International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.8, No.3 (2015), pp.65-74
http://dx.doi.org/10.14257/ijsip.2015.8.3.06

Region-based Moving Shadow Detection using Affinity Propagation

Jiangyan Dai[1] and Dianyuan Han[2]

[1,2]*School of Computer Engineering, Weifang University, Weifang, China, 261061*
daijyan@163.com

Abstract

Moving shadow detection is a challenging task in computer vision applications, such as surveillance, video conference, visual tracking, object recognition, and many other important applications. In this paper, region-based moving shadow detection using affinity propagation (RMSDAP) is presented, which detects shadows in terms of texture similarity. Firstly, we divide foreground image into no overlapping blocks and extract color features from each block. Secondly, affinity propagation is utilized to cluster foreground blocks adaptively and sub regions are generated after coarse segmentation. Specially, each sub region has the characteristics of regional uniformity. Finally, we extract texture feature from irregular sub regions while calculate texture similarity and normalized correlation coefficient simultaneously for each sub region to classify moving shadows. Extensive experiments demonstrate that RMSDAP is superior to some well-known methods especially pixel-based methods. In particular, our method exhibits much better performance compared with fixed block method, which can maintain the texture consistency in one region adequately.

Keywords: *Moving shadow detection, region-based detection, affinity propagation, texture consistency*

1. Introduction

Shadows often share the same movement characteristic and have similar intensity change with that of moving objects, which will give rise to detect shadows as parts of moving objects. Therefore, moving cast shadow detection is a critical step for improving accuracy of moving object detection in computer vision.

In the past decade, many moving shadow detection methods have been studied, most of which are based on pixels [1-7]. Sanin *et al.* [1] proposed a comprehensive review about pixel-based methods and analyzed the advantages and disadvantages for each method. Cucchiara *et al.* [2] pointed that hue and saturation components varied a little for moving shadows which were darker than background in luminance component. Hence, they detected shadows with color information in HSV color space. Song and Tai [4] suggested color ratio model to detect shadow pixels and presented two kinds of spatial processing operations to enhance shadow detection accuracy. Khan *et al.* [6] constructed a framework to automatically detect shadows, which learned the most relevant features at super-pixel level and along the object boundaries in supervised manner. Wang *et al.* [7] employed online sub-scene shadow modeling learned by Gaussian functions and object inner-edges analysis to detect moving shadows adaptively. These methods achieved better performance and outperformed some state-of-the-art methods.

However, most of pixel-based methods are easily influenced by noise or uncertain factors which will lead to lower shadow detection accuracy. To solve this issue, many

ISSN: 2005-4254 IJSIP

researchers utilized local spatial correlation of pixels and a series of morphological operations to refine shadow detection results further. Nevertheless, with the increase of the number of pixels in foreground image, the computational complexity for pixel-based methods will be increased accordingly. Recently, region-based methods have been attracting much attention in moving shadow detection field. In terms of the nonlinear tone mapping existed between shadow patches and that of corresponding background patches, Bullkich *et al*. [8] adopted Matching by Tone Mapping as the metric to distinguish shadows from suspected foreground patches.

Inspired by the problem, we suggest a region-based moving shadow detection method which has better robustness compared to pixel-based methods in this paper. After dividing foreground image into no overlapping blocks, we extract color information in HSV color space for clustering. Subsequently, due to the complexity and uncertainty of video scenes, we employ affinity propagation to cluster foreground blocks adaptively and obtain coarse segmentation regions. Meanwhile, Gabor feature is extracted from each irregular sub region. At last, texture similarity based on Gabor feature and normalized correlation coefficients are calculated simultaneously to classify for each sub region. Aiming to evaluate the performance of RMSDAP, we test it on several different surveillance videos and compare it with some well-known method, which is demonstrated that RMSDAP exhibits excellent performance and outperforms some existing methods especially fixed block method.

The reminding parts of this paper are organized as follows. Section 2 describes the proposed method in detail. Experiments and comparisons are given in Section 3, and Section 4 is the conclusions.

2. The Proposed Method

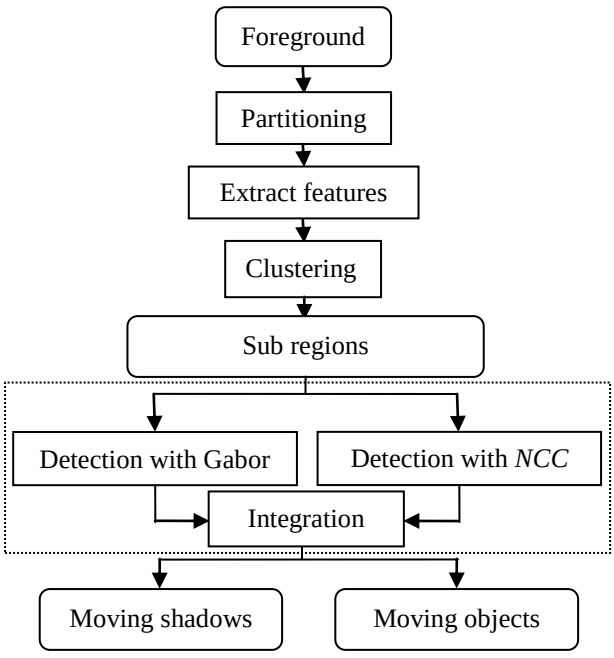


Figure 1. The Flowchart of Proposed Method

In this section, we describe the region-based moving shadow detection method using affinity propagation (RMSDAP) in detail. The flowchart of RMSDAP is illustrated in Figure 1. On the basis of foreground extraction, RMSDAP is composed of two parts: adaptive foreground segmentation and region-based shadow detection.

2.1. Adaptive Foreground Segmentation

To maintain consistency attributes in one region, adaptive foreground segmentation is taken into account in our study. Moreover, due to the complexity of video scenes, it is hard to determine the number of sub regions. Consequently, unsupervised clustering method named affinity propagation (AP) [9] is adopted to achieve foreground segmentation adaptively. Given foreground image *F*, the corresponding binary mask image is M^F, and the block size is set to $b \times b$.

2.1.1. Dividing Foreground into Blocks: Before adaptive segmentation, the foreground *F* is divided into regular blocks with size of $b \times b$, and each block is denoted as B_i^F . Meanwhile, the binary mask image M^F is also divided into blocks with the same size. According to the following rule, we extract blocks which contain foreground pixels from all blocks and put them into the set S_F. Initially, $S_F = \varnothing$.

$$S_F = S_F \cup B_i^F \quad s.t. \sum_{n_1=0}^{n} \sum_{n_2=0}^{n} B_i^M (n_1, n_2) > 0 \tag{1}$$

Where $n_1$and n_2 are coordinates of pixels in *i*th block B_i^M .

2.1.2. Feature Extraction: In HSV color space, color feature *fea* are extracted from blocks in S_F. Subsequently, mean and variance are employed to describe color information.

$$m_i^c = Mean(B_i^{F^c}) \tag{2}$$

$$v_i^c = Var(B_i^{F^c}) \tag{3}$$

Where $i = 1, \cdots, N$, N is the number of blocks in S_F , m_i^c and v_i^c are the mean and variance of *i*th B_i^F at channel *c* in HSV color space, $c \in \{H, S, V\}$. Therefore, fea_i of *i*th block is defined:

$$fea_i = [m_i^1, m_i^2, m_i^3, v_i^1, v_i^2, v_i^3] \tag{4}$$

Clearly, the dimension of fea_i is 6.

2.1.3. Adaptive Clustering: As a result of complexity and uncertainty for video scenes, AP is introduced to cluster foreground blocks automatically. The advantage of AP is that it does not need to determine the clustering number beforehand. The feature similarity of blocks in S_F is:

$$sim(fea_i, fea_j) = -\left(\sqrt{\left| fea_i - fea_j \right|^2} + \sqrt{(fea_i.x - fea_j.x)^2 + (fea_i.y - fea_j.y)^2} \right) \tag{5}$$

Where fea_i and fea_j are the features of *i*th and *j*th blocks, respectively. *Sim* (fea_i, fea_j) is similarity derived from *i*th and *j*th blocks. $fea_i.x$、 $fea_i.y$ and $fea_j.x$、 $fea_j.y$ are coordinates of *i*th and *j*th blocks. During computing the similarity, not only the similarity between features is taken into account, but also position information of blocks is considered. The larger *sim* (fea_i, fea_j) is the higher similarity between fea_i and fea_j and the closer between *i*th block and *j*th block is.

After implementing AP algorithm for foreground blocks, we can obtain several irregular sub regions which can be regarded as coarse segmentation result of foreground. Each sub region is denoted as Sub_i^F. And then, for each region Sub_i^F, we use texture similarity and normalized correlation coefficient to determine it whether is shadow or not.

2.2. Region-based Shadow Detection

In the process of region-based shadow detection, texture similarity obtained by Gabor feature and normalized correlation coefficients are used to calculate the similarity between foreground sub region Sub_i^F and background sub region Sub_i^B. Subsequently, we determine whether the sub region belongs to shadow or not.

2.2.1. Shadow Detection based on Texture Similarity: Gabor filter [10] is widely applied in the fields of image processing, computer vision and pattern recognition. The Gabor feature of foreground sub region Sub_i^F is calculated by:

$$g_{mn}^F(x,y) = \sum_{w}^{W-1}\sum_{h}^{H-1} Sub_i^F(x-w, y-h) G_{mn}(w,h) \tag{6}$$

Where W´ H the size of Gabor is kernel G_{mn}, m and n denote the scale and orientation of G_{mn}, respectively. Likewise, Gabor feature of background sub region Sub_i^B is denoted as g_{mn}^B.

Thus, Gabor energy of ith region Sub_i^F at scale m with orientation n is:

$$E_{Sub_i^F}(m,n) = \sum\sum g_{mn}^F(x,y) \tag{7}$$

Then, we adopt mean and variance to describe Gabor energy for each sub region, as follows:

$$\mu_{mn}^{Sub_i^F} = \frac{E_{Sub_i^F}(m,n)}{\left|Sub_i^F\right|} \tag{8}$$

$$\sigma_{mn}^{Sub_i^F} = \frac{\sqrt{\sum\sum\left(\left|g_{mn}^F(x,y)\right| - \mu_{mn}^{Sub_i^F}\right)^2}}{\left|Sub_i^F\right|} \tag{9}$$

Where $\left|Sub_i^F\right|$ denotes the number of pixels in Sub_i^F. In the same way, the mean $\mu_{mn}^{Sub_i^B}$ and variance $\sigma_{mn}^{Sub_i^B}$ of background sub region Sub_i^B are also can be computed. To compare the texture difference between Sub_i^B and Sub_i^F, Euclidian distance is used to calculate the similarity:

$$dist(Sub_i^F, Sub_i^B) = \sum_{m}^{M-1}\sum_{n}^{N-1}\sqrt{\left(\mu_{mn}^{Sub_i^B} - \mu_{mn}^{Sub_i^F}\right)^2 + \left(\sigma_{mn}^{Sub_i^B} - \sigma_{mn}^{Sub_i^F}\right)^2} \tag{10}$$

Where $dist(Sub_i^F, Sub_i^B)$ is the texture similarity. The smaller $dist(Sub_i^F, Sub_i^B)$ is, the higher similarity between Sub_i^B and Sub_i^F is.

Finally, according to the following rule, we determine each region Sub_i^B whether belongs to shadow or not.

$$M_i^1 = \begin{cases} 1 & dist(Sub_i^F, Sub_i^B) < T_1 \\ 0 & otherwise \end{cases} \tag{11}$$

Where $M_i^1 = 1$ indicates that the ith sub region Sub_i^F belongs to shadow whereas $M_i^1 = 0$ implies that the region is object. T_1 is the threshold determined empirically.

2.2.2. Shadow Detection based on Normalized Correlation Coefficient: Normalized correlation coefficient (*NCC*) has better performance for evaluating the similarity between foreground and background. In our work, *NCC* is computed as:

$$ncc(Sub_i^F, Sub_i^B) = \frac{\sum_{(x,y)\in\Omega_i} Sub_i^F(x,y) \cdot Sub_i^B(x,y)}{\sqrt{\sum_{(x,y)\in\Omega_i} Sub_i^F(x,y) \cdot Sub_i^F(x,y) \cdot \sum_{(x,y)\in\Omega_i} Sub_i^B(x,y) \cdot Sub_i^B(x,y)}} \quad (12)$$

Where Ω_i is the set of pixel coordinates in ith sub region, $ncc(Sub_i^F, Sub_i^B)$ is similarity coefficient of ith foreground sub region Sub_i^F and background sub region Sub_i^B. The larger $ncc(Sub_i^F, Sub_i^B)$ is the higher similarity between Sub_i^B and Sub_i^F is and vice versa. Therefore, we can judge ith sub region is shadow or not according to the following rule:

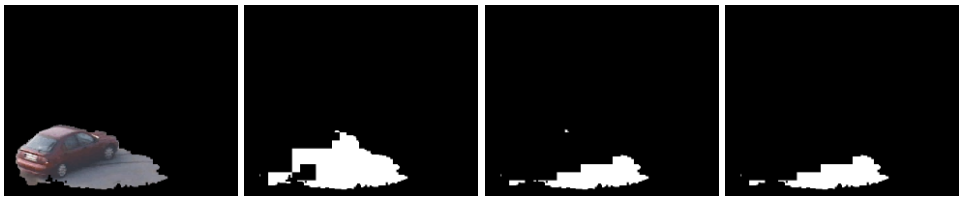
$$M_i^2 = \begin{cases} 1 & ncc(Sub_i^F, Sub_i^B) > T_2 \\ 0 & otherwise \end{cases} \quad (13)$$

Where $M_i^2 = 1$ states that Sub_i^F is shadow. T_2 is the threshold determined empirically.

Figure 2 shows one foreground image from Campus sequence, and shadow detection results obtained from texture similarity and *NCC*, respectively. Obviously, single feature cannot fully classify all shadow regions correctly. In order to obtain a more complete result, the two results are integrated effectively, which is shown in Figure 2(d). The classification decision is:

$$M_i^S = M_i^1 \wedge M_i^2 \quad (14)$$

Where M_i^S is the binary shadow mask corresponding to Sub_i^F . $\wedge$ denotes logical and operation.



(a) (b) (c) (d)

Figure 2. Shadow Detection Results (a) Foreground, (b) Result Obtained by Texture Similarity, (c) Result Obtained by NCC (d) Result from (b) and(c)

3. Experiments and Comparisons

In this section, to evaluate the performance of RMSDAP, experiments are tested on several challenging video sequences [11-13] known in the literatures. One side, we discuss and determine the suitable block size for different videos. One the other side, we compare our shadow detection results with results from fixed block method and some well-known methods from the quantitative and qualitative aspects.

The process of fixed block method is: firstly, dividing foreground image into blocks with $b \times b$ and then extracting Gabor and *NCC* features for shadow detection, respectively. Lastly, we integrate the results effectively for the final result. Different

from RMSDAP, fixed block method do not utilize affinity propagation for coarse segmentation regions before shadow detection.

In order to verify the effectiveness of RMSDAP, three metrics are applied to analyze shadow detection performance, named as shadow detection rate (η), shadow discrimination rate (ζ) and weighted accuracy (γ) [14]:

$$\eta = \frac{TP_S}{TP_S + FN_S} \times 100\%, \ \zeta = \frac{TP_O}{TP_O + FN_O} \times 100\%, \ \gamma = \frac{GT_S}{GT} \cdot \eta + \frac{GT_O}{GT} \cdot \zeta \quad (15)$$

Where subscripts S and O denote shadow and object, respectively. TP_S and TP_O are the number of shadow and object pixels classified correctly. FN_S and FN_O are the number of shadow and object pixels detected incorrectly. $GT = GT_S + GT_O$, $GT_S = TP_S + FN_S$ and $GT_O = TP_O + FN_O$.

3.1. Parameter Selection

During adaptive image segmentation, it is necessary to determine the block size primarily. Considering the shadow size in different videos and clustering speed, block size is set to 8×8, 9×9, 10×10, 11×11 and 12×12, respectively. Figure 3 indicates the weighted accuracy trends derived from different block size in various videos. Particularly, in Hallway and Highway, shadow size is small. In these sequences, choosing relatively small block size can better maintain the consistency attributes in one region. While CAVIAR and Campus have larger shadow, in which choosing relatively larger block size not only can retain the consistency attributes in one region, but also can improve the clustering speed. Obviously, the similar trends are shown in Figure 3. In our experiments, the weighted accuracy reaches the top when block size of Hallway, Highway, Campus and CAVIAR is set to 8×8, 9×9, 12×12 and 11×11, respectively.

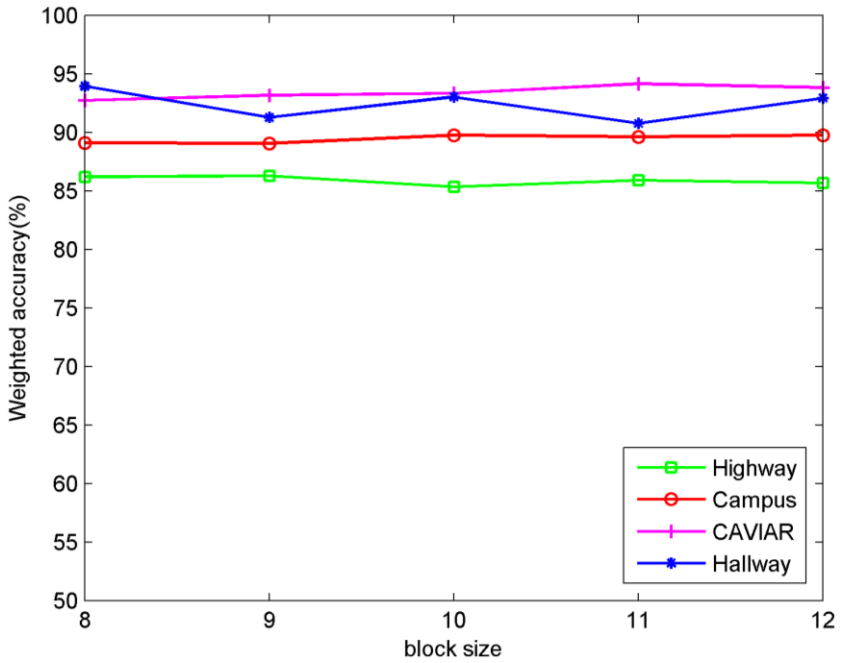


Figure 3. The Weighted Accuracy Trends with Various Block Sizes

3.2. Quantitative Comparisons

Table 1 lists the shadow detection rate η, shadow discrimination rate ζ and weighted accuracy γ resulting from RMSDAP, fixed block method and traditional methods (DNM [2], ICF [3], SNP [4], CCM [5], and MTM [8]). From the aspects of weighted accuracy to analyze, it can be seen that RMSDAP is superior to fixed block method and traditional methods in Hallway, Highway and Campus. In CAVIAR, the weighted accuracy γ obtained from RMSDAP is only lower than MTM with 0.30%. Specifically, RMSDAP has higher weighted accuracy than fixed block method about 2.61%, 17.67%, 5.44% and 0.56%, which demonstrates that the validity of adaptive segmentation for foreground image.

Table 1. Quantitative Comparison Results

Videos	Metric	DNM	ICF	SNP	CCM	MTM	Fixed block	RMSDAP
Hallway	η	83.73	95.32	83.93	96.46	85.16	83.73	88.29
	ζ	79.61	83.01	98.10	68.55	83.18	94.71	96.36
	γ	81.77	88.19	92.94	79.92	85.79	91.33	**93.94**
Highway	η	86.84	84.54	86.18	87.65	70.14	30.87	73.38
	ζ	58.31	60.93	58.07	36.37	95.14	91.16	94.69
	γ	69.36	0.19	68.71	56.18	84.60	68.61	**86.28**
CAVIAR	η	93.24	92.76	88.12	87.45	92.16	91.14	89.68
	ζ	78.80	88.56	97.60	94.77	95.96	88.08	96.42
	γ	84.22	90.01	94.07	91.64	**94.45**	88.71	94.15
Campus	η	52.84	56.22	65.85	62.56	53.23	42.56	33.78
	ζ	90.44	82.74	75.36	43.07	81.36	93.42	96.52
	γ	87.67	82.36	78.49	52.24	81.95	89.18	**89.76**

3.3. Qualitative Comparisons

To validate the effectiveness and superiority of RMSDAP intuitively, comparison results are displayed in Figure 4, with the detected moving shadows and moving objects being marked in blue and white, respectively. Figure 4 (a) and Figure 4 (b) show original frames and their ground truths. For sake of indicating the superiority of RMSDAP shown in Figure 4 (g), we compare the results with well-known pixel-based methods such as ICF and SNP shown in Figure 4(c)-(d) and patch-based method MTM shown in Figure 4 (e). In addition, results in Figure 4(f) derived from fixed block method verify the performance of RMSDAP further. Moreover, we can see that MTM, fixed block method and RMSDAP have a certain resistance to noise or uncertain factors. However, the detected results from fixed block method have more incorrect regions. In contrast, RMSDAP can guarantee the region consistency adequately and achieve well results, which is owing to RMSDAP adopts adaptive clustering method for coarse segmentation.

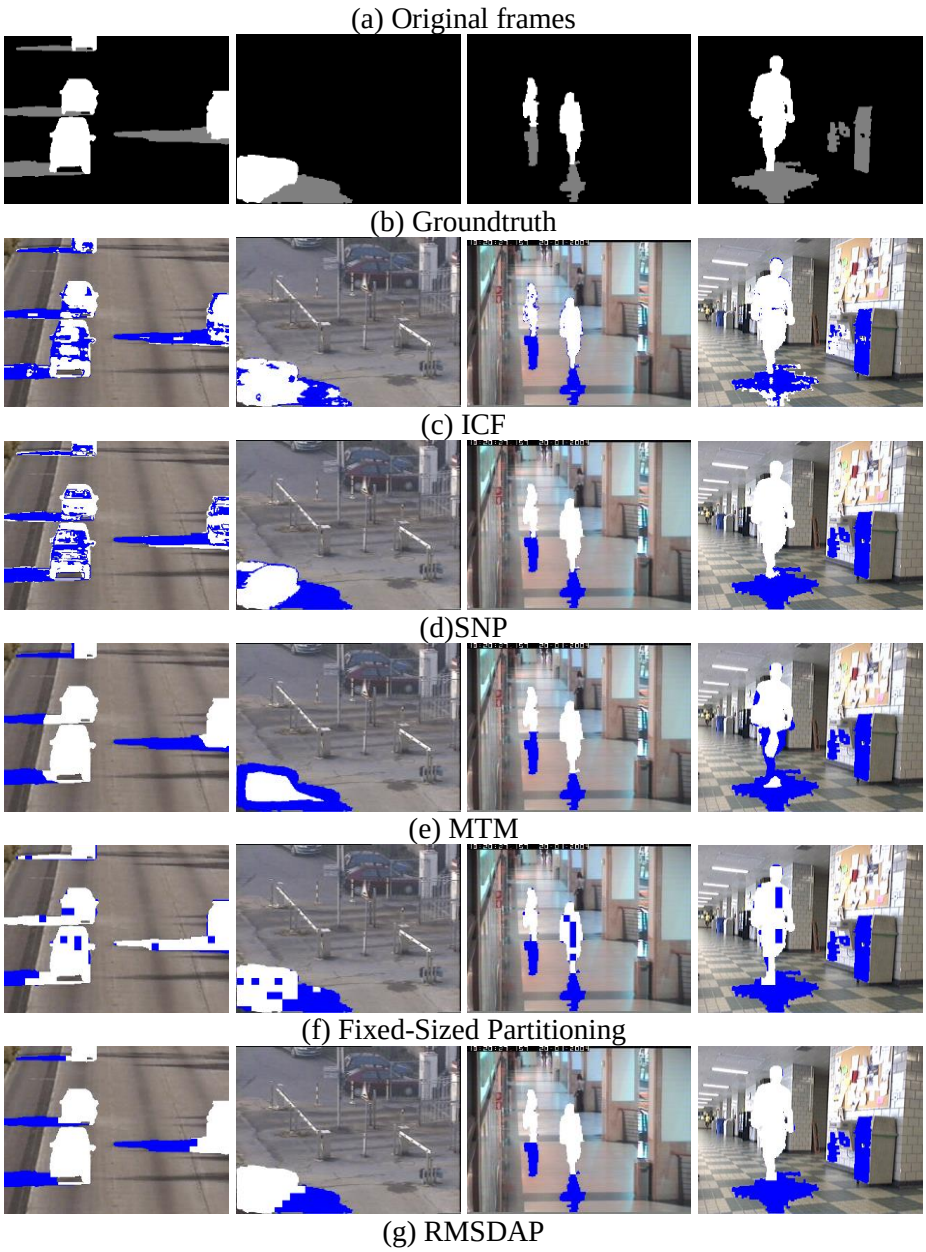


Figure 4. The Qualitative Comparison Results

4. Conclusions

Different from pixel-based methods, RMSDAP detect moving shadows on the basis of adaptive segmentation regions derived from affinity propagation. The proposed RMSDAP can reduce the influence of noise or uncertain factors for detection results, which has good robustness for surveillance videos. Besides, experiment results demonstrate that RMSDAP is superior to fixed block method and some well-known method.

Acknowledgements

This research is partially supported by National Natural Science Foundation of China (61403077), the Project of Shandong Province Higher Educational Science and Technology Program (NO.J13LN39), the Project of Shandong Province Science and Technology Development Program (NO.2011YD01047) and the Project of Shandong Province Spark Program (2013XH06031).

References

[1] A. Sanin, C. Sanderson and B. C. Lovell, "Pattern Recognition", vol. 45, no. 4, (**2012**).
[2] R. Cucchiara, C. Grana, M. Piccardi and A. Prati, "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 25, no. 10, (**2003**).
[3] E. Salvador, A. Cavallaro and T. Ebrahimi, "Computer Vision and Image Understanding, vol. 95, no. 2, (**2004**).
[4] K. T. Song and J. C. Tai, "Proceedings of the IEEE", vol. 95, (**2007**) February 2.
[5] R. Qin, S. Liao, Z. Lei and S. Li, "Moving cast shadow removal based on local descriptors", in: *International Conference on Pattern Recognition*, (**2010**), pp. 1377-1380.
[6] S H Khan, M Bennamoun and F. Sohel, "Automatic feature learning for robust shadow detection", 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (**2014**), pp. 1939-1946.
[7] J. Wang, Y. Wang and M. Jiang, "Journal of Visual Communication and Image Representation", vol. 25, no. 5, (**2014**).
[8] E. Bullkich, I. Ilan, Y. Moshe, Y. Hel-Or, and H. Hel-Or. Proc. of 19th Intl. Conf. on Systems, Signals and Image Processing (IWSSIP 2012), (2012) April;Vienna.
[9] B. J. Frey and D. Dueck, "Science", vol. 315, (**2007**).
[10] J. G. Daugman, "Optical Society of America", Journal, A: Optics and Image Science, vol. 2, no. 7, (**1985**).
[11] http://vision.gel.ulaval.ca/~CastShadows/
[12] http://cvrr.ucsd.edu/aton/shadow/
[13] http://homepages.inf.ed.ac.uk/rbf/CAVIAR/
[14] J. Y. Dai, M. Qi and J. Z. Wang, "Optics & Laser Technology", vol. 54, (**2013**).

Authors

Jiangyan Dai, she was born in Weifang, Shandong province, China, in 1985. She received her BS, MS and PhD degrees from Computer School of Northeast Normal University, China, in 2008, 2010 and 2014, respectively. She now works in school of computer engineering, Weifang University. Her main research interests are digital image processing, computer vision, biometrics and information security.

Dianyuan Han, he was born in Weifang, Shandong province, China, in 1971. He received his BS degree from Information School of Shandong Normal University, China, in 1997 and received his MS degrees from Information School of Shandong University of Science and Technology, China, in 2004. In 2013, He received his PhD degree from Information School of Beijing Forestry University, China. His main research interests are digital image processing and computer vision