Adaptive Moving Cast Shadow Detection

Guizhi Li[1], Lei Qin[2,3], Qingming Huang[1,2,3],

[1]Graduate University of Chinese Academy of Sciences, Beijing, 100049, China
[2]Key Lab of Intell. Info. Process., Chinese Academy of Sciences, Beijing 100190, China
[3]Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
E-mail: {gzli, lqin, qmhuang}@jdl.ac.cn

Abstract. Moving object detection is an important task in real-time video surveillance. However, in real scenario, moving cast shadows associated with moving objects may also be detected, making moving cast shadow detection a challenge for video surveillance. In this paper, we propose an adaptive shadow detection method based on the cast shadow model. The method combines ratio edge and ratio brightness, and reduces computation complexity by the cascading algorithm. It calculates the difference of ratio edge between the shadow region and the background according to the invariability of the ratio edge of object in different light. Experimental results show that our approach outperforms existing methods.

Keywords: Shadow Detection, Ratio Edge, Ratio Brightness, Video Surveillance.

1. Introduction

In recent years, moving objects detection is a hot topic in computer vision, with a lot of applications such as surveillance system, vehicle tracking and video conferences. However, when detecting moving objects in real world scenes, the cast shadow associated with the moving object is also detected. For instance, cars may cause the sun to cast shadows on the road. Labeling the cast shadow is favorable to detect the accurate shape of object, thus shadow detection is a key issue unavoidable.

There are many approaches for shadow detection. Usually, they can be divided into three categories: color-based, statistic-based, and texture-based.

The color-based methods attempt to describe the color feature change of shadow pixels. Cucchiara et al. [4] operate brightness, saturation, and hue properties in the HSV color space. For avoiding using the time consuming HSV color space transformation, Schreer et al. [5] adopt the YUV color space, and distinguish the shadow regions from the foreground regions according to the observation that the YUV pixel value of shadows is lower than the linear pixels. According to the shadow model, Salvador et al. [3] identify an initial set of shadow pixels basing on RGB color space, according to the fact that shadow region darkens the surface; they combined color invariance with geometric properties of shadow. Though color-based methods have shown its power in shadow detection, they may not be reliable to detect moving shadows when the moving objects have similar color with moving shadows,

The principle of statistic-based methods is to build pixel-based statistical models detecting cast shadows. In [6], Zivkovic et al. using Gaussian Mixture Model (GMM) to detect moving cast shadows. The method consists of building a GMM for moving objects, identifying the distribution of moving objects and shadows, and modifying the learning rates of the distributions. In [8], Nicolas et al. propose Gaussian Mixture Shadow Model (GMSM). The algorithm models moving cast shadows of non-uniform and varying intensity, and builds statistical models to segment moving cast shadows by using the GMM learning ability. Claudio et al. [9] also use a statistical approach combined with geometrical constraints for detecting and removing shadows, but it is only for gray-scale video sequences. The statistic-based methods identify distribution of shadow pixel value and are robust in different scenes. Though the methods reduce false hits of property descriptions (i.e. color or intensity) of shadow, they cannot eliminate them. Generally, they need to be initialized by property descriptions.

The texture-based methods are based on the fact that the texture of shadow region remains same as that of the background, while the texture of moving object is different with that of the background. Zhang et al. [11] explore ratio edges for shadow detection. They prove that the ratio edge is illumination invariant. The local ratios are modeled as a chi-squared distribution in shadow districts. In addition to using scene brightness distortion and chromaticity distortion, Choi et al. [1] proposes three estimators which use the properties of chromaticity, brightness, and local intensity ratio. The chromaticity difference obeys a standard normalize distribution between the shadow region and the background. The texture-based methods may be the most promising technique for shadow detection, the textual information of different scenes can be captured by the above methods. However, the above methods have complex progress about computation.

In this paper, we propose an adaptive shadow detection method based on the cast shadow model. The method is made up of ratio edge and ratio brightness, and reduces complexity by the cascading algorithm. Though this paper uses ratio edge that has been defined in [11], it is different from calculating the chi-square distribution of ratio edge in [11]. In this paper, we calculate the difference of ratio edge between the shadow region and the background according to the invariability of the ratio edge of object in different lighting condition. Experimental results show that our approach outperforms existing methods.

The paper is organized as follows. Section 2 describes the cast shadow model. Section 3 presents the overall process, which includes the ratio edge, ratio brightness, and spatial adjustment. Section 4 shows the experimental results, and the conclusion is presented in section 5.

2. Cast Shadow Model

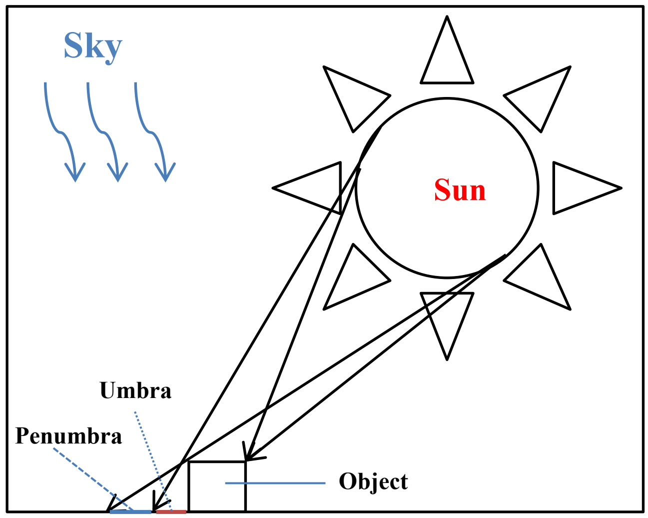

Fig.1. Cast Shadow Model

In the real world, shadows can be generally divided into static shadow and dynamic shadow. Static shadows are cast by static object such as buildings, trees, parked cars etc. Dynamic shadows in video are cast by moving objects such as moving vehicles, pedestrians, etc. Generally speaking, the illumination can be divided into two kinds: a light source (i.e. Sun) and an ambient light (i.e. Sky) , as shown in Fig. 1. Shadow is deprived the direct light source by the foreground object. Each pixel of shadow has the same ambient light; the pixels which do not belong to the shadow not only have the ambient light, but also have the same light source.

The color intensity of a pixel can be given by

$$f(x,y) = r(x,y) * p(x,y) \tag{1}$$

where $f(x,y)$is the color intensity, $r(x,y)$ is the illumination function and $p(x,y)$ is the reflectance factor. The illumination function $r(x,y)$ is expressed as a function of the intensity of the light source c_l, the intensity of the ambient light c_a, the direction L of the light, and the object surface normal [14]:

$$r(x,y) = \begin{cases} c_a + c_l * \cos(L, n(x,y)) & illuminated\ area \\ c_a + i(x,y) * c_l * \cos(L, n(x,y)) & penumbra\ area \\ c_a & umbra\ area \end{cases} \tag{2}$$

where $i(x,y)$ represents the transition inside the penumbra between shaded and illuminated regions and $i(x,y) \in [0,1]$.

We define the neighboring region of a pixel as follows:

$$\Omega(x,y) = \{f(x+i, y+i) | 0 < i^2 + j^2 \le r^2\} \tag{3}$$

where r represents the radius of the neighboring region. In this paper, the ratio edge of pixel $f(x,y)$ is then defined as follows [11]:

$$R(x,y) = \sum_{(i,j)\in\Omega(x,y)} \frac{f(x,y)}{f(i,j)} \tag{4}$$

We consider $i(x,y)$, L, $n(x,y)$ and c_a to be constant in one neighboring region. Let R_l , R_p, and R_u be the ratio edge results of the same object when the region is respectively in illuminated area, penumbra area, and umbra area. By [11], we can get :

$$R_l(x,y) = R_p(x,y) = R_u(x,y). \tag{5}$$

Equation (5) represents that the ratio edge of the object keeps unchanged in different lighting conditions. Then we define ratio edge difference as follows:

$$C(x,y) = R_F(x,y) - R_B(x,y) \tag{6}$$

where $R_F(x,y)$ is the ratio edge result of source frame, $R_B(x,y)$ is the ratio edge of background image. According to equation (4) and (5), if a pixel belongs to the shadow region and there is noise in an image, we can get $C_s(x,y) \sim N(0, \varepsilon(x,y)^2)$.

We assume that the shadow pixels are corrupted with the illuminated pixels and Gaussian white noise, according to equation (1) and (2) :

$$f_S(x,y) = \alpha(x,y) f_B(x,y) + \lambda(x,y),\ \lambda(x,y) \sim N(0, \sigma(x,y)^2) \tag{7}$$

where $f_B(x,y)$ is the intensity of a pixel in the background image, $f_S(x,y)$ is the intensity of a pixel in the shadow region of the source frame, $0 \le \alpha(x,y) \le 1$ relates to the intensity of shadow pixels, $\lambda(x,y)$ is Gaussian noise. Then we can define ratio brightness as follows:

$$B(x,y) = \frac{f_F(x,y)}{f_B(x,y)} \tag{8}$$

where $f_F(x,y)$ is pixel of the source frame, $f_B(x,y)$ is pixel of the background image. If we consider that $\alpha(x,y) \approx \alpha$ is constant within a neighboring region, the ratio brightness of shadow pixel becomes

$$B_s(x,y) = \frac{f_S(x,y)}{f_B(x,y)} = \theta(x,y) \tag{9}$$

where $\theta(x,y) = \alpha + \lambda(x,y)/f_B(x,y) \sim N(\alpha, (\sigma(x,y)/f_B(x,y))^2)$.

3. The Procedure of Shadow Detection

In this section, we present the detailed procedures of shadow detection. The proposed method is a multi-stage approach and the flow chart is shown in Fig.2.

3.1. Moving Object Detection

In this step, GMM [10] is used to estimate background image. We detect the foreground by subtracting the estimated background image from the current frame. In the set of foreground pixels, there are moving objects pixels as well as the shadow pixels. In this paper, we use the following notations: in frame n, F_n is the binary mask of foreground

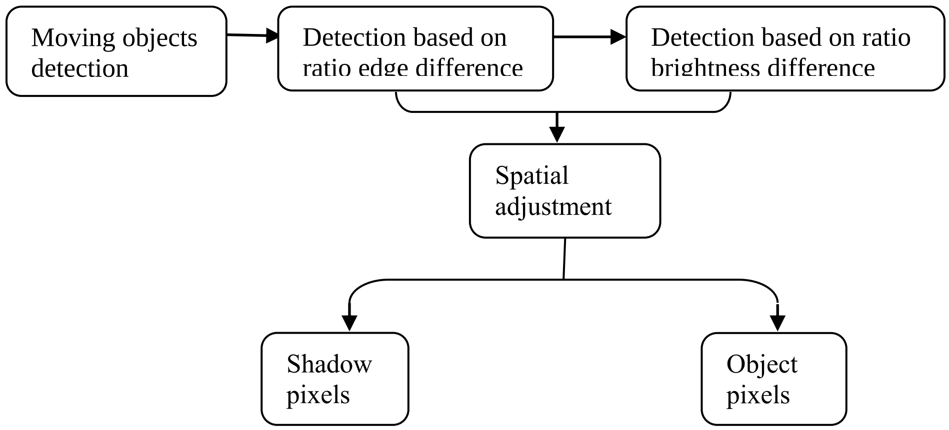

Fig.2. Flow Chart of the Algorithm

pixels to be detected, S_n is the binary mask of moving shadows to be detected, and O_n is the binary mask of moving objects.

At first, we assume all the foreground pixels are shadows pixels, and will discriminate the shadow pixels and object pixels in following steps:

$$S_n = F_n = 1 \ , O_n = 0 \tag{10}$$

3.2. Moving Shadow Detection Based on Ratio Edge Difference

We use the equation (6) for determining whether a pixel is a moving object pixel or shadow pixel. Generally, if a pixel belongs to the shadow, the probability that the neighboring pixels of the pixel belong to shadow region is higher, and the ratio edge differences are near zero. We can define $\mu_{C(x,y)}$ as follows:

$$\mu_{C(x,y)} = \frac{1}{N_S} \sum_{(i,j)\epsilon\Omega_S(x,y)} C(i,j) \tag{11}$$

where N_S is the number of pixels in Ω_S. According to the equation (4), $\mu_{C(x,y)}$ of the shadow pixel has a normalized Gaussian distribution. However, we cannot confirm the distribution of moving object pixel.

We can initially determine shadow pixels by $\mu_{C(x,y)}$, because $\mu_{C(x,y)}$ of shadow pixel is nearby zero. Therefore, a pixel can be identified as an object pixel as follows:

$$O_n(x,y) = 1 \ ; \quad S_n(x,y) = 0$$
$$if \ \left(\left(\left(\left|\mu^R_{C(x,y)}\right| > \mu^R_S\right)\middle|\left(\left|\mu^G_{C(x,y)}\right| > \mu^G_S\right)\middle|\left(\left|\mu^B_{C(x,y)}\right| > \mu^B_S\right)\right) \& S_n(x,y) = 1\right) \tag{12}$$

where $\mu^K_{C(x,y)}$ is $\mu_{C(x,y)}$ of each channel in RGB color space; μ^K_S is the threshold of each channel, and $0 < \mu^K_S < 1$.

We can determine object pixels from the scale of the whole foreground. To reduce computation, we only use pixels which satisfy

$$\left|\mu_{C(x,y)}^{K}\right| \leq \mu_{S}^{K} \tag{13}$$

Then, we estimate $m_{\mu_C}^{K}$ and $\sigma_{\mu_C}^{K}$ using max likelihood estimation (MLE) [7] as follows:

$$m_{\mu_C}^{K} = \frac{1}{N_C}\sum_{(x,y)\epsilon C}\mu_{C(x,y)}^{K} \quad ; \quad \left(\sigma_{\mu_C}^{K}\right)^2 = \frac{1}{N_C}\sum_{(x,y)\epsilon C}\left(\mu_{C(x,y)}^{K} - m_{\mu_C}^{K}\right)^2 \tag{14}$$

where C is the set pixels that satisfy (13) in the set of ($S_n(x,y) = 1$), and N_C is the number of pixel in C.

We compute the threshold α_h^K and α_l^K using the estimated $m_{\mu_C}^{K}$ and $\sigma_{\mu_C}^{K}$ as follows:

$$\alpha_h^K = m_{\mu_C}^{K} + 1.96 * \sigma_{\mu_C}^{K} \; ; \; \alpha_l^K = m_{\mu_C}^{K} - 1.96 * \sigma_{\mu_C}^{K} \tag{15}$$

with reliability of 95%; then the pixel is determined as an object pixel as follows:

$$O_n(x,y) = 1 \; ; \quad S_n(x,y) = 0$$

$$if\left(\begin{pmatrix}\left(\mu_{C(x,y)}^{R} > \alpha_h^R\right)\left|\left(\mu_{C(x,y)}^{R} < \alpha_l^R\right)\right|\left(\mu_{C(x,y)}^{G} > \alpha_h^G\right)\right| \\ \left(\mu_{C(x,y)}^{G} < \alpha_l^G\right)\left|\left(\mu_{C(x,y)}^{B} > \alpha_h^B\right)\right|\left(\mu_{C(x,y)}^{B} < \alpha_l^B\right)\end{pmatrix} \& S_n(x,y) = 1\right) \tag{16}$$

3.3. Moving Shadow Detection Based on Ratio Brightness Difference

In this part, we use the equation (9) for determining whether a pixel is a moving object pixel or shadow pixel. An analogous process about ratio edge difference can be applied on ratio brightness difference. We define $\mu_{B(x,y)}$ as follows:

$$\mu_{B(x,y)} = \frac{1}{N_S}\sum_{(i,j)\epsilon\Omega_S(x,y)} B(i,j) \; , \tag{17}$$

According to the equation (9), $\mu_{B(x,y)}$ of a pixel in shadow region is Gaussian distribution. We can find that $\mu_{B(x,y)}$ generally should be less than 1.0 by (7), thus a pixel is estimated as an object pixel as follows:

$$O_n(x,y) = 1 \; ; \quad S_n(x,y) = 0$$

$$if\left(\begin{pmatrix}\left(I_{low} < \mu_{B(x,y)}^{R} < I_{high}\right)\left|\left(I_{low} < \mu_{B(x,y)}^{G} < I_{high}\right)\right| \\ \left(I_{low} < \mu_{C(x,y)}^{B} < I_{high}\right)\end{pmatrix} \& S_n(x,y) = 1\right) \tag{18}$$

where $\mu_{B(x,y)}^{K}$ represent $\mu_{B(x,y)}$ in R, G or B; I_{low} and I_{high} represent the threshold, and $0 < I_{low} < 1,\; 1 < I_{high}$.

Just like ratio edge difference process, we determine whether a pixel is a shadow pixel or moving object pixel by the whole mean and variation. To reduce computation complexity, we just use pixels satisfy

$$0 < \mu_{B(x,y)}^{K} < I_{low} \tag{19}$$

Then we estimate $m_{\mu_B}^{K}$ and $\sigma_{\mu_B}^{K}$ by MLE as follows:

$$m_{\mu_B}^{K} = \frac{1}{N_B}\sum_{(x,y)\epsilon B}\mu_{B(x,y)}^{K} \quad ; \quad \left(\sigma_{\mu_B}^{K}\right)^2 = \frac{1}{N_B}\sum_{(x,y)\epsilon B}\left(\mu_{B(x,y)}^{K} - m_{\mu_B}^{K}\right)^2 \tag{20}$$

where B is the set pixels that satisfying (19) in the set of ($S_n(x,y)=1$), and N_B is the number of pixel in B .

We compute the threshold β_h^K and β_l^K using the estimated $m_{\mu_B}^K$ and $\sigma_{\mu_B}^K$ as follows:

$$\beta_h^K = m_{\mu_B}^K + 1.96 * \sigma_{\mu_B}^K \; ; \; \beta_l^K = m_{\mu_B}^K - 1.96 * \sigma_{\mu_B}^K \tag{21}$$

within the reliability of 95%; then the pixel is determined as an object pixel as follows:

$$O_n(x,y) = 1 \; ; \quad S_n(x,y) = 0$$

$$if\left(\begin{pmatrix}(\mu_{B(x,y)}^R > \beta_h^R)|(\mu_{B(x,y)}^R < \beta_l^R)|(\mu_{B(x,y)}^G > \beta_h^G)| \\ (\mu_{B(x,y)}^G < \beta_l^G)|(\mu_{B(x,y)}^B > \beta_h^B)|(\mu_{B(x,y)}^B < \beta_l^B)\end{pmatrix} \& S_n(x,y) = 1\right) \tag{22}$$

3.4. Spatial Adjustment

We will adjust the property of some small isolated regions. For example a moving object pixel that is surrounded by shadow pixels will be determined to be a shadow pixel and the adjustment will be carried out to the shadow pixels in a similar way.

4. Experimental Results

In this section, we present some experimental results obtained by the proposed approach. And we compare the results with other algorithm. For the proposed approach, the choices of parameters (i.e. $r, \mu_S^K, I_{low}, I_{high}$) influence the accuracy rate of the results. In our experiments, we choose $r = 3$, $\mu_S^K = 2.0$, $I_{low} = 0.9, I_{high} = 1.5$. Fig.3 shows two frames of the video sequences (*HighwayI* and *Intelligentroom*) along with the final shadow detection results by the proposed algorithm. In Fig.3, the green region is object pixels and the red region is shadow pixels.

We also give the quantitative evaluation results about the proposed method. The performance is expressed by using the shadow detection rate η and shadow discrimination rate ξ [2], and they are defined as follows:

$$\eta = \frac{TP_S}{TP_S + FN_S} \times 100\% \; ; \quad \xi = \frac{\overline{TP_F}}{TP_F + FN_F} \times 100\% \tag{23}$$

where subscript S and F respectively represent shadow and foreground; TP_F are the number of correctly identified object pixels, and FN_F is the number of incorrectly identified object pixels; TP_S and FN_S denote analogous parameters regarding shadow identification; $\overline{TP_F}$ is the number of ground-truth foreground object pixels minus the number of pixels marked as shadows, but belonging to foreground object. Therefore, η indicates how well the algorithm detects shadows, and ξ describes how well shadows are discriminated from actual foreground pixels by the method.

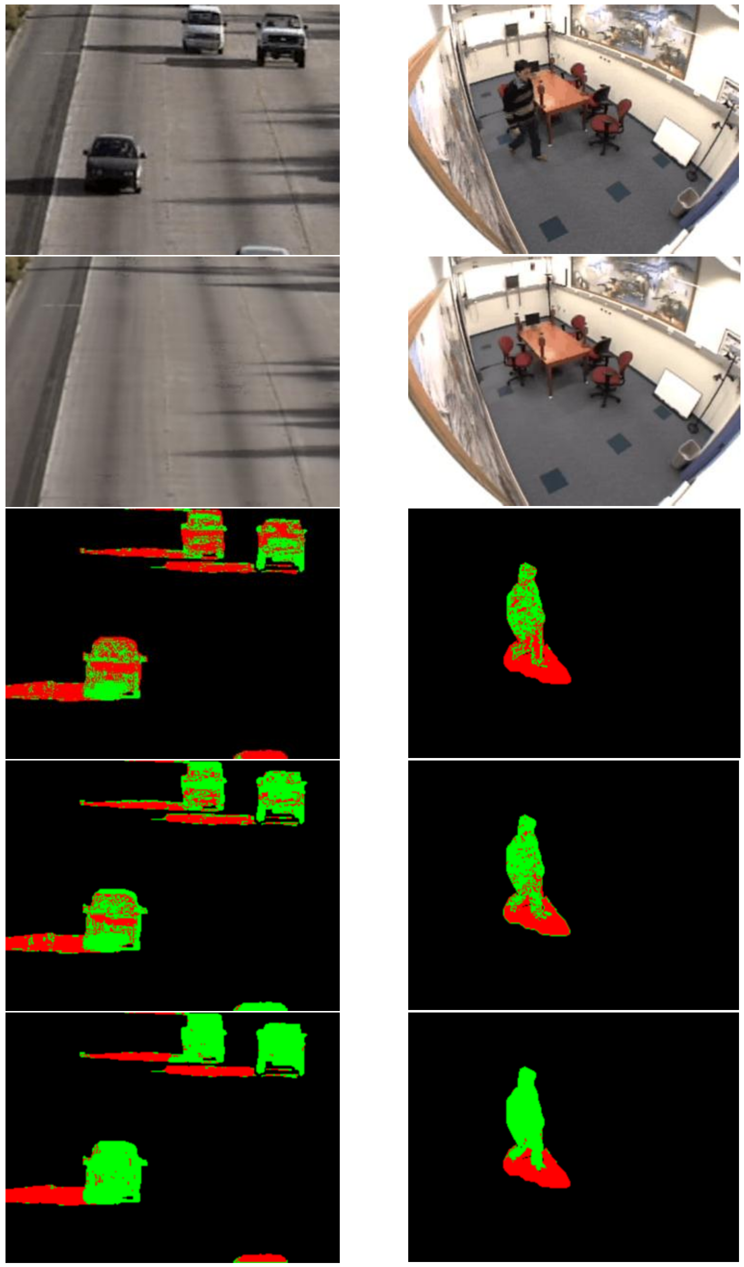

Fig.3. Shadow Detection Results

	Highway		**Intelligent room**	
	η	ξ	η	ξ
SP	**58.55%**	**87.13%**	**79.85%**	**87.82%**
SNP	**76.71%**	**66.60%**	**85.07%**	**76.60%**
DNM1	**60.23%**	**86.63%**	**82.2%**	**88.2%**
DNM2	**72.17%**	**77.86%**	**78.60%**	**86.8%**
ICF	**71.82%**	**79.29%**	**73.45%**	**86.52%**
GMSM	**75.43%**	**74.67%**	**73.6%**	**79.1%**
ER	**67.17%**	**90.19%**	**88.63%**	**88.91%**
Proposed	**74.76%**	**91.75%**	**86.27%**	**89.17%**

Table. 1. Comparison with other algorithms

We compare our proposed approach with the following approaches: the statistical parametric approach (SP [15]), statistical nonparametric approach (SNP [12]), deterministic non-model based approaches (DNM1 [4] and DNM2 [14]), the Gaussian mixture shadow model (GMSM [7]), and the shadow suppression algorithm based on edge ratio (ER [11]). In our analysis, we use two video sequences (*HighwayI* and *Intelligentroom*) for shadow detection. Table 1 shows the comparison results. From the table, we can see that our proposed method achieves better performances than SP, DNM1, DNM2, ICF, GMSM, and is comparable with SNP, ER. The proposed method reaches the highest shadow discrimination rate in *HighwayI* and *Intelligentroom* sequences, a little lower shadow detection rate than SNP in *HighwayI* sequence, and a little lower shadow detection rate than ER in *Intelligentroom* sequence.

5. Conclusion

In this paper, we propose an approach to automatically detect shadow using ratio edge and ratio brightness. The approach can be operated very simply because it determines the shadow threshold automatically without an additional training step. Experimental results show the proposed approach is effective. The performance of our approach is better or comparable with other methods.

Acknowledgements

This work was supported in part by National Basic Research Program of China (973 Program): 2009CB320906, in part by National Natural Science Foundation of China: 61025011, 61035001 and 61003165, and in part by Beijing Natural Science Foundation: 4111003.

References

[1] JinMin Choi, YungJun Yoo, and JinYoung Choi.: Adaptive Shadow Estimator for Removing Shadow of Moving Object, Computer Vision and Image Understanding, pp.1017-1029, 2010.
[2] A. Prati, I. Mikic, M.M. Trivedi, and R. Cucchiara.: Detecting Moving Shadows: Algorithms and Evaluation, IEEE Trans.Pattern Analysis and Machine Intelligence, vol. 25, no. 7, pp. 918-923, July 2003.
[3] E. Salvador, A. Cavallaro, and T. Ebrahimi.: Cast Shadow Segmentation Using Invariant Color Features," Computer Vision and Image Understanding, pp.238-259, 2004.
[4] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving Shadow Suppression in Moving Object Detection with HSV Color Information," Proc. Intelligent Transportation Systems Conf., pp. 334-339, 2001.
[5] O. Schreer, I. Feldmann, U. Goelz, and P. Kauff.: Fast and Robust Shadow Detection in Videoconference Applications, Proc. Fourth IEEE Int'l Symp. Video Processing and Multimedia Comm., pp. 371-375, 2002.
[6] Zoran Zivkovic, Ferdinand van der Heijden.: Efficient adaptive density estimation per image pixel for the Task of Background Subtraction, Pattern Recognition Letters, 2006.
[7] Richard O. Duda, Peter E. Hart, David G. Stork.: Pattern Classification, 2nd ed.,Wiley-Interscience, 2000.
[8] N. Martel-Brisson, and A. Zaccarin.: Learning and Removing Cast Shadows through a Multidistribution Approach, IEEE Trans.Pattern Analysis and Machine Intelligence, vol. 29, no. 7, pp. 1133-1146, July 2007.
[9] Claudio Rosito Jung.: Efficient Background Subtraction and Shadow Removal for Monochromatic Video Sequences, IEEE Transactions on Multimedia, vol. 11, no. 3, pp. 571-577, April 2009.
[10] C. Stauffer and W. E. L. Grimson.: Learning patterns of activity using real-time tracking, IEEE Trans. Pattern Anal. Machine Intell., vol. 22, pp. 747–757, 2000.
[11] W. Zhang, X. Z. Fang, and X. K. Yang.: Moving cast shadows detection using ratio edge, IEEE Trans. Multimedia, vol. 9, no. 6, pp.1202–1214, Oct. 2007.
[12] Thanarat Horprasert, David Harwood, Larry S. Davis.: A statistical approach for real-time robust background subtraction and shadow detection, in: Proceedings of the IEEE International Conference on Computer Vision, 1999.
[13] Kunfeng Wang, Qingming Yao, Xin Qiao, Shuming Tang, Fei-Yue Wang.: Moving object refining in traffic monitoring applications, in: Proceedings of the 2007, IEEE Conference on Intelligent Transportation Systems, pp. 540–545, October 2007.
[14] J. Stauder and R. M. Ostermann.: Detection of moving cast shadows for object segmentation, IEEE Trans. Multimedia, vol. 1, no. 1, pp.65–76, Mar. 1999.
[15] I. Mikic, P. C. Cosman, G. T. Kogut, and M. M. Trivedi.: Moving shadow and object detection in traffic scenes, Proc. Int. Conf. Pattern Recognition, pp. 321–324, Washington, DC, 2000.