

A PHYSICAL APPROACH TO MOVING CAST SHADOW DETECTION

Jia-Bin Huang and Chu-Song Chen

Academia Sinica, Institute of Information Science
Taipei, Taiwan, R.O.C.

ABSTRACT

This paper presents a physics-based approach capable of detecting cast shadows in video sequence effectively. We develop a new physical model of cast shadows without making prior assumption of the spectral power distribution (SPD) of the light sources and ambient illumination in the scene. The background appearance variation caused by cast shadows is characterized as the interaction of the blocked light sources and the background surface reflectance. We then take advantage of the statistical prevalence of cast shadows to learn and update the shadow model parameters using the Gaussian mixture model (GMM) over time. The proposed algorithm is completely unsupervised and can adapt to specific environment with complex illumination condition as well as changing shadow conditions. Experimental results on three challenging sequences demonstrate the effectiveness of the proposed method.

Index Terms— Visual surveillance, Object detection, Moving Cast Shadow Detection

1. INTRODUCTION

Detection of moving objects is an important low-level vision task for many vision-based applications. One of the most challenging problems in these applications is identifying cast shadows in video. Cast shadows are often misclassified as parts of foreground regions, resulting in distortion of the true shape and color properties of the target objects. An effective shadow detection method is thus necessary for accurate foreground segmentation.

Moving cast shadows are caused by the occlusion of light sources. Shadows reduce the total energy incident at the background surfaces where the light sources are partially or totally blocked by the foreground objects. Hence, points under shadows have lower luminance values but similar chromaticity values compared to their reference background points.

Many works have been proposed for cast shadow detection. Most of them are based on the assumption that the RGB values of shadow pixels will fall on the line between the illuminated values and the origin in the RGB color space (linear model). This linear model has been employed in different color spaces [1, 2, 3]. One main limitation of these methods is that they require explicit tuning in a large set of parameters for each new scene and are difficult to work for on-line applications. To adapt to environment changes, statistical learning methods have been employed to describe the cast shadows [4, 5, 6]. However, the linear model assumption may not be fulfilled in a real-world environment. For instance, in an outdoor scene the illumination may consist of direct sunlight, the diffused

light scattered by the sky, and other colored light from nearby surfaces (i.e. color bleeding). These lights can have significantly different spectral power distributions (SPD). Therefore, the RGB value of a shaded pixel may not attenuate linearly along the line between the value of the corresponding background and the origin.

Little attention has been directed toward this problem before. Nadami et al. [7] addressed this non-linearity by dichromatic reflection model for both the sun and the sky illuminations in an outdoor scene. However, the method is designed only for an outdoor environment. Recently, a more general shadow model was presented in [8]. This model introduced an ambient illumination term which determines a direction in the RGB space where the background values under cast shadows can be found. Since the ambient term may have spectrum which differs from the incident light sources, the values of the shaded pixels may not be proportional to the values under direct light. Nonparametric density estimation is used to model the surface variation under cast shadows in an unsupervised way. By providing a better description of cast shadows on background surfaces, the shadow model presented in [8] considerably improved the previous approaches based on linear model. However, in [8] the SPDs of all light sources are assumed having the same profile but with different power factor. This assumption may not hold for many of the real-world cases. In addition, the ambient illuminations are assumed to be constant, but they may be slightly changed due to circulation of foreground objects.

In this paper, we extend the shadow model presented in [8] by releasing these assumptions. The proposed shadow model is more general and can be applied to real-world cases more appropriately. However, a more general model may suffer from a longer period of time for learning the model parameters. To address this problem, we introduce the confidence-rated Gaussian mixture learning which accelerates the parameters convergent rate. Moreover, we observe that shadows on the light/shadow borders show different behavior from shadows inside the shaded region. The shadows detection rate can be further improved by addressing this issue.

The remainder of this paper is organized as follows. We first propose in Section 2 a new physical model for describing appearance variation of a background value under cast shadows. Based on the proposed model, we present an unsupervised learning approach to learn the behavior of cast shadows on surfaces by taking advantage of the statistical prevalence of shadows in Section 3. In Section 4, we derive the posterior probability of cast shadows and foreground given an observed sample. Experimental results are presented in Section 5. Section 6 concludes the paper.

2. PHYSICAL MODEL FOR CAST SHADOWS

According to the Lambertian model, the camera sensor response $g_k(p)$, $k \in \{R, G, B\}$ at pixel p depends on the intensity func-

This research was supported by the National Science Council of Taiwan under Grant No. NSC 95-2221-E-001-028-MY3 and 97-EC-17-A-02-S1-032.

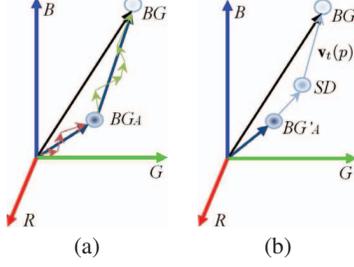


Fig. 1. Physical cast shadow model. (a) Linear attenuation model and the proposed model. The linear model will fail when the SPD of ambient illumination does not consistent with that of the light sources. (b) Background appearance variation due to cast shadows.

tion of the light sources $E(\lambda, p)$, the reflectance of an object surface $\rho(\lambda, p)$, and sensor spectral sensitivity $S_k(\lambda)$:

$$g_k(p) = \int E(\lambda, p)\rho(\lambda, p)S_k(\lambda)d\lambda. \quad (1)$$

We describe the illumination incident to a surface point p by N light sources and lights reflected from M ambient objects (the position p is ignore):

$$E(\lambda) = \sum_{n=1}^N E_{incident,n}(\lambda) + \sum_{m=1}^M E_{ambient,m}(\lambda). \quad (2)$$

For a color camera with a linear response, the response to a specific illumination follows a line in the RGB space. We show our model in Fig. 1(a), where BG_A is the ambient term from the contribution of all ambient lightings and BG is the background surface value. The zigzag lines starting from BG_A and ending at BG represent the response from the incident light sources. Based on the additive color model, their net effect forms the vector from BG_A to BG . Note that in this model, the ambient illumination can have different SPDs from that of the total contribution of incident light sources (i.e. the direction of BG_A may not be consistent with the direction from BG_A to BG). In addition, the SPD of individual light source may differ from that of other light sources.

In comparison, in [8] the SPDs of the incident light sources are assumed sharing the same profile. That is, the first term in (2) degenerates into $E_{incident}(\lambda) \sum_{n=1}^N \gamma_n$, where γ_n is the scaling factor of the n_{th} light sources. Moreover, the ambient illumination is assumed to be constant. Under these two assumptions, all shadow values are expected to be found on the line between background BG and the ambient value BG_A . This approach may fail to deal with environment with complex illumination conditions.

Given a surface point, the background surface value variation under cast shadows is resulted from the occlusion of lights by foreground objects. The blocked lights may contain partial or total direct incident light sources and some reflected lights from nearby ambient objects. Here, some of the incident lights will disappear since they are blocked by foreground objects. The first term in (2) is then reduced to $\sum_{n=1}^N E_{incident,n}(\lambda) - \sum_{n \in blocked} E_{incident,n}(\lambda)$. The ambient lights are also changed slightly due to the circulation of foreground objects. Therefore, according to (1), the vector from the shadow value (i.e. SD in Fig. 1) to its corresponding background value can be interpreted as the interaction between the blocked lights and surface reflectance (see Fig. 1(b)).

For a pixel p at time t , the vector from shadow to background value is denoted as $\mathbf{v}_t(p)$. The vector $\mathbf{v}_t(p)$ is the net response

from the contribution of the blocked incident and ambient illuminations. We then develop a 3D color feature of cast shadows $\mathbf{x}_{s,t}(p) = [\alpha_t(p), \theta_t(p), \phi_t(p)]^T$ to encode this information, where $\alpha_t(p)$ represents the illumination attenuation and $\theta_t(p)$ and $\phi_t(p)$ indicate the direction of the vector $\mathbf{v}_t(p)$ in spherical coordinate. Similar to Horprasert et al. [1], we define the illumination attenuation $\alpha_t(p)$ by

$$\alpha_t(p) = \frac{\|\mathbf{v}_t(p)\|}{\|BG_t(p)\|}, \quad (3)$$

where $BG_t(p)$ is the background value at the pixel p . The illumination attenuation, which is smaller than one, describes the strength of brightness of the blocked lights with respect to that of the background value. The direction of the vector $\mathbf{v}_t(p)$, the chromaticity line direction, is determined by $\theta_t(p)$ and $\phi_t(p)$:

$$\theta_t(p) = \arctan\left(\frac{\mathbf{v}_t^G(p)}{\mathbf{v}_t^R(p)}\right), \phi_t(p) = \arccos\left(\frac{\mathbf{v}_t^B(p)}{\|\mathbf{v}_t(p)\|}\right), \quad (4)$$

where the superscript R, G, B indicate the components in the RGB space. We can use this 3-D color feature vector $\mathbf{v}_t(p)$ to describe the appearance variation induced by the blocked light sources on shaded regions. Unlike [8], our model takes the variation of chromaticity direction into consideration and thus better describes the cast shadows on background surface.

Since different foreground objects often block the light sources in a similar way, cast shadows on background surfaces are thus similar and uncorrelated with foreground objects. We present in the following section how to learn the cast shadow model in a completely unsupervised way.

3. LEARNING AND DETECTING CAST SHADOWS

Shadow detection methods with fixed parameters can not adapt to time-varying illumination. In this paper, we use Gaussian mixture model (GMM) [9] to model cast shadows by using scene activities. We first apply weak shadow detector to moving pixels detected by background model to select possible shadow samples. The color features of these possible shadow pixels are then learned by GMM over time. Then, we can detect shadows by using the learned shadow model. Note that our approach is pixel-based, model the SPDs of ambient lighting and light sources locally, and thus can deal with environment with complex illumination conditions.

3.1. Weak Shadow Detector

To model the cast shadows, impossible shadow samples that belong to background and foreground should be excluded. Therefore, we apply a weak shadow detector that evaluates every moving pixels detected by the background model to filter out some impossible shadow samples. Since cast shadows on surface reduce luminance values and change the saturation, we define the potential shadow values fall into the conic volume around the corresponding background color [4].

The weak shadow detector is applied to moving regions detected by the background model to obtain possible shadow points. We call these possible shadow points as ‘‘shadow candidates’’, which form the set Q . At the pixel level, we model $\mathbf{x}_{s,t}(p), p \in Q$ using the GMM to learn the color feature of shadows.

3.2. Spatial Information

With the proposed physical shadow model, we can accurately describe the appearance variation of background surface under shadows at the pixel level. However, two shortcomings remain. One

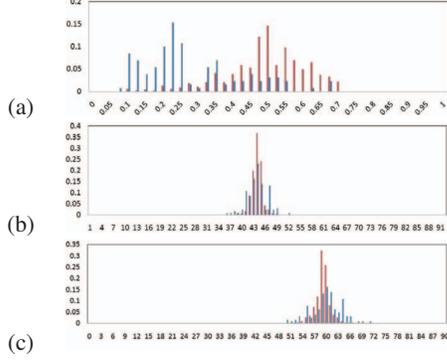


Fig. 2. Color feature value distribution of shadows inside (in red) and at the light/shadow border (in blue). Samples are taken from the sequence “Highway I”. (a) Illumination attenuation $\alpha_t(p)$ (b)(c) angle information: $\theta_t(p)$, and $\phi_t(p)$, respectively.

is that the proposed shadow model may suffer from slow learning of model parameters if motion is rare on a pixel. With the conventional Gaussian mixture learning [9], training the pixel-based shadow model requires a long time while the illumination conditions should remain stable. The other problem arises from foreground objects having similar colors as background, which may be misclassified as cast shadows. Therefore, we propose to incorporate spatial information for improving the shadow discriminative ability and the convergent rate of the model parameters.

In addition to the illumination attenuation, when a surface point is shaded, the gradient intensity around the point will also be smaller than that of the illuminated surface. Consequently, points having larger gradient intensities than background are more likely to be considered as foreground objects. We thus penalize these samples by lessening the learning rate in this situation. We define a confidence value of shadow using spatial gradient intensities of background $|\nabla(B_t)|$ and current image $|\nabla(I_t)|$:

$$\omega_t(p) = \frac{\varepsilon + |\nabla(B_t(p))|}{\varepsilon + \max\{|\nabla(I_t(p))|, |\nabla(B_t(p))|\}}, \quad (5)$$

where ε is a smooth term to alleviate the effect of noise in textureless regions. Thus, we can discriminate foreground objects with similar color as background from shadow with the aid of gradient information. To improve the convergent rate, we use $\omega_t(p)$ as the confidence value to update the GMM at pixel p through confidence-rated learning. In [10], Lee has shown that fast updating can be achieved by combining recursive filter type of learning with incremental EM in background modeling. We propose to weight these two learning scheme using the confidence value $\omega_t(p)$ extracted from spatial domain [11]. If $\omega_t(p)$ is large, the learning rate depends more on the formulation of incremental EM algorithm, which converges quickly with few observations. On the other hand, if $\omega_t(p)$ is small, the learning rate follows the recursive type of learning to maintain model stability.

3.3. Shadows at Light/Shadow Border

Once the shadow model for a pixel p is built, we can estimate the likelihood of the observation belong to shadows by checking whether the 3-D color feature vector $\mathbf{x}_{s,t}(p)$ is associated with one of the working Gaussian distribution in the mixture. However, if the light sources are clustered in the scene and are relatively brighter

than ambient illumination, blocking them induces a sharp shadow. In this case, shadows at light/shadow border show different behavior from shadows inside the shaded region. This is caused by the insufficient resolution of CCD camera. Pixels at light/shadow border may receive part of energy from light sources. We demonstrate this observation in Fig. 2. We first collected shadow samples throughout the sequence “Highway I”. Shadow border pixels can be detected from background posterior image where large gradient intensities are present. We classify these shadow samples into two groups: “inside” and “at border”. Fig. 2(a) presents that the illumination attenuation have significantly different distribution for shadows at border and inside. On the other hand, Fig. 2(b)(c) show that the distribution on angle information are very similar. The reason is that neighboring pixels often receive incident energies with similar SPDs. Consequently, we propose to address shadows at border by evaluating only their angle information.

4. POSTERIOR PROBABILITIES

In this section, we present how to derive the posterior probabilities of cast shadows and foreground given the observed sample $x_t(p)$ using the proposed shadow model and spatial information.

4.1. Cast Shadow Posterior

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

$$P(SD|x_t(p)) = P(SD|x_t(p), FS)P(FS|x_t(p)), \quad (6)$$

where $P(FS|x_t(p)) = 1 - P(BG|x_t(p))$ can be directly computed from the background model. Second, we remove pixels that are definitely foreground and consider only shadow candidate (SC). Also, we penalize pixels whose gradient intensities are amplified instead of reduced by multiplying $\omega_t(p)$: $P(SD|x_t(p), FS) = P(SD|x_t(p), FS, SC) * \omega_t(p)$. Here, the color feature vector $\mathbf{x}_{c,t}(p)$ is the sufficient statistics for the observed sample $x_t(p)$. Therefore, we have $P(SD|x_t(p), FS, SC) = P(SD|\mathbf{x}_{c,t}(p), FS, SC)$. As mentioned in section 3, we learn the cast shadows using the GMM [9]. The posterior probability can be expressed in terms of the mixture components in GMM:

$$P(SD|\mathbf{x}_{c,t}(p), FS, SC) = \sum_k P(SD|G_k)P(G_k|\mathbf{x}_{c,t}(p)), \quad (7)$$

where G_k is the k_{th} Gaussian state in the mixture. Given the observation that color change due to objects has higher variance compared to the color change due to cast shadows, Gaussian states with higher prior probabilities and smaller variances would be considered as shadows. Therefore, we approximate $P(SD|G_k)$ using logistic regression similar to that in [10] for background subtraction.

4.2. Foreground Posterior

Computing foreground posterior probability is much easier. Given a pixel p at time t , we first compute the background posterior $P(BG|x_t(p))$ from the background model. Then, we can obtain shadow posterior probability $P(SD|x_t(p))$ using the learned shadow model. Based on the probability theory, the foreground posterior can be obtained as:

$$P(FG|x_t(p)) = 1 - P(BG|x_t(p)) - P(SD|x_t(p)). \quad (8)$$

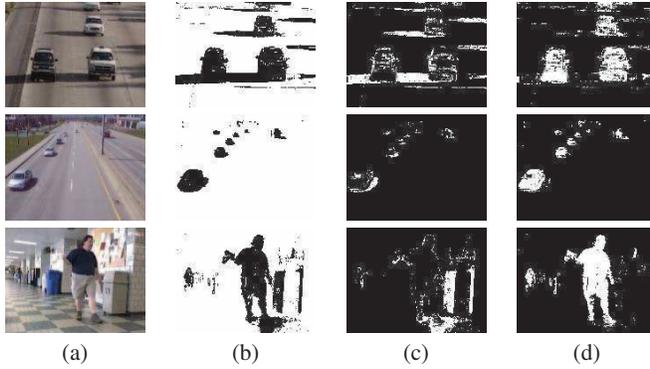


Fig. 3. Visual results of the proposed algorithm. (a) Original frame. (b) Background posterior $P(BG|x_t(p))$. (c) Shadow posterior $P(SD|x_t(p))$. (d) Foreground posterior $P(FG|x_t(p))$.

Table 1. Quantitative results on surveillance sequences

Sequence	Highway I		Highway II		Hallway	
	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$
Proposed	72.34	84.98	72.70	79.89	71.69	88.25
Kernel [8]	70.50	84.40	68.40	71.20	72.40	86.70
LGf [6]	72.10	79.70	-	-	-	-
GMSM [5]	63.30	71.30	58.51	44.40	60.50	87.00

5. EXPERIMENTAL RESULTS

We validate the proposed method on three challenging video sequences. Both visual results and quantitative measurement are presented. We compare our algorithm quantitatively with previous shadow detection methods when results are available. The effect of considering shadows at light/shadow border is also present.

5.1. Assessing Shadow Detection Accuracy

We show in Fig. 3 the visual results of three video sequences from the proposed algorithm. Fig. 3(a) shows one frame from the sequence. The background, shadow, and foreground posterior are given in Fig. 3 (b)(c)(d), respectively. We can see that in all three test sequences the cast shadows are correctly detected.

We evaluate our method through the quantitative metrics (shadow detection $\eta\%$ and discrimination rate $\xi\%$) proposed in [12]. For exact equations, please refer to [12]. We report our results in Table 1 and compare them with recent works using statistical method to detect shadows. Our method can achieve higher shadow detection and discrimination rate than the state-of-the-art approaches, where the quantitative results of previous works are directly taken from [8] and [6].

5.2. Effect of Shadows at Shadow/Light border

We show the visual results with and without considering shadows at shadow/light border. Fig. 4(a) is a traffic scene where sharp shadows are induced when blocking strong light sources (e.g. sun). Fig. 4(b) shows the foreground posterior without taking shadows at border into consideration. Shadows at border are misclassified as foreground due to the distinct distribution of color feature values. On the other hand, in Fig. 4(c), shadows at border can be correctly detected when evaluating their angle information only.

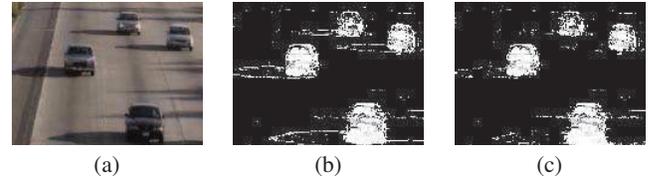


Fig. 4. Effect of shadows at shadow/light border (a) Original frame of sequence "Highway I". (b)(c) Foreground posterior without/with considering shadows at shadow/light border.

6. CONCLUSION

In this paper, we have introduced a novel physical model for cast shadow detection in video sequences. Qualitative and quantitative evaluation of the physical shadow model validate that our approach is more accurate in describing background surface variation under cast shadows. Spatial information is used to help discriminating foreground with similar chroma from cast shadows as well as improving the convergent rate of the shadow model through confidence-rated learning. We also address the problem of shadows at light/shadow border where these shadows have different behavior from other regions. Note that the proposed approach is pixel-based; a contextual model that incorporate spatial and temporal coherence can further enhance the performance.

7. REFERENCES

- [1] 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.
- [2] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. PAMI*, vol. 25, no. 10, pp. 1337–1342, 2003.
- [3] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," *CVIU*, vol. 95, no. 2, pp. 238–259, 2004.
- [4] F. Porikli and J. Thornton, "Shadow flow: a recursive method to learn moving cast shadows," in *ICCV*, Oct. 2005, pp. 891–898 Vol. 1.
- [5] N. Martel-Brisson and A. Zaccarin, "Learning and removing cast shadows through a multidistribution approach," *IEEE Trans. PAMI*, vol. 29, no. 7, pp. 1133–1146, 2007.
- [6] Z. Liu, K. Huang, T. Tan, and L. Wang, "Cast shadow removal combining local and global features," in *CVPR*, 2007, pp. 1–8.
- [7] S. Nadimi and B. Bhanu, "Physical models for moving shadow and object detection in video," *IEEE Trans. PAMI*, vol. 26, no. 8, pp. 1079–1087, 2004.
- [8] 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.
- [9] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. PAMI*, vol. 22, no. 8, pp. 747–757, 2000.
- [10] D. S. Lee, "Effective gaussian mixture learning for video background subtraction," *IEEE Trans. PAMI*, vol. 27, no. 5, pp. 827–832, 2005.
- [11] J. B. Huang and C. S. Chen, "Learning moving cast shadows for foreground detection," in *The Eighth International Workshop on Visual Surveillance (VS), in conjunction with ECCV*, 2008.
- [12] A. Prati, I. Mikic, M.M. Trivedi, and R. Cucchiara, "Detecting moving shadows: algorithms and evaluation," *IEEE Trans. PAMI*, vol. 25, no. 7, pp. 918–923, 2003.