BACKGROUND SUBTRACTION USING A PIXEL-WISE ADAPTIVE LEARNING RATE FOR OBJECT TRACKING INITIALIZATION

Ka Ki Ng and Edward J. Delp
Video and Image Processing Laboratories (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA

ABSTRACT

In this paper we present a new method for object tracking initialization using background subtraction. We propose an effective scheme for updating a background model adaptively in dynamic scenes. Unlike the traditional methods that use the same "learning rate" for the entire frame or sequence, our method assigns a learning rate for each pixel according to two parameters. The first parameter depends on the difference between the pixel intensities of the background model and the current frame. The second parameter depends on the duration of the pixel being classified as a background pixel. We also introduce a method to detect sudden illumination changes and segment moving objects during these changes. Experimental results show significant improvements in moving object detection in dynamic scenes such as waving tree leaves and sudden illumination change, and it has a much lower computational cost compared to Gaussian mixture model.

Keywords: background subtraction, adaptive thresholding, pixel-wise adaptive learning rate, moving object detection, object tracking, visual surveillance, sudden illumination changes, shading model

1. INTRODUCTION

Visual surveillance has been a very active research topic in the last few years due to the growing importance for security in public places [1]. A typical automated visual surveillance system consists of moving object detection, object classification, object tracking, activity understanding, and semantic description.

Moving object detection is not only useful for object tracking initialization for a visual surveillance system, it is also the first step of many image analysis applications, for example video indexing and human machine interaction [2]. Since subsequent processes are greatly dependent on the performance of this step, it is important that the classified foreground pixels accurately correspond to the moving objects of interests.

Background subtraction segments foreground objects more accurately in most cases compared to other common moving object detection methods, and detects foreground objects even if they are motionless. However, one drawback of traditional background subtraction methods is that they are susceptible to environmental changes, for example, gradual or sudden illumination changes. The reason for this drawback is that most methods assume a static background, and hence one needs to update the background model for dynamic backgrounds. The update of the background model is one of the major challenges for background subtraction methods.

When the change of background dynamics becomes smaller (implying the background model is "converging"), the rate for updating the background model using the current video frame should decrease. This rate of background model updating is sometimes referred to as the learning rate. This is useful in order not to miss any foreground pixels. When the change of background dynamics increase (changes not induced by foreground objects), the learning rate should increase for the background model to converge more rapidly in order to reduce the number of false alarms. However, various parts of a frame may have different dynamics, hence different learning rates are required. The typical background subtraction method [3] uses a constant learning rate for the

This material is based upon work supported by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0001. Address all correspondence to E. J. Delp (ace@ecn.purdue.edu).

entire sequence. In [4], the authors use two learning rates for the background update, one in the initial part of the sequence (before the L^{th} frame is available) and a different rate for the rest of the sequence (when the L^{th} frame is available) using a Gaussian mixture model (GMM) for the background. They estimate the GMM using the Expectation-Maximization (EM) algorithm by for the initial part and then switch to L-recent window version for the rest of the sequence. Using this method, the GMM learns faster and more accurately at the beginning, it does not improve the convergence rate of the background model if the background changes at normal phase.

We propose an effective scheme for updating the background and adaptively model dynamic scenes. Unlike the traditional methods that use the same learning rate for the entire frame or sequence, our method assigns a learning rate for each pixel using two parameters. The first parameter depends on the difference between the pixel intensities of the background model and the current frame. The second parameter depends on the duration of the pixel being classified as a background pixel.

Another contribution of this paper is to introduce a method to detect sudden illumination changes and segment moving objects during these changes. Sudden illumination change is still a very challenging problem for foreground segmentation. Many state-of-the-art methods [3, 5] can handle gradual illumination changes but remain susceptible to sudden changes. Some examples of sudden illumination changes are turning on/off light sources in a room, or open/close window curtains or doors. These situations alter the background model and make the color or intensity-based subtraction methods fail (false positive i.e. detecting background pixels as foreground pixels). In one of the popular approaches, Gaussian Mixture Models (GMM) proposed by Stauffer and Grimson [3] are robust to gradual illumination changes as well as moving background regions. However, it is not robust to sudden illumination changes, foreground objects could be merged into the background model if they remain static for a long period of time. The method also has a relatively higher computational cost. Xie et. al. [6] suggest that the ordering among pixels is preserved even in the presence of large photometric distortions. They proposed a method using the Phong's shading model [7] with slowly spatially varying illumination, the sign of the difference between two pixels is robust to sudden illumination changes. Pilet et al. [8] propose a method that replaces the statistical background model by a statistical illumination model. They model the ratio of intensities between a stored background image and an input image in all three channels as a GMM to account for illumination effects. We propose a method to detect sudden illumination changes based on the frame level maintenance mechanism of Wallflower [9] and segment moving objects using a shading model. The details will be described in the next section.

2. PROPOSED METHOD

2.1. Background Model Initialization

To detect moving objects using background subtraction, the first step is to construct the background model at the beginning of a video sequence. We assume that the sequence starts with the background in the absence of moving objects. We use the selective averaging method [10] to construct the initial background model as:

$$BM_N(x,y) = \frac{\sum_{m=1}^{N} I_m(x,y)}{N}, \tag{1}$$

where $BM_N(x,y)$ is the intensity of pixel (x,y) of the background model, $I_m(x,y)$ is the intensity of pixel (x,y) of the m^{th} frame, and N is the number of frames used to construct the background model. It has been shown that $N = 100$ is a good choice for N [10].

2.2. Background Subtraction

After obtaining the initial background model, we need to obtain the difference between the current frame and the background model,

$$D_t(x,y) = \left| I_t(x,y) - BM_{t-1}(x,y) \right|, \tag{2}$$

where $BM_t(x,y)$ is the intensity of pixel (x,y) of the background model at time t, and $I_t(x,y)$ is the intensity of pixel (x,y) in the current frame at time t.

The difference $D_t(x, y)$ is then compared to an adaptive threshold Th_{ad} for foreground background pixel classification, the computation for Th_{ad} is described in [11]. Th_{ad} is obtained iteratively using the histogram of the difference frame $D_t(x, y)$ to account for frame-to-frame changes in the background [11]. If $D_t(x, y) < Th_{ad}$, the pixel is classified as a background pixel. If $D_t(x, y) \geq Th_{ad}$, then it is a foreground pixel.

$$(x, y) \subset \begin{cases} foreground & \text{if } D_t(x, y) \geq Th_{ad} \\ background & \text{if } D_t(x, y) < Th_{ad} \end{cases} \tag{3}$$

2.3. Background Model Update

The background model must be updated for every frame in order to accommodate for background dynamics such as illumination changes and waving tree leaves. We proposed a method to update the background model in a pixel-by-pixel fashion, each pixel has an adaptive learning rate, $\alpha_{ad,t}(x, y)$ defined below. Let,

$$BM_t(x, y) = \alpha_{ad,t}(x, y)I_t(x, y) + (1 - \alpha_{ad,t}(x, y))BM_{t-1}(x, y), \tag{4}$$

where $0 \leq \alpha_{ad,t}(x, y) \leq 1$. The learning rate $\alpha_{ad,t}(x, y)$ depends on two weighted parameters, α_1 and α_2 (note that the subscript t and (x, y) are dropped for simplicity),

$$\alpha_{ad,t}(x, y) = w_1\alpha_1 + w_2\alpha_2, \tag{5}$$

where w_1 and w_2 are the weights for α_1 and α_2 respectively, and $w_1 + w_2 \leq 1$.

The first parameter α_1 depends on the magnitude of $D_t(x, y)$. A larger value for α_1 is assigned for a smaller $D_t(x, y)$.

$$\alpha_1 = \begin{cases} e^{-\frac{1}{2}\frac{D_t(x,y)^2}{\sigma_1^2}} & \text{if } D_t(x, y) < Th_{ad}, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where σ_1 is $\frac{Th_{ad}}{5}$. σ_1 is a function of Th_{ad} to ensure that the portion of pixels with the smallest $D_t(x, y)$ have higher α_1. Figure 1 shows the histogram of $D_t(x, y)$ (using the test sequence shown in Figure 2), the threshold based on the histogram, the σ_1 which is a function of the threshold, and the α_1 based on σ_1.

The second parameter α_2 is obtained depending on the temporal duration of a pixel in the background. We assume the longer a pixel is in the background, the more stable and reliable a background pixel is. The stability and reliability is measured by the temporal background count C_{bg}, according to:

$$\alpha_2 = \begin{cases} e^{-\frac{1}{2}\frac{(\zeta_{max} - C'_{bg})^2}{\sigma_2^2}} & \text{if } C_{bg} \geq \zeta_{min}, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where $\sigma_2 = 15$, $\zeta_{max} = 150$, and $\zeta_{min} = 30$ have been empirically determined, and $C'_{bg} = min(\zeta_{max}, C_{bg})$. If a pixel remains as a background pixel for more than ζ_{min} frames, a non-zero α_2 is assigned to that pixel. The parameter α_2 increases with C_{bg} until C_{bg} is greater than ζ_{max}. This parameter eliminates insignificant objects with small and repetitive motion such as waving tree leaves and water ripples which can result in false alarms. This works under the assumption that if the pixel switches between foreground and background states frequently (that is the pixel does not retain its background status for more than ζ_{min} frames), then it is more likely to be a background pixel. Note that the temporal background count C_{bg} is reset to zero if the pixel is detected as a foreground pixel.

3. SUDDEN ILLUMINATION CHANGE USING A SHADING MODEL

When there is a sudden illumination change, it induces photometric distortion. The increase of intensities is not a constant even when the illumination change is global. This results in a drastic increase in false positive detections. Many times the entire image is detected as foreground. A simple method based on the frame level maintenance mechanism of Wallflower [9] is used to detect sudden illumination changes. A shading model is imposed to our algorithm to segment moving objects under sudden illumination changes.

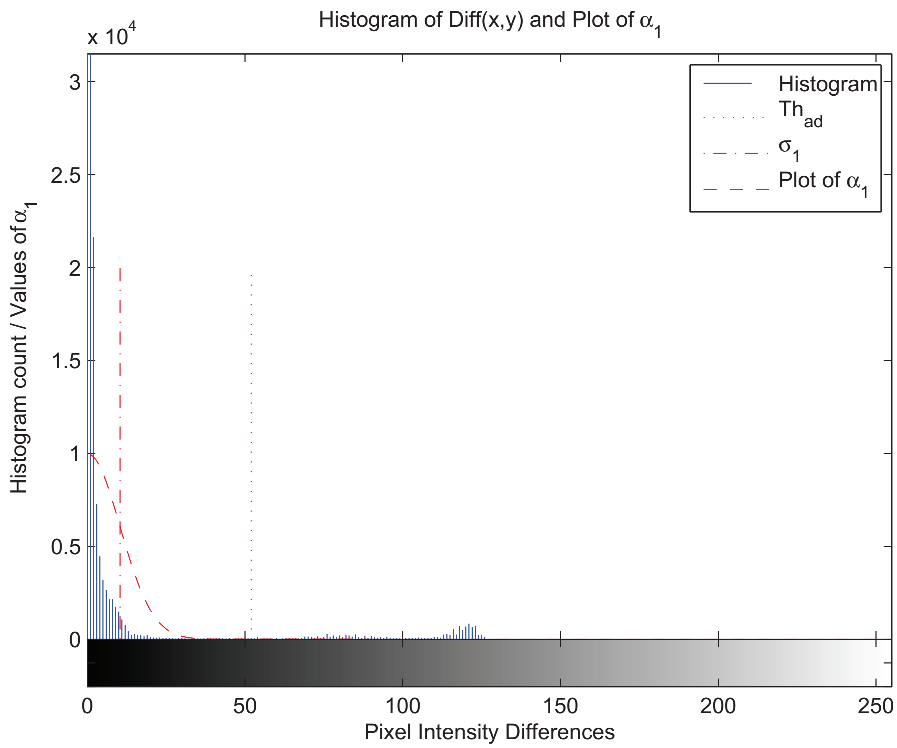

Figure 1. Histogram of $D_t(x, y)$ (using the test sequence shown in Figure 2) and the plot of σ_1 (scaled by 10^4) overlaid in the same diagram.

3.1. Sudden Illumination Change and Shading Model

The number of foreground pixels in a frame (foreground count) $N_{fg,t}$ is obtained for each frame. When there is a drastic increase in the number of foreground pixels in a frame relative to the size of foreground object in the previous frame, i.e. $N_{fg,t} \geq N_{fg,t-1}$, we assume that there is a global illumination change. In this case, Equation 2 will not be used to find the foreground pixels, a shading model will be used instead. Note that in this case the background model will not update in Equation 4 because both α_1 and $\alpha_2 = 0$.

Skifstad and Jain [12] propose an illumination-invariant method for change detection using a shading model. The model relates the intensity at a pixel I_p to the illumination $\pounds_i$ of the pixel by a shading coefficient S_p. The model is defined by:

$$I_p = \pounds_i S_p. \tag{8}$$

The assumption of the model is that if there is no physical change between two regions of two images, then the ratio $\Re$ expressed in Equation (8) remains constant. That implies:

$$\Re = \frac{I_{p,1}}{I_{p,2}} = \frac{\pounds_{i,1}}{\pounds_{i,2}}, \tag{9}$$

where the subscript 1 and 2 indicate two different regions and this expression is independent of the shading coefficients S_p. Skifstad and Jain's model uses the ratio of intensities at a pixel of the input image and the reference frame. In our proposed method, we use this intensity ratio for each pixel of the input image and the background model. Since this ratio is constant if the pixel is a background pixel, we obtain the sample mode of ratio of pixels between input image and the background model $\bar{\Re}$. Then we obtain the $N_{fg,t-1}$ (the number of foreground pixels in the last frame) pixels that have the closest $\Re$ values to $\bar{\Re}$ using the k-nearest neighbors [13, 14]. These $N_{fg,t-1}$ pixels are classified as the foreground pixels of the current frame. It is assumed that the object remains the same size compared to the last frame.

After computing the motion mask, the position and shape of the moving objects can be extracted. The position is computed as the centroid of the moving segment in the motion mask, and the shape is modeled by the smallest ellipse that encloses the moving segment in the motion mask. This information will be used to initialize object tracking. Details of the object tracking method used in our experiments is based on a particle filter method and is described in more detail in [15].

4. EXPERIMENTAL RESULTS

In our experiments, the test sequences include outdoor scenes under strong sunlight, waving tree leaves, walking pedestrians, fast moving bicycles, and indoor scenes with sudden illumination change.

First we observe that our method has a much lower computational time compared to a typical GMM-based method [3]. Table 1 shows the execution time of our method and a GMM-based method [3]. It should be noted that while the actual time durations depend on the computing platform, their relative magnitudes provide a direct comparison of their complexity.

Figure 2 shows a comparison of the updated background models between the same GMM-based method using a constant learning rate and our method. The regions covered by foreground objects are not updated, i.e. $\alpha_{ad} = 0$ and the background model is "cleaner" because it is not contaminated by foreground pixels.

Table 1. Table Execution Time

Sequence	GMM	Our Proposed Method
A	1008	191
B	983	189
C	1881	331
D	2150	395
E	2029	393

Figure 2. Comparison of the updated background models of frame 138 between a GMM-based method[3] (left) using a constant learning rate and the proposed method (right), the proposed background model reduces the learning rate to zero as the foreground object enters the scene.

Sample results are shown in Figures 3 - 5. The top row in each figure shows the original frames, the second row shows the background models, the third row shows the foreground masks using our proposed method, the fourth row shows the background models using the same GMM method of the same sequence, and the last row shows the foreground masks using the GMM method. The GMM method will need $1/\alpha$ number of frames to construct a background that does not have the foreground contamination. Figure 3 shows an indoor scene with a person walking. The background model using our proposed method is not contaminated by foreground pixels because the learning rates for background update for the uncovered region have very small values. This is because the pixels are just uncovered, α_2 in Equation 7 has a very small value, but this value will increase as the uncovered background pixels become more stable over time. Figure 4 shows the result of a sequence of an indoor scene with sudden illumination change with $\Delta I = 22.14$ that occurs between frame 240 and 242. The GMM method fails when there is a sudden illumination change, the false alarm rate is very high in frame 242 when the sudden illumination change occurs. Figure 5 shows the result of a sequence with waving tree leaves. Our proposed temporal background pixel count eliminates most, but not all, of the false alarms from the moving tree leaves, as shown in frame 139. This is because there are pixels of tree leaves that remain as foreground pixels for more than ζ_{min} frames. The GMM method works well even when there are moving tree leaves in the background.

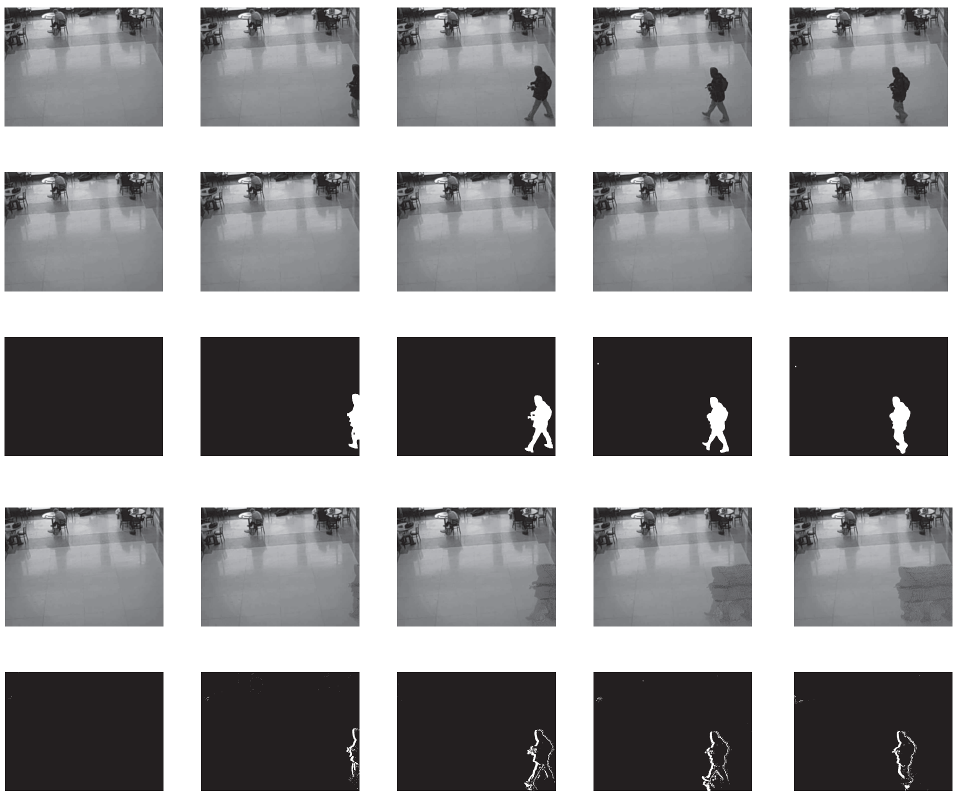

Figure 3. The top row shows the original frames, the second row shows the background models, and the last row shows the foreground masks. Observe that the background model is not contaminated by foreground pixels. The fourth row shows the background models using the GMM method of the same sequence, and the last row shows the foreground masks using the GMM method. The GMM method will need $1/\alpha$ number of frames to construct a background that does not have the foreground contamination.

Figure 6 is a test sequence under sudden illumination change. The illumination of current frame is very low, the right frame shows that our algorithm can segment the foreground mask successfully. The result can be refined using post-processing such as morphological filters.

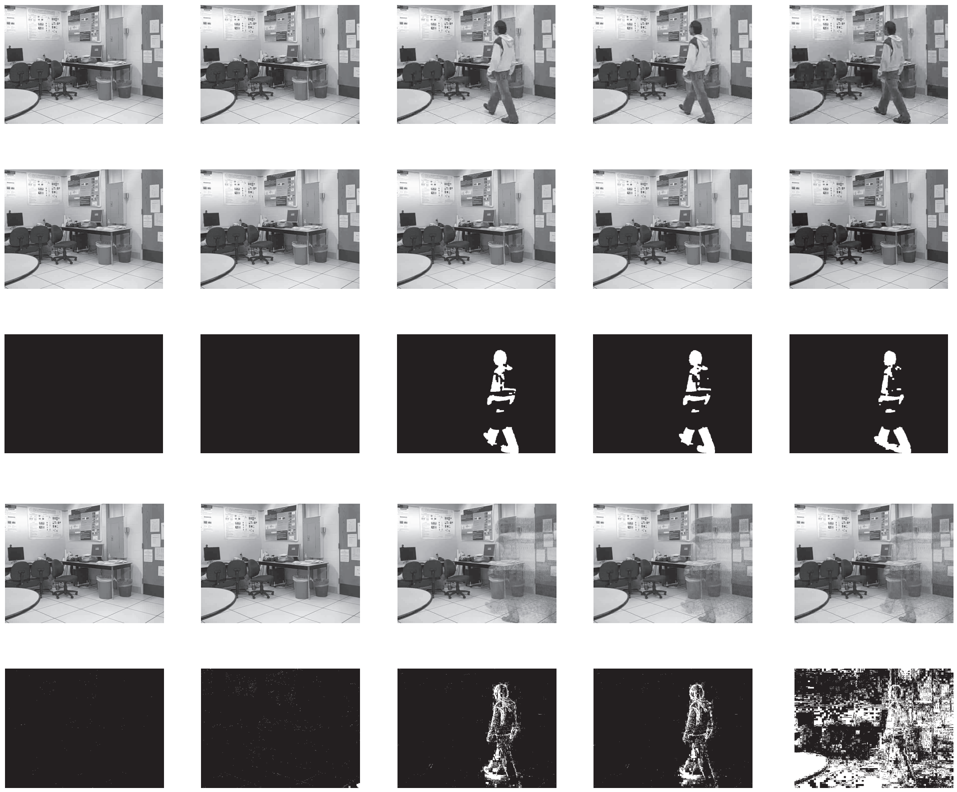

Figure 4. Frame 150, 185, 239, 240, 242 of a sequence with sudden illumination change that occurs between frame 239 and 242, with $\Delta I = 22.14$. The GMM method fails when there is a sudden illumination change.

5. CONCLUSIONS

In this paper we have proposed a new approach to background subtraction. In contrast to the traditional background subtraction techniques, our method can converge to the current background dynamics more rapidly and accurately. With the shading model, it is very robust to illumination changes. Our proposed algorithm performs as well as typical GMM methods in most cases but it also has a much lower computational complexity.

REFERENCES

1. W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, vol. 34, no. 3, pp. 334–352, August 2004.
2. R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 294–307, March 2005.
3. C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, August 2000.
4. P. Kaewtrakulpong and R. Bowden, "An improved adaptive background mixture model for realtime tracking with shadow detection," *Proceedings of European Workshop on Advanced Video Based Surveillance Systems*, London, September 2001, pp. 1–5.
5. A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1151–1163, July 2002.
6. B. Xie, V. Ramesh, and T. Boult, "Sudden illumination change detection using order consistency," *Image and Vision Computing*, vol. 22, no. 2, pp. 117–125, February 2004.

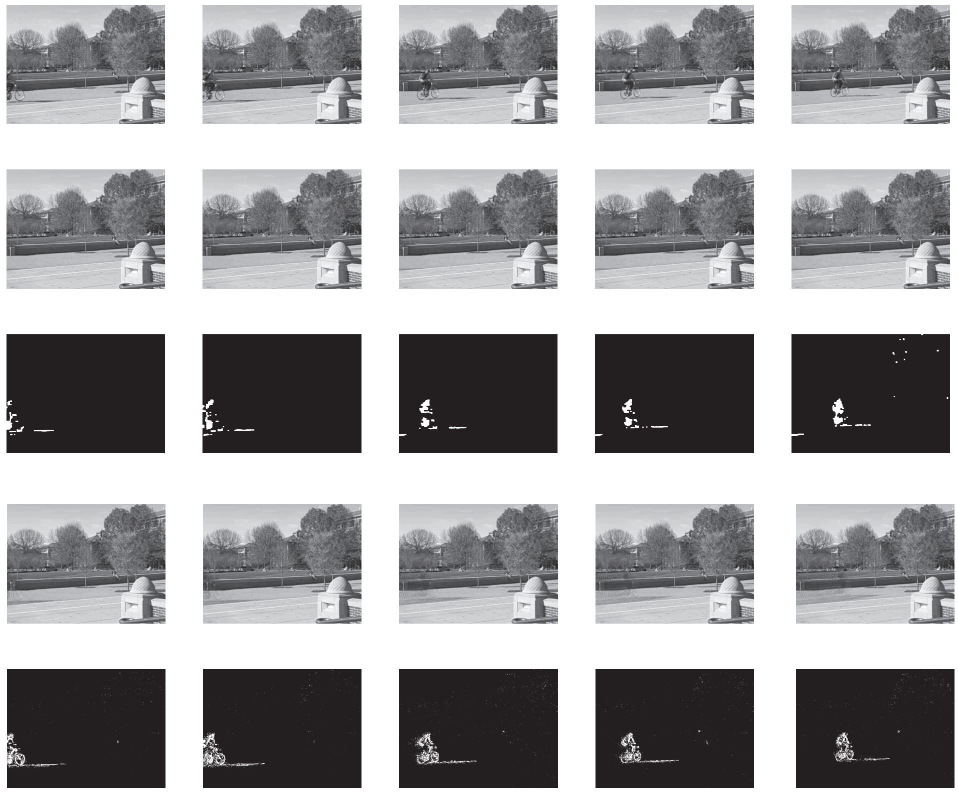

Figure 5. Frame 113, 115, 126, 130, 139 of a sequence with waving tree leaves. Our proposed temporal background pixel count eliminates most, but not all, of the false alarms from the moving tree leaves, as shown in frame 139. The GMM method works well even when there are moving tree leaves in the background.

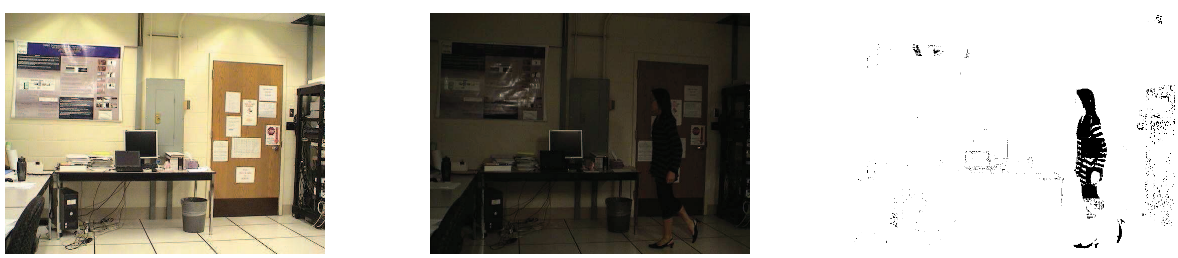

Figure 6. A sequence under sudden illumination change. The left frame is the background mode, the middle one is the current frame with a person walking, the right frame is the foreground mask using our proposed method.

7. B. T. Phong, "Illumination for computer generated pictures," *Communications of the ACM*, vol. 18, no. 6, pp. 311–317, 1975.
8. J. Pilet, C. Strecha, and P. Fua, "Making background subtraction robust to sudden illumination changes," *Proceedings of the European Conference on Computer Vision*, Marseille, France, October 2008, pp. 567–580.
9. K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 1, p. 255, 1999.
10. S.-K. Wang, B. Qin, Z.-H. Fang, and Z.-S. Ma, "Fast shadow detection according to the moving region," *Proceedings of the International Conference on Machine Learning and Cybernetics*, vol. 3, August 2007, pp. 1590–1595.
11. K. K. Ng and E. J. Delp, "Object tracking initialization using automatic moving object detection," *Pro-*

ceedings of SPIE/IS&T Conference on Visual Information Processing and Communication, vol. 7543, San Jose, CA, January 2010.

12. K. Skifstad and R. Jain, "Illumination independent change detection for real world image sequences," *Computer Vision, Graphics, and Image Processing*, vol. 46, no. 3, pp. 387–399, June 1989.
13. E. Fix and J. L. Hodges, "Discriminatory analysis: Nonparametric discrimination: Consistency properties," USAF School of Aviation Medicine, Randolph Field, Texas, Tech. Rep. 4, February 1951.
14. ——, "Discriminatory analysis: Nonparametric discrimination: Small sample performance," USAF School of Aviation Medicine, Randolph Field, Texas, Tech. Rep. 11, August 1952.
15. K. K. Ng and E. J. Delp, "New models for real-time tracking using particle filtering," *Proceedings of SPIE/IS&T Conference on Visual Communications and Image Processing*, vol. 7257, San Jose, CA, January 2009.