Dual-mode Detection for Foreground Segmentation in Low-contrast Video Images

Du-Ming Tsai and Wei-Yao Chiu
Department of Industrial Engineering and Management, Yuan-Ze University, Taiwan, Taiwan

Keywords: Background Updating, Foreground Segmentation, Object Detection, Mode Estimation.

Abstract: In video surveillance, the detection of foreground objects in an image sequence from a still camera is critical for object tracking, activity recognition, and behavior understanding. In this paper, a dual-mode scheme for foreground segmentation is proposed. The mode is based on the most frequently occurring gray level of observed consecutive image frames, and is used to represent the background in the scene. In order to accommodate the dynamic changes of a background, the proposed method uses a dual-mode model for background representation. The dual-mode model can represent two main states of the background and detect a more complete silhouette of the foreground object in the dynamic background. The proposed method can promptly calculate the exact gray-level mode of individual pixels in image sequences by simply dropping the last image frame and adding the current image in an observed period. The comparative evaluation of foreground segmentation methods is performed on the Microsoft's Wallflower dataset. The results show that the proposed method can quickly respond to illumination changes and well extract foreground objects in a low-contrast background.

1 INTRODUCTION

In video surveillance, foreground object segmentation is very important for the success of object tracking, incident detection, activity recognition, and behavior understanding. The complete silhouette of a segmented object can provide more detail about posture and position. General video surveillance systems should be robust to any changes of indoor/outdoor environments or lighting conditions. For the widely-used background updating models with a single Gaussian or a mixture of Gaussians, the main idea is based on the gray-level mean and the variance of each pixel in the background image sequence. When the background is represented by the mean pixel value for each pixel over a short period of time in the image sequence, it fails to detect a camouflaged foreground object against a low-contrast background or a moving object under sudden illumination changes of the observed scene because the mean value is inevitably disturbed by noise and outliers in the scene. If the outlier is a low-contrast background or a sudden illumination change, the mean value will be pulled closer to the value of the outlier. Thus, the mean background model cannot detect the low-contrast object and respond to sudden light changes. In this paper, we propose a mode-based background modeling method for foreground segmentation. It is well suited for scenes with under-exposure, variations in lighting and low-contrast objects. The proposed method can quickly calculate the exact gray-level modes of individual pixels in the image sequence by simply deleting the last image scene and adding the current image scene into the image sequence. It therefore involves only two arithmetic operations to update the intensity statistics, regardless of the number of consecutive image frames. In order to accommodate dynamic changes in a background, the proposed method uses a dual-mode model, instead of a single-mode model, for background representation.

2 DUAL-MODE BACKGROUND MODEL

2.1 Single-mode Background Modeling

In a surveillance system, the background needs to promptly respond to changes and resist noise in the

Tsai D. and Chiu W..
Dual-mode Detection for Foreground Segmentation in Low-contrast Video Images.
DOI: 10.5220/0004216004310435
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2013), pages 431-435
ISBN: 978-989-8565-47-1

environment. Mode detection assumes that the number of image frames of the background is larger than that of the foreground in a fixed duration of observed frames.

Let $\boldsymbol{F}_T(x, y) = \{f_t(x, y), t = T, T-1, \ldots, T-N+1\}$ be a series of N consecutive image frames, where T denotes the current time frame and $f_t(x, y)$ the gray-level of a pixel at coordinates (x, y) in image frame t. The single-mode background model is given by

$$M_{T-1}^{1st}(x, y) = \arg\max_{z} \{h[z \mid \boldsymbol{F}_{T-1}(x, y)], \forall z\} \quad (1)$$

where

$h[z \mid \boldsymbol{F}_{T-1}(x, y)] =$ Frequency of gray level z at pixel location (x, y) in $\boldsymbol{F}_{T-1}(x, y)$.

The gray-level mode background modeling formulas in eq. (1) involve one parameter value, the number of image frames *N*, to be determined. This value could be equivalent to the learning rate α in the Gaussian mixture model that must be determined beforehand. A very small number of *N* produces a fast update of the background statistics. The mode $M_T(x, y)$ obtained from a small *N* can then be interpreted as a short-term background. It can quickly absorb non-stationary background objects such as a moving chair and opening/closing of a curtain as parts of the background. It is therefore very responsive to dynamic changes in the environment. In contrast, a very large number of *N* produces a slow update of the background statistics. The mode $M_T(x, y)$ obtained from a large *N* can be considered as a long-term background. It can extract more accurate silhouette of a moving object.

2.2 Dual-mode Background Modeling

In a surveillance system, a complex environment sometimes has repeated dynamic interference in the scene. The single mode model may not be able to handle a dynamic background with only one background image. For example, two gray levels of a pixel may present the open/closed states of an automatic door. The single mode background can indicate only one of the two states of the door. Therefore, we propose a dual-mode background model for detecting foreground objects in a dynamic background.

Denote by $M_{T-1}^{1st}(x, y)$ and $M_{T-1}^{2nd}(x, y)$ the first and the second modes, i.e., the largest and the second largest frequencies of occurrence, in an observed duration. The formulas for calculating the first and the second modes are, thus, given by

$$M_{T-1}^{1st}(x, y) = \arg\max \{h[z \mid \boldsymbol{F}_{T-1}(x, y)], \forall z\}$$

$$M_{T-1}^{2nd}(x, y) = \arg\max \{h[z \mid \boldsymbol{F}_{T-1}(x, y)], \forall z \neq M_{T-1}^{1st}(x, y)\}$$

where

$h[z \mid \boldsymbol{F}_{T-1}(x, y)] =$ Frequency of gray level z at pixel location (x, y) in $\boldsymbol{F}_{T-1}(x, y)$.

$\mu_{D_{T-1}}^{1st}(x, y)$ and $\sigma_{D_{T-1}}^{1st}(x, y)$ are the mean and standard deviations of the distance $D_{T-1}(x, y)$ with respect to the first mode $M_{T-1}^{1st}(x, y)$ for all *N* image frames, which are adaptively updated with $M_{T-1}^{1st}(x, y)$ at the current time frame *T*. Likewise, statistics $\mu_{D_{T-1}}^{2nd}(x, y)$ and $\sigma_{D_{T-1}}^{2nd}(x, y)$ are updated with respect to the second mode $M_{T-1}^{2nd}(x, y)$. The foreground detection results are represented as two binary images, $B_T^{1nd}(x, y)$ and $B_T^{2nd}(x, y)$, for the dual-mode backgrounds $M_{T-1}^{1st}(x, y)$ and $M_{T-1}^{2nd}(x, y)$, where

$$B_T^{1st}(x, y) = \begin{cases} 1, if \left| f_T(x, y) - M_{T-1}^{1st}(x, y) \right| > \mu_{D_{T-1}}^{1st}(x, y) + K \cdot \sigma_{D_{T-1}}^{1st}(x, y) \\ 0, \text{otherwise} \end{cases}$$

and

$$B_T^{2nd}(x, y) = \begin{cases} 1, if \left| f_T(x, y) - M_{T-1}^{2nd}(x, y) \right| > \mu_{D_{T-1}}^{2nd}(x, y) + K \cdot \sigma_{D_{T-1}}^{2nd}(x, y) \\ 0, \text{otherwise} \end{cases}$$

The union of the two binary images $B_T^{1st}(x, y)$ and $B_T^{2nd}(x, y)$ is given by $B_T^{dual}(x, y)$, where

$$B_T^{dual}(x, y) = \begin{cases} 1 \,(foreground), if\ B_T^{1st} = 1\ and\ B_T^{2nd} = 1 \\ 0 \,(background), otherwise \end{cases}$$

The pixel (x, y) is identified as a foreground point only if it is a foreground point with respect to both background modes $M_{T-1}^{1st}(x, y)$ and $M_{T-1}^{2nd}(x, y)$. The dual-mode background model allows the background pixels with repeated changes of two gray-levels. When the background pixel is static over the entire observed duration, the two estimated modes will converge into a single one. Therefore, the dual-mode model can be used to represent a background with static and dynamic changes of gray levels.

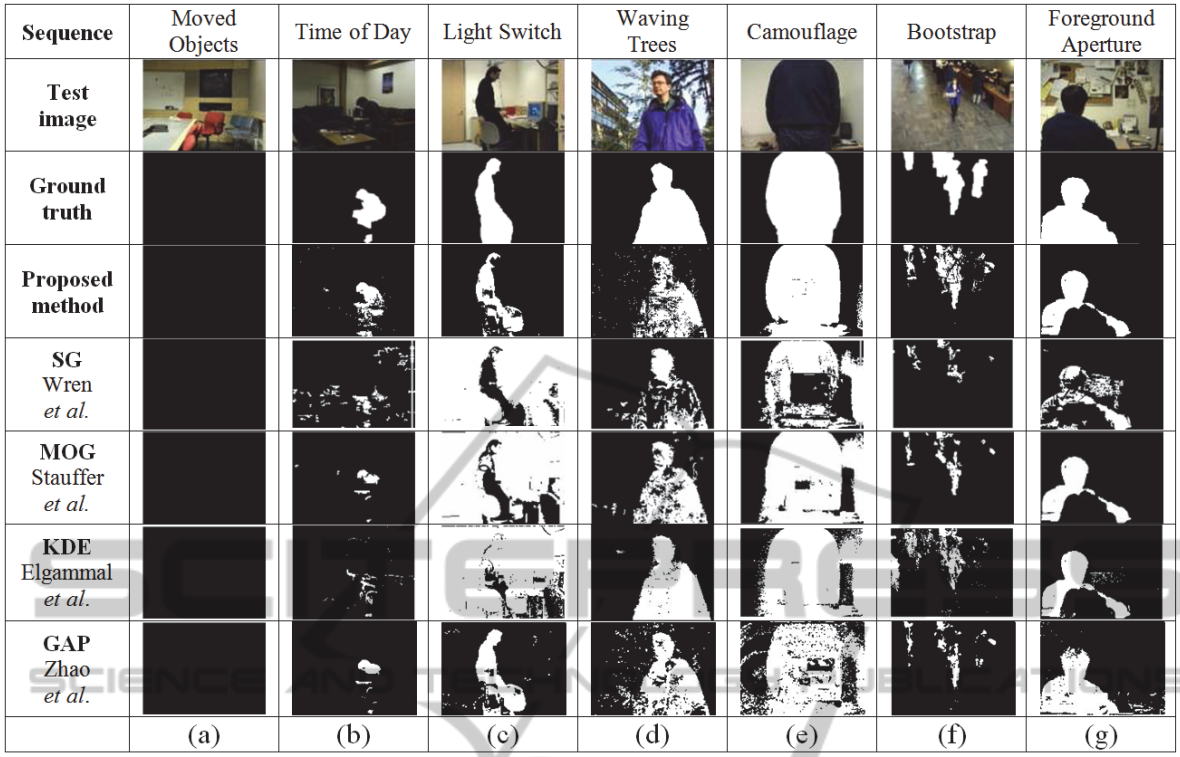

Figure 1: Comparative results of foreground segmentation methods on the Microsoft's Wallflower dataset.

2.3 Fast Updating of Dual-mode Background

The updating of the mode in each new image frame using Eq. (1) is computationally expensive. In this study, we propose a fast computation of the mode by updating the frequency of image frames without recounting all N image frames. Once the foreground objects in image frame $f_T(x, y)$ have been segmented from the background, the frequency $h[z|\boldsymbol{F}_{T-1}(x, y)]$ at the current time frame T can be quickly updated by deleting the presence of gray level z in the last frame $T-N$ and adding the presence of gray level z in the current image frame T. That is,

$$h\left[z|\boldsymbol{F}_T(x, y)\right] = h\left[z|\boldsymbol{F}_{T-1}(x, y)\right] - h\left[z|f_{T-N}(x, y)\right] + h\left[z|f_T(x, y)\right] \quad (2)$$

where

$$h\left[z|f_T(x, y)\right] = \begin{cases} 1, \text{ if gray level } z \text{ is present at } (x, y) \text{ in } f_T(x, y) \\ 0, \text{ otherwise } (f_T(x, y) \neq z) \end{cases}$$

and

$$h\left[z|f_{T-N}(x, y)\right] = \begin{cases} 1, & \text{if } f_{T-N}(x, y) = z \text{ in time frame } T-N \\ 0, & \text{otherwise} \end{cases}$$

The frequency of observed N image frames $h\left[z|\boldsymbol{F}_T(x, y)\right]$ can be updated by only two arithmetic operations, regardless of the number of consecutive image frames N. The first and second modes $M_T^{1st}(x, y)$ and $M_T^{2nd}(x, y)$ at current time frame T can then be updated from the frequency $h\left[z|\boldsymbol{F}_T(x, y)\right]$.

3 EXPERIMENTAL RESULTS

This section presents the experimental results from Microsoft's Wallflower dataset (Toyama et al., 1999) for the performance evaluation and the surveillance scenarios with low-contrast scenes to demonstrate the effectiveness of the proposed dual-mode model. The proposed algorithms were implemented using the C++ language and run on an INTEL Core2 2.53GHz 2046 MB personal computer. The test images in the experiments were 120×160 pixels wide with 8-bit gray levels. The computation time per frame of size 120×160 was 0.0065 seconds (i.e., 153 fps).

The Microsoft's Wallflower dataset, containing seven scenarios of image sequences, is used as the

benchmark for comparison. Each sequence shows a different type of challenge to foreground segmentation applications. The Wallflower dataset also provides a hand-segmented ground truth for evaluation.

Figure 1 shows the comparative results on the Microsoft benchmark dataset.

The experiment results show that the proposed method has overall the best performance on the Microsoft's Wallflower dataset, as seen in Figure 1. In terms of the capability for detecting low-contrast objects, the experimental results of the "Time-of-Day" video sequence in the Wallflower dataset show that the proposed dual-mode method significantly outperforms the GAP method as seen in column (b) in Figure 1. In terms of computational efficiency, the GAP method is computationally very expensive. It takes 25 seconds to process just an image frame of size 160×120. It can only be used for off-line applications, such as video retrieval.

The outdoor scenario entails monitoring of a parking lot, in which a person with an umbrella passes a clump of bushes and a light pole on a dark, rainy night. The path is poorly lit and the man is hardly observable. Rain and wind produce dynamic changes in the background. The image sequence is also influenced by the rain in the environment. The video sequence was filmed at 15 fps. The first column (a) of Figure 2 shows the original video sequence at varying time frames. The proposed foreground segmentation scheme can effectively extract the person under the very low-contrast background, as seen in the second column (b) of Figure 2. The detection results from the single Gaussian model are presented in the third column (c) of Figure 2. That model failed to detect the low-contrast person against the noisy background. The last column (d) of Figure 5 presents the detection results from the Gaussian mixture model. This column shows that the mixture of Gaussians model was robust to the noise of moving foliage and rain in the background. However, the foot of the person and the umbrella were mis-detected, and the shape is not complete.

4 CONCLUSIONS

In this paper, we have presented a dual-Mode model for foreground detection from a static camera. It detects the most frequently occurring gray level of each pixel, instead of the mean, in the image sequence. It can quickly respond to changes in illumination and accurately extract foreground objects against a low-contrast background. The variance of gray-level distance between pixel value and mode indicates the degree of scene changes. The dual-mode model can handle noise and repetitive movements such as an opening and closing door and flashes on a monitor. It can accurately extract the silhouette of a foreground object in a low-contrast scene, and it is very responsive to both gradual and radical changes in the environment. A process frame rate of 153 fps can be achieved for images of size 120×160 on an INTEL Core2 2.53GHz 2046 MB personal computer. Experimental results have revealed that the proposed method can be applied to monitoring of both indoor and outdoor scenarios with under-exposed environments and low-contrast foreground objects.

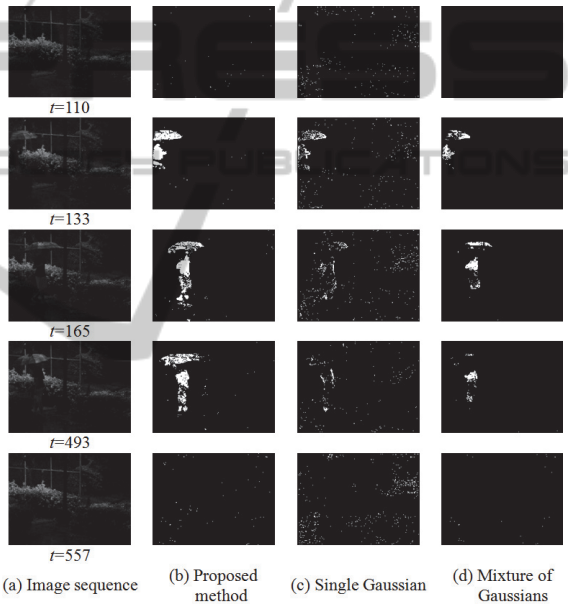

Figure 2: Experimental results of an outdoor parking lot on a rainy night: (a) discrete scene images in a video sequence filmed at 15 fps; (b) detected foreground objects from the proposed method; (c) detection results from the single Gaussian model; (d) detection results from the Gaussian mixture model. (The symbol *t* indicates the frame number in the sequence).

REFERENCES

Toyama K, Krumm J, Brumitt B, Meyers B. 1999, Wallflower: Principles and practice of background maintenance. *Inter Conf on Computer Vision 1999*: 255-261, Corfu, Greece, September.

Wren, C. R., Azarbayejani, A., Darrell, T., Pentland, A. P., 1997, Pfinder: real-time tracking of the human body, *IEEE Trans. Pattern Analysis and Machine*

Intelligence, 19 (7), pp. 780-785.

Stauffer, C., Grimson, W. E. L., 2000, Learning patterns of activity using real-time tracking, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22 (8), pp. 747-757.

Elgammal, A., Duraiswami, R., Davis, L., 2003, Efficient kernel density estimation using the Fast Gauss Transform with applications to color modeling and tracking, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25 (11), pp. 1499-1504.

Zhao, X., Satoh, Y., Takauji, H., Kaneko, S., Iwata, K., Ozaki, R., 2011, Object detection based on a robust and accurate statistical multi-point-pair model, *Pattern Recognition*, 44(6), pp. 1296-1311.