffic monitoring applications," *Journal of WSCG*, vol. 11, no. 1, pp. 57–64, 2003.
- [17] C. Benedek and T. Szirányi, "Study on color space selection for detecting cast shadows in video surveillance," *International Journal of Imaging Systems and Technology*, vol. 17, no. 3, pp. 190–201, 2007.
- [18] W. W. L. Lam, C. C. C. Pang, and N. H. C. Yung, "Highly accurate texture-based vehicle segmentation method," *Optical Engineering*, vol. 43, no. 3, pp. 591–603, 2004.
- [19] M. Heikkilä and M. Pietikäinen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 657–662, 2006.
- [20] W. Zhang, X. Z. Fang, X. K. Yang, and Q. M. J. Wu, "Moving cast shadows detection using ratio edge," *IEEE Transactions on Multimedia*, vol. 9, no. 6, pp. 1202–1214, 2007.
- [21] Y.-P. Guan, "Spatio-temporal motion-based foreground segmentation and shadow suppression," *IET Computer Vision*, vol. 4, no. 1, pp. 50–60, 2010.
- [22] M. Kampel, H. Wildenauer, P. Blauensteiner, and A. Hanbury, "Improved motion segmentation based on shadow detection," *Electronic Letters on Computer Vision and Image Analysis*, vol. 6, no. 3, pp. 1–12, 2007.
- [23] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [24] W. Niblack, *Introduction to Image Processing*, Prentice-Hall International, Englewood Cliffs, NJ, USA, 1986.
- [25] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern Recognition*, vol. 33, no. 2, pp. 225–236, 2000.
- [26] F. George and B. M. G. Kibria, "Confidence intervals for signal to noise ratio of a Poisson distribution," *The American Journal of Biostatistics*, vol. 2, no. 2, pp. 44–55, 2011.
- [27] I. Huerta, D. Rowe, M. Viñas, M. Mozerov, and J. González, "Background subtraction fusing colour, intensity and edge cues," in *Proceedings of the 2nd Computer Vision: Advances in Research and Development (CVCRD '07)*, pp. 105–110, Bellaterra, Spain, 2007.
- [28] M. Nawaz, J. Cosmas, P. I. Lazaridis, Z. D. Zaharis, Y. Zhang, and H. Mohib, "Precise foreground detection algorithm using motion estimation, minima and maxima inside the foreground object," *IEEE Transactions on Broadcasting*, vol. 59, no. 4, pp. 725–731, 2013.
- [29] J. V. de Weijer and R. V. den Boomgaard, "Local mode filtering," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 2, pp. II-428–II-433, Kauai, Hawaii, USA, December 2001.
- [30] C. Wang, X. Zhang, C. Yuan, L. Zhou, and Y. Liu, "Integrating local correlation and interframe continuity for robust foreground object segmentation," in *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology*, pp. 458–461, Vancouver, Canada, August 2006.
- [31] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 251–255, 2008.

