Learning and Removing Cast Shadows through a Multidistribution Approach

Nicolas Martel-Brisson, *Student Member*, *IEEE*, and André Zaccarin, *Member*, *IEEE*

Abstract—Moving cast shadows are a major concern for foreground detection algorithms. The processing of foreground images in surveillance applications typically requires that such shadows be identified and removed from the detected foreground. This paper presents a novel pixel-based statistical approach to model moving cast shadows of nonuniform and varying intensity. This approach uses the Gaussian mixture model (GMM) learning ability to build statistical models describing moving cast shadows on surfaces. This statistical modeling can deal with scenes with complex and time-varying illumination, including light saturated areas, and prevent false detection in regions where shadows cannot be detected. The proposed approach can be used with pixel-based descriptions of shadowed surfaces found in the literature. It significantly reduces their false detection rate without increasing the missed detection rate. Results obtained with different scene types and shadow models show the robustness of the approach.

Index Terms—Shadow detection, GMM, GMSM, background subtraction, multidistribution, segmentation, image models, pixel classification.

1 Introduction

Moving cast shadows are a major concern for foreground detection algorithms. Image pixels representing cast shadows differentiate themselves from the background and generally fall within the group of pixels associated with foreground objects. Labeling cast shadows as foreground objects induces silhouette distortions and object fusions, thus reducing vision algorithm aptitudes in scene monitoring, target counting, and object recognition.

Shadow detection algorithms can be classified as property-based or model-based algorithms. Property-based approaches use features like geometry, brightness, or color to identify shadowed regions. Unlike model-based techniques, they do not use any a priori knowledge of scene geometry, foreground objects, or light sources. Well suited to particular situations, model-based approaches have shown less robustness than property-based algorithms when used in different scene and illumination conditions [1]. A review of shadow detection techniques can be found in [2].

A surface's appearance depends on its reflectivity properties and the total energy incident at the surface. When an object casts a shadow on a surface, it deprives the latter of direct illumination from a light source, hence inducing a variation of its appearance. This variation is more or less severe as a function of the scene composition, such as the presence of other light sources and the reflectivity properties of other scene objects. In the particular case of an indoor scene, the light sources are generally white and numerous, the objects populating the scene are closely distributed, and they diffuse photons in every direction, leading to good ambient lighting. Thus, when an object occludes a surface from direct illumination generated by one of many light sources, the total energy incident to this surface varies slightly. The perceived shadow is light and the surface appearance is relatively maintained.

In these particular conditions, cast shadows on a surface reduce luminance values while maintaining chromaticity values. In an RGB color space, this means that background values under a cast shadow are proportional to background values under direct lighting. This behavior is exploited in many property-based algorithms. For example, in [3], Salvador et al. use the fact that a shadow darkens the surfaces on which it is cast to identify an initial set of shadowed pixels, that is then pruned by using color invariance and geometric properties of shadows. In [4], Cucchiara et al. use the hypothesis that shadows reduce surface brightness and saturation while maintaining hue properties in the HSV color space. Schreer et al. [5] adopt the YUV color space to avoid using the time consuming HSV color transformation and segment shadows from foreground objects based on the observation that shadows reduce the YUV pixel value linearly. In [6], Horprasert et al. build a model in the RGB color space to express normalized luminance variation and chromaticity distortions. In addition to using scene brightness properties, [7] uses edge width information to differentiate penumbra regions from the background. The algorithm in [8] combines luminance, chrominance, and gradient density scores in a shadow confidence score function for segmentation. In [9], Finlayson et al. use shadow edges along with illuminant invariant images to recover full color shadow-free images. Finally, [10] computes the luminance ratio between lighted and shadowed surfaces in a training set and uses this ratio as a constant to segment shadows in real images.

• *The authors are with the Computer Vision and Systems Lab, Department of Electrical and Computer Engineering, Université Laval, Quebec City, Quebec, Canada, G1K-7P4. E-mail: {nmartel, zaccarin}@gel.ulaval.ca.*

Manuscript received 28 Oct. 2005; revised 12 May 2006; accepted 29 Aug. 2006; published online 18 Jan. 2007.
Recommended for acceptance by G. Finlayson.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0584-1005.
Digital Object Identifier no. 10.1109/TPAMI.2007.1039.

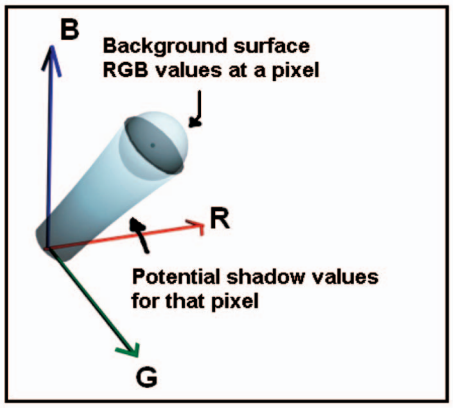

Fig. 1. Background RGB values (sphere) centered at the mean pixel value and associated shadow volume (tapered cylinder) assuming chroma constancy under cast shadows.

Even in controlled environments, these algorithms based on chroma constancy under cast shadows can falsely label pixels as cast shadows. The following illustrates this problem. Foreground detection algorithms using pixel statistics, such as Gaussian Mixture Models (GMM) [11], represent the background value at a pixel with one (or several) Gaussian distributions. The likely background values at a pixel can then be represented as a volume. In the simplified case where we assume equal variances and thresholds along the RGB axis, this volume is reduced to a sphere, as depicted in Fig. 1. The RGB values associated with a shadowed background also form a volume whose shape and size are functions of the properties and parameters used to describe the shadow. Typically, this volume is substantial and is illustrated as a tapered cylinder in Fig. 1.

Even if it correctly captures the possible range of RGB values of a cast shadow, this volume will generally be too large for a given scene at a given time and can lead to false detections of shadows. Fig. 2 shows a clear example of a foreground object which falls in the shadow volume for three different shadow models. Since the person's shirt has a lower luminance value than the background but similar chromaticity values, algorithms based on chromaticity consistency of shadowed background falsely label the shirt as a shadowed surface. Without a priori knowledge of the scene, the shadow volume defined by property-based shadow detection algorithms cannot be reduced without increasing the number of missed detections. With this large shadow volume, however, foreground objects with chromaticity values similar to that of the background may also fall into this volume and will be wrongly labeled as shadows.

In a more general situation, light sources are not exclusively white and the objects' chromatic properties diffuse on their neighbors (color bleeding). Consequently, RGB background values under a cast shadow will not necessarily be proportional to RGB values under direct light. This situation occurs, for example, in outdoor scenes when blocking a surface from direct sunlight since the light scattered by the sky has a spectrum which differs from that of the sun. Hence, the cast shadow is deep and bluish [2].

Fig. 2. Foreground detection and pixels labeled as cast shadows by three different shadow models. First row: YUV model of Section 5.2. Second and third rows: models of [6] and [4]. First column: frame from the sequence. Second column: foreground detection. Third column: pixels detected as shadows.

To address these problems, shadow volumes must be defined locally to reflect scene lighting conditions. Our approach allows this kind of parameterization by using a multidistribution statistical learning process.

2 Proposed Approach and Underlying Assumption

The appearance of a shadowed surface shows a certain regularity even in scenes with complex illumination conditions. This regularity is caused by several factors: The light sources are generally stable and fixed, the foreground objects circulating in the scene have a similar scale factor, and they move following physical constraints like walls, ground, roads, hallway, etc. Since different foreground objects block light sources in a similar manner, the shadows cast on the background surfaces are relatively similar at the pixel level. This phenomenon is particularly strong in busy hallways or highways where different people or different cars induce the same intensity variation on a surface when blocking a light source. We exploit the repetitiveness of the appearance of cast shadows to learn shadowed surface values. This is done by parameterizing probability density functions representing these shadowed surfaces.

Using a property-based algorithm, we first detect pixels capturing shadowed surfaces by verifying if their RGB values fall within the shadow volume of the background. This shadow volume is determined by the property-based algorithm. As we illustrated in the previous section, there will be many false detections. However, at this stage, they

are not an issue. It is important, however, to minimize missed detections.

All pixel values are then fed to a multidistribution learning algorithm. In our implementation, we use a GMM. Values that are frequently seen by a pixel, like background values, are captured into stable Gaussian distributions, while values that are infrequently seen are quickly discarded. Shadow values lie between both situations: They are not as frequent as background values but their rate of appearance is higher than random foreground values. Hence, shadow values are associated with frequently seen distributions which are labeled as foreground by the GMM. To prevent them from being quickly discarded, we increase their learning rate, leading to stable shadow distributions.

However, because shadows are not constantly cast on surfaces, the GMM may eventually discard shadow distributions just as other foreground distributions can be discarded. We therefore need to store the parameters of the shadow distributions to enable shadow detection based on learned distributions. Also, complex and changing scene conditions may require learning and storing the parameters of more than one shadow distribution per pixel in order to correctly capture the RGB values of cast shadows. To do so, we use a second multidistribution learning process, also based on a GMM, that we call the Gaussian Mixture Shadow Model (GMSM). The GMSM is composed of learned distributions representing background surfaces when shadows are cast on them. We periodically identify stable shadow distributions in the GMM and we use their parameters to update those of shadow distributions stored in the GMSM. A distribution in the GMM is identified as capturing cast shadows if its mean value falls within the shadow volume of the background value.

This learning process is continuous. The detection or labeling process is independent. For each frame of the image sequence, a pixel is labeled as background, foreground, or as a moving cast shadow. A pixel is labeled as a moving cast shadow if its value can be associated with one of the distributions stored in the GMSM at that time.

This approach differs from existing algorithms in four distinct aspects. First, the properties of shadowed surfaces are learned and described by probability density distributions. There is no local supervision to determine whether a pixel represents a shadowed surface or not. A global property-based condition is only used to identify pixels from which the shadow distributions are learned. Second, the shadow descriptions are pixel-based and time varying. They can therefore adapt to a scene's nonuniform and time-varying lighting conditions. Third, the approach can also learn shadow distributions whose RGB values deviate slightly from the hypothesis that they are proportional to the background RGB values, without increasing false detections. Finally, regions where moving cast shadows cannot be detected are excluded from the shadow model. These regions are generally regions where a shadow is already cast from a background object or dark surfaces where cast shadows are not visible. By excluding these regions, we further reduce false shadow detection when a foreground object, with chromaticity properties like those of the background, crosses the scene. Test results show that our proposed framework clearly enhances the segmentation quality of property-based shadow detection algorithms described in the literature.

In Section 3, we begin by summarizing the GMM. Our approach for learning and detecting moving cast shadow is explained in detail in Section 4. Results showing the robustness and advantages of our approach are provided in Sections 5, 6, and 7. Section 8 illustrates how the GMSM can be extended to deal with shadows on light saturated surfaces. Concluding statements are given in Section 9.

3 Gaussian Mixture Models

Our approach for modeling and segmenting cast shadows was developed to be integrated into a background detection algorithm based on a GMM. In this section, we summarize the main elements of this approach as described in [11] and modified for online implementation in [12].

For each pixel, a fixed number of states K, typically between 3 and 5, is defined. Some of these states will model the YUV values of background surfaces and the others, foreground surfaces. Each pixel value X_t is a sample in a color space of a random variable $\mathbf{X}$. A Gaussian probability density function $f_{\mathbf{X}|k}$ and a priori probability ω_k are associated with each state $k \in \{1, 2, \ldots, K\}$. The Gaussian probability density function with parameters $\theta_k = \{\mu_k, \sigma_k\}$ describes the color components of a surface that comes into the pixel's view,

$$f_{\mathbf{X}|k}(X|k, \theta_k) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu_k)^T \Sigma_k^{-1}(X-\mu_k)}, \quad (1)$$

where μ_k is the mean and Σ_k is the covariance matrix. We assume that the components of X_t are uncorrelated so that Σ_k is diagonal and may be represented by the n-dimensional variance σ_k^2. The parameter ω_k is the a priori probability that the surface modeled by $f_{\mathbf{X}|k}$ will come into the pixel view in the next frame and $\sum_{k=1}^{K} \omega_k = 1$. Assuming that each pixel views background states more often than foreground ones, the K states are ordered by decreasing values of the $\omega_k/\|\sigma_k\|$ ratio. This ratio is an indicator of the state's stability. A state converges or becomes more stable when $\omega_k/\|\sigma_k\|$ increases, hence increasing its ranking. Background states have a higher a priori probability and a lower variance and, hence, larger $\omega_k/\|\sigma_k\|$ than foreground states.

The first B states whose combined a priori probability of appearing is greater than a threshold T, i.e.,

$$B = \underset{b}{argmin}\left(\sum_{k=1}^{b} \omega_k > T\right), \quad (2)$$

are labeled as background states and the other states are labeled as foreground states. Consequently, the GMM performs better when the foreground is a small fraction of the frame most of the time. Usually, only one state per pixel will be sufficient to parameterize the background adequately. This reduces the relative importance of setting a good threshold T since a small T will guarantee that only one state is considered as background. In situations where the background is multimodal, a large threshold like $T = 0.8$ will allow the GMM to model the background with

more than one state if necessary. Particular attention, however, should be given to the choice of T in situations where the background is multimodal and the activity in the scene is strong. In those situations, T should be large enough to characterize the background with more than one state but small enough to prevent foreground states from being labeled as background.

When a new frame is acquired, each pixel value is associated with a state and the pixel is labeled background or foreground according to that state. Pixel value X_t is associated to the state k with the smallest label among the states satisfying

$$d_{k,t}^T d_{k,t} < \lambda^2, \tag{3}$$

where

$$d_{k,t} = \big(diag(\sigma_{k,t,1}, \sigma_{k,t,2}, \sigma_{k,t,3})\big)^{-1}(X_t - \mu_{k,t}). \tag{4}$$

If we cannot associate a pixel value to an existing distribution, a new state k is created around this value with a priori probability w_{init} and the less probable state is dropped. At each time instant, the a priori probability ω_k of each state is updated: It is increased for the state to which the pixel value is associated and reduced for all of the other states. We also update the distribution parameters of the state capturing the pixel value. The equations describing the GMM update are [12]:

$$\omega_{k,t+1} = (1 - \alpha_t)\omega_{k,t} + \alpha_t P(k|X_t, \Phi), \tag{5}$$

$$\mu_{k,t+1} = (1 - \rho_{k,t})\mu_{k,t} + \rho_{k,t} X_t, \tag{6}$$

$$\begin{aligned}\sigma_{k,t+1}^2 = (1 - \rho_{k,t})\big((\sigma_{k,t}) \circ (\sigma_{k,t})\big) + \rho_{k,t}\big((X_t - \mu_{k,t+1}) \\ \circ (X_t - \mu_{k,t+1})\big),\end{aligned} \tag{7}$$

where

$$\rho_{k,t} = \frac{\alpha_t P(k|X_t, \Phi)}{\omega_{k,t+1}}. \tag{8}$$

Typically, $P(k|X_t, \Phi)$ is set to 1 if there is a match and 0 elsewhere. Φ represents the total set of parameters, $\Phi = \{\omega_1 \ldots \omega_k, \theta_1 \ldots \theta_k\}$, $\circ$ is the element-wise (Hadamard) multiplication operator, and α_t is the learning parameter related to the convergence rate of the GMM states.

4 Learning Shadow Distributions

In this section, we describe our approach for learning cast shadow distributions as integrated with a GMM for background modeling and detection. As we mentioned in Section 2, our approach exploits the repetitiveness of shadow values to learn shadow distributions. There are three steps in the learning process: 1) identification of pixel values that could represent cast shadows, 2) facilitating the generation of stable shadow distributions within the GMM, and 3) learning and storing parameters of shadow distributions with the GMSM. These steps are described in the following subsections as well as the detection or labeling process, which is performed independently of the learning process.

4.1 Shadowed Surface Properties

Our learning process first requires the identification of pixels whose values could describe a shadowed surface. This identification is done using a property-based description of a shadowed surfaces, like the ones discussed in the introduction. Formally, if we match a pixel value X_t to a nonbackground state of the GMM, following (3), we then verify if that pixel value matches the description of a shadowed surface for one of the background states $k = 1, \ldots, B$. We symbolize this by testing if:

$$f_s\big(X_t, \theta_{1..B,t}\big) = 1, \tag{9}$$

where $f_s(.,.)$ represents the test associated to a given property-based description of shadowed surfaces. Any property-based description of shadowed surface can be used with our algorithm. In the following section, we introduce a YUV-based description that we have implemented. Other shadowed surface descriptions [4], [6] were also implemented and results are presented in Sections 6 and 7.

The pixels meeting the test of (9) form the pool of information from which we learn the values of surfaces with cast shadows. It does not mean, however, that these pixels will be labeled as shadows. As shown in Fig. 2, many of these pixel values are, in fact, foreground objects wrongly labeled as shadowed surfaces. We next explain how we take advantage of the learning property of the GMM to filter out any false shadow values.

4.2 GMM Convergence of Shadowed Surfaces

The second step toward achieving shadowed background parametrization is to exploit the convergence of frequently seen distributions within a GMM. Using one of our test sequences, Fig. 3 illustrates how distributions capturing cast shadows converge in the GMM. The right-hand side of the figure shows the distributions of the states created through the sequence for one pixel (circled in red). These distributions are normalized by $log\, \omega_k / \|\sigma_k\|$ to illustrate the state stability and only one color channel is shown. When a static scene is perturbed by an event (Fig. 3a), a new state is initialized for each pixel representing the foreground object with a distribution centered on the pixel value. Since, in this example, the GMM had only one initialized state for that pixel, the algorithm creates a second one which is the most stable foreground state, illustrated by a red distribution in Fig. 3b. If a person casts a shadow on that pixel (Fig. 3c), a new state representing the value of this shadow will be created with a small a priori probability ω_{init} (green distribution in Fig. 3d). When a shadow is cast again in similar conditions on the pixel (Fig. 3e), the algorithm will associate the pixel values to the same new state describing the shadow value and the a priori probability of this state, ω_k, will increase (red distribution in Fig. 3f). As more people walk into the scene (Fig. 3g), the shadow state will become the more stable foreground state (illustrated by the red distribution's height in Fig. 3h) and the state parameters (mean and variance) will converge around the sample values. If the scene has multiple light sources and cast shadows by foreground objects are light and deep, one Gaussian distribution may not be adequate to cover the

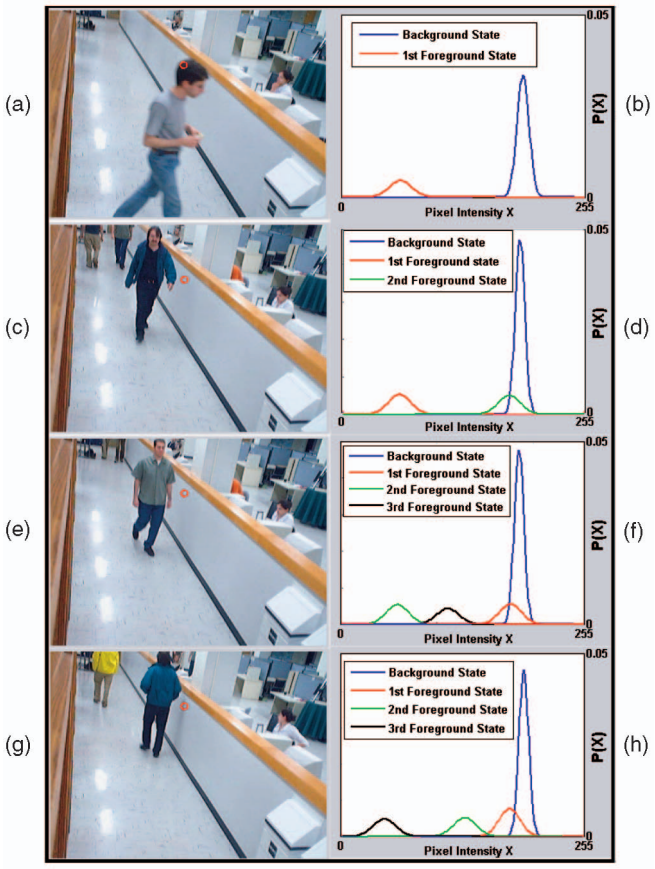

Fig. 3. Convergence of shadow distributions within the GMM. As more people cast shadows, shadow states are created and their stability (and stability ranking) increases. The colors used to represent the state distributions from highest ranking to lowest ranking are blue, red, green, and black.

entire range of shadow values. Thus, two or more states will represent shadow values. Within the GMM, the ranking of these states will fluctuate as a function of the scene dynamics.

When a pixel value is associated to a state, the a priori probability of the state increases as

$$\omega_{k,t+1} = \omega_{k,t} + \alpha_{s,t}\, M_{k,t}, \tag{10}$$

where $\alpha_{s,t}$ is the learning parameter, $M_{k,t}$ is equal to 1 for the state that is associated to the pixel value and zero for the other states, hence increasing only the a priori probability of the associated state. With this formulation, which differs from (5), the a priori probability of an unmatched state only decreases by normalization.

Moreover, unlike the update equation originally proposed in [11], the learning parameter $\alpha_{s,t}$ is not only a function of time but also of the pixel value X_t. When the pixel value could describe a shadow over the background surface, we increase the learning rate of the state associated to this pixel value. This modification allows a state representing a shadow on a background surface to rapidly become a stable foreground state (i.e., its ranking based on $\omega_k/\|\sigma_k\|$ will increase).

To do so, we define

$$\alpha_{s,t} = \alpha_t S, \tag{11}$$

where the parameter S is greater than 1 when X_t matches the description of a shadowed background surface (see (9)) and $S = 1$ in the other cases. In our implementation, we use a GMM with four states $(K = 4)$ and relative learning rate $S = 3$ when the shadow test (9) is positive. We also tested values of $S = 2$, 2.5, and 4 without much effect on the results. S should be relatively small (≤ 4), especially if the scene shows significant activity, so that the stability ranking of "accelerated" shadow states remains lower than the ranking of a background state. The GMMs described in the literature usually have between three and five states. In our case, we recommend at least four states since the GMM models not only background states but also cast shadows.

When there are no people or objects crossing the scene for a long time, the a priori probability w_k of the foreground states modeling the cast shadows will tend toward zero. Since w_k may become smaller than w_{init}, any future detection of a foreground event will result in a new state and the destruction of foreground states capturing the values of shadowed surfaces. By imposing a maximum value $w_{1,max} = 0.95$ on the a priori probability of state $k = 1$, we conserve the most frequently appearing foreground states, which are most likely to be shadow states.

The pixels meeting test (9) form the pool of information from which we learn the luminance and chromaticity values of surfaces with cast shadows. It does not mean, however, that these pixels will be labeled as shadows. As shown in Fig. 2, many of these pixel values are, in fact, foreground objects wrongly labeled as shadowed surfaces. We stress that the learning process of the GMM filters out false shadow detections. Foreground pixels that meet the shadow test (9) are far more infrequent than cast shadows and their values are scattered in the entire shadow volume. These values lead to the creation of multiple foreground states that do not become stable within the GMM and are therefore quickly destroyed over time since the total number of states K is finite.

4.3 Gaussian Mixture Shadow Models

The next and final step in the learning process is to periodically transfer to a second mixture model, the GMSM, the parameters of shadow distributions learned in the GMM. As we explained in Section 2, this is necessary because shadow distributions can be destroyed in the GMM and because complex and changing scene conditions often require more than one Gaussian distribution to capture the range of shadow values. The GMSM is like a GMM except that its input values are Gaussian probability density functions $f_{\mathbf{X}|k}$ with parameters $\theta_k = \{\mu_k, \sigma_k\}$ instead of realizations of a random variable $\mathbf{X}$. Gaussian distributions that are transferred to the GMSM are those of the GMM states identified as describing cast shadows on background surfaces.

At each time interval (Δt), we process the most stable GMM foreground state, i.e., state $k = B + 1$. We test if the mean of this distribution could describe a shadowed surface, using the same property-based model as in (9):

$$f_s\left(\mu_{B+1,t}, \theta_{1..B,t}\right) = 1. \tag{12}$$

If this test is true, the parameters of the $B+1$ GMM state distribution $f_{\mathbf{X}|B+1}$ are transferred to the GMSM. They are then compared to the existing GMSM distributions:

$$d_{k,t}^T d_{k,t} < \lambda_{s,a}^2, \tag{13}$$

where

$$d_{k,t} = \left(diag(\sigma_{k,t,1}^s, \sigma_{k,t,2}^s, \sigma_{k,t,3}^s)I\right)^{-1}(\mu_{B+1,t} - \mu_{k,t}^s) \tag{14}$$

and where we use the superscript s when referring to the GMSM. If there is a match, the parameters θ_k^s are updated:

$$\mu_{k,t+1}^s = (1-\alpha^s)\mu_{k,t}^s + \alpha^s \mu_{B+1,t}, \tag{15}$$

$$\mu_{k,t+1}^s = (1-\alpha^s)\sigma_{k,t}^s + \alpha^s \sigma_{B+1,t}, \tag{16}$$

where α is a constant. Also, we set $M_{k,t}^s = 1$ in the following equation:

$$\omega_{k,t+1}^s = \omega_{k,t}^s + \alpha^s M_{k,t}^s. \tag{17}$$

If there is no match, a new state is added in the GMSM, up to a maximum of K^s states. For this new state,

$$\omega_{k,t+1}^s = \omega_{init}^s, \tag{18}$$

$$\mu_{k,t+1}^s = \mu_{B+1,t}, \tag{19}$$

$$\sigma_{k,t+1}^s = \sigma_{B+1,t}. \tag{20}$$

The a priori probabilities $\omega_{k,t}^s$ are then normalized and the states sorted in decreasing order of $\omega_{k,t}^s$. We do not use the ratio $\omega_k/\|\sigma_k\|$ since the variance is relatively constant for all states in the GMSM. Within the GMSM, the first B^s states, where

$$B^s = \underset{b}{argmin}\left(\sum_{k=1}^{b}\omega_{k,t}^s > T^s\right), \tag{21}$$

are used to detect moving cast shadows on a background surface. T^s is chosen relatively large (0.9 in our simulations) to include most of the states defined in the GMSM, but it is not equal to one to ensure that states with small relative probability $\omega_{k,t}^s$ are not considered. Our hypothesis is that relatively infrequent states with shadow-like characteristics can find their way in the GMSM and would result in false detections if they were considered.

Typically, a scene with complex illumination will have two to three valid GMSM states at each pixel for describing cast shadows. If the background changes, new shadow surfaces will be learned by the model and old states describing old shadowed surfaces will be discarded with time.

4.4 Shadow Detection

The shadow detection process uses the GMM and GMSM data, but does not interfere with the learning process. First, using the GMM, a pixel is labeled as background or foreground. If it is labeled as foreground, this pixel value, X_t, is then compared to the shadow states of the GMSM:

$$d_{k,t}^T d_{k,t} < \lambda_{s,b}^2, \tag{22}$$

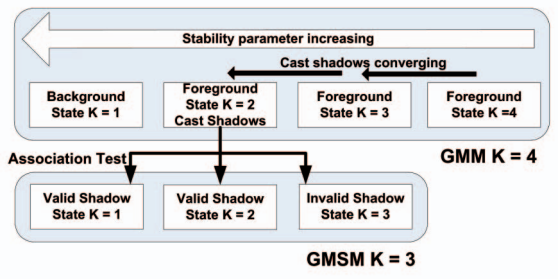

Fig. 4. This figure illustrates, at the distribution level, the relation between the GMM and GMSM. For the purpose of the illustration, we assume a GMM with four states, of which only one represents the background, and a GMSM with three states, of which only the first two are used to label pixels as shadows.

where

$$d_{k,t} = \left(diag(\sigma_{k,t,1}^s, \sigma_{k,t,2}^s, \sigma_{k,t,3}^s)I\right)^{-1}(X_t - \mu_{k,t}^s). \tag{23}$$

If this condition is met for a state k with $k \leq B^s$, the pixel is labeled as a moving cast shadow. In order to be considered as a shadow, a pixel value X_t must fall into the volume defined by the distributions of valid GMSM states, a volume which is smaller than the one given by the property-based description of shadows. For all of our scenarios, we set $\lambda = 3$ for both GMM and GMSM since, for a Gaussian distribution, more than 99.7 percent of the samples lie within this threshold.

4.5 Summary of GMM/GMSM Algorithm

Algorithms 1 and 2 summarize in pseudolanguage the learning and labeling processes of the GMM/GMSM algorithm. Fig. 4 illustrates the relation between the GMM and the GMSM at the level of state distributions. The figure shows that distributions created within the GMM and that converge, i.e., distributions whose ranking based on stability increases, are transferred to GMSM if they capture cast shadows. This figures emphasizes that distributions are learned in the GMM, that only stable shadow distributions are transferred to the GMSM, and that the input to the GMSM are distributions or, more specifically, their parameters.

Algorithm 1: Learning Process

```
for each image pixel X_t at each time t do
    Set S = 1;
    if pixel value can be associated to a GMM state (eq. 3) then
        if pixel value respects shadow properties on surfaces (eq. 9) then Set S = 3;
        Update learning rate (eq. 11) and other parameters of the associated state (eq. 10, 6, 7);
    else
        Destroy state K;
        Create new state in GMM using the pixel value;
    end
    if time = nΔt and mean of state B + 1 respect shadow properties (eq. 12) then
        if mean can be associated to a shadow state in GMSM (eq. 13) then
            Update shadow state parameters (eq. 17, 15, 16);
        else
            Destroy state K^s  Create new shadow state in GMSM (eq. 18, 19, 20);
        end
    end
end
```

Algorithm 2: Labeling Process

```
for each image pixel X_t at each time t do
    if the rank of the pixel associated state is greater than B then
        if pixel value can be associated to GMSM state (eq. 22) then
            if rank of associated state is less than or equal to B^s then
                Label pixel as cast shadow;
            else Label pixel as foreground;
        else Label pixel as foreground;
    else Label pixel as background;
end
```

5 GMM/GMSM Performance Based on a YUV Description of Shadowed Surfaces

The results presented in this section and in Sections 6 and 7 illustrate the properties and performance of the approach we have described for shadow detection. We begin by describing our experimental conditions. We follow by giving the YUV-based characterization of shadowed surfaces used in the GMM/GMSM to generate the results presented in this section.

5.1 Experimental Conditions

All results shown in this paper have been obtained with a stationary low cost webcam, a Logitech QuickCam Pro 4000 with a resolution of 320 $\times$ 240 pixels, connected via a USB port to a laptop with a 2.2 GHz Intel Celeron processor. The frame rate of our software implementation was approximately 10 Hz. Results shown here are raw results, without any postprocessing. For each environment, parameters were set once. Results selected are typical of the algorithm's performance throughout the sequences.

As explained in Section 4, a property-based shadow description is used by the learning process. For each frame, pixels that satisfy (9), i.e., pixels that would be labeled as shadows by the property-based model, are assigned to foreground states which may eventually become shadow states. During any run of our algorithm, we can therefore extract, for each frame, the pixels that are labeled as shadows by the property-based model and compare this to the GMSM's result. This will show that, for the same set of parameters and the same scene conditions, a property-based description fails to properly identify shadowed pixels but can be used to build a GMSM that correctly labels shadowed pixels.

We have implemented and tested in the GMM/GMSM algorithm three different property-based descriptions of shadowed surfaces. Our objective is not to compare the performance of these three models against each other. Instead, we want to demonstrate that the GMM/GMSM improves the performance of a variety of property-based shadow models.

5.2 YUV-Based Description of Shadowed Surfaces

Like others, this description is also based on the widely used simplification that a shadow cast on a surface will equally attenuate the value of its three color components. We first estimate this attenuation ratio using the luminance component Y and we then verify that both the U and V components are also reduced by a similar ratio.

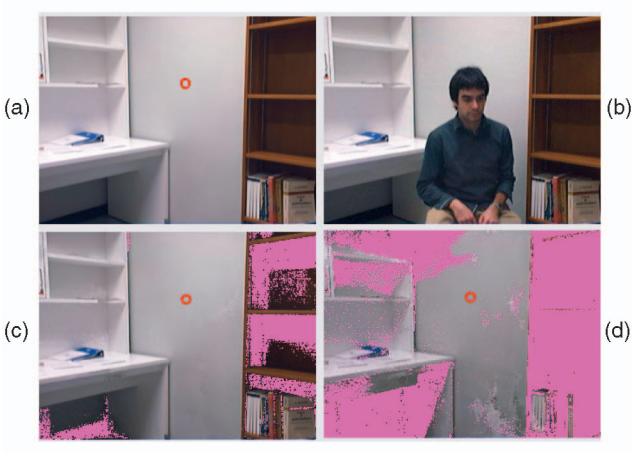

Fig. 5. Office scene with complex illumination. (a) Mean value of the first GMM background state. (b) A frame in the sequence. (c) Mean value of the first GMSM state, $\mu^s_{1,t}$. (d) Mean value of the second GMSM state, $\mu^s_{2,t}$.

More specifically, the color vector X represents the shadow cast on a surface whose average color value is μ with variance σ_μ if:

$$\alpha_{min} < \alpha_Y < 1 \quad \text{with } \alpha_Y = X_Y/\mu_Y, \tag{24}$$

$$(1/\sigma_{\mu_U})|X_U - \alpha_Y \mu_U| < \Lambda_U, \tag{25}$$

$$(1/\sigma_{\mu_V})|X_V - \alpha_Y \mu_V| < \Lambda_V. \tag{26}$$

This set of conditions defines the function $f_s(X, \theta_{1,t})$, where X stands for X_t or $\mu_{B+1,t}$, depending on whether we are testing a pixel value or the mean of the first nonbackground distribution from the GMM, as given by (9) and (12). The variable α_{min} is a threshold on maximum luminance reduction. $\Lambda_{U,V}$ represents the tolerable chromaticity fluctuation around the surface value $\mu_{U,V}$. Setting α_{min} requires a fast scene analysis. For example, in outdoor situations, cast shadows could be light or deep. Thus, the threshold must be set to zero since the luminance of a bright surface could fall near dark values under a cast shadow. In this case, the number of false shadow detections will be high with the property-based algorithm, especially for foreground objects having luminance properties near the black vertex. However, distribution convergence within the GMM filters out these cases. In indoor situations, where shadows are light, setting a nonzero α_{min} improves the convergence rate of true shadow distributions since it prevents dark foreground values on bright surfaces from being mistaken for potential shadow values.

5.3 Shadow Description Using Multiple States

The first set of results demonstrates that multiple shadow states within the GMSM are necessary and appropriate to capture and model the dynamic range of cast shadows. These results were obtained by creating shadows in a static scene (Fig. 5a) with complex illumination. Because of the multiple light sources on the ceiling and the high reflectivity of the walls, shadows cast on the background by moving objects have a large variation in intensity. At a given pixel, the shadow will go from being fairly light to

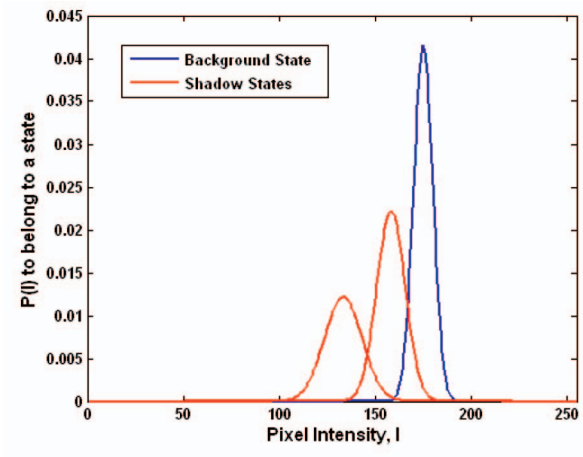

Fig. 6. Distributions of normalized background state (blue) and shadow states (red).

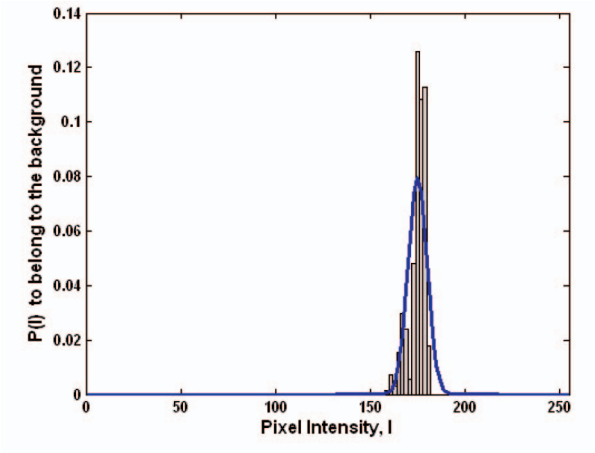

Fig. 7. Histogram of the background pixels acquired during the video sequence and the computed background distribution.

being fairly deep as a function of the position of the moving object. A snapshot of these shadows can be seen in Fig. 5b.

Figs. 5c and 5d, respectively, show the mean intensity of the first two GMSM states, i.e., $\mu^s_{1,t}$ and $\mu^s_{2,t}$. We can observe that the pixel mean values are darker than those of the background state and that we can visually discriminate the mean values of the two shadow states, illustrating that different levels of shadows were well captured by the GMSM. In these figures, pink pixels are those for which the GMSM was never initialized because shadows were not cast or were not detectable. Shadows are not detectable if the value of a shadowed pixel falls within the distribution of a background state or if the stability of its state never increased sufficiently to be transferred to the GMSM. In Fig. 5c, these areas are those with static shadows like under the desk or in the bookshelves.

In more complex scenes, a simple Gaussian kernel is not the best approximation of shadow values. Indeed, for a surface point, lights could be blocked partially and the transition from a background state to a shadow state would not be binary. Also, the overall distribution of shadow values does not necessarily match a parametric distribution, unlike video aquisition noise which is often assumed to be Gaussian. However, unlike a standard GMM, where only one state is generally labeled as the background state, we label almost all states in the GMSM as shadows. As said in Section 4.3, the threshold T^s of (21) is set high to only filter out states with a very low probability of occurrence. Therefore, more than one Gaussian kernel is used to model the distribution of cast shadow values for a single surface point.

The following refers to the pixel circled in red in Fig. 5. For this pixel, the GMM/GMSM architecture led to the creation of two shadow states. The Gaussian kernels associated with the background and shadow states for one color channel are illustrated in Fig. 6. The distributions of the background state and the shadow states are normalized by $\omega_k/\|\sigma_k\|$ and $\omega^s_k/\|\sigma^s_k\|$, respectively. The relative amplitude between the background state and the shadow state distributions is irrelevant since these distributions originate from two different models, the GMM and the GMSM. However, we can see for this situation that the transition from a background kernel to a shadow kernel is continuous.

For this pixel, we first manually identified the frames where it represented the background or a cast shadow, i.e., we excluded frames where it represented a foreground object. Since, for the naked eye, it is impossible to differentiate the lower end of background values and the upper end of shadow values, we then used the labeling done by the GMM/GMSM to separate frames showing a background from frames showing a cast shadow. Fig. 7 presents the histogram of the background values superimposed with the background Gaussian kernel computed during the sequence. The histogram of the shadow values is shown in Fig. 8 superimposed with the global shadow distribution, f_G, which we define as

$$f_G = \sum_{k=1}^{B^s} \left[\omega_k f_{\mathbf{I}|k}(I|k, \theta_k)\right] \tag{27}$$

since the first B^s states of the GMSM capture the shadow values. In this example, only two states had been defined for this pixel. Fig. 8 shows that, although the histogram of shadow values is not Gaussian, the histogram can be adequately modeled using more than one Gaussian distribution.

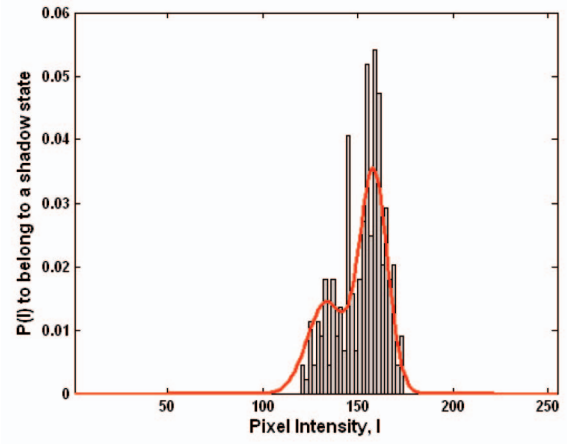

Fig. 8. Histogram of the cast shadow pixels acquired during the video sequence and the computed shadow distribution.

Fig. 9. Foreground and shadow detection for Fig. 5. (a) Foreground detection from GMM. (b) Shadow detection with the first GMSM state. (c) Shadow detection with any of the B^s GMSM states (maximum of three in this case).

Fig. 9 shows the segmentation results obtained for the frame of Fig. 5b. The detected foreground, including cast shadows, is presented in Fig. 9a. Figs. 9b and 9c clearly demonstrate the value of using many states to capture the shadow values. A single shadow state would fail to detect some shadowed pixels, as shown on the right of the image in Fig. 9b.

5.4 Hallway 1

The following set of results were obtained in noncontrolled environments. Fig. 10a shows a hallway where light shadows are cast by people circulating (Fig. 10b). Fig. 10e shows the pixels labeled as shadow by the YUV model, (24), (25), and (26), while Fig. 10f shows the pixels labeled as shadow by the GMSM. Fig. 10f shows that the GMSM detects the cast shadows without all of the false detections of the YUV model.

Using the parameters extracted from the GMM and GMSM, we have computed background and shadow volumes for the pixel identified by the red circle in Fig. 10a. The volume given by (24), (25), and (26) is shown in gray in Fig. 11. This volume follows the hypothesis of proportionality between the background and its shadowed values (illustrated by the blue line). The shadow volume does not extend to the dark values of the color space because α_{min} was set to 0.4 since cast shadows in this scenario are light. The red circle within the shadow volume is the YUV value of a foreground object (here, the color of someone's shirt, see Fig. 10b) observed at the same location. The volume defined by the distribution of the first and only GMSM shadow state for the same pixel is shown in blue in the image. We can see that the space occupied by the GMSM shadow volume is much smaller, thus reducing the number of false shadow detections. In this example, the chromaticity values of the shirt are similar to those of the background and fall within the shadow volume (gray), but outside the GMSM shadow volume (blue). This example clearly shows the advantage of learning the distribution of shadow values. Although the volume defined by the GMSM is much smaller, it does not result in missed detections, as shown in Figs. 10f and 10h. Note that we could not reduce the number of false hits with the YUV model by constraining its parameters without dramatically increasing the number of misses.

5.5 Hallway 2

The next set of results, shown in Fig. 12, were also acquired in a hallway. Here, the wall and the floor high reflectivity combined with the multiple light sources (windows and fluorescent lamps) caused the GMSM to converge to two

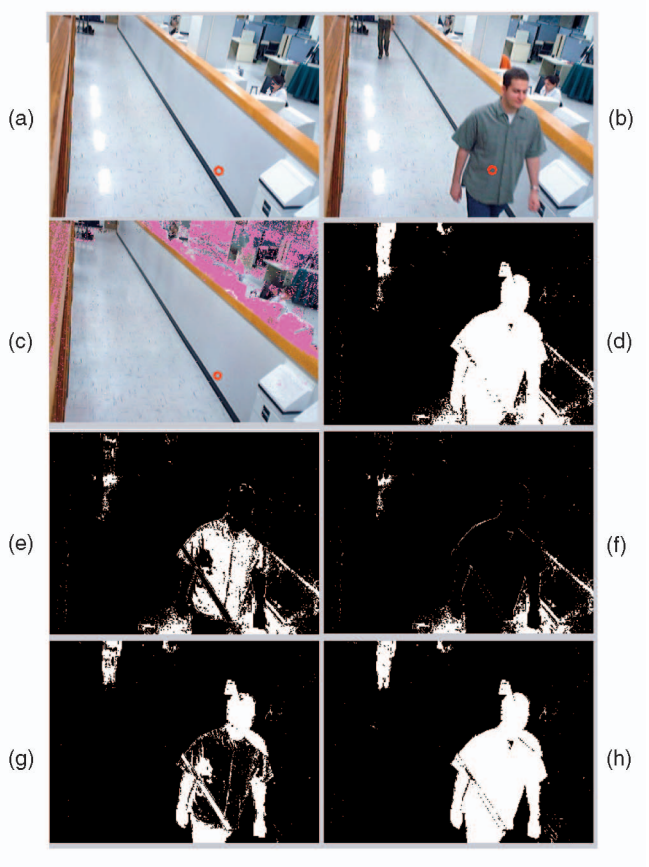

Fig. 10. Hallway 1. (a) Mean value of first background state of GMM. (b) Frame in the sequence. (c) Mean value of the first GMSM state. (d) Foreground detection from GMM. (e) Shadow detection from the YUV description. (f) Shadow detection from the GMSM. (g) Foreground detection: Img (d) - Img (e). (h) Foreground detection: Img (d) - Img (f).

states for a significant number of pixels in order to adequately model cast shadows. We note here that the first state of the GMSM was not initialized on the upper part of the walls. Shadows were not cast on these surfaces due to the position of the light sources. Moving objects with chromaticity values like those of the walls, but with lower luminance values, cannot be falsely detected as shadow

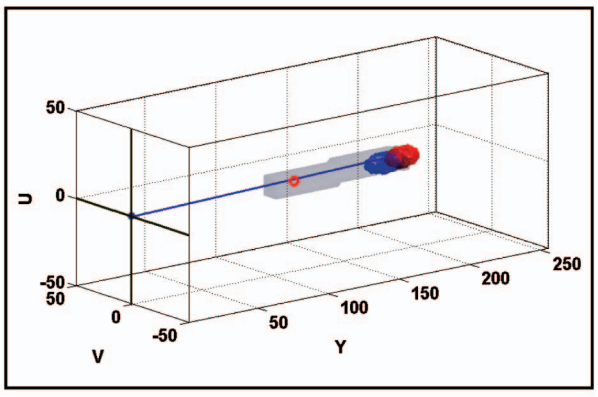

Fig. 11. Shadow volumes from the YUV (gray) and GMSM (blue) models and background volume (red) for the pixel identified in Fig. 10a. The red circle represents the YUV value of the pixel identified in Fig. 10b, which belongs to the foreground.

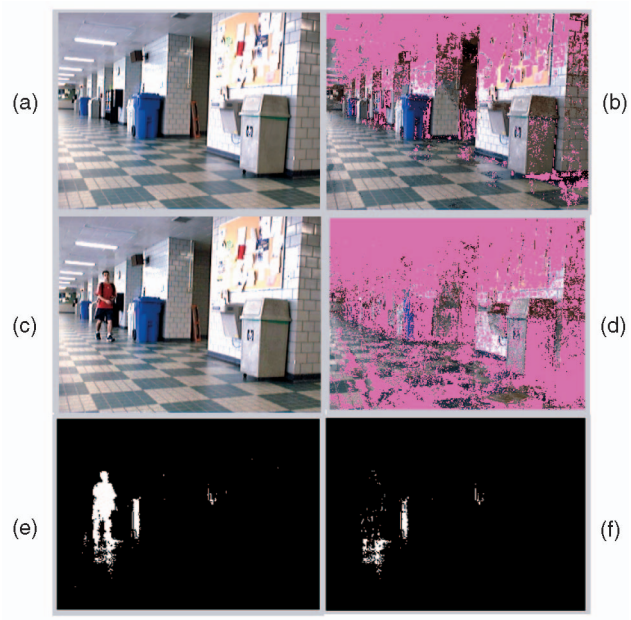

Fig. 12. Hallway 2. (a) Mean value of the first background state of GMM. (b) Mean value of the first GMSM state. (c) Frame in the sequence. (d) Mean value of the second GMSM state. (e) Foreground detection from GMM. (f) Shadow detection from the GMSM.

pixels since no shadow state is defined for these pixels. This property of our approach contributes to reducing false detection of shadow pixels.

6 Brightness and Chromaticity Distortion Model

Horprasert et al. proposed in [6] a pixel-based segmentation model in RGB color space which decomposes each background value into its brightness (α) and chromaticity distortion (CD). This approach performed relatively well in the shadow detection comparative study [1]. Results in this section show that better shadow detection can be obtained from this model if integrated into our GMM/GMSM.

6.1 The Shadow Model [6]

For a given pixel, the expected background value $E_t = [\mu_r, \mu_g, \mu_b]$ is computed from N training frames representing the static background. For each subsequent frame $X_t = [X_R(t), X_G(t), X_B(t)]$, brightness (α_t) and chromaticity distortions (CD_t) from the background value are given by:

$$\alpha_t = \frac{\left(\frac{X_R(t)\mu_R}{\sigma_R^2} + \frac{X_G(t)\mu_G}{\sigma_G^2} + \frac{X_B(t)\mu_B}{\sigma_B^2}\right)}{\left[\frac{\mu_R}{\sigma_R}\right]^2 + \left[\frac{\mu_G}{\sigma_G}\right]^2 + \left[\frac{\mu_B}{\sigma_B}\right]^2}, \tag{28}$$

$$CD_t = \sqrt{\left[\frac{X_R(t) - \alpha_t\mu_R}{\sigma_R}\right]^2 + \left[\frac{X_G(t) - \alpha_t\mu_G}{\sigma_G}\right]^2 + \left[\frac{X_B(t) - \alpha_t\mu_B}{\sigma_B}\right]^2}. \tag{29}$$

In the RGB space, the chromaticity distortion is the length of the perpendicular vector between a pixel value X_t and the line joining the zero intensity point and the background value μ. It is an indicator of how much the pixel color differs from the background color. During the training phase, the variation b of the chromaticity distortion is evaluated,

$$b = \sqrt{\frac{\sum_{t=0}^{N}(CD_t)^2}{N}}, \tag{30}$$

and used to compute a normalized chromaticity distortion, $\hat{CD}$:

$$\hat{CD}_t = \frac{CD_t}{b}. \tag{31}$$

This procedure is similar to (25) or (26), where the variation between a pixel value and the background value is normalized by the pixel noise. Hence, only one variation threshold, T_{CD}, can be applied for the entire statistical model. In the original approach, the threshold is set by plotting the histogram of the normalized chromaticity distortion during the training sequence, excluding foreground perturbations. For a given scene, the threshold T_{CD} is chosen according to a successful detection rate determined by the user.

Pixels are then labeled background, foreground, cast shadow, or highlight. Background pixels have small normalized brightness distortion, and small normalized chromaticity distortion. A pixel is labeled as cast shadow or highlight if it has a small normalized chromaticity distortion and a lower (cast shadow) or higher (highlight) brightness value than the background value. Unclassified pixels are labeled foreground. More specifically, a pixel is labeled as cast shadow if these two conditions are respected:

$$\hat{CD}_t < T_{CD}, \tag{32}$$

$$\alpha_{min} < \alpha_t < 1. \tag{33}$$

6.2 GMSM Implementation

We used this brightness and chromaticity distortion (BCD) model as the property-based description of shadowed surfaces in our GMM/GMSM algorithm. Equations (32) and (33) (and (28), (29), (30), and (31) for the parameters) define the function $f_s(X, \theta_{1,t})$ used to test if a color vector X (pixel or mean value of a distribution) can describe a shadowed surface.

In [6], the background average value, E_t, and the chromaticity distortion variations, b, are learned through an initialization sequence of N frames. In our scenario, since the background evolves over time, we set $E_t = \mu_{1,t}$ and we update b every N frames. However, for a given pixel, b will be computed using only the frames where the pixel value X_t is associated to a background state since we do not want foreground objects contributing to the computation of the background chromaticity distortion.

6.3 Hallway 1

Fig. 13 shows results for the same hallway scene of Fig. 10 where light shadows are cast by people circulating. Fig. 13e shows the pixels labeled as shadow by the BCD model. We can see that a large section of the person's shirt was labeled

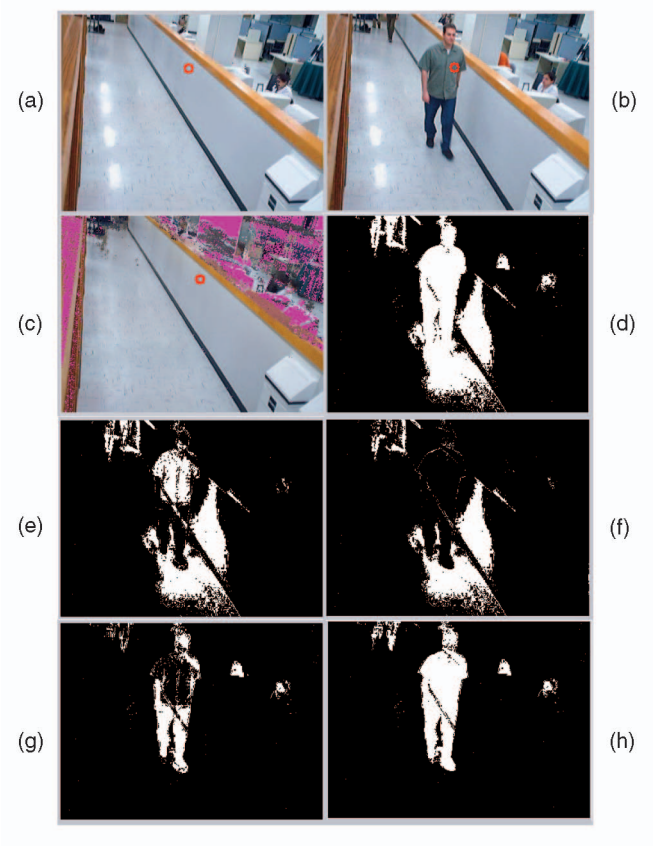

Fig. 13. Hallway 1. (a) Mean value of the first background state of GMM. (b) Frame in the sequence. (c) Mean value of the first GMSM state. (d) Foreground detection from GMM. (e) Shadow detection from the brightness and chromaticity distortion model. (f) Shadow detection from the GMSM. (g) Foreground detection: Img (d) - Img (e). (h) Foreground detection: Img (d) - Img (f).

as cast shadow, while, in Fig. 13f, the GMSM does not label it as shadow.

Using the parameters extracted from the GMM and GMSM, we computed shadow volumes for the pixel identified by the red circle in Fig. 13a. In Fig. 14, we show the gray cylinder-shaped volume defined by the BCD model. The cylinder radius is a function of the chromaticity distortion threshold T_{CD}. This threshold depends on the acquisition noise and the scene illumination complexity as it must be adjusted to obtain a desired detection rate (smaller thresholds would increase missed detections). The cylinder's length depends on the maximum brightness reduction threshold, α_{min}, which is set to prevent dark foreground objects being labeled as shadows when they occlude light background surfaces. Setting α_{min} improves the convergence rate of shadow distributions within the GMM. The volume defined by the GMSM, shown in blue, is bigger than the background volume (in red) since shadows in this scene showed more intensity fluctuations for all three color channels than the still background, hence increasing the variance of the associated shadow distributions. The GMSM shadow volume is, however, still more constrained than the volume of the BCD model and more appropriately characterizes true shadow values. The red circle in Fig. 14 represents the RGB value of the foreground pixel identified by a red circle in Fig. 13b. This foreground pixel value falls within the

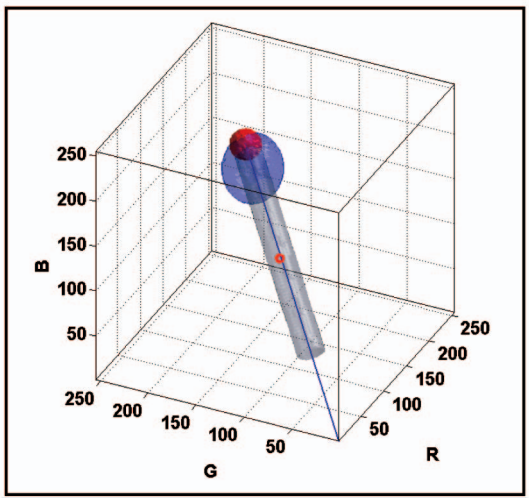

Fig. 14. Hallway: Background (red), shadow volumes using the BCD model (gray) and the GMSM (blue). The red circle represents the RGB value of the pixel identified by the red circle in Fig. 13b and belonging to the foreground person.

BCD shadow volume, but not into the GMSM shadow volume.

6.4 Highway

This set of results was obtained outdoors. The scene, Fig. 15, shows a highway where there is typically a steady stream of vehicles. Results are shown for three different frames through the video sequence. Unlike indoor sequences, cast shadows here are fairly deep, as can be seen in Fig. 15b. With these strong shadows, the intensity threshold α_{min} of (33) has to be set to zero. Therefore, the cylinder volume of the BCD model will extend all the way to the black vertex and a considerable number of foreground pixels will be labeled as shadows by the BCD model. More specifically, for this traffic scene, it mislabels any car that is a shade of gray darker than the road, as shown in Figs. 15e, 15i, and 15m. Even in this difficult environment, the GMSM performed better, although it also suffers from false detections caused by the chromaticity and luminance features of the car windshields (Figs. 15f, 15j, and 15n).

Fig. 16 shows the computed shadow volumes for the red circled pixel in Fig. 15a. As before, the computed GMSM volume (blue) is more constrained than the BCD shadow volume (gray), reducing the false shadow detection rate. Due to lighting and video acquisition conditions, the cast shadows induce a small color shift toward the red spectrum. This shift is visible in the first GMSM state pictured in Fig. 15b and also in the position of the GMSM shadow volume in Fig. 16, shown by the perpendicular offset red line in the figure. This demonstrates that the GMM/GMSM can capture shadow distributions whose values deviate from the proportional relationship with the background values. In order to correctly label these pixels as cast shadows, the thresholds of the property-based shadow models have to be very permissive, generating false detections (Figs. 15e, 15i, and 15m). The GMM/GMSM confines the volumes around the real shadow values, reducing the number of false detections (Figs. 15f, 15j, and 15n).

In this outdoor sequence, the light source could be considered a point source and the majority of the photons vectors are parallel. Hence, the process of blocking the light

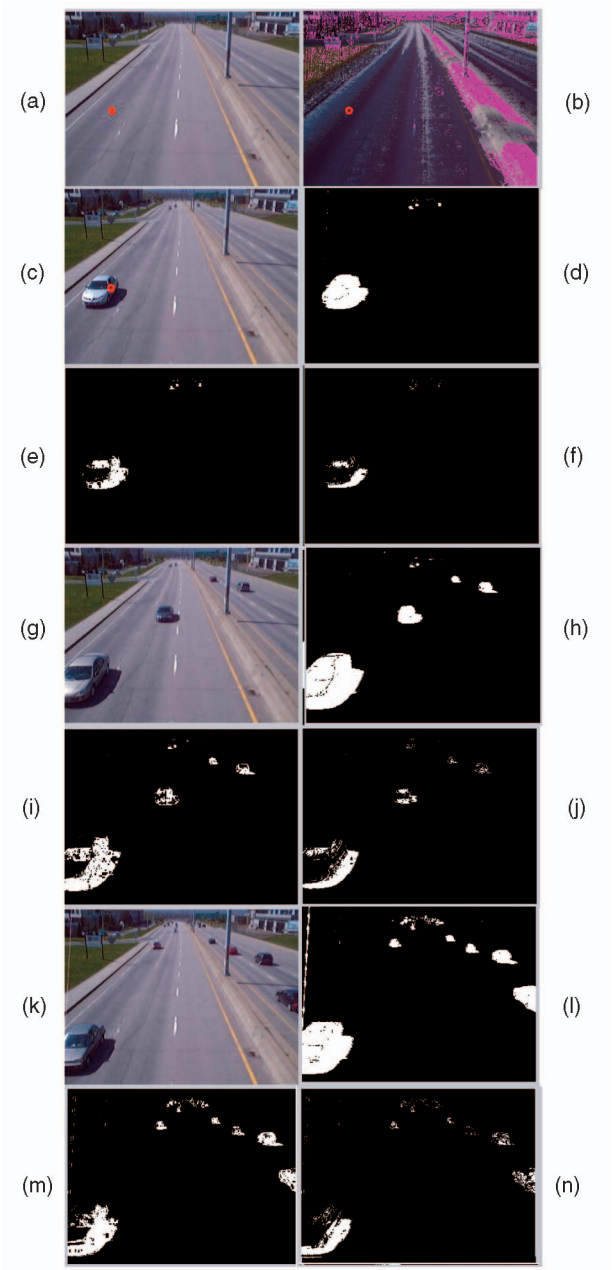

Fig. 15. Highway. (a) Mean value of the first background state of GMM. (b) Mean value of the first GMSM state. (c), (g), and (k) Frames from the sequence. (d), (h), and (l) Foreground detection from the GMM. (e), (i), and (m) Shadow detection from the brightness and chromaticity distortion model. (f), (j), and (n) Shadow detection from the GMSM.

and casting a shadow on a surface could be seen as a binary event, especially for rapidly moving foreground objects. In this situation, the penumbra region is quasi-inexistent and the GMSM converges to a single state. The fluctuations for this shadow state are mainly caused by the acquisition noise; thus, the volume is similar in size to that of the background volume.

7 HSV Shadow Model

7.1 The Shadow Model [4]

In [4], Cucchiara et al. proposed a pixel-based shadow model that exploits the brightness and color separation

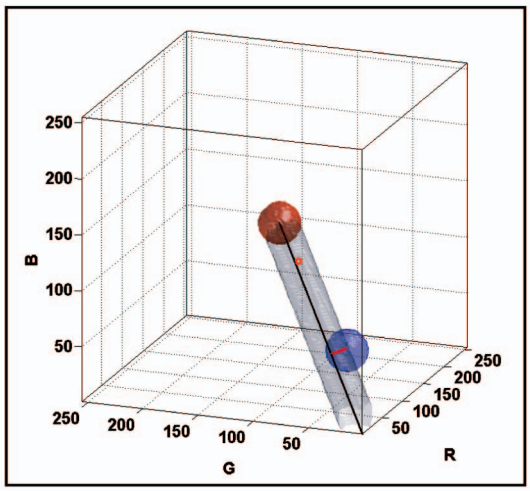

Fig. 16. Highway: Background (red), shadow volumes using the BCD model (gray), and the GMSM (blue). The red circle represents the RGB value of the pixel identified by the red circle in Fig. 15c and belonging to a foreground car.

inherent of the HSV color space. Based on the observation that shadows cast on a surface reduce the brightness value while maintaining chromaticity properties, cast shadows are defined by a diminution of the luminance (V) and saturation (S) values while keeping the hue value variation (ΔH) below a threshold parameter.

A pixel $X = [X_H, X_S, X_V]$ is labeled as a shadow over the background surface μ_1 if the three following conditions are met:

$$\alpha \leq \frac{X_V}{\mu_{1,V}} \leq 1, \tag{34}$$

$$\left(X_S - \mu_{1,S}\right) \leq \tau_S, \tag{35}$$

$$\left|\left(X_H - \mu_{1,H}\right)\right| \leq \tau_H, \tag{36}$$

where α, τ_S, and τ_H are thresholds on the intensity (V), saturation (S), and hue (H) values, respectively.

7.2 GMSM Implementation and Results

The shadow model proposed in [4] achieved the best all-around results in the shadow detection comparative study [1]. However, the photometric color invariant hue (H) is unstable near the black vertex of the RGB space in the presence of noise [3], making it difficult to know if background hue values are conserved. For this reason, instead of using the HSV color space, we achieve better results by using the RGB color space and performing a RGB to HSV color transformation for all shadow model related tests. The color transformation causes the implementation to be computationally more expensive than for the two previous shadow models. The function $f_s(X, \theta_{1,t})$ used in the GMM/GMSM is defined by (34), (35), and (36).

Fig. 17 shows results for the hallway and highway scenes. As with other property-based shadow models, the HSV falsely labels a number of pixels as shadows, while the GMM/GMSM reduces the number of false hits without increasing the number of misses.

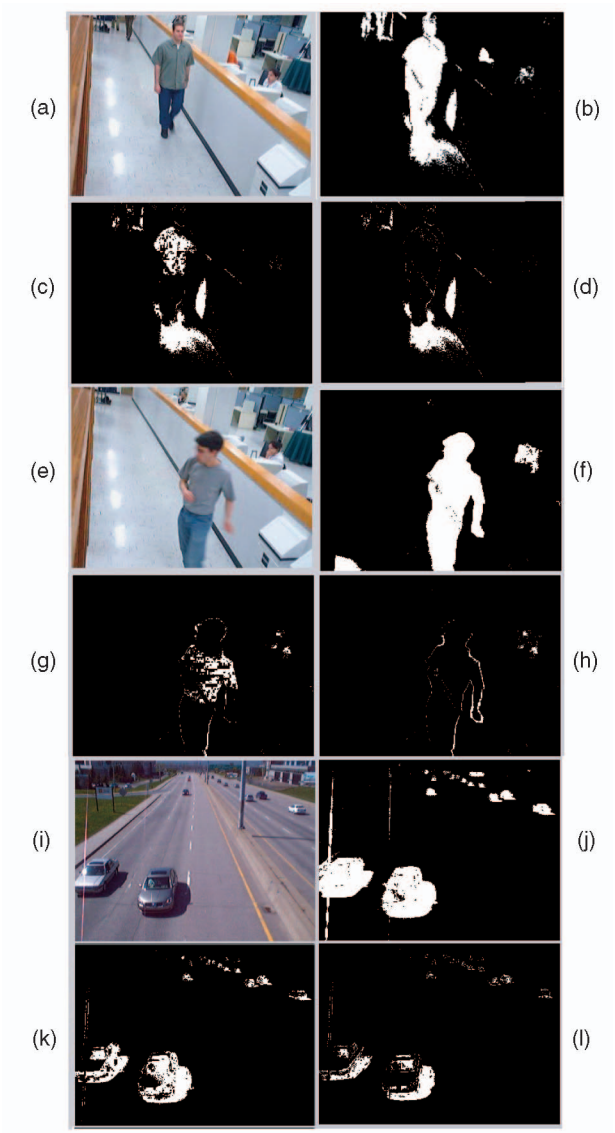

Fig. 17. Hallway. (a), (e), and (i) Frame in the sequence. (b), (f), and (j) Foreground detection from GMM. (c), (g), and (k) Shadow detection from the HSV shadow model. (d), (h), and (l) Shadow detection from the GMSM.

8 Shadow Detection in Light Saturated Areas

The dynamic range of video acquisition devices is often problematic and causes color saturation in scenes where outdoor and indoor lighting are combined. The flexibility of the GMSM allows a slight color saturation since the distance between the linear relation and the shadow values will not be excessive. However, in a case like Fig. 18a, a large portion of the background is strongly saturated due to exterior illumination through windows. In such conditions, it is impossible for property-based algorithms to identify shadows over the background surfaces since the cast shadows have color properties which do not preserve those of the saturated background.

In Fig. 18c, a person blocks the exterior light and a deep shadow is cast, revealing the floor's color. The foreground detection is shown in Fig. 18d and Fig. 18e shows that cast shadows over the light saturated areas have not been

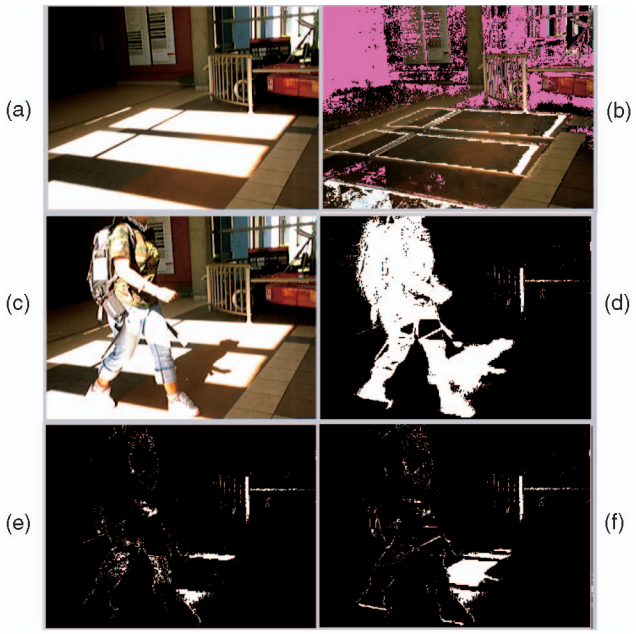

Fig. 18. Shadow detection in light saturated areas. (a) Mean value of first background state of GMM. (b) Mean value of first GMSM state. (c) Frame in the sequence. (d) Foreground detection. (e) Shadow detection from YUV shadow model. (f) Shadow detection from GMSM.

labeled as shadows by our YUV-based shadow model. This implies that the states associated to these pixels will not be identified as shadow states and the GMSM will fail. It is possible, however, to correctly model the shadowed background (as shown in Fig. 18b) with a slight modification to the algorithm presented in this paper. To do so, we only need to add a new set of conditions to (12) to allow states to be transferred to the GMSM. In this scenario, we can also transfer the more stable GMM foreground state to the GMSM if the background state is light saturated. It is the higher frequency of shadow observations, compared to other foreground distributions, that allows this state to converge within the GMM. By modifying (12) for saturated light, the GMSM could show more instability, requiring the use of a threshold $\omega^s_{k,t} > \omega_{min}$ to consider a GMSM state to be valid. Fig. 18f shows the pixels labeled as shadow using the proposed approach. These results could not be achieved by any property-based algorithm.

9 Summary and Conclusions

We have presented a novel pixel-based statistical approach to model and detect moving cast shadows. The proposed approach uses a GMM to learn from repetition the properties of shadowed background surfaces. The algorithm identifies distributions of pixel values that could represent shadowed surfaces, modifies their learning rates to allow them to become rapidly stable within the GMM, and then uses them to build a second mixture model for moving shadows on background surfaces, the GMSM. The GMM/GMSM approach can be used with different property-based descriptions of shadowed surfaces and we showed that it reduces their false detection rate without increasing the miss detection rate. In color space, this was illustrated by looking at the volume of pixels labeled as shadows: By learning, the GMM/GMSM reduces the shadow volume of

property-based descriptions. The framework presented in this paper could also be used with more complex multi-distribution approaches.

The approach presented is pixel-based. A contextual constraint like Markov Random Field could improve the quality of a background/shadow/foreground segmentation algorithm. However, a contextual constraint strongly benefits from a good local likelihood model. Our results, shown without any spatial postprocessing, demonstrates that our approach could be used as this likelihood model of shadowed values.

Our approach differs from previous work in several aspects. Shadow properties (mean and variance) are defined at each pixel and several distributions can be used at a pixel depending on the complexity of scene illumination and scene activity. Shadow properties evolve over time and regions where shadows cannot be detected or cast are identified, further reducing the false detection rate. We have also shown that the GMSM can capture shadows in light saturated areas, where property-based descriptions cannot detect them, as well as capture shadows whose values deviate from the proportional relationship with background values.

We have shown the robustness of the approach in different indoor and outdoor scenes with complex illumination. Using very few parameters, the GMSM is self-constructed, evolves over time, and clearly captures shadowed background surfaces.

The GMM/GMSM reduces false hits of property-based descriptions but does not eliminate them. A foreground pixel representing a person may still fall within the shadow volume and be mislabeled. However, by significantly reducing shadow volumes, the likelihood of these occurrences is much smaller and it will be easier for a tracking/foreground detection algorithm to handle these situations. Also, since the approach builds shadow models by learning, the first cast shadows will not be labeled as such. This can happen at start-up or during sudden changes in scene conditions. We can initialize the system by using the property-based model while the GMSM self-constructs and switch to the GMSM when it can be established that its states are stable, for example, based on computed scene activity [13].

Acknowledgments

The authors would like to thank PRECARN and NSERC for their financial support, and to thank the associate editor and reviewers for their comments, which helped improve this paper.

References

[1] A. Prati, I. Mikic, M.M. Trivedi, and R. Cucchiara, "Detecting Moving Shadows: Algorithms and Evaluation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 7, pp. 918-923, July 2003.

[2] S. Nadimi and B. Bhanu, "Physical Models for Moving Shadow and Object Detection in Video," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 8, Aug. 2004.

[3] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast Shadow Segmentation Using Invariant Color Features," *Computer Vision and Image Understanding,* pp. 238-259, 2004.

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

[5] O. Schreer, I. Feldmann, U. Goelz, and P. Kauff, "Fast and Robust Shadow Detection in Videoconference Applications," *Proc. Fourth IEEE Int'l Symp. Video Processing and Multimedia Comm.,* pp. 371-375, 2002.

[6] T. Horprasert, D. Hardwood, and L.S. Davis, "A Statistical Approach for Real-Time Robust Background Subtraction and Shadow Detection," *Proc. Int'l Conf. Computer Vision FRAME-RATE Workshop,* 1999.

[7] J. Stauder, R. Mech, and J. Ostermann, "Detection of Moving Cast Shadows for Object Segmentation," *IEEE Trans. Multimedia,* vol. 1, no. 1, pp. 65-76, Jan.-Mar. 1999.

[8] G.S.K. Fung, N.H.C. Yung, G.K.H. Pang, and A.H.S. Lai, "Effective Moving Cast Shadows Detection for Monocular Image Sequences," *Proc. 11th Int'l Conf. Image Analysis and Processing,* pp. 404-409, 2001.

[9] G.D. Finlayson, S.D. Hordley, C. Lu, and M.S. Drew, "On the Removal of Shadows from Images," *IEEE Trans. Pattern Analysis and Machince Intelligence,* vol. 28, no. 1, pp. 59-68, Jan. 2006.

[10] J.-M. Pinel and H. Nicolas, "Shadows Analysis and Synthesis in Natural Video Sequences," *Proc. IEEE Int'l Conf. Image Processing,* vol. 3, pp. 285-288, June 2002.

[11] C. Stauffer and W.E.L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 747-757, Aug. 2000.

[12] P.W. Power and J.A. Schoonees, "Understanding Background Mixture Models for Foreground Segmentation," *Proc. Image and Vision Computing,* pp. 267-271, 2002.

[13] T. Thongkamwitoon, S. Aramvith, and T.H. Chalidabhongse, "An Adaptive Real-Time Background Subtraction and Moving Shadows Detection," *Proc. IEEE Int'l Conf. Multimedia and Expo,* vol. 2, pp. 1459-1462, 2004.

Nicolas Martel-Brisson received the BS degree in engineering physics (2003) and the MS degree in electrical engineering (2005), both from Université Laval, Québec, Canada. Currently, he is working toward the PhD degree in the Computer Vision and Systems Lab at the same university. From 2000 to 2002, he worked part-time in medical imagery in the Oncology Department at the Centre de Recherche de l'Hôtel-Dieu de Québec. His current research interests include nonparametric statistical learning, video analysis for low-level segmentation, and medical imagery. He is a student member of the IEEE.

André Zaccarin received the BScA and MSc degrees in electrical engineering from Université Laval, Québec, Canada, and the MA and PhD degrees in electrical engineering from Princeton University, Princeton, New Jersey. He is currently a professor in the Department of Electrical and Computer Engineering at Université Laval and is a member of the Computer Vision and Systems Lab. In 2000 and 2001, he was a senior researcher with Intel Corporation, Santa Clara, California, where he conducted research in video coding. He was with ABB Bomem Inc. (Quebec City) during the fall of 2004 and with the Università degli Studi di Parma (Italy) as a visiting professor in the spring of 2005. Jointly with Bede Liu in 1994, he received the best paper award from the *IEEE Transactions on Circuits and Systems for Video Technology* for their work on motion estimation for video compression. His current research interests are in the areas of video processing, including compression, computer vision, scene segmentation, and hyperspectral imaging. He is a member of the IEEE.