An Unsupervised Shadow Detection Algorithm

Chung-Hsien Huang and Ruei-Cheng Wu
Advanced Technology Center, Information & Communications Research Laboratories, Industrial Technology Research Institute, Hsinchiu, Taiwan
E-mail:{DavidCHHuang, AllenRCWu}@itri.org.tw

Abstract *- Shadow detection is a critical issue for most applications of video surveillance. In this study, we proposed an unsupervised method to detect shadow caused by moving objects without setting any priori parameters or threshold values. The color histograms of moving objects' image patch are collected to learn an accumulated distribution in each color domain. The accumulating strategy strengthens the impact of shadow parts in the image patch but reduces the effects of other non-shadow parts. By taking the advantage of robust estimation, the most significant peak, i.e. the shadows, in the accumulated distribution is fitted by a Gaussian function. As a result, if a pixel's feature value is within the 2* σ *range of the Gaussian function,, it can thus be determined as shadows.*

Keywords: Shadow Detection, Video Surveillance, Robust Estimation.

1. Introduction

Video surveillance technology has witnessed increased importance in recent years, especially in the post 9/11 era. Numerous video cameras have been deployed in a number of major cities worldwide. However, conventional video surveillance systems need heavy human monitoring and attention. The more cameras deployed, the more inspection personnel employed. In addition, attention of inspection personnel is decreased over time, resulting in lower effectiveness at recognizing events while monitoring real-time surveillance videos. As a result, research in the field of intelligent video surveillance seems blooming in recent years.

In order to detect moving objects in the surveillance video for further analyses, a conventional way is to maintain a background model and then subtracting the background from the current frame to get the moving foreground patch. However, the casting shadows caused by moving objects will be also segmented from the background, resulting in inaccurate object description. Therefore, shadow detection is a critical but necessary preprocess in most video surveillance applications.

Numerous studies based on color feature transformation have been revealed. Cucchiara et al. [1] transformed video frames from RGB space to Hue-Saturation-Intensity (HSI) space. Shadows show two physical characteristics in HSI domain. In the intensity component, the value of the shadow pixel is lower than that in background model. In the hue and saturation components, the value of the shadow pixel shows slightly different from that in the background model. Therefore, shadows can be detected by using thresholding technique to obtain the pixels which are satisfied the physical characteristics. Nadimi et al. [2] proposed an approach based on a spatio-temporal albedo test to account for both the sun and the sky illuminations. Bunyak et al. [3] combined the features of intensity, chromaticity, and reflectance ratio to detect shadow. Zhang [4] proposed a new feature called ratio edge which shows strong invariance to illumination. In the work of Shan et al. [5], they evaluated the performance of thresholding-based shadow detection approach on different color spaces such as HSI, YCrCb, c1c2c3, L*a*b. To sum up, the mentioned approaches are all based on transforming the RGB features to another color domain or features, which have better characteristics to represent shadows. But no matter what kind of color spaces or features is adopted, users all need to set one or more threshold values to rule shadows out.

On the other hand, Joshi et al. [6] modeled four features such as edge magnitude error, edge gradient direction error, intensity ratio, and color error by probability density functions. The shadow can thus be detected by applying Bayesian decision without selecting any threshold values. However, this approach still needs the user to label shadow pixels manually on couple frames to learn the probability density functions of shadows.

Recently, Nicolas et al. [7] proposed an approach named Gaussian Mixture Shadow Model (GMSM) for shadow detection. The GMSM utilities two Gaussian mixture models (GMM) to model the static background and casting shadows. Afterward, Tanaka et al. [8] used the same idea but modeled the distributions of background and

shadows non-parametrically by Parzon windows. It is faster than GMSM but costs more storage space. Both of them are based on statistical analysis and have better discriminability especially when the color of moving object shows similar attribute to the shadow pixel. Nevertheless, both approaches still need to set some threshold values to detect the pixels which are then regarded as shadow candidates.

In summaries, the past studies on shadow detection put most attention on color space transformation or feature selection, but rarely discussed how to choose suitable threshold values automatically to rule shadows out. The difficulty is that the illuminative effects changes dramatically among different environments such as indoor or outdoor. Even in the same environment, the illumination also shows much different at different time. Therefore, a user-defined threshold may work well at a specific environment but useless at others environments or at different time.

In this study, we propose an unsupervised method to detect shadows without setting any priori parameters or threshold values. Thus, our proposed method can be applied to different environments without manually tuning the thresholds of the shadow detection. In this paper, we describe the proposed shadow detection algorithm in Section 2. Section 3 reveals the experimental results and discusses. Section 4 presents the conclusions and future works.

2. Proposed Method

Figure 1 illustrates the idea of the proposed algorithm. After performing background subtraction, the foreground, i.e. moving objects with casting shadows are segmented. Each connected object of foreground is called an observation herein. We then transform the observed foreground in the video frames from RGB to HSI. Take the intensity component for example, the intensity difference between background and foreground pixels of an observation is computed separately and then forms as a histogram to represent the observation. As shown in Fig. 1, although the histograms are gathered statistically from different observations, they have a similar part in the distribution on common, i.e. the bins within the blue rectangle. Since shadows are caused by blocking the light, the shadow pixels show similar decrease in the intensity. This physical characteristic shows statistical meaning to differ the shadow pixels from the object itself. Therefore, our strategy is to strengthen the impact of shadow parts in the image patch but reduces the effects of other non-shadow parts by accumulating the histograms of observations. Next, taking the advantage of robust estimation, the most significant peak in the accumulated histogram, which can be modeled as a Gaussian, is then extracted as a shadow function. As a result, if a pixel's intensity feature is located within a statistical meaning range (say 2σ) of the shadow function, the pixel can be determined as a shadow pixel.

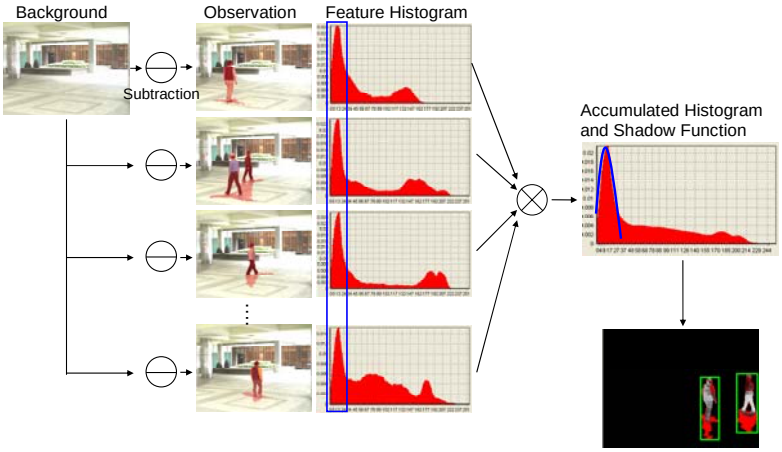

Figure 1. The concept of the proposed unsupervised algorithm for shadow detection

Figure 2 shows the flowchart of the training stage of the proposed algorithm. First, moving objects are detected by background subtraction. Connected component algorithm is then applied to group connected pixels to form a connected object. Pixels which located at the connected object in the current frame and background are transformed from RGB to HSI domain. The differences

between foreground and background in H, S, and I components are then computed, separately. We gather the difference values to form a statistical feature distribution, the difference histograms. Meanwhile, an accumulated distribution which integrates previously consecutive feature distributions is updated by accumulating the current feature distribution. The accumulation is referred to use a mathematical way to emphasis the statistical impact of shadows and suppress the others. Finally, we use a robust estimation technique to fit the most significant peak as a Gaussian function in the accumulated distribution. The fitted Gaussian function is, therefore, the shadow function.

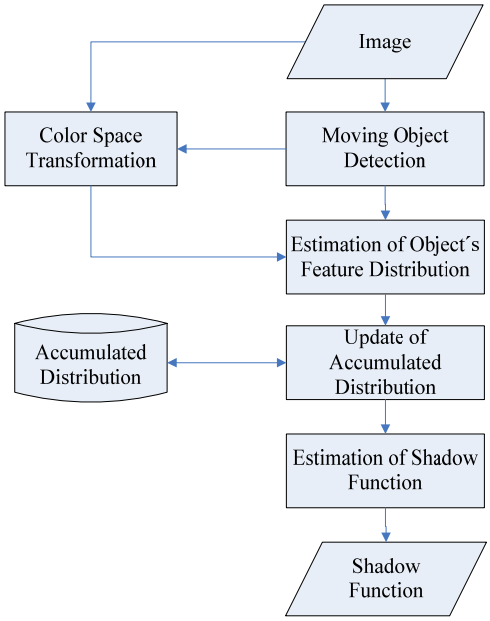

Figure 2. The flowchart of the learning stage.

2.1. Preprocessing

The Gaussian mixture model (GMM)-based background subtraction approach presented by Stauffer and Grimson [9] is a commonly used tool for extracting the moving objects. Basically, it uses couples of Gaussian to model the reasonable variation of the background pixels. Therefore, a pixel will be considered as foreground/moving object if the variation is larger than a threshold. In our system, each pixel in RGB color space is modeled as four Gaussians for foreground detection.

After extracting the foreground, a morphological smoothing procedure is adopted to remove the noises (small-size foreground regions) by taking the morphological opening followed by a morphological closing operator with a 3 × 3 rectangular structural element. In order to represent the spatial information of the extracted moving object candidates, a bounding box containing all the connected foreground pixels is recorded. The bounding box can be considered as a valid observation if and only if it satisfies the following two constrains. First, the size of a valid box should be larger than a certain pixels since the small-size box usually contain part of a pedestrian. Second, the boundary of the bounding box keeps at least five pixels away from the boundary of the input frame. This constrain ensures the object is fully observed in the filed of view.

2.2. Color Space Transformation

After extracting the connected moving object $C_{object} = \{x_i \mid i = 1,2,...,k\}$ in a valid bounding box, all pixels $\boldsymbol{x_i}$ belonged to the object are transformed from RGB to HSI domain by using the following equations.

$$H(x) = \begin{cases} 0, & if \quad \max = \min \\ (60^\circ \times \frac{G(x) - B(x)}{\max - \min} + 0^\circ) \bmod 360^\circ, & if \quad \max = R(x) \\ 60^\circ \times \frac{B(x) - R(x)}{\max - \min} + 120^\circ, & if \quad \max = G(x) \\ 60^\circ \times \frac{R(x) - G(x)}{\max - \min} + 240^\circ, & if \quad \max = B(x) \end{cases} \tag{1}$$

$$S(x) = \begin{cases} 0, & if \quad \max = 0 \\ 1 - \frac{\min}{\max}, & otherwise \end{cases} \tag{2}$$

$$I(x) = \max \tag{3}$$

where

$$\max = MAX(R(x), G(x), B(x)) \tag{4}$$

$$\min = MIN(R(x), G(x), B(x)) \tag{5}$$

Meanwhile, the pixels on the same position, which the object covers on the background image, are transformed to HSI domain, too. We then compute the difference between foreground and background pixels on each component separately as the feature for the following probability distribution calculation.

The shadow pixels show some physical characteristics on HSI color space. When the pixel is covered by shadows, its intensity decreases but the hue and saturation are similar. As a result, the differences of the values in hue, saturation, and intensity components between the current frame and the background are adopted to represent a pixel as features, shown in Eq. (6)-(8).

$$\Delta H(x) = \left| H_b(x) - H_c(x) \right| \tag{6}$$

$$\Delta S(x) = S_b(x) - S_c(x) \tag{7}$$

$$\Delta I(x) = I_b(x) - I_c(x) \tag{8}$$

where $|\bullet|$ stands for computing the absolute

angle degrees between two hue features and the subscript ***b*** and ***c*** indicate the pixel ***x*** on the background image and on the current frame, respectively.

2.3. Estimation and Accumulation of Object's Feature Distribution

After computing the differences between the background image and the current frame in hue, saturation, and intensity domains, the feature values in each domain are then normalized linearly to form a 256-bin histogram (hue: $[0\ 180] \rightarrow [0\ 255]$; saturation: $[-1\ 1] \rightarrow [0\ 255]$; intensity: $[0\ 1] \rightarrow [0\ 255]$). Notice that only the positive part of the difference of intensity is considered. The reason is that when a pixel is covered by shadows, its intensity is supposed to be darker than it is in the background image. Each histogram is then converted to a probability distribution by dividing the value of each bin by the summations of bin values. As a result, each connected object is represented by three probability distributions, the hue-feature, saturation-feature, and intensity-feature distributions, as shown in Fig. 3 (b), (c), and (d), respectively.

The system also maintains three distributions named accumulated distributions corresponding to the hue, saturation and intensity features. The accumulated distributions are updated by accumulating the current observation to the previous observations, as shown in Eq. (9).

$$\hat{h}(i) = (h(i))^{1-\alpha} \cdot (f(i))^{\alpha} \tag{9}$$

where $f(i)$, $h(i)$ and $\hat{h}(i)$ stand for the feature distributions of the current observation, the accumulated distribution of the previous observations and the current accumulated distribution, respectively. The variable *i* is the index of bins. The learning rate α reveals the impact of the current observation to the accumulated distribution. It is set to 0.01 in this study. Figure 3 (e), (f), and (g) show the accumulated distributions of hue, saturation and intensity feature, respectively.

2.4. Estimation of Shadow Function

As mentioned above, the most significant peak of the accumulated distribution is usually caused by shadows. Therefore, we use a parameterized function to fit the peak as the shadow function. The parameterized function is modeled as a Gaussian function, shown in Eq. (10):

$$g(i,m,H,\sigma) = H \cdot \exp(-\frac{|i-m|}{2\sigma^2}) \tag{10}$$

where ***H*** and ***m*** are the probability (height) and the mean of the peak, respectively. The fitting problem, therefore, becomes a minimization problem of estimating the variance σ of Gaussian, as shown in Eq. (11).

$$\tilde{\sigma} = \arg\min_{\sigma} \sum_{i} (g(i,m,H,\sigma) - h(i))^2 \tag{11}$$

However, the non-shadow data, which will affect the fitting accuracy, are regarded as outliers. We apply a robust estimation technique to get a better fitting result. The least-square-estimation term in Eq. (11) is modified as Eq. (12) by using an influence function ρ. The turkey's bi-weight function is adopted as the influence function, which is widely used in the standard robust estimation [10]. The turkey's bi-weight function is shown in Eq. (13). The blue lines in Fig. 3 (e), (f) and (g) show the estimated shadow functions.

$$\hat{\sigma} = \arg\min_{\sigma,c} \sum_{i} \rho(g(i,m,H,\sigma) - h(i), c) \tag{12}$$

$$\rho(x,c) = \begin{cases} \frac{c^2}{6}(1-[1-((x/c)^2)]^3) & if \quad |x| \le c \\ \frac{c^2}{6} & if \quad |x| > c \end{cases} \tag{13}$$

When the shadow function is obtained, a pixel is then detected as a shadow pixel if its feature values are all within a 2σ range (95% of confidence interval) of the estimated Gaussian. Figure 3 (h), (i) and (j) show the pixels (colored in red) are within the 2σ range constrain and (k) shows the pixels which are satisfied to the shadow constrains in all feature domains.

3. Experimental Results

To evaluate the performance of the proposed algorithm, two videos captured using Sony DCR-PC 110 DV HandyCam camcorder with 320 × 240 resolution were used. The videos are obtained with frame rate at 25 frames/sec. The first 10-minute video was captured at an outdoor scene but under a roof. The light source is from sky illuminations and couple fluorescent lamps under the roof. The second 2-minute video was captured at an outdoor walkway. The light source is from the sun and sky illuminations but the weather is partially cloudy.

In order to evaluate the detection rate and false alarm rate of shadows, the ground truth was labeled manually first. Forty-five and sixteen frames were selected as ground truth images from video one and video two respectively. Frames in each video were selected in approximately equal time interval. The reason we did not select them in

exactly equal time interval is that in these frames there was maybe no object showing up or objects just partial appeared. The pixels defined as foreground in those selected frames were labeled in red by using GMM background subtraction algorithm. Then we labeled the shadow pixels manually in blue from the foreground pixels. Figure 4 (a) and (b) shows the labeled ground truth of a frame in testing video one and video two, respectively.

(a) (b)

Figure 4. The labeled ground truth in the testing videos (a) the frame #1197 in the testing video one (b) the frame #552 in the testing video two.

Table 1 shows the conditions of detection results. For example, a true positive (TP) means the testing pixel which is labeled as a shadow pixel and is correctly detected as a shadow, a false positive (FP) stands for the pixel is not labeled as a shadow pixel but is detected as shadows, and so on. Accuracy of the shadow detection is then evaluated by calculating detection rate (DR = TP/(TP+FN)) and false alarm rate (FAR = FP/(TP+FP)). Higher DR means the shadows can be detected more correctly, while lower FAR shows the objects will not be misclassified as shadows. A good detection algorithm is to make a compromise between DR and FAR since higher DR usually results in higher FAR and vice versa.

Comparing to previous studies of shadow detection, the contribution of this study is to propose an unsupervised approach to learn the thresholds automatically. As a result, we compared the performance of the proposed algorithm to a threshold-based method [1]. The threshold-based method is by manually setting several thresholds in hue, saturation, and intensity domains to rule out the shadow pixels. The experimental results are shown in Table 2. The third column shows the DR and FAR of the proposed algorithm. In the fourth column, we manually selected the thresholds appropriately by visual inspection from five frames of video one, and then used them to detect shadows in both videos. It is no doubt that the selected thresholds work well on video one but quite unacceptable on video two (DR=0.56). Similarly, the fifth column means we chose the thresholds by visually inspecting video two and then applied them to detect shadows on both videos. The results show the selected thresholds also not suitable on video one (FAR=0.30). However, since the proposed algorithm utilizes an unsupervised way to learn the shadow distribution, it does not need any human assistance on the threshold selection and has a good compromise between DR and FAR.

Table 1. The conditions of the results of shadow detection.

		Is labeled as shadow?	
		1	0
Is detected as shadow?	true	TP	FP
	false	FN	TN

(T: true, F: false, P: positive, N: negative)

Table 2. The result of shadow detection by the proposed method and by manually selecting threshold values.

		Proposed method	Threshold set 1	Threshold set 2
Video one	DR	0.88	0.92	0.92
	FAR	0.20	0.26	0.30
Video two	DR	0.86	0.56	0.88
	FAR	0.22	0.10	0.28

4. Conclusion

This study presents an unsupervised algorithm to detect shadows for the applications of video surveillance. Our proposed algorithm can automatic learn the distribution of the shadow without manually setting any threshold values. The objects appear in the video are collected as observations to train a shadow function. The distributions of hue-difference, saturation-difference, and intensity-difference are obtained from each observation. Then the distributions obtained from consecutive observations are then accumulated to increase the impact of shadows but decrease the others. Therefore, the shadow function, which is further applied to detect shadows, can be obtained by fitting a Gaussian function on the significant peak in the accumulated distribution.

So far, the proposed algorithm works well in indoor scenes and in outdoor scenes under the cloudy weather (the shadows are only caused by sky illuminations). But when the scene is under a strong sun light situation, the shadows show two unstable peaks in the accumulated distributions. Therefore, a single Gaussian does not afford to model the shadows. In addition, since those two peaks are quite unstable, using two Gaussians to model the shadows does not work out, too. The

feature work will address on how to deal with the outdoor cases under strong sun light. It may be conducted by modifying the accumulating strategy.

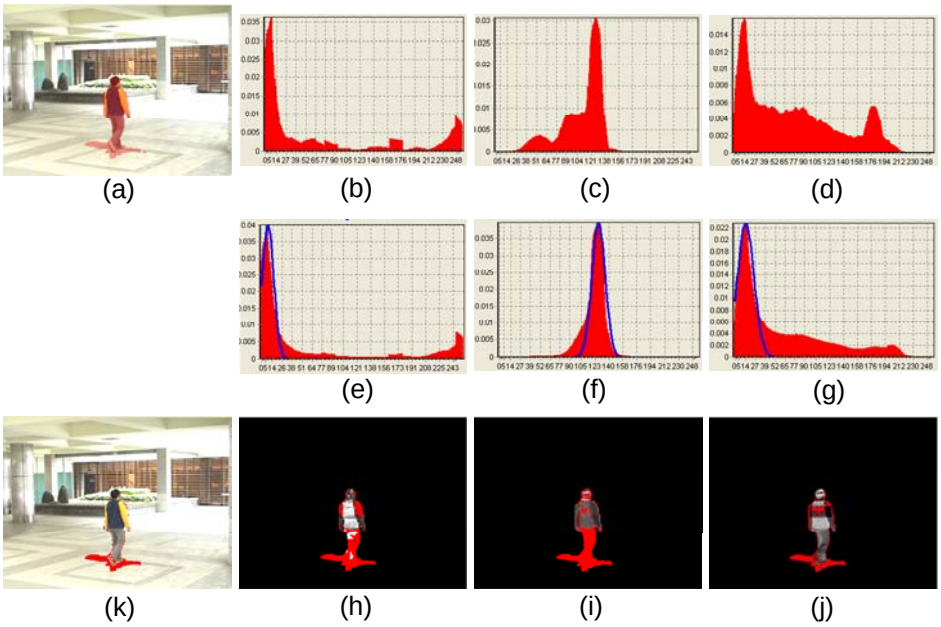

Figure 3. The results of the proposed shadow detection algorithm. (a) the current frame, (b) the hue-feature distribution, (c) the saturation-feature distribution, (d) the intensity-feature distribution, (e) the accumulated hue-feature distribution and its corresponding shadow function, (f) the accumulated saturation-feature distribution and its corresponding shadow function, (g) the accumulated intensity-feature distribution and its corresponding shadow function, (h) the detected shadow candidates in hue domain, (i) the detected shadow candidates in saturation domain, (j) the detected shadow candidates in intensity domain, (k) the detected shadows.

References

[1] R. Cucchiara, C. Grana, M. Piccardi, A. Prati and S. Sirotti, "Improving Shadow Suppression in Moving Object Detection with HSV Color Information," in *Proceedings of 2001 IEEE Intelligent Transportation Systems Conference*, pp. 334-339, 2001.

[2] S. Nadimi and B. Bhanu, "Physical Models for Moving Shadow and Object Detection in Video," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1079-1087, 2004.

[3] F. Bunyak, I. Ersoy, and S. R. Subramanya, "Shadow Detection by Combined Photometric Invariants for Improved Foreground Segmentation," in *Proceedings of the 7th IEEE Workshop on Applications of Computer Vision*, pp. 510-515, 2005.

[4] W. Zhang, X. Z. Fang, X. K. Yang, and Q. M. J. Wu, "Moving Cast Shadow Detection Using Ratio Edge," *IEEE Transactions on Multimedia*, vol. 9, no. 6, pp. 1202-1214, 2007.

[5] Y. Shan, F. Yang, and R. Wang, "Color Space Selection for Moving Shadow Elimination," in *Proceedings of 4th International conference on Image and Graphics*, pp. 496-501, 2007.

[6] A. J. Joshi, S. Atev, O. Masoud, and N. Papanikolopoulos, "Moving Shadow Detection with Low- and Mid-Level Reasoning," in *Proceedings of 2007 IEEE International Conference on Robotics and Automation*, pp. 4827-4832, 2007.

[7] M.-B. Nicolas and A. Zaccarin, "Learning and Removing Cast Shadows through a Multidistribution Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligent*, vol. 29, no. 7, pp. 1133-1146, 2007.

[8] T. Tanaka, A. Shimada, D. Arita, and R. Taniguchi, "Non-parametric Background and Shadow Modeling for Object Detection," *Lecture Notes in Computer Science*, no. 4843, pp. 159-168, 2007.

[9] C. Stauffer and W. E. L. Grimson, "Adaptive Background Mixture Models for Real-time Tracking," *in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 246-252, 1999.

[10] P. J. Huber, *Robust Statistic*, Wiley, 1981.