Moving Cast Shadow Detection from a Gaussian Mixture Shadow Model

Nicolas Martel-Brisson and André Zaccarin*
Computer Vision and Systems Laboratory
Department of Electrical and Computer Engineering
Laval University, Québec, QC, Canada G1K 7P4
(nmartel,zaccarin)@gel.ulaval.ca

Abstract

Moving cast shadows are a major concern for foreground detection algorithms. Processing of foreground images in surveillance applications typically requires that such shadows have been identified and removed from the detected foreground. This paper presents a novel pixel-based statistical approach to model moving cast shadows of non-uniform 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, and prevent false detection in regions where shadows cannot be detected. Gaussian mixture shadow models (GMSM) are automatically constructed and updated over time, are easily added to a GMM architecture for foreground detection, and require only a small number of parameters. Results obtained with different scene types show the robustness of the approach.

1 Introduction

Detection of moving foreground objects generally includes their cast shadow as a foreground object since the shadow intensity differs from the background and the shadow moves with the foreground object. This inclusion of shadows as foreground objects can cause various unwanted behavior such as object shape distortion and object merging, affecting surveillance capability like target counting and identification. To obtain a better segmentation quality, detection algorithms must correctly separate foreground objects from the shadows they cast.

Shadow detection techniques can be classified in two groups. Model-based techniques are usually used for specific situations such as traffic monitoring and aerial imaging, where *a priori* knowledge of scene geometry and foreground objects can be incorporated into a model. Property-based approaches, for which features like geometry and brightness or color are used to identify shadow regions, are more robust to different scene and illumination conditions.

In [1], cast shadows are first detected based on the fact that a shadow darkens the surfaces on which it is cast. The validity of these detected shadows is then verified by using color invariance and geometric properties of shadows in moving videos and still images. In [2], the proposed method uses assumptions that shadows reduce surface brightness and saturation while maintaining chromaticity properties in the HSV color space. The authors in [3] adopt the YUV color space to avoid using the time consuming HSV color transformation. They were able to segment shadows and foreground objects following their observation that shadows reduced the pixel value linearly to its prior value in the YUV space. [4] builds a model from the RGB color space to express normalized luminance variation and chromaticity distorsions. In addition to using the scene brightness properties, [5] uses edge width information to differentiate penumbra regions from the background. A comparative study of many cast shadow segmentation algorithms can also be found [6].

In this paper, we present an approach for detecting and modeling moving cast shadows from a Gaussian mixture model (GMM) in a background subtraction algorithm. With assumptions based on the properties of shadowed surfaces in the YUV color space, we successfully model the behavior of cast shadows within the GMM for different scenes with complex illumination. These properties are then used to build a Gaussian mixture shadow model (GMSM) which represents the background surfaces when different types of shadows are cast on them, depending on the illumination complexity.

This approach differs from what can be found in the literature in three distinct aspects. Firstly, shadowed surfaces are defined spatially. Each pixel models its own shadow values (mean and variance) based on the background surface properties. Penumbra and deeper shadows can be modeled

[1] Authors thank PRECARN and NSERC for their financial support.

at the same time for the same pixel. Secondly, the properties of shadowed surface are learned through a modified GMM. There is no local parametrization nor ground truth examples selected manually to determine whether a region is a shadowed surface or not. There is, however, a general condition in the YUV space helping the hypothetically shadowed surfaces to be learned by the model. Thirdly, our approach identifies regions where moving cast shadows cannot be detected. These regions are generally regions where a shadow is already cast from a background object, or very dark surfaces where cast shadows are not visible. By identifying these regions, it is possible to reduce false shadow detection when a foreground object with similar brightness properties crosses the scene.

In section 2, we summarize the Gaussian mixture models. Our moving cast shadow detection approach is explained in detail in section 3. Results showing the robustness of the approach are shown in section 4, while concluding statements are found in section 5.

2 Gaussian Mixture Models

Our approach for modeling and segmenting cast shadows uses Gaussian mixture models. It was developed to be integrated into a background detection algorithm also based on a Gaussian mixture model. In this section, we summarize the main elements of this approach as described in [7] and modified for online implementation in [8].

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 YUV color space of a random variable **X**. A Gaussian density probability function $f_{\mathbf{X}|k}$ and *a priori* probability ω_k are associated to each state $k \in \{1, 2..K\}$. The Gaussian probability density function (pdf) with parameters $\theta_k = \{\mu_k, \sigma_k\}$ describes the YUV components of a surface that comes into the pixel's view,

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

where μ_k is the mean and Σ_k is the covariance matrix. In a YUV color space, we can assume that the components of X 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 $\sum_{k=1}^{K} \omega_k = 1$. The GMM is based on the assumption that each pixel views background states more often than foreground ones. Based on this premise, the K states are ordered decreasingly by their $\omega_k/\|\sigma_k\|$ ratio and the first B states whose combined probability of appearing is greater than the threshold T, i.e.,

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

are labeled as background states, and the other states are labeled as foreground states. 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 state k if

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

where

$$d_{k,t} = (\sigma_{k,t} \, I)^{-1}(X_t - \mu_{k,t}) \tag{3}$$

If we cannot associate a pixel value to an existing pdf, a new state k is created around this value with probability of w_{init}, and the less probable state is dropped. The GMM is updated over time using these equations[8]:

$$\omega_{k,t+1} = (1 - \alpha_t)\omega_{k,t} \ + \ \alpha_t \, P(k|X_t, \Phi) \tag{4}$$

$$\mu_{k,t+1} = (1 - \rho_{k,t})\mu_{k,t} \ + \ \rho_{k,t}X_t \tag{5}$$

$$\sigma_{k,t+1}^2 = (1-\rho_{k,t})\sigma_{k,t}^2 + \rho_{k,t}((X_t-\mu_{k,t+1})\circ(X_t-\mu_{k,t+1})) \tag{6}$$

where

$$\rho_{k,t} = \frac{\alpha_t \, P(k|X_t, \Phi)}{\omega_{k,t+1}} \tag{7}$$

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

3 Shadow modeling

Our approach takes advantage of the learning properties of the GMM in order to model surface values when shadows are cast by foreground objects. It is based on the hypothesis that, for a given pixel, the shadow cast by different moving foreground objects is relatively similar. The shadow values will therefore converge to one of the K states of a GMM describing the scene. The states that capture cast shadows are identified, and their pdfs are used to build a second GMM, which we call the Gaussian mixture shadow model (GMSM). In this GMSM, the pdfs $f_{\mathbf{X}|k}$ of the states describe background surface values when shadows are cast by moving objects or persons.

3.1 GMM Convergence of shadowed surfaces

When using a GMM to build a background model for foreground detection, it is possible to observe that some of the states model the shadows cast by persons moving across the scene. This can be explained as follows. A first person crosses the scene, and on pixels where its shadow is cast, a new state representing the value of this shadow will be created with a small *a priori* probability ω_{init}. When the next person crosses the scene and casts a shadow on the same pixels, the algorithm will associate the pixel values to the same new state describing the shadow value. The *a priori* probability of this state will then increase as

$$\omega_{k,t+1} \;=\; \omega_{k,t} \;+\; \alpha_{s,t}\, M_{k,t} \tag{8}$$

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. With this formulation, which differs from equation (4), the *a priori* probability of an unmatched state only decreases by normalization.

Also, unlike the update equation originally proposed in [7], 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 surface background, we increase the learning rate of the pixel associated state parameter ω_k. 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), and this without preventing the creation of states describing the background. To do so, we define

$$\alpha_{s,t} = \alpha_t\, S \tag{9}$$

where the parameter $S > 1$ when X_t correspond to a shadowed background surface, and $S = 1$ in the other cases. We define a shadowed surface in the following section.

In our implementation, we use a GMM for foreground detection with four states ($K = 4$) and a relative learning rate $S = 3$ for surfaces with cast shadows. If the scene shows significant activity, the weight of the parameter S must be reduced in order to not interfere with states describing the background.

When there are no persons 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 will 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 shadow 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 states describing cast shadows on background surfaces.

3.2 Shadowed surfaces properties

We have described in section 3.1 how we modify the GMM to increase the convergence of states that represent shadows. To do this, we need a model to identify pixels in shadowed surfaces. Our shadow model is described in this section. It is important to stress, however, that the GMSM can be used with other shadow models, such as the ones in [4] or [2].

As in [3], we adopted the YUV color space and our model is also based on the observation that a shadow cast on a surface will equally attenuate the value of its three components. We first estimate this attenuation ratio using the luminance component Y, and we then verify that both U and V components are also reduced by a similar ratio. More specifically, if color vector X represents the shadow cast on a surface whose average color value is μ with variance σ_μ, we have

$$\alpha_{min} \;<\; \alpha_Y \;<\; 1 \quad with\, \alpha_Y \;=\; X_Y/\mu_Y \tag{10}$$

$$(1/\sigma_{\mu_U})|X_U \;-\; \alpha_Y\;\mu_U| \;<\; \Lambda_U \tag{11}$$

$$(1/\sigma_{\mu_V})|X_V \;-\; \alpha_Y\;\mu_V| \;<\; \Lambda_V \tag{12}$$

where α_{min} is a threshold on maximum luminance reduction. This threshold is important when the U and V components of a color are small, in which case any dark pixel value would be labeled as a shadow for a light color surface. $\Lambda_{U,V}$ represent the tolerable chromaticity fluctuation around the surface value μ_U, V

This shadow model is similar but computationally less expensive than the model in [4]. However, such models cannot always adequately label pixels on surfaces with cast shadows. By not using all of the information given by the luminance value (the first condition only tests if the pixel luminance value is lower than that of the surface), our models will wrongly label, as shadows, foreground objects whose luminance values are lower than that of the background, and whose chromaticity values are similar to that of the background.

It is impossible to know *a priori* the luminance value of a cast shadow since it changes with surfaces, and can vary with time for a same surface. Our approach uses the shadow model as a starting point to learn in the GMM the luminance and chromaticity values of surfaces with cast shadows. If we match a pixel value X_t to a non-background state of the GMM, following equation (3), we then verify if that pixel value could be the shadowed surface of the most frequent background state $k = 1$:

$$\alpha_{min} \;<\; \alpha_Y \;<\; 1 \;\; with \;\; \alpha_Y \;=\; X_{t,Y}/\mu_{1,t,Y} \tag{13}$$

$$(1/\sigma_{1,\mu,t,U})\,|X_{t,U} \;-\; \alpha_Y\mu_{1,t,U}| \;<\; \Lambda_U \tag{14}$$

$$(1/\sigma_{1,\mu,t,V})\,|X_{t,V} \;-\; \alpha_Y\mu_{1,t,V}| \;<\; \Lambda_V \tag{15}$$

If these three conditions are met, we increase the learning rate $\alpha_{s,t}$ of the *a priori* probability ω_k of that state so that it rapidly becomes a stable foreground state within the GMM, as discussed in section 3.1.

3.3 Gaussian Mixture Shadow Models

Updating the Gaussian mixture shadow model is the next step. Unlike the GMM [7], the GMSM filters the input values, and these 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 **X**. These Gaussian pdf are those of the states which have been identified as describing cast shadows on background surfaces.

At each time interval (Δt), we process the foreground state with the largest *a priori* probability, $w_{k,t}$, i.e, state $k = B + 1$. We use the following conditions on $f_{\mathbf{X}|B+1}$ to determine if this state describes a cast shadow on the background surface of state $k = 1$:

$$\alpha_{min} \; < \; \alpha_Y \; < \; 1 \;\; with \;\; \alpha_Y \; = \; \mu_{B+1,t,Y}/\mu_{1,t,Y} \tag{16}$$

$$(1/\sigma_{1,\mu,t,U}) \, |\mu_{B+1,t,U} \; - \; \alpha_Y \mu_{1,t,U}| \; < \; \Lambda_U \tag{17}$$

$$(1/\sigma_{1,\mu,t,V}) \, |\mu_{B+1,t,V} \; - \; \alpha_Y \mu_{1,t,V}| \; < \; \Lambda_V \tag{18}$$

If these conditions are met, pdf $f_{\mathbf{X}|B+1}$ is transfered to the GMSM. It is then compared to the existing GMSM pdfs:

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

where

$$d_{k,t} = (\sigma_{k,t}^s \; I)^{-1}(\mu_{B+1,t} - \mu_{k,t}^s) \tag{19}$$

and where we use the superscript ^{s} when referring to the Gaussian mixture shadow model. If there is a match, the parameters θ_k^s are updated:

$$\mu_{k,t+1}^s = (1 - \alpha) \; \mu_{k,t}^s + \alpha \; \mu_{B+1,t} \tag{20}$$

$$\sigma_{k,t+1}^s = (1 - \alpha) \; \sigma_{k,t}^s + \alpha \; \sigma_{B+1,t} \tag{21}$$

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

$$\omega_{k,t+1}^s \; = \; \omega_{k,t}^s \; + \; \alpha \; M_{k,t}^s \tag{22}$$

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,

$$\omega_{k,t+1}^s = \omega_{init}^s \tag{23}$$

$$\mu_{k,t+1}^s = \mu_{B+1,t} \tag{24}$$

$$\sigma_{k,t+1}^s = \sigma_{B+1,t} \tag{25}$$

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

$$B^s = \underset{b}{argmin} \left(\sum_{k=1}^{b} \omega_{k,t}^s \; > T^s \right) \tag{26}$$

are used to model moving cast shadows on a background surface. T^s is chosen relatively large to include most of the states defined in the GSMS, but it is not equal to one so that states with small relative probability $\omega_{k,t}^s$ are not used to model shadows. 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 may have two to three valid states for describing cast shadows for each pixel in the GMSM. 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.

Figure 1 shows the relation between the GMM and the GMSM. A pdf representing a cast shadow on the background surface converges to more stable foreground states in the GMM. At each time interval Δt, the pdf from the most stable foreground state is transfered to the appropriate GMSM state if it is identified as representing a cast shadowed surface.

3.4 Removing Shadows

In a GMM, a pixel is labeled as foreground if its value X_t is matched to a foreground state. This detected pixel is then compared to the shadow states of the GMSM, states described by their pdf:

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

where

$$d_{k,t} = (\sigma_{k,t}^s \; I)^{-1}(X_t - \mu_{k,t}^s) \tag{27}$$

If this condition is met for a state ω_k^s with $k \leq B^s$, the pixel is labeled as representing a moving cast shadow. It is then simply removed from the detected pixels of the GMM. We are then left with detected pixels representing foreground objects without their cast shadows. It is important to stress that even if we use a small value for the parameter $\lambda_{s,b}$, for example $\lambda_{s,b} = 1.5$, we are able to adequately identify cast shadows while limiting false detections. In order to be considered as a shadow, a pixel value X_t must fall into the narrow distribution of a valid GMSM pdf, which reduces the false detections.

4 Results

Results shown in this section have been obtained with a low cost web cam, a Logitech QuickCam Pro 4000 with a resolution of 320x240 pixels, connected via a USB port to a laptop with an Intel Celeron processor 2.2GHz. The frame rate of our software implementation was approximately at

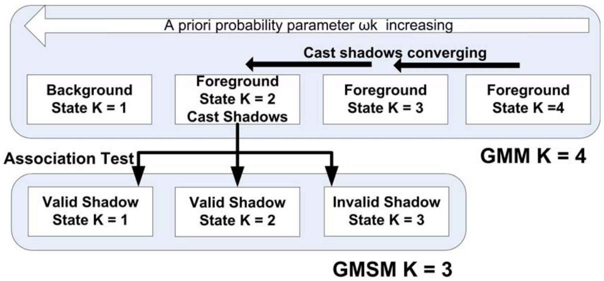

Figure 1. Relation between the GMM and the GMSM

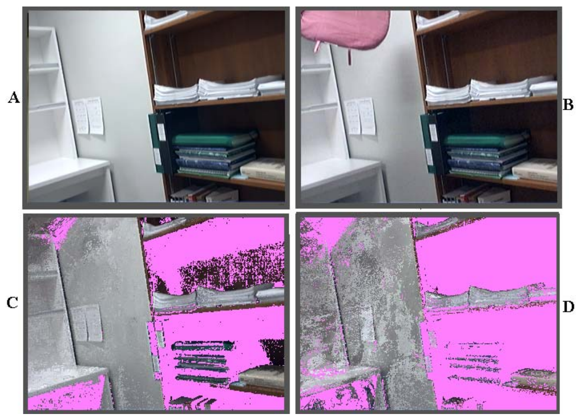

Figure 2. Complex Background Illumination

10Hz. Results shown here are raw results, without any post-treatment. For each environment, parameters were set once. Results we have selected represent a snapshot of the algorithm results and are typical of the performance throughout the sequences.

Figure 2A shows a scene 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 being fairly deep as a function of the moving object position. A snapshot of these shadows can be seen in Figure 2B. Figures 2C and D show the mean intensity of the first two states of the GMSM, i.e., $\mu^s_{1,t}, \mu^s_{2,t}$. Pink pixels are pixels for which the GMSM was never initialized. Shadows were either not cast on these pixels, or if a shadow was cast, it was not detectable, either because it was associated to a background state, or because its probability of occurence never increased sufficiently. In Figure 2C, note that background areas with static shadows, like under the desk or under one of the bookshelves, are not included in the GMSM as expected. The GMSM is defined only for surfaces on which shadows are cast by moving objects.

Figure 3 shows the segmentation results obtained for the same images. Figure 3A shows in white all the pixels that were detected as being foreground pixels, based on the background state values shown in Figure 2A. Figure 3B shows the pixels identified as belonging to the first state of the GMSM, while Figure 3C shows the pixels identified as belonging to any of the B^s GMSM states (3 states in this case). This clearly shows the value of using many states to describe the shadows. With one shadow state only, there would have either been some misses, as shown by Figure 3B, or the parameter $\lambda_{s,b}$ would have had to be set to a larger value, which would have led to false detections of shadowed surfaces.

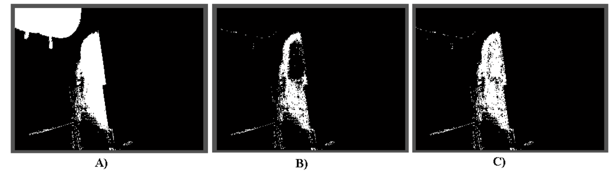

Figure 3. Raw and Shadow Detection

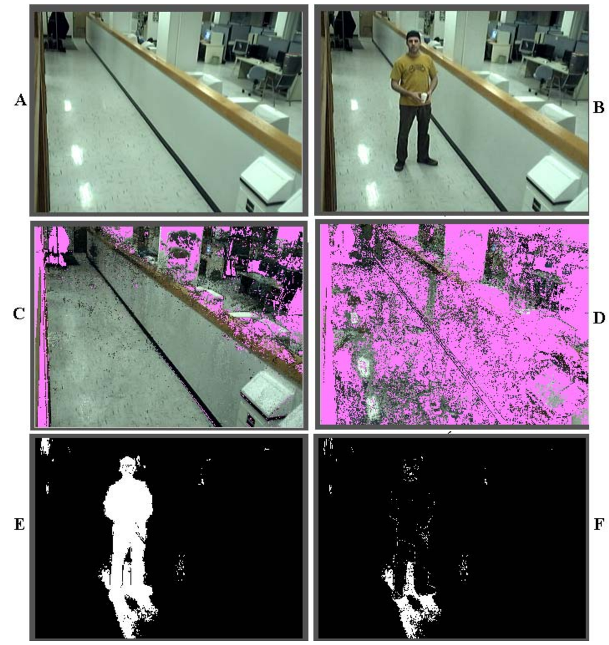

Figure 4. Hallway 1 (A,B), GMSM states (C,D), Raw and shadow detection (E,F)

Figure 4A shows a hallway where light shadows are cast by circulating persons, as in Figure 4B. Figure 4C and Figure 4D show respectively the mean value of the first and second GMSM states, $\mu^s_{1,t}$ and $\mu^s_{1,t}$, for each pixel. We can see in Figure 4A that a section of the floor is presenting complex reflected highlights. As a result, the GMSM has used two states to adequately model the shadows that were cast on this part of the floor, where on other parts of the floor, one state was sufficient. This shows the adaptability of the proposed model. Since shadows cast on the

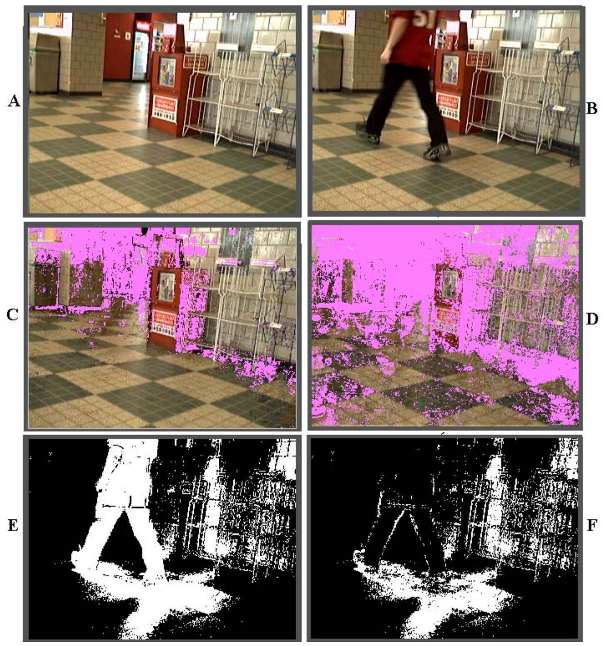

Figure 5. Hallway 2 (A,B), GMSM states (C,D), Raw and shadow detection (E,F)

dark clothes in the upper left portion of Figure 4A were undetectable, the GMSM was not initialized (Figure 4C), preventing false identification of future dark foreground objects as cast shadow for this region. Raw foreground detection by the GMM is shown in Figure 4E while pixels identified as showing a moving cast shadow are shown in Figure 4F. In this example, pixels of some image regions (shoe tips, pants by the coffee cup, face), have chroma values similar to the background but lower luminance values. In these cases, a simple test using equations (10,11) and (12) would have wrongly labeled, as shadows, many of those pixels. Because we model both the luminance and chroma values of the shadow in the GMSM, we did not have these false detections of shadow pixels.

The next example, Figure 5A, shows a relatively busy hallway. Figure 5B shows a person with a red shirt occluding a wall with a similar color. Here also, Figure 5C and Figure 5D show respectively the average $\mu^s_{k,t}$ of the first and second states of the GMSM for each pixel. It is worth noting that the GMSM was not initialized on the side of the newspaper box in the center of the image. Although shadows were cast on that surface by circulating persons, they were not detectable, and the algorithm behaved as expected. The GMSM was not initialized either in the center of the image since shadows were not cast on the red wall in the back. Figure 5E shows the raw foreground detection, and Figure 5F, the pixels identified as cast shadows. The pixels representing the person's red shirt were correctly labeled as foreground.

5 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 the shadowed background surfaces. The algorithm identifies distributions that could represent shadowed surfaces, modify their learning rates to allow them to converge within the GMM, and then uses them to build a GMM for moving shadows on background surfaces. This approach can be used with various shadow models in different color spaces.

Our approach differs from previous work in many aspects: shadow properties (mean and variance) are defined locally for more than one shadow type depending on scene illumination complexity, shadow properties are learned through a GMM used for background subtraction, and regions where shadows cannot be cast are identified, minimizing false shadow detections from a foreground object. We have shown the robustness of the approach in different indoor scenes with complex illumination. With very few parameters, the GMSM builds itself and evolves over time, and results clearly showed that it captures shadowed background surfaces.

References

[1] E.Salvador et al., "Cast shadow segmentation using invariant color features", *Computer Vision and Image Understanding*, pp. 238-259, 2004.

[2] R. Cucchiara et al., "Improving Shadow Suppression in Moving Object Detection with HSV Color Information" *Proc of Intel. Transportation Systems Conf.*, pp. 334-339, 2001.

[3] O. Schreer et al., "Fast and robust shadow detection in videoconference applications", *4th IEEE Intern. Symposium on Video Proces. and Multimedia Comm.*, pp. 371-375, 2002.

[4] T. Horprasert et al., "A Statistical Approach for Real-Time Robust Background Subtraction and Shadow Detection", *ICCV FRAME-RATE Workshop*, 1999.

[5] J. Stauder et al., "Detection of Moving Cast Shadows for Object Segmentation", *IEEE Transactions on Multimedia*, vol. 1, no. 1, pp. 65-76, Mar. 1999.

[6] A. Prati et al., "Detecting moving shadows: algorithms and evaluation", *IEEE Trans. Pattern Anal. Mach. Intell.* vol. 25 no.7, pp. 918-923, 2003

[7] C. Stauffer and W. Grimson, "Adaptive background mixture models for real time tracking", *Proc. of IEEE Int. Conf. on CVPR*, Vol.2,pp.246-252, June 1999.

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