Learning Moving Cast Shadows for Foreground Detection

Jia-Bin Huang, Chu-Song Chen

Learning Moving Cast Shadows for Foreground Detection

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

Abstract

We present a new algorithm for detecting foreground and moving shadows in surveillance videos. For each pixel, we use the Gaussian Mixture Model (GMM) to learn the behavior of cast shadows on background surfaces. The pixel-based model has the advantages over regional or global model for their adaptability to local lighting conditions, particularly for scenes under complex illumination conditions. However, it would take a long time for convergence if motion is rare on that pixel. We hence build a global shadow model that uses global-level information to overcome this drawback. The local shadow models are updated through confidence-rated GMM learning, in which the learning rate depends on the confidence predicted by the global shadow model. For foreground modeling, we use a nonparametric density estimation method to model the complex characteristics of the spatial and color information. Finally, the background, shadow, and foreground models are built into a Markov random field energy function that can be efficiently minimized by the graph cut algorithm. Experimental results on various scene types demonstrate the effectiveness of the proposed method.

1. Introduction

Moving object detection is at the core of many computer vision applications, including video conference, visual surveillance, and intelligent transportation system. However, moving cast shadow points are often falsely labeled as foreground objects. This may severely degrade the accuracy of object localization and detection. Therefore, an effective shadow detection method is necessary for accurate foreground segmentation.

Moving shadows in the scene 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, shadow points have lower luminance values but similar chromaticity values. Also, the texture characteristic around the shadow points remains unchanged since shadows do not alter the background surfaces.

Much research has been devoted to moving cast shadow detection. Some methods are based on the observation that the luminance values of shadowed pixels decrease respect to the corresponding background while maintaining chromaticity values. For examples, Horprasert et al. [6] used a computational color model which separates the brightness from chromaticity component and define brightness and chromaticity distortion. Cucchiara et al. [3] and Schreer et al. [15] addressed the shadow detection problem in HSV and YUV color space respectively and detected shadows by exploiting the color differences between shadow and background. Nadimi et al. [10] proposed a spatial-temporal albedo test and a dichromatic reflection model to separate cast shadows for moving objects.

Texture features extracted from spatial domain had also been used to detect shadows. Zhang et al. [20] proved that ratio edge is illumination invariant and the distribution of normalized ratio edge difference is a chi-square distribution. A significance test was then used to detect shadows. Fung et al. [4] computed a confidence score for shadow detection based on the characteristics of shadows in luminance, chrominance, gradient density, and geometry domains. However, the above-mentioned methods require to set parameters for different scenes and can not handle complex and time-varying lighting conditions. A comprehensive survey of moving shadow detection approaches was presented in [13].

Recently, the statistical prevalence of cast shadows had been exploited to learn shadows in the scenes. In [9], Martel-Brisson et al. used the Gaussian mixture model (GMM) to describe moving cast shadows. Proikli et al. [11] proposed a recursive method to learn cast shadows. Liu et al. [8] presented to remove shadow using multi-level information in HSV color space. The major drawback in [9] and [11] is that their shadow models need a long time to converge while the lighting conditions should remain stable. Moreover, in [9], the shadow model is merged into the background model. The Gaussian states for shadows will be discarded in the background model when there are no

shadows for a long period. Therefore, these two approaches are less effective in a real-world environment. Liu et al. [8] attempted to improve the convergence speed of pixel-based shadow model by using multi-level information. They used region-level information to get more samples and global-level information to update a pre-classifier. However, the method still suffers from the slow learning of the conventional GMM [17], and the pre-classifier will become less discriminative in scenes having different types of shadows, e.g. light or heavy.

Another foreground and shadow segmentation approaches are through Bayesian methods, which construct background, shadow, and foreground model to evaluate the data likelihood of the observed values [19], [1]. These two approaches used global parameters to characterize shadows in an image, which are constant in [19] and probabilistic in [1]. Although the global models are free from the convergence speed problem, they lose the adaptability to local characteristics and cannot handle scenes with complex lightings.

In this paper, we propose a Bayesian approach to moving shadow and object detection. We learn the behavior of cast shadows on surfaces by GMM for each pixel. A global shadow model (GSM) is maintained to improve the convergence speed of the local shadow model (LSM). This is achieved by updating the LSM through confidence-rated GMM learning, in which the learning rates are directly proportional to the confidence values predicted by the GSM. The convergence speed of LSM is thus significantly improved. We also present a novel description for foreground modeling, which uses local kernel density estimation to model the spatial color information instead of temporal statistics. The background, shadow, and foreground are then segmented by the graph cut algorithm, which minimizes a Markov Random Field (MRF) energy. A flow diagram of the proposed algorithm is illustrated in Fig. 1.

The remainder of this paper is organized as follows. The formulation of the MRF energy function is presented in Section 2. In Section 3, we describe the background and shadow model, primarily focusing on how to learn cast shadows. Section 4 presents a nonparametric method for estimating foreground probability. Experimental results are presented in Section 5. Section 6 concludes the paper.

2. Energy Minimization Formulation

We pose the foreground/background/shadow segmentation as an energy minimization problem. Given an input video sequence, a frame at time t is represented as an array $\mathbf{z}_t = (z_t(1), z_t(2), ...z_t(p), ...z_t(N))$ in RGB color space, where N is the number of pixels in each frame. The segmentation is to assign a label l_p to each pixel $p \in P$, where P is the set of pixels in the image and the label

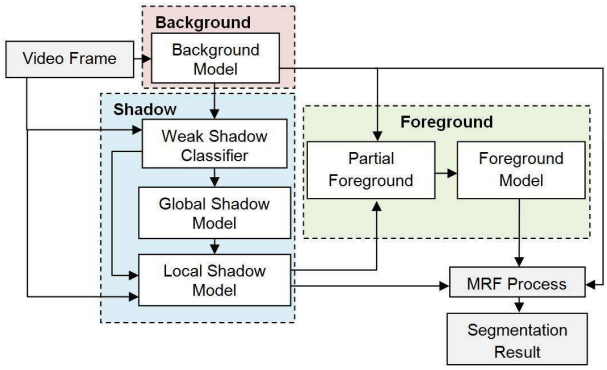

Figure 1. Flow chart of the proposed algorithm

$l_p \in \{BG, SD, FG\}$ corresponds to three classes: background, shadow, and foreground, respectively.

The global labeling field $L = \{l_p | p \in P\}$ is modeled as MRF [5]. The first order energy function of MRF can be decomposed as a sum of two terms:

$$\begin{aligned} E(L) = & \ E_{data}(L) + E_{smooth}(L), \\ = & \sum_{p \in P} [D_p(l_p) + \sum_{q \in N_p} V_{p,q}(l_p, l_q)] \end{aligned} \quad (1)$$

in which the first term $E_{data}(L)$ evaluates the likelihood of each pixel belonging to one of the three classes, the second term $E_{smooth}(L)$ imposes a spatial consistency of labels through a pairwise interaction MRF prior, $D_p(l_p)$ and $V_{p,q}(l_p, l_q)$ are the data and smoothness energy terms respectively, and the set N_p denotes the 4-connected neighboring pixels of point p. The energy function in (1) can be efficiently minimized by the graph cut algorithm [2], so that an approximating maximum *a posteriori* (MAP) estimation of the labeling field can be obtained.

The Potts model [12] is used as a MRF prior to stress spatial context and the data cost $E_{data}(L)$ is defined as

$$E_{data}(L) = \sum_{p \in P} D_p(l_p) = \sum_{p \in P} -\log p(z_p | l_p), \quad (2)$$

where $p(z_p|l_p)$ is the likelihood of a pixel p belonging to background, shadow, or foreground. In the following sections, we will show how to learn and compute these likelihoods.

3. Background and Shadow Model

3.1. Modeling the Background

For each pixel, we model the background color information by the well-known GMM [17] with K_{BG} states in RGB color space. The first B states with higher weights and smaller variances in the mixture of K_{BG} distributions are considered as background model. The index B is determined by

$$B = argmin_b \sum_{k=1}^{b} \omega_{BG,k} > T_{BG}, \tag{3}$$

where T_{BG} is the pre-defined weight threshold, and $\omega_{BG,k}$ is the weight of the k_{th} Gaussian state in the mixture model.

The likelihood of a given pixel p belonging to the background can be written as (the subscript time t is ignored):

$$p(z(p)|l_p = BG) = \frac{1}{W_{BG}} \sum_{k=1}^{B} \omega_{BG,k} G(z(p), \mu_{BG,k}, \Sigma_{BG,k}), \tag{4}$$

where $W_{BG} = \sum_{j=1}^{B} \omega_{BG,j}$ is a normalization constant, and $G(z(p), \mu_{BG,k}, \Sigma_{BG,k})$ is the probability density function of the k_{th} Gaussian with parameters $\theta_{BG,k} = \{\mu_{BG,k}, \Sigma_{BG,k}\}$, in which $\mu_{BG,k}$ is the mean vector and $\Sigma_{BG,k}$ is the covariance matrix.

3.2. Learning Moving Cast Shadow

This subsection presents how to construct the shadow models from the background model. Since different foreground objects often block the light sources in a similar way, cast shadows on background surfaces are thus similar and independent of foreground objects. The ratio of shadowed and illuminated value of a given surface point is considered to be nearly constant. We exploit this regularity of shadows to describe the color ratio of a pixel under shadow and normal illumination. We first use the background model to detect moving pixels, which might consist of real foregrounds and cast shadows. The weak shadow detector in Section 3.2.1 is then employed as a pre-filter to decide possible shadow points. These samples are used to train the LSM in Section 3.2.2 and a GSM in Section 3.2.3 over time. The slow convergence speed of LSM is refined by the supervision of the GSM through confidence-rated learning in Section 3.3.

3.2.1 Weak Shadow Detector

The weak shadow detector evaluates every moving pixels detected by the background model to filter out some impossible shadow points. The design principle of the weak shadow classifier is not to detect moving shadows accurately, but to determine whether a pixel value fits with the property of cast shadows. For simplicity, we pose our problem in RGB color space. Since cast shadows on a surface reduce luminance values and change the saturation, we define the potential shadow values fall into the conic volume around the corresponding background color [11], as illustrated in Fig. 2. For a moving pixel p, we evaluate the relationship of the observation pixel values $z_t(p)$ and the corresponding background model $b_t(p)$. The relationship can be characterized by two parameters: luminance ratio $r_l(p)$ and angle variation $\theta(p)$, and can be written as:

$$r_l(p) = \frac{\|b_t(p)\|}{\|z_t(p)\| \cos(\theta(p))}, \tag{5}$$

$$\theta(p) = \arccos\left(\frac{\langle z_t(p), b_t(p)\rangle}{\|z_t(p)\|\|b_t(p)\|}\right), \tag{6}$$

where $\langle \cdot, \cdot \rangle$ is the inner product operator, and $\| \cdot \|$ is the norm of a vector. A pixel p is considered as a potential cast shadow point if it satisfies the following criteria:

$$r_{min} < r_l(p) < r_{max}, \theta(p) < \theta_{max}, \tag{7}$$

where $r_{max} < 1$ and $r_{min} > 0$ define the maximum brightness and darkness respectively, and θ_{max} is the allowed maximum angle variation.

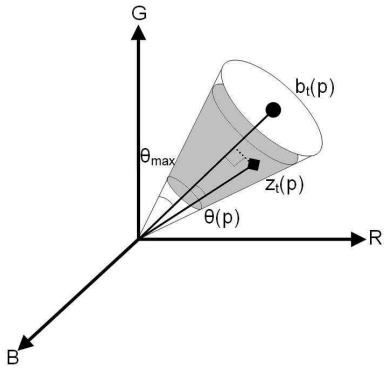

Figure 2. The weak shadow detector. The observation value $z_t(p)$ will be considered as shadow candidate if it falls into the shaded area

3.2.2 Local Shadow Model (LSM)

The color ratio between shadowed and illuminated intensity values of a given pixel p is represented as a vector of random variables:

$$r_t(p) = \left(\frac{z_t^r(p)}{b_t^r(p)}, \frac{z_t^g(p)}{b_t^g(p)}, \frac{z_t^b(p)}{b_t^b(p)}\right), \tag{8}$$

where $b_t^i(p), i = r, g, b$ are the means of the most probable Gaussian distributions (with the highest $\omega/||\Sigma||$ value) of the background model in three channels, and $z_t^i(p), i = r, g, b$ are the observed pixel values. The weak shadow detector is applied to moving regions detected by 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 $r_t(p), p \in Q$ using the GMM as the LSM to learn and describe the color feature of shadows.

Once the LSM is constructed at pixel p, it can serve as the shadow model to evaluate the likelihood of a local observation $z_t(p)$. Since we describe shadow by its color ratio of the pixel intensity values between shadow and background, the likelihood of an observation $z(p)$ can be derived by the transformation from "color ratio" to "color". Let $s(p)$ be a random variable describing the intensity values of a pixel p when the pixel is shadowed. Then, $s(p)$ can be characterized by the multiplication of two random variables: $s(p) = r(p)b(p)$, where $b(p)$ is the most probable Gaussian distribution with parameter $\{\mu_{BG,1}, \Sigma_{BG,1}\}$, and $r(p)$ is the color ratio modeled as GMM with parameter settings $\{\omega_{CR,k}, \mu_{CR,k}, \Sigma_{CR,k}\}$, in which CR denotes "color ratio". When the pixel p is shadowed, the likelihood can be written as:

$$p(z(p)|l_p = SD) = \frac{1}{W_{SD}} \sum_{k=1}^{S} \omega_{SD,k} G(z(p), \mu_{SD,k}, \Sigma_{SD,k}), \tag{9}$$

where normalization constant W_{SD} and S are determined in similar ways as in (3) with threshold T_{SD}; the weight $\omega_{SD,k}$, mean vector $\mu_{SD,k}$ and the diagonal covariance matrix $\Sigma_{SD,k}$ of the k_{th} are:

$$\omega_{SD,k} = \omega_{CR,k} \tag{10}$$

$$\mu_{SD,k} = \mu_{BG,1} * \mu_{CR,k} \tag{11}$$

$$\Sigma_{SD,k} = \mu_{BG,1}^2 \Sigma_{CR,k} + \mu_{CR,k}^2 \Sigma_{BG,1} + \Sigma_{CR,k} \Sigma_{BG,1} \tag{12}$$

Benedek et al. [1] also defined color features in CIELuv color space in a similar form and performed a background-shadow color value transformation. However, they did not take the uncertainty of the background model into consideration. Thus, the parameter estimation of the shadow model might become inaccurate for scenes with high noise level.

As gathering more samples, the precision of the LSM will become more accurate. With the online learning ability, the LSM can not only adapt to local characteristics of the background, but also to global time-varying illumination conditions. However, the LSM requires several training samples for the estimated parameters to converge. In other words, a single pixel should be shadowed many times while under the same illuminating condition. This assumption is not always fulfilled, particularly on the regions where motion is rare. Thus, the LSM on its own is not effective enough to capture the characteristics of shadows.

3.2.3 Global Shadow Model (GSM)

To increase the convergence rate, we propose to update the LSM using samples along with their confidence values. This is achieved by maintaining a global shadow model, which adds all shadow candidates in an image ($r_t(p), p \in Q$) into the model. In contrast to the LSM, the GSM obtains much more samples (every pixel $p \in Q$), and thus does not suffer from slow learning. We characterize shadows of GSM in an image also by GMM.

To evaluate the confidence value of a pixel $p \in Q$, we first compute the color feature $r_t(p)$ by (8) and check against every states in the GSM. If $r_t(p)$ is associated with the m_{th} state in the mixture of GSM, described as G_m, the confidence value can be approximated by the probability of G_m being considered as shadows:

$$C(r_t(p)) = p(l_p = SD|r_t(p)) = p(l_p = SD|G_m). \tag{13}$$

In GMM, Gaussian states with higher prior probabilities and smaller variances would be considered as shadows. Therefore, we approximate $p(SD|G_m)$ using logistic regression similar to that in [7] for background subtraction. On the other hand, if there are no states in GSM associated with the input color feature value $r_t(p)$, the confidence value is set to zero.

For each pixel $p \in Q$, we evaluate the confidence value $C(r_t(p))$ predicted by the GSM and then update the LSM through confidence-rated learning (presented in Section 3.3). With the proposed learning approach, the model needs not to obtain numerous samples to converge, but a few samples having high confidence value are sufficient. The convergence rate is thus significantly improved.

3.3. Confidence-rated Gaussian Mixture Learning

We present an effective Gaussian mixture learning algorithm to overcome some drawbacks in conventional GMM learning approach [17]. Let α_ω and α_g be the learning rates for the weight and the Gaussian parameters (means and variances) in the LSM, respectively. The updating scheme follows the the formulation of the combination of incremental EM learning and recursive filter:

$$\alpha_\omega = C(r_t) * (\frac{1 - \alpha_{default}}{\Sigma_{j=1}^{K} c_{j,t}}) + \alpha_{default}, \tag{14}$$

$$\alpha_g = C(r_t) * (\frac{1 - \alpha_{default}}{c_{k,t}}) + \alpha_{default}, \tag{15}$$

where $c_{k,t}$ is the number of match of the k_{th} Gaussian state, and $\alpha_{default}$ is a small constant, which is 0.005 in this paper. In the initial learning stage, the total number of match of the pixel $\Sigma_{j=1}^{K} c_{j,t}$ is small and thus the learning rate for weight is relatively large. As time goes on, the pixel becomes stable and the learning rate approaches to the type of recursive filter. Similarly, the learning rate for Gaussian parameters α_g is relatively large for newly-generated states. Instead of blind update scheme, which treats each sample in the same way, we propose to update a sample with its confidence value. The two learning rates are controlled by a confidence value $C(r_t)$, which indicates how confident the sample belongs to the class. Observations with higher confidence values will converge faster than those with low ones. The confidence-rated learning procedure in one dimension is described in Algorithm 1.

Here, we describe the drawbacks of the conventional GMM learning [17] for background modeling and how does the proposed learning approach overcome these disadvantages. Firstly, the method suffers from slow learning in the initial stage. If the first value of a pixel is a foreground object, the background model will have only one Gaussian state with weight equaling to unity. It will take a long time for the true pixel values to be considered as part of background. In the initial learning stage, our method will follow the incremental EM learning and approaches to recursive filter over time. Therefore, we do not suffer from the slow learning in the initial stage. Secondly, the method faces the trade-off problem between model convergence speed and stability. In order to maintain the system stability, a very small learning rate will be chosen to preserve a long learning history. However, a small learning rate results in slow convergence speed. While larger learning rate improves the convergence speed, the model becomes unstable. In the proposed method, the adaptability of the learning rate for Gaussian parameters allows fast convergence for the newly-generated states without degrading the stability. Lastly, there are tradeoffs regarding where to update in the image. Typically, there are two ways to update the model: selective update and blind update Selective update only adds samples being considered as background and thus enhances the detection of foreground. Unfortunately, this approach would fall into a deadlock situation whenever the incorrect update decision was made. On the other hand, the blind update adds all samples into the model, thus it might result in more false negatives. By updating all samples with corresponding confidence, the trade-off regarding how to update in the image can be avoid.

Algorithm 1: Confidence-Rated Learning Procedure in One Dimension

User-defined Variables : K, ω_{init}, σ_{init}, $\alpha_{default}$

Initialization : $\forall_{j=1...K}\ \omega_j = 0$, $\mu_j = Inf$, $\sigma_j = \sigma_{initial}$, $c_j = 0$

while *new observed data* r_t *with confidence* $C(r_t)$ **do**

$\alpha_\omega = C(r_t) * (\frac{1-\alpha_{default}}{\Sigma_{j=1}^{K} c_{j,t}}) + \alpha_{default}$

if r_t *is associated with the* k_{th} *Gaussian* **then**

for $j \leftarrow 1$**to** K **do**

if *j*=*k* **then** $M_{j,t}$**=1 else** $M_{j,t} = 0$

$\omega_{j,t} = (1 - \alpha_\omega) \cdot \omega_{j,t-1} + \alpha_\omega \cdot M_{j,t}$

end

$c_{k,t} = c_{k,t-1} + 1$

$\alpha_g = C(r_t) * (\frac{1-\alpha_{default}}{c_{k,t}}) + \alpha_{default}$

$\mu_{k,t} = (1 - \alpha_g) \cdot \mu_{k,t-1} + \alpha_g \cdot r_t$

$\sigma_{k,t}^2 = (1 - \alpha_g) \cdot \sigma_{k,t-1}^2 + \alpha_g \cdot (r_t - \mu_{k,t-1})^2$

else

for $j \leftarrow 1$**to** K **do**

$\omega_{j,t} = (1 - \alpha_\omega) \cdot \omega_{j,t-1}$

$\mu_{j,t} = \mu_{j,t-1}$, $\sigma_{j,t}^2 = \sigma_{j,t-1}^2$

end

$k = \text{argmin}_j \{\frac{\omega_{j,t}}{\sigma_{j,t}}\}$

$c_{k,t} = 1$

$\omega_{,t} = C(z_t) * (\frac{1-\omega_{init}}{\Sigma_{j=1}^{K} c_{j,t}}) + \omega_{init}$

$\mu_{k,t} = r_t$, $\sigma_{k,t}^2 = \sigma_{initial}^2$

end

end

4. Foreground Model and Segmentation

The foreground model has been described as a uniform distribution [19], providing a weak description of the foreground. In [16], they exploited the temporal persistence as a property to model foreground using joint domain-range nonparametric density function. However, the foreground model in the previous frame cannot provide an accurate description for foreground in the current frame when large motion occurs or new moving objects enter into the scene. Therefore, instead of exploring temporal statistics, we turn to spatial color information in the current frame. We use the background and shadow model to generate a partial foreground mask, i.e. a set of pixels that can be confidently labeled as foreground. Nonparametric method is used to estimate the foreground probability because the spatial augmented GMM might face the problem of choosing the correct number of modes. Details are introduced below.

In the partial foreground generation stage, we aim not to find all foreground pixels, but some samples from foreground. This is achieved by finding pixels whose intensity values are impossible to be generated from the existing background and shadow model. In other words, the partial foreground mask F can be generated by simple thresholding:

$$F = \{p | E_{BG}(z_t(p)) > \kappa, E_{SD}(z_t(p)) > \kappa\}, \quad (16)$$

where E(.) is the minus log of the likelihood of a pixel belonging to background or shadow, κ is a threshold.

We can now estimate the foreground probability of a query pixel p using samples from foreground. We use the neighborhood pixels around a query point p to gather M relevant samples, which forms the set $G_p = \{g_1, g_2, \cdots, g_M\}$.

When a pixel p is in the foreground, the probability can be estimated using kernel density estimation method [18]:

$$p(z_p | l_p = FG) = \frac{1}{M} \sum_{i=1}^{M} K_{\mathbf{H}}(z_p - g_i), \quad (17)$$

where $K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{1/2}\mathbf{x})$. The kernel, K, is taken to be a 3-variate density function with $\int K(\mathbf{w}) d\mathbf{w} = 1$, $\int \mathbf{w} K(\mathbf{w}) d\mathbf{w} = 0$, and $\int \mathbf{w}\mathbf{w}^T K(\mathbf{w}) d\mathbf{w} = I_3$, and $\mathbf{H}$ is a symmetric positive definite 3x3 bandwidth matrix. We choose a common Gaussian density as our kernel K,

$$K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2} (2\pi)^{-d/2} exp(-\frac{1}{2}\mathbf{x}^{\mathbf{T}}\mathbf{H}^{-1}\mathbf{x}). \quad (18)$$

Although variable bandwidth kernel density estimators can lead to improvement over kernel density estimators using global bandwidth [14], the computational cost of choosing an optimal bandwidth is expensive for a real-time surveillance system. Hence, the bandwidth matrix $\mathbf{H}$ is assumed to be diagonal, in which only two parameters are defined: variances in spatial σ_s, and in color domain σ_c.

After computing the data likelihood of the local observation $z_t(p)$ of three classes, we can perform maximum *a posteriori* (MAP) estimator for the label field L. The energy function of Eq. 1 can be efficiently minimized by the graph cut algorithms.

A concrete description of the learning process and detection process are shown in Algorithms 2 and 3, respectively.

Algorithm 2: Learning Process

At time t, with segmentation fields S_t from the detection process
for *each pixel* $p \in P$ **do**
  Update the background model
  if $S_t(p) \neq BG$ **then**
    if *pixel p satisfies shadow property (Eq. 7)* **then**
      Compute confidence value $C(r_t(p))$ from global shadow model (Eq. 13)
      Update global shadow model
      Update local shadow model at pixel p with confidence $C(r_t(p))$
    end
  end
end

Algorithm 3: Detection Process

At time t,
for *each pixel* $p \in P$ **do**
  Compute background likelihood $P(z_t(p)|l_p = BG)$ (Eq. 4)
  Compute shadow likelihood $P(z_t(p)|l_p = SD)$
  Generate partial foreground mask F (Eq. 16)
  Find samples respect to pixel p from the neighboring pixels of $p \in F$
  Compute foreground likelihood $P(z_p|l_p = FG)$ (Eq. 17)
end
Construct the graph to minimize Eq. 1
Generate the segmentation fields S_t

5. Experimental Results

We have implemented and test the proposed approach on various scene types. In the LSM and GSM, three and five Gaussian states are assigned, respectively. The initial variance of both LSM and GSM are set as 0.0025. We use three Gaussian models for the background, and the covariance matrix is assumed diagonal.

Here, we describe the test sequences we used in this paper.

- "Highway" video is a sequence from the benchmark set [13] . This sequence shows a traffic environment. The shadows appearing in the video are dark and cover large area in the scene.
- "Laboratory" video is a typical indoor sequence with light shadow. But, the lighting conditions are more complex than the outdoor scenes.
- "Intelligent Room" [13] is an indoor sequence with low shadow strength.
- "Entrance" [1] are typical surveillance videos captured in different time of a day, thus with different lighting conditions.

Figure. 3 and 4 demonstrate the effectiveness of the proposed method in both outdoor and indoor scenes. (a) shows the original image in the sequence. (b) shows the moving pixels detected by the background model and the gray regions of (b) is labeled as shadow points by the weak shadow detector. Note that the weak shadow detector has

a very high detection rate, but with some false detections. (c) is the confidence map predicted by GSM. The lighter these regions are, the higher confidence value the model predicts. (d) shows the segmentation result after performing the graph cut algorithm.

The effect of GSM is illustrated by Figure 5. The detection result without using the GSM is presented in Figure 5(b), in which some portions of real foreground are miss labeled as shadows. This is due to the slow learning of the LSM. In Figures 5(c), the foreground objects are accurately detected. Figure. 6 demonstrates the adaptability of the proposed algorithm to different lighting conditions. Figure. 6(a),(c), and (e) show the same scene captured in different periods of the day: morning, noon, and afternoon, respectively. As shown in the Figure. 6(b),(d), and (f), our method can detect the foreground and shadows correctly. Table 5 and 5 show the quantitative comparison with previous approaches. The shadow detection and discriminative rate η and ξ follow the metrics described in [13]. The readers should refer to [13] for exact equations.

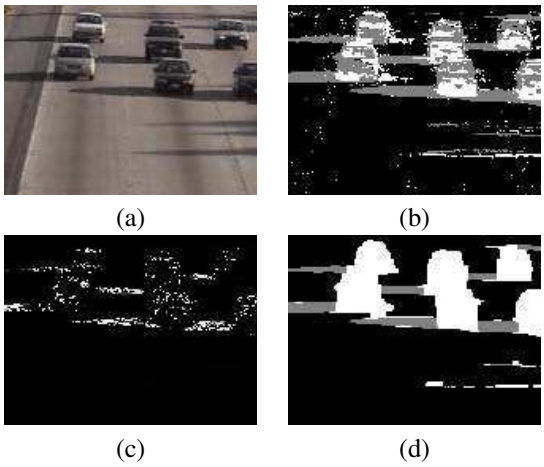

Figure 3. Outdoor sequence: "Highway". (a) Frame in the sequence. (b) Detection result by the weak shadow detector. (c) The confidence map by GSM. (d) Foreground and Segmentation result

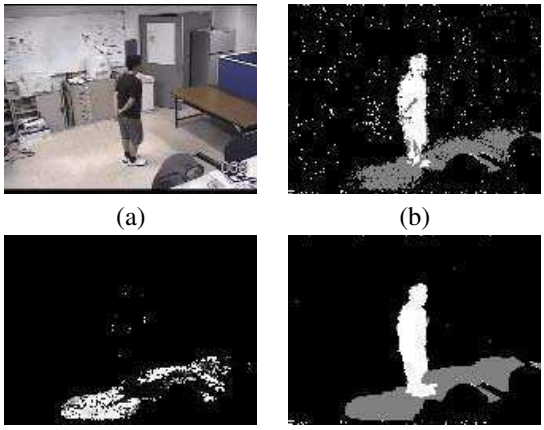

Figure 4. Indoor sequence: "Laboratory". (a) Frame in the sequence. (b) Detection result by the weak shadow detector. (c) The confidence map by GSM. (d) Foreground and Segmentation result

Figure 5. Effect of global shadow model in the sequence "Intelligent Room". (a) Frame in the sequence. (b) Detection result without using GSM (c) Detection result with GSM

6. Conclusion and Future Work

In this paper, we have presented a general model for foreground/shadow/background segmentation. There are several novelties in this work. For cast shadow detection, we have proposed a pixel-based model to describe the properties of shadows. The pixel-based model has the advantage over the global model that it can adapt to the local illumination conditions, particularly for the background under complex lighting conditions. To solve the slow convergence speed of local shadow model, we maintain a global shadow model to predict the confidence value of a given sample, and update the local shadow model through confidence-rated learning. We have developed a nonparametric foreground model that exploits the spatial color characteristics, free from the assumption that the object motion is slow. The likelihoods of background, shadow, and foreground are built into MRF energy function in which an optimal global inference can be achieved by the graph cut algorithm. The effectiveness and robustness of the proposed method have been validated on both indoor and outdoor surveillance videos. We have only introduced the color feature in this work. In the future, other features such as edge or motion can also be easily incorporated into our framework.

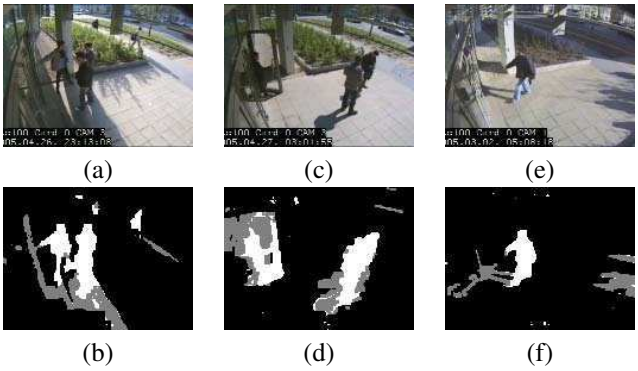

Figure 6. Different periods of the day in the "Entrance" sequence. (a) In the morning. (c) At noon. (e) In the afternoon. (b)(d)(f) Detection results of frame in (a),(c), and (e), respectively

Table 1. Quantitative comparison on "Highway" sequence.

Method	η%	ξ%
Proposed	76.76%	95.12%
SNP [13]	81.59%	63.76%
SP [13]	59.59%	84.70%
DNM1 [13]	69.72%	76.93%
DNM2 [13]	75.49%	62.38%

Acknowledgment

This research was supported by the National Science Council of Taiwan under Grant No. NSC 96-3113-H-001-011 and NSC 95-2221-E-001-028-MY3.

References

[1] C. Benedek and T. Sziranyi. Bayesian foreground and shadow detection in uncertain frame rate surveillance videos. *IEEE Trans. Image Processing*, 17(4):608–621, April 2008.

[2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, Nov 2001.

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

[4] G. S. K. Fung, N. H. C. Yung, G. K. H. Pang, and A. H. S. Lai. Effective moving cast shadow detection for monocular color traffic image sequences. *Optical Engineering*, 41(6):1425–1440, 2002.

Table 2. Quantitative comparison on "Intelligent Room" sequence.

Method	η%	ξ%
Proposed	83.12%	94.31%
SNP [13]	72.82%	88.90%
SP [13]	76.27%	90.74%
DNM1 [13]	78.61%	90.29%
DNM2 [13]	62.00%	93.89%

[5] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *PAMI*, pages 721–741, Nov. 1984.

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

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

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

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

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

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

[12] R. Potts. Some generalized order-disorder transformations. In *Proc. of the Cambridge Philosoph. Soc.*, volume 48, page 81, 1952.

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

[14] S. R. Sain. Multivariate locally adaptive density estimation. *Comput. Stat. Data Anal.*, 39(2):165–186, 2002.

[15] 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.

[16] Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *PAMI*, 27(11):1778–1792, Nov. 2005.

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

[18] M. Wand and M. Jones. *Kernel Smoothing*. Chapman & Hall/CRC, 1995.

[19] Y. Wang, K.-F. Loe, and J.-K. Wu. A dynamic conditional random field model for foreground and shadow segmentation. *PAMI*, 28(2):279–289, Feb. 2006.

[20] Z. Wei, F. Xiang Zhong, X. K. Yang, and Q. M. J. Wu. Moving cast shadows detection using ratio edge. *IEEE Trans. Multimedia*, 9(6):1202–1214, 2007.