Moving Cast Shadow Detection using Physics-based Features

Jia-Bin Huang and Chu-Song Chen
Institute of Information Science, Academia Sinica, Taipei, Taiwan
jbhuang0604@gmail.com, song@iis.sinica.edu.tw

Abstract

Cast shadows induced by moving objects often cause serious problems to many vision applications. We present in this paper an online statistical learning approach to model the background appearance variations under cast shadows. Based on the bi-illuminant (i.e. direct light sources and ambient illumination) dichromatic reflection model, we derive physics-based color features under the assumptions of constant ambient illumination and light sources with common spectral power distributions. We first use one Gaussian Mixture Model (GMM) to learn the color features, which are constant regardless of the background surfaces or illuminant colors in a scene. Then, we build up one pixel-based GMM for each pixel to learn the local shadow features. To overcome the slow convergence rate in the conventional GMM learning, we update the pixel-based GMMs through confidence-rated learning. The proposed method can rapidly learn model parameters in an unsupervised way and adapt to illumination conditions or environment changes. Furthermore, we demonstrate that our method is robust to scenes with few foreground activities and videos captured at low or unsteady frame rates.

1. Introduction

Extracting moving objects from video sequences is at the core of various vision applications, including visual surveillance, contend-based video coding, and human-computer interaction, etc. One of the most challenging problems of extracting moving objects is detecting and removing moving cast shadows. When performing background subtraction, cast shadows are often misclassified as parts of foreground objects, distorting the estimation of shape and color properties of target objects. The distortion caused by cast shadows may hinder subsequent vision algorithms, such as tracking and recognition.

Cast shadows are caused by the occlusion of light sources. When foreground objects cast shadows on background surfaces, the light sources are partially or entirely blocked, and thus the total energy incident at the background regions is reduced. Hence, shadow points are expected to have lower luminance but similar chromaticity values.

There have been many works dedicated to detecting cast shadows. Most of them are based on the assumption that shadow pixels should have lower luminance and the same chrominance as the corresponding background (i.e. the RGB values of shadow pixels will fall on the line between the illuminated value and the origin in the RGB color space). This linear attenuation property has been employed in different colors spaces like RGB [2], HSV [1], YUV [12], and $c_1c_2c_3$ [11]. Besides, other shadow-induced features like edge or gradient information extracted from the spatial domain have also been used to detect cast shadows [14, 16]. The major limitation of these algorithms is that they often require explicit tuning of a large set of parameters for each new scene. Thus, they are inappropriate for on-line applications.

To adapt to environment changes, statistical learning-based approaches have been developed to learn and remove cast shadows [9, 5, 4]. However, the linear proportionality assumption may not always hold in a real-world environment. For instance, in an outdoor scene, the light sources may consist of direct sunlight, diffused light scattered by the sky, and other colored light from nearby surfaces (i.e. color bleeding). These light sources may have different spectral power distributions (SPDs). Therefore, the RGB values of a shadow pixel may not attenuate linearly.

Little attention has been paid to the non-proportionality attenuation problem before. Nadami and Bhanu [8] addressed the non-linearity by using a dichromatic reflection model to account for both the sun and the sky illuminations in an outdoor environment. Recently, a more general shadow model was presented in [6], which introduced an ambient illumination term that determines the direction in the RGB color space along which the shaded background values can be found. Since the ambient term may have a different SPD from the incident light sources, the values of shadow pixels may not decrease proportionally. Nonparametric density estimation was used to model surface variation under cast shadows in an unsupervised way. By provid-

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

ing a better description of cast shadows, the shadow model in [6] provided improved performance over the previous approaches which used linear models.

However, these learning-based approaches [9, 5, 4, 6] may suffer from insufficient training samples since the statistical models are learned from background surface variation under cast shadows. Unlike obtaining samples in every frame in background modeling, shadows may not appear at the same pixel in each frame. A single pixel should be shadowed many times till its estimated parameters converge, while the illumination conditions should be stable. Therefore, this kind of pixel-based shadow models require a longer period of training time when foreground activities are rare. This problem becomes more serious when video sequences are captured at a low or unsteady frame rate that depends on the transmission conditions.

In this paper, we characterize cast shadows with "global" parameters for a scene. Based on the bi-illuminant dichromatic reflection model (BIDR) [7], we first derive normalized spectral ratio as our color features under the assumptions of constant ambient illumination and direct light sources with a common SPD. The normalized spectral ratio remains constant regardless of different background surfaces and illumination conditions. We then model the color features extracted from all moving pixels using a single Gaussian Mixture Model (GMM). To further improve the differentiating ability for cast shadows having similar colors to background, we use a pixel-based GMM to describe the gradient intensity distortion for each pixel. We update the pixel-based GMMs using the confidence predicted from the global GMM through confidence-rated learning to accelerate convergence rates. Contributions are presented in two key aspects. Firstly, with the global shadow model learned from physics-based features, our approach does not require numerous foreground activities or high frame rates to learn the shadow model parameters. This makes the proposed method more practical than existing works using only pixel-based models. Secondly, the proposed confidence-rated learning can be used for fast learning of local features in pixel-based models. We provide a principled scheme for the local and global features to collaborate with each other.

The remainder of this paper is organized as follows. We briefly describe in Section 2 the dichromatic reflection model [13] and its extension BIDR. In Section 3, we present the proposed learning approach. The posterior probability of cast shadows and foreground are developed in Section 4. Both visual and quantitative results are shown in Section 5 to verify the performance of our method and the robustness to few foreground activities. Section 6 concludes this paper.

2. Physics-Based Shadow Model

2.1. Bi-illuminant Dichromatic Reflection Model

There are three terms in the Shafer's model [13]: body reflection, surface reflection, and a constant ambient term. Each of the two reflection types can be decomposed into chromatic and achromatic parts: 1) composition: a relative SPD c_b or c_s which depends only on wavelength, 2) magnitude: a geometric scale factor m_b or m_s which depends only on geometry. Given a scene geometry, the radiance in the direction (θ_e, ϕ_e) can be expressed as

$$I(\theta_e, \phi_e, \lambda) = m_b(\theta_e, \phi_e)c_b(\lambda) + m_s(\theta_e, \phi_e)c_s(\lambda) + c_d(\lambda), \tag{1}$$

where c_d is the constant ambient term.

While this model included a term to account for ambient illumination, the model did not separate it into body and surface reflection. Recently, Maxwell et al. [7] proposed the BIDR model, which contains four terms: two types of reflection for both the direct light sources and ambient illumination. Then, the BIDR model is of the form:

$$\begin{aligned} I(\theta_e, \phi_e, \lambda) &= m_b(\theta_e, \phi_e, \theta_i, \phi_i)c_b(\lambda)l_d(\theta_L, \phi_L, \lambda) \\ &\quad + m_s(\theta_e, \phi_e, \theta_i, \phi_i)c_s(\lambda)l_d(\theta_L, \phi_L, \lambda) \\ &\quad + c_b(\lambda)\int_{\theta_i,\phi_i} m_b(\theta_e, \phi_e, \theta_i, \phi_i)l_a(\theta_L, \phi_L, \lambda)d\theta_i d\phi_i \\ &\quad + c_s(\lambda)\int_{\theta_i,\phi_i} m_s(\theta_e, \phi_e, \theta_i, \phi_i)l_a(\theta_L, \phi_L, \lambda)d\theta_i d\phi_i, \end{aligned} \tag{2}$$

where (θ_L, ϕ_L) is the direction of the direct light source relative to the local surface normal, and (θ_e, ϕ_e) and (θ_i, ϕ_i) are the angles of emittance and incidence, respectively. In this reflection model, the range of m_b, m_s, c_b, and c_s all lie in $[0, 1]$ (we refer reader to [7] for further details of the deviation of BIDR model.)

With a specific geometry and representing the two ambient integrals as $M_{ab}(\lambda)$ and $M_{as}(\lambda)$, we can simplify the BIDR model to

$$\begin{aligned} I(\lambda) &= c_b(\lambda)[m_b l_d(\lambda) + M_{ab}(\lambda)] \\ &\quad + c_s(\lambda)[m_s l_d(\lambda) + M_{as}(\lambda)]. \end{aligned} \tag{3}$$

Considering only matte surfaces, we can ignore the latter part of (3). To describe the appearance of cast shadows on a background surface, we multiply the direct illumination with an attenuation factor $\alpha \in [0, 1]$, which indicates the unoccluded proportion of the direct light. We assume that all direct light sources have a common SPD with different power factor and the ambient illumination is constant over lit and shaded regions (see Fig. 1). This gives us the simplified form of the BIDR model

$$I(\lambda) = \alpha m_b c_b(\lambda) l_d(\lambda) + c_b(\lambda) M_{ab}(\lambda). \tag{4}$$

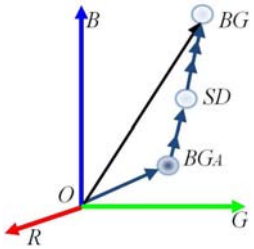


Figure 1. The contribution of all direct light sources and ambient illuminance. The shadow values SD are expected to be observed along the line between background value BG and the constant ambient term BG_A. Note that the shadow values do not necessary to be proportional to the direction of background values.

The camera sensor response g_i at the pixel level can be obtained through the spectral projection $g_i = \int F_i(\lambda)I(\lambda)d\lambda$, where $F_i(\lambda), i \in \{R, G, B\}$ is the sensor spectral sensitivities and λ denotes the wavelength. By applying the linearity of spectral projection, we have

$$g_i = \alpha F_i m_b c_b^i l_d^i + F_i c_b^i M_{ab}^i, i \in \{R, G, B\} \quad (5)$$

The formulation of g_i defines a line segment in the RGB color space varied between two ends: shadowed pixel ($\alpha = 0$) to fully lit pixel ($\alpha = 1$).

2.2. Extracting Useful Features

2.2.1 Spectral Ratio

To extract color features that are constant and independent to different background surfaces, we need to identify measurements that are invariant to illumination attenuation factor α, geometry shading factor m_b, and the chromatic aspect of body reflection c_b^i. We calculate the ratio of illuminants to be the spectral ratio $\vec{S} = [S_R, S_G, S_B]^T$ using

$$S_i = \frac{SD^i}{BG^i - SD^i} = \frac{\alpha F_i m_b c_b^i l_d^i + F_i c_b^i M_{ab}^i}{(1-\alpha) F_i m_b c_b^i l_d^i} = \frac{\alpha}{1-\alpha} + \frac{M_{ab}^i}{(1-\alpha) m_b l_d^i}, i \in \{R, G, B\}. \quad (6)$$

If the shaded regions received only the ambient illumination (i.e. all direct light sources are blocked: $\alpha = 0$), the first term in (6) disappears and $S_i = \frac{M_{ab}^i}{m_b l_d^i}$. We can then derive features invariant to m_b by normalizing S_i with its length $|S|$ since the m_b term can be extracted from the normalization constant. However, this assumption does not hold in real-world environments. Take an indoor scene as an example, where there are usually multiple light sources. When a foreground object occludes one or some of the light sources, there is still energy from the remaining light sources incident to this surface. Consequently, assuming the attenuation factor to be zero will induce bias in estimating the ratio between two illuminants.

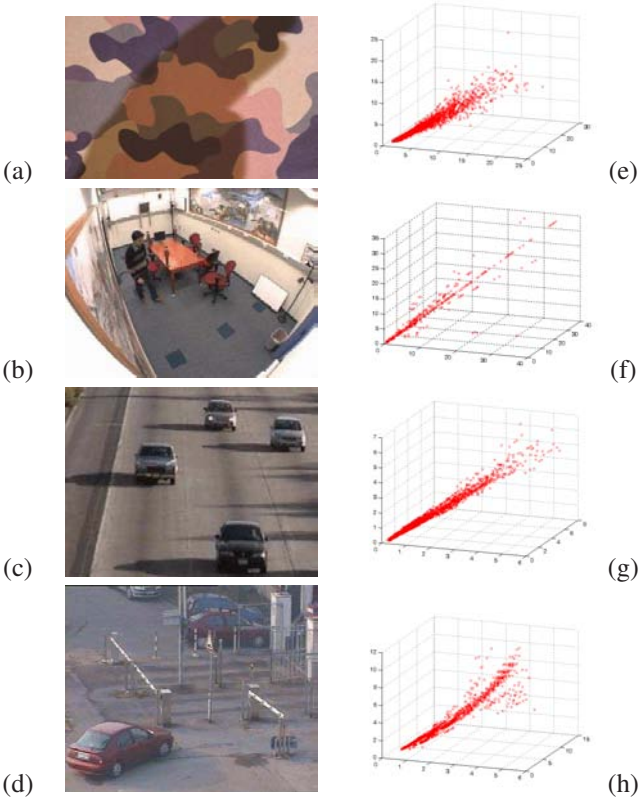


Figure 2. The color feature value distribution in various environments and illumination conditions. (a)-(d) Frame with cast shadows. (e)-(h) The corresponding feature value distribution in the $S_1S_2S_3$ space. Note that the feature values extracted from different background surface generally follow a line.

To address this problem, we introduce $\vec{\gamma} = [\gamma_R, \gamma_G, \gamma_B]^T$ by subtracting each element of S by $\frac{\alpha}{1-\alpha}$:

$$\gamma_i = S_i - \frac{\alpha}{1-\alpha} = \frac{M_{ab}^i}{(1-\alpha) m_b l_d^i}. \quad (7)$$

Similarly, we can obtain the normalized spectral ratio $\hat{\gamma}$ with higher accuracy by factoring out $(1-\alpha)m_b$ through normalization:

$$\hat{\gamma}_i = \frac{M_{ab}^i}{(1-\alpha) m_b l_d^i} \left(\frac{1}{|\gamma|}\right) \quad (8)$$

$$|\gamma| = \frac{1}{(1-\alpha) m_b} \sqrt{\left(\frac{M_{ab}^R}{l_d^R}\right)^2 + \left(\frac{M_{ab}^G}{l_d^G}\right)^2 + \left(\frac{M_{ab}^B}{l_d^B}\right)^2} \quad (9)$$

We validate the proposed physics-based color feature by observing its distributions from cast shadows in various environments and illumination conditions. Fig. 2 (a) and its reference background images are by courtesy of Maxwell et al. [7]. Fig. 2 (b)-(d) show frames from the benchmark sequences provided in the Prati et al.'s survey paper [10]. We manually label shadow pixels in the given images, and then extract the spectral ratio $S_i, i \in \{R, G, B\}$ for each

pixel in shadow regions using the given image and the reference background models. The resultant feature distributions are presented in Fig. 2 (e)-(h). The feature values extracted from different background surfaces roughly follow a straight line in the $S_R S_G S_B$ space. Therefore, we can use the direction of the line as our color feature, which is roughly the same for all shadow pixels. The direction of the line in the $S_R S_G S_B$ space can be characterized with two angles, the zenith and azimuth in the spherical coordinate, which correspond to the normalized spectral ratio. From the feature distributions in Fig. 2 (e)-(h), we also observe that larger feature values tend to be unstable and deviate from the major direction of most feature values. This is because there is not sufficient difference between the shaded and lit pixel value to robustly measure the orientation. In addition, the value of $\frac{\alpha}{(1-\alpha)}$ in the scene can be easily estimated by intersecting the fitted line with the line passing through (1,1,1) and the origin.

2.2.2 Gradient Intensity Distortion

Now we have derived color features that are invariant to different background surfaces. This low dimensional (2D) features, however, might fail to distinguish foreground with colors similar to background from cast shadows. Thus, other shadow-induced properties like edge or gradient information may be used to further describe the background appearance variation under cast shadows. In this paper, we just use a simple gradient intensity distortion as our local features to demonstrate the improvement by incorporating additional local features.

For a given pixel p, we define the gradient intensity distortion ω_p as

$$\omega_p = |\nabla(BG)_p| - |\nabla(F)_p|, \tag{10}$$

where BG and F are luminance channels of the background image and current frame, and $\nabla(\cdot)$ is the gradient operator.

3. Learning Cast Shadows

In this section, we show how to build models for cast shadows in an unsupervised way. Here, we use GMM to learn the background surface variation over time. It is also possible to use other statistical learning method such as kernel density estimation.

3.1. Weak Shadow Detector

To model the cast shadows, impossible shadow samples that belong to background or foreground (e.g. color values that are the same as or brighter than background values) should be excluded. Therefore, we apply a weak shadow detector that evaluates every moving pixel to filter out some impossible samples. Since cast shadows reduce the luminance values, the potential shadow values should fall into the conic volume around the corresponding background color. The weak shadow detector is illustrated in Fig 3, where values of cast shadows are expected to fall into the gray conic region. Pixel values that fall into the gray conic region are considered as potential shadow samples. These samples are then used to learn the global shadow model for the scene and the local shadow model for each pixel.

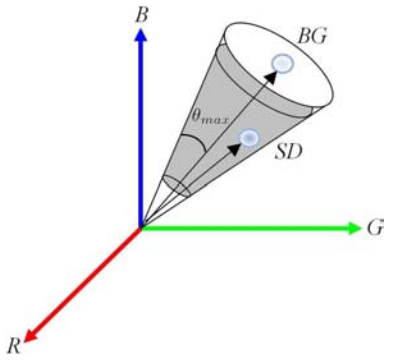


Figure 3. The weak shadow detector. The observation will be considered as potential shadow point if it falls into the gray area. The weak shadow detector contains three parameters: maximum allowed color shift, and minimal and maximal illumination attenuation.

3.2. Global Shadow Model

Using the background surface invariant color features, a global shadow model is learned for the whole scene. Here, we model the background color information by the well-known GMM [15] in the RGB color space. For every frame, we obtain potential shadow points by applying the weak shadow detector on moving pixels, which are identified via background subtraction. We then use one GMM to learn the normalized spectral ratio $\hat{r} = [\hat{\gamma_R}, \hat{\gamma_G}]^T$ in the scene. Note that the reason why we only use two of the three dimensional features is that the third component is redundant since $\hat{\gamma_R}^2 + \hat{\gamma_G}^2 + \hat{\gamma_B}^2 = 1$. The normalized spectral ratio in the whole scene is modeled by K Gaussian distributions with mean vector μ_k and full covariance matrix Σ_k. Then, the probability of the normalized spectral ratio $\hat{r}$ is given by:

$$p(\hat{r}|\mu, \Sigma) = \sum_{k=1}^{K} \pi_k G_k(\hat{r}, \mu_k, \Sigma_k), \tag{11}$$

where μ, Σ denote all parameters of the K Gaussians, π_k is the mixing weight, and G_k is the k_{th} Gaussian probability distribution. We use the Expectation-Maximization (EM) algorithm to estimate the parameters in the GMM. The estimated parameters in the current frame are propagated to next frame, so that the EM algorithm can converge quickly. Since the light sources are usually stable in the scene, we

find that it is sufficient for the estimated parameters to converge with a single EM iteration at each frame.

3.3. Local Shadow Model

Besides using color features in the global shadow model, we build GMM for each pixel to learn the gradient intensity distortion under cast shadows similar to background modeling [15]. For a given pixel p, its gradient feature value ω_p is sampled and learned whenever it is a potential shadow pixel. However, as we mentioned before, pixel-based models often suffer from insufficient training data because the samples are not available at the pixel every frame.

To address this problem, we adopt the confidence-rated learning to improve the convergence rate of the local model parameters. The basic idea is that each sample is weighted with different importance computed from the global shadow model. For example, if we update the model using a potential shadow point whose color features matched with the global shadow model, then we think this sample is relatively more important than others. In this way, the learning process of the local shadow model is guided by the global shadow model.

3.4. Confidence-Rated Gaussian Mixture Learning

We present an effective Gaussian mixture learning algorithm to overcome some drawbacks in conventional GMM learning approach. Let ρ_π and ρ_G be the learning rates for the mixing weight and the Gaussian parameters (means and covariances) in the local shadow model, respectively. The updating scheme follows the the formulation of the combination of incremental EM learning and recursive filter [3]:

$$\rho_\pi = C(\hat{\gamma}) * (\frac{1-\rho_{default}}{\sum_{j=1}^{K} c_j}) + \rho_{default} \quad (12)$$

$$\rho_G = C(\hat{\gamma}) * (\frac{1-\rho_{default}}{c_k}) + \rho_{default}, \quad (13)$$

where c_k is the number of matches of the k_{th} Gaussian state, and $\rho_{default}$ is a small constant, which is 0.005 in our experiments. The two types of learning rates are controlled by a confidence value $C(\hat{\gamma})$, which indicates how confident the sample belongs to the shadow. Observations with higher confidence will then converge faster than those with lower confidence.

3.5. Attenuation Factor Estimation

From the RGB values we observed in the current frame and the reference background image, we can only compute the value of S, which may introduce bias in estimating normalized spectral ratios. Therefore, the estimation of attenuation factor α is required for accurate shadow modeling. We can see that in (7) the value of $\gamma_i, i \in \{R, G, B\}$ is obtained by subtracting $\frac{\alpha}{1-\alpha}$ from S_i. From the feature distribution of S, our aim is to find the location of point t such that t passes through both the lines passing through the origin with direction vector (1,1,1) and the line that fits the observations S. This estimation can be achieved using the robust fitting method that is less sensitive than ordinary least squares to large changes (outliers). In addition, we perform the recursive linear regression to update the estimated $\frac{\alpha}{1-\alpha}$ value adaptively. For simplicity, the attenuation factor α is assumed the same for every pixel.

4. Cast Shadow and Foreground Posterior Probabilities

In this section, we present how to derive the posterior probabilities of cast shadows and foreground given the observed sample x_p in the RGB color space by using the proposed global and local shadow models.

4.1. Cast Shadow Posterior

The shadow posterior is first computed by decomposing $P(SD|x_p))$ over the (BG, FS) domain, where FS indicates moving pixels (real foreground and cast shadows). Since $P(SD|x_p, BG) = 0$, the decomposition gives

$$P(SD|x_p) = P(SD|x_p, FS)P(FS|x_p), \quad (14)$$

where $P(FS|x_p) = 1 - P(BG|x_p)$ can be directly computed from the background model. Second, we remove pixels that are definitely foreground (i.e. pixels that are rejected by the weak shadow detector) and consider only potential shadow points (PS): $P(SD|x_p, FS) = P(SD|x_p, FS, PS)$. Then, we decompose $P(SD|x_p, FS, PS)$ into two parts: N_a, and N_{na}, which stand for color features that are associated with the normalized spectral ratio or not, respectively. If the color features do not associate with the working states of the GMM, then the probability of belonging to shadow equals to zero. Therefore, we have

$$P(SD|x(p), FS, PS) = P(SD|x_p, FS, PS, N_a) * P(N_a|x_p, FS, PS) \quad (15)$$

Here, the gradient intensity distortion ω_p and the color feature $\hat{\gamma}_p$ are the sufficient statistics for x_p in the first and second part of (15). The posterior probability of cast shadow can thus be computed.

4.2. Foreground Posterior

Computing foreground posterior probability is much easier. Given a pixel p, we first compute the background posterior $P(BG|x_p)$ from the background model. Then, we can obtain shadow posterior probability $P(SD|x_p)$ using

the learned shadow models. Based on the probability theory, the foreground posterior can be obtained as:

$$P(FG|x_p) = 1 - P(BG|x_p) - P(SD|x_p). \quad (16)$$

4.3. Summary

Algorithms 1 and 2 summarize in pseudocode the learning and detection processes of the proposed algorithm. Our method can be attached to other moving object detection programs as an independent module. Shadow detection process is only applied to moving pixels detected by the background model and the learning process occurs only when these moving pixels are considered as shadow candidates. Consequently, the proposed algorithm is practical, it does not introduce heavy computational burden and can work effectively to detect shadows.

Algorithm 1: Learning Process

At time t,
for *each pixel* p *in the frame* **do**
  if $P(BG|x_t(p)) < 0.5$ **then**
    if *pixel* p *satisfies shadow property* **then**
      -Compute normalized spectral ratio $\hat{\gamma}_p$
      -Compute gradient intensity distortion ω_p
      -Update local shadow model at pixel p using confidence value $C(\hat{\gamma})$ through confidence-rated learning
    end
  end
end
Run one EM iteration to estimate the parameters of global shadow model using the collected color features.

Algorithm 2: Detection Process

At time t,
for *each pixel* $p \in P$ **do**
  -Obtain background posterior $P(BG|x_p)$ from background modeling
  -Compute shadow posterior $P(SD|x_p)$(eq. 14)
  -Compute foreground posterior $P(FG|x_p)$(eq. 16)
  if $P(FG|x_p) > P(SD|x_p) \& P(FG|x_p) > P(BG|x_p)$ **then**
    Label pixel p as foreground
  else
    Label pixel p as background
  end
end

5. Experimental Results

We present the visual results from challenging video sequences captured in various environments, including both indoor and outdoor scenes. We also compare the quantitative accuracy of the proposed method in several videos with other approaches when the results are available. Previous approaches using statistical model have higher success rate in detecting cast shadow when numerous foreground activities are present. However, we show that our method can deal with the situation that cast shadows first appear in complex scenes and unknown illumination conditions as well as rare foreground activity.

5.1. Qualitative Results

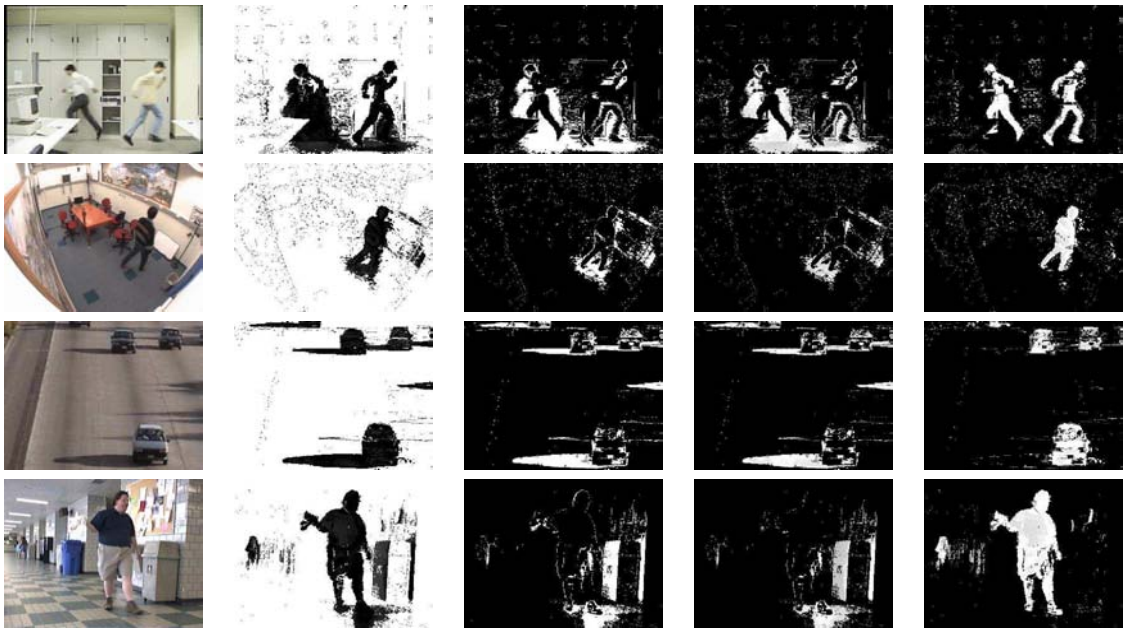
In Figure 4, we show sample cast shadow detection results from four video sequences. The first three sequence: Laboratory, Intelligent Room, and Highway I are part of the benchmark sequences for validating shadow detection algorithm. The last one Hallway is taken from [6]. Figure 4 (a) shows one frame selected from the video, where cast shadows are present in the scene. The background posterior probability is presented in Figure 4(b), where the dark region indicates the less probability of belonging to background. From Figure 4(c)(d), we show the confidence map of global shadow model and the posterior probability of cast shadows, respectively. In Fig. 4(e) we show the probability values of belonging to foreground objects. We can see that in these video sequences, the proposed algorithm is capable of detecting cast shadows without misclassifying foreground as shadows. Note that in the second video, Intelligent Room, the man just walks in the room by once. Thus, there is no chance for the pixel-based shadow model to learn its parameters. The use global shadow model enables us to detect shadows first appear in the scene.

To verify the effectiveness of the proposed method, the results presented here are raw data and without any post-processing. We can obtain binary results simply with thresholding the foreground posterior values $P(FG|x_p)$. The posterior probabilities can also be incorporated with context model that use spatial and temporal coherence to improve the segmentation accuracy.

5.2. Quantitative Results

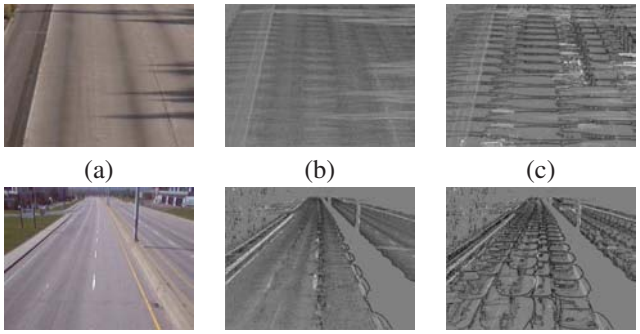
The quantitative evaluation follows the method proposed by Prati et al. [10]. There are two defined metrics for evaluating the performance of cast shadow detection algorithm: shadow detection rate η and shadow discrimination rate ξ. The formulations of the two metrics are as follows:

$$\eta = \frac{TP_S}{TP_S + FN_S}; \xi = \frac{\overline{TP_F}}{TP_F + FN_F}, \quad (17)$$



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

Figure 4. Sample visual results of detecting cast shadows in various environment. (a) Frame from video sequence. (b) Background posterior probability $P(BG|x_p)$. (c) Confidence map predicted by the global shadow model. (d) The shadow posterior probability $P(SD|x_p)$. (e) The foreground posterior probability $P(FG|x_p)$.



(a) (b) (c) (d) (e) (f)

Figure 5. The effect of confidence-rated Gaussian mixture learning. The mean maps of local shadow model are taken at the 100_{th} frame (the first row) and 1000_{th} frame (the second row) (a)(d)The background image. (b)(e) The mean map of the most important Gaussian in the mixture with confidence-rated learning. (c)(f)The mean map without using confidence-rated learning.

where the subscript S stands for shadow and F for foreground, and TP and FN denote true positive and false negative, respectively. The $\overline{TP_F}$ is the number of ground-truth points of the foreground objects minus the number of points detected as shadows, but belonging to foreground objects.

We show the quantitative results in Table 1. Note that results of other's approaches are taken directly from [6][4].

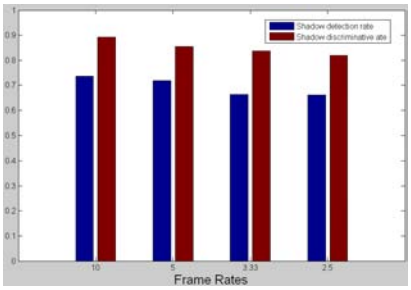


Figure 6. Quantitative results of the Intelligent Room sequence. Shadow detection and shadow discriminative rate are calculated under different frame rates settings.

5.3. Fast Learning of local Shadow Model

We demonstrate the effect of using confidence-rated learning in Figure 5. In this experiment, we learn the local shadow model in two traffic scenes: Highway I and Highway II. Figure 5(a)(d) show the background model of these two outdoor scene. With confidence-rated learning, we obtain the mean value of the most important Gaussian (i.e. with highest mixing weight and smallest variance) in Figures 5(b)(e). We can see that the local models of gradient intensity distortion under cast shadows are well constructed. On the other hand, if we learn the local shadow model following conventional Gaussian mixture learning method, then we obtain the results in Figure 5(c)(f), in which the

Table 1. Quantitative results on surveillance sequences

Sequence	Highway I		Highway II		Hallway	
Method	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$
Proposed	70.83	82.37	76.50	74.51	82.05	90.47
Kernel [6]	70.50	84.40	68.40	71.20	72.40	86.70
LGf [4]	72.10	79.70	-	-	-	-
GMSM [5]	63.30	71.30	58.51	44.40	60.50	87.00

models are still not built due to the long training time and disturbance by foreground objects.

5.4. Handling Scene with Few Foreground Activities

Here we use the benchmark sequence, Intelligent Room, to demonstrate the robustness of our approach to videos captured at low frame rates and scenes with few foreground activities. Downsampled sequences with lower frame rates are obtained by taking a sample from the original image sequence for every $M \in \{2, 3, 4\}$ samples. Thus, we have sequences with 10, 5, 3.33, and 2.5 frame rates, respectively. Figure 6 shows the quantitative results on these sequences. We can see that the performance on sequences lower frame rates degraded slightly, demonstrating that our approach can still learn and removing cast shadows even in such low frame rates.

6. Conclusion

In this paper, we have presented a novel algorithm capable of detecting cast shadows in various scenes. Qualitative and quantitative evaluation of the physics-based shadow model validated that our approach is more effective in describing background surface variation under cast shadows. The physics-based color features can be used to learn a global shadow model for a scene. Therefore, our method does not suffer from the problem of insufficient training data as in pixel-based shadow models. Moreover, with the aid of the global shadow model, we can update the local shadow models through confidence-rated learning, which is significantly faster than conventional online updating. To further improve the detection accuracy, more discriminative features or the spatial and temporal smoothness constraints can also be incorporated into the detection process in the future.

Acknowledgement

This work was supported in part by the National Science Council of Taiwan, R.O.C., under Grant NSC95-2221-E-001-028-MY3.

References

[1] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Trans. PAMI*, 25(10):1337–1342, 2003. 1

[2] T. Horprasert, D. Harwood, and L. S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *ICCV Frame-rate Workshop*, 1999. 1

[3] D. S. Lee. Effective gaussian mixture learning for video background subtraction. *IEEE Trans. PAMI*, 27(5):827–832, 2005. 5

[4] Z. Liu, K. Huang, T. Tan, and L. Wang. Cast shadow removal combining local and global features. In *CVPR*, pages 1–8, 2007. 1, 2, 7, 8

[5] N. Martel-Brisson and A. Zaccarin. Learning and removing cast shadows through a multidistribution approach. *IEEE Trans. PAMI*, 29(7):1133–1146, 2007. 1, 2, 8

[6] N. Martel-Brisson and A. Zaccarin. Kernel-based learning of cast shadows from a physical model of light sources and surfaces for low-level segmentation. In *CVPR*, June. 2008. 1, 2, 6, 7, 8

[7] B. Maxwell, R. Friedhoff, and C. Smith. A bi-illuminant dichromatic reflection model for understanding images. In *CVPR*, pages 1–8, June 2008. 2, 3

[8] S. Nadimi and B. Bhanu. Physical models for moving shadow and object detection in video. *IEEE Trans. PAMI*, 26(8):1079–1087, 2004. 1

[9] F. Porikli and J. Thornton. Shadow flow: a recursive method to learn moving cast shadows. In *ICCV*, pages 891–898 Vol. 1, Oct. 2005. 1, 2

[10] A. Prati, I. Mikic, M. Trivedi, and R. Cucchiara. Detecting moving shadows: algorithms and evaluation. *IEEE Trans. PAMI*, 25(7):918–923, 2003. 3, 6

[11] E. Salvador, A. Cavallaro, and T. Ebrahimi. Cast shadow segmentation using invariant color features. *CVIU*, 95(2):238–259, 2004. 1

[12] O. Schreer, I. Feldmann, U. Golz, and P. A. Kauff. Fast and robust shadow detection in videoconference applications. In *IEEE Int'l Symp. Video/Image Processing and Multimedia Communications*, pages 371–375, 2002. 1

[13] S. Shafer. Using Color to Separate Reflection Components. *Color Research Applications*, pages 210–218, 1985. 2

[14] J. Stander, R. Mech, and J. Ostermann. Detection of moving cast shadows for object segmentation. *IEEE Trans. Multimedia*, 1(1):65–76, Mar 1999. 1

[15] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. CVPR*, volume 2, pages –252, 1999. 4, 5

[16] W. Zhang, X. Fang, X. Yang, and Q. M. J. Wu. Moving cast shadows detection using ratio edge. *IEEE Trans. Multimedia*, 9(6):1202–1214, 2007. 1