Large Area Video Surveillance System with Handoff Scheme among Multiple Cameras

Cheng-Chang Lien
Department of CSIE, Chung Hua University, Taiwan, ROC
cclien@chu.edu.tw

Yue-Min Jiang
Industrial Technology Research Institute, ISTC, Taiwan, ROC
jongfat@itri.org.tw

Lih-Guong Jang
Industrial Technology Research Institute, ISTC, Taiwan, ROC
lihgong@itri.org.tw

Abstract

Conventional video surveillance systems often have several shortcomings. First, object detection can't be accurate under the illumination variation environment or clustering background. Second, some dynamic textures e.g., moving leaves, clouds, can reduce the reliability of object detection. Third, when an object is tracked with multiple cameras, a handoff scheme is seldom considered. In this study, the background model fusing for temporal and texture models, multi-mode target tracking, and hand-off between two cameras are proposed to improve the abovementioned problems. Experimental results show that the targets on the crowd scene may be detected and tracked accurately with the rate above 10 fps.

Keywords: *dynamic textures, multi-mode target tracking temporal probability background model, handoff scheme.*

1. Introduction

In the conventional target detection systems, some typical methods are applied to extract the moving objects, e.g., background subtraction, and pixel-wise temporal difference analysis [2]. However, these methods are extremely sensitive to the variation of lighting or the dynamic background changing. Applying pixel-wise temporal differencing [3] may reduce the influence of the dynamic illumination change, but the regions of the moving objects are extracted incompletely when the background variation occurs. All the abovementioned methods do not utilize the motion information including the object and camera motions. By applying the method of optical flow, the moving objects may be detected even in the presence of camera motion. However, the high computation complexity makes the real-time object detection difficult. In this study, we apply the pixel-wise temporal statistical model [5], voting rule for the Y, C_r, and C_b Bayesian classifiers, foreground verification with the dynamic texture modeling to detect the targets on the crowd scene accurately.

In most target tracking systems, central point on the target is used as the reference location to prediction the position at the next frame. However, central point can be influenced easily by the inaccurate foreground detection. Here, the methods of principle-axis detection [8] is applied to extract the ground point of each target to serve as the reference point in the target tracking algorithms, e.g., Kalman filter [6] and Particle filter [7].

Most of the surveillance systems equipped with multiple cameras doesn't consider the construction of efficient and precise handoff scheme among cameras. Recently, the method proposed in [1] focused the study of object handoff between two overlapping views. In this study, we consider object handoff between two non-overlapping views by applying the Bayesian network [9] to establish a probability model to describe the probability of the position prediction when the target crosses different views. Fig. 1 illustrates the block diagram of the proposed system and Fig. 2 shows the block diagram of handoff scheme.

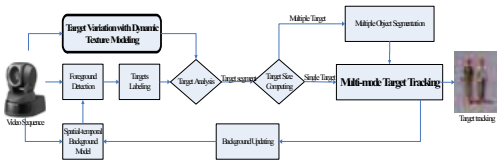


Fig. 1. The block diagram of the proposed system.

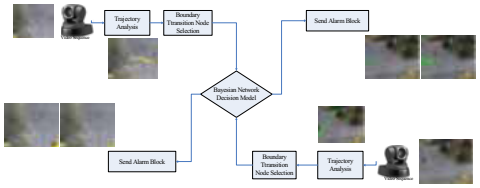


Fig. 2. The block diagram of the handoff Scheme.

2. Temporal Probability Background Model

In this section, the background model fusing with temporal and texture models is established to segment the foreground and background on a light variant or clustering background.

2.1 Pixel-Wise Temporal Probability Model

In an image sequence, the intensity variation within a time period for each pixel can be modeled by the Gaussian distribution function [4]. The pixel-based MOG function is defined as:

$$p(I) = p(I \mid B)P(B) + \sum_{j=1}^{c-1} p(I \mid \omega_j)p(\omega_j), \quad (1)$$

where, I is the intensity value, B denotes the background, ω_j denotes the moving object and c denotes the number of

Gaussians. The intensity distribution of the background pixel at a certain position x_b can be expressed as:

$$p(x_b \mid B) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{(I-\bar{I}(x_b))^2}{2\sigma^2}\right), \quad (2)$$

where $\bar{I}(x_b)$ and σ are the mean and standard deviation of the pixel intensity at x_b. According to the Bayesian decision rule [9], whether the pixel belongs to the background or the foreground (the moving objects) can be determined by the following likelihood inequality.

$$\frac{p(I \mid B_{x_b})}{p(I \mid T_{x_b})} \geq \frac{P(T)}{P(B)} = \lambda \quad (3)$$

where $P(T)$ and $P(B)$ are the prior probabilities for the background and moving objects respectively. By replacing $p(I \mid B_{xb})$ with Eq. (2), the likelihood ratio can be further simplified as:

$$\left|I-\bar{I}(x_b)\right| \leq k\sigma, \quad (4)$$

where $k = \sqrt{-2\ln(\sqrt{2\pi}\,\sigma\lambda/L)}$. If $|\, I - \bar{I}(x_b)\,| \leq k\sigma$, then the pixel is categorized as the background, otherwise, the pixel is categorized as the foreground.

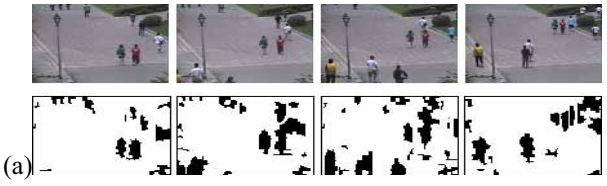
2.2 Foreground Detection Rule

Because it is difficult to detect objects when the intensity distribution is closed to the background model, the fusion of likelihood ratios of three color components (RGB or YC_rC_b) are proposed to overcome this problem. In general, there are two fusion rules to detect the foreground about linear combination and voting rules as:

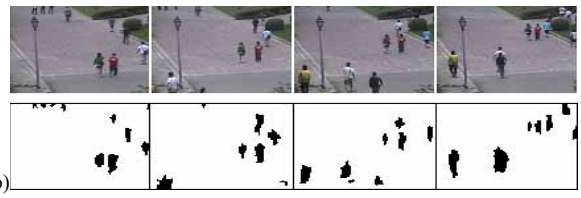
Linear combination rule:
If $w_y\, p_Y(u_y|B_x)+w_{cr}\, p_{Cr}(u_{Cr}|B_x)+w_{cb}\, p_{Cb}(u_{Cb}|B_x) > T$
pixel u is classified as background,
otherwise,
pixel u is classified as foreground,
where, w_y, w_{cr}, w_{cb} are weighting factors and sum of

Voting rule:
Given $p_Y(u_y|B_x)$, $p_{Cr}(u_{Cr}|B_x)$, $p_{Cb}(u_{Cb}|B_x)$,
If a pixel is classified as background with more than two components' background models,
Pixel u is classified as background,
Otherwise
Pixel u is classified as foreground.

By comparing several fusion rules, we apply the voting rule to cope with the illumination variation problem. If a pixel is classified as background with more than two components' background models, then this pixel is classified as background, otherwise, it is classified as foreground. Fig. 3 illustrates the foreground detection using the voting rule. It is obvious that the foreground detection using voting rule outperform the one using linear combination rule. Hence, we apply the voting rule to detect the objects on the outdoor crowd scene to cope with the illumination variation problem.



(a)



(b)

Fig. 3. (a) Foreground detection using linear combination rule. (b) Foreground detection using voting rule.

3. Foreground Verification using Texture Modeling

Many environmental dynamic textures such as leaves, fire, smoke, and sea waves may reduce the accuracy of target detection. Here, the dynamic texture will be modeled by using the modified local binary pattern (LBP)[11] and then the target can be detected without the influence of dynamic textures in the crowd scene. Here, a local texture pattern T [11] centering the pixel g_c and having P neighboring pixels is defined as:

$$T \approx t(s(g_0-g_c), s(g_1-g_c), ..., s(g_{P-1}-g_c)), \quad (5)$$

where

$$s(x) = \begin{cases} 1, & |\, x\,| \geq threshold \\ 0, & |\, x\,| < threshold \end{cases}. \quad (6)$$

Then, we transform the modified LBP in (5) to an integer value with the formula in Eq. (7).

$$LBP_{PR} = \sum_{0}^{P-1} s(g_p - g_c)2^P. \quad (7)$$

3.1 Foreground Detection using the Modified LBP

Here, we apply the modified LBP to perform the dynamic texture background modeling and remove the false foreground detection. In the LBP-based foreground detection, two threshold values are required to estimate the bit difference η between the captured scene and LBP-based background model. The LBP-based foreground detection rule is defined as:

$$P^{frame(t+1)}(\eta) = \begin{cases} foreground, & if \quad \eta \geq \eta\_th \\ background, & if \quad \eta < \eta\_th \end{cases}. \quad (8)$$

The bit difference η is calculated as:

$$\eta = \sum_{p=0}^{8} (LBP_p^{frame(t+1)}\ XOR\ LBP_p^{frame(t)}) \quad (9)$$

where, p is the index of the pixel on the circular chain.

3.2 Foreground Variation

In this study, both the pixel-wise temporal probability model and LBP texture model are constructed to detect the foreground, but how to integrate both background models to reduce the false detection is a very important issue. Based on the careful observation of foreground detections, the foreground detection rule is then designed as:

If $R(O_c) \in$ foreground
count $R(F_c^{LBP}|O_c)$,
If $N(R(F_c^{LBP}|O_c)) > N_{th}$,
$O_c \in$ True foreground,
update $R(O_c)$,
else
$O_c \in$ False foreground,
clear $R(O_c)$,
Endif

where, $R(O_c)$ denotes the region of a detected object using the pixel-wise temporal probability model on the current frame c, $R(F^c_{LBP}|O_c)$ denotes the region of the foreground detected by pixel-wise LBP texture model on the current frame c around the region of O_c. In order to correct the false detection, we propose the update/clear method as follow:

$$R'(O_c) = \begin{cases} R(O_c) \text{ Y } R(F_c^{LBP} \mid O_c), \text{if Update}_{foreground} \\ Null \quad \text{, if Clear}_{foreground} \end{cases}, \qquad (10)$$

Consequently, we can not only correct the detected regions of objects, but also can remove the false foreground. In the experimental results, the regions of moving leaves are removed by applying the proposed detection algorithm.

4. Multi-Mode Target Tracking

The targets tracking become complex on a crowd scene because the split and merging or occlusion conditions among the tracked targets occur frequently. In addition, the targets appear on the scene at the first time may be a single target or a merged multiple targets. In this study, the bottom-up tracking scheme is applied to develop the multi-modes tracking scheme. Each detected image blob is classified into single or multiple targets according to its area and then the tracked targets are examined whether they belong to the targets appear on previous frames. Based on the careful observation of target tracking on a crowd scene, there are six target tracking modes [10] described as follows.

Mode 1: An image blob is detected as a single target and its location is not predicted by other tracked targets from previous frames, i.e., the target is appeared on the scene at first time.

Mode 2: An image blob is detected as a single target and its location is predicted by one of the tracked targets from previous frames.

Mode 3: An image blob is detected as a single target and its location is predicted by a multiple target from previous frames, i.e., the object occlusion is occurred.

Mode 4: An image blob is detected as merged multiple targets and its location is not predicted by other tracked targets from previous frames, i.e., the merged multiple targets is appeared on the scene at first time.

Mode 5: An image blob is detected as a merged multiple target and its location is predicted by one of the tracked merged multiple targets from previous frames.

Mode 6: An image blob is detected as a merged multiple target and its location is predicted by a single target from previous frames, i.e., the objects' separation is occurred.

The detailed description of the multi-mode target tracking can be seen in [10].

When the tracked targets are slightly occluded it is possible to separate these targets. Then, the separated target can be tracked according to rules of modes 1 or 2. In general, color feature is effective to separate the merger targets. However, to develop robust target segmentation we apply the color-based difference projection to separate the targets from the merged targets. By observing the color difference projection histogram, we found the peak is not distinct for each color feature. To overcome this problem, the correlation for the color feature is used to find the segmentation line. In Fig. 4, the position of the prominent correlation peak is obtained to extract the segmentation line.

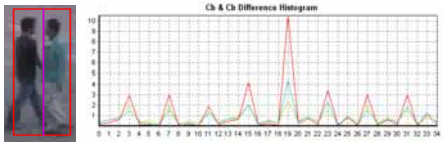


Fig. 4 the correlation for the color feature is used to find the segmentation line

5. Camera Handoff Scheme

To develop surveillance system over a large open space, multiple cameras are demanded to be integrated to monitor the targets on a large scene. Furthermore, the handoff scheme should be developed to track the targets across the cameras. The proposed handoff scheme constructed by using the Bayesian network [9] can predict the trajectory across multiple cameras. We divided the boundary of the camera scene into several regions which is regarded as the transition nodes in the Bayesian network. Therefore, the transition among the nodes will be described with the probabilities of trajectory prediction, which is defined as:

$$P(s_1, s_2, ...s_d) = \prod_{i=0}^{d} P(s_i \mid parents(s_i)), \qquad (11)$$

where s denotes the status nodes, d denotes the partition region index shown in Fig. 6-(b)(c). Eq. (11) can be further factorized as:

$$P(s_1, s_2, ...s_d) = \prod_{i=0}^{d} \frac{P(parents(s_i) \mid s_i) P(s_i)}{P(parents(s_i))}. \qquad (12)$$

By training the posterior probability in (12) with large training data we can predict the trajectory of a tracked target across multiple cameras. Fig. 5 illustrates the handoff scheme.

(a) (b) (c)

Fig. 5 Example of target tracking using the proposed handoff scheme. (a) Panoramic view. (b) Zoomed image of the yellow area in (a). (c) Zoomed image of the green area in (a).

6. Experimental Results

6.1 Moving Object Filtering using Dynamic Texture Model

In Fig. 6-(a), it shows an outdoor scene. Fig. 6-(b) represents the foreground with the pixel-wise temporal probability model, and the dynamic texture detection model is described in Fig. 6-(c). By using the dynamic detection model, the targets will be separated into the truly foreground target and the constant texture object. If the

object with too many constant textures, we will define the target as the noise, and it then will be removed, i.e. in Fig. 6-(d). Finally, we can improve the accuracy about the detected target.

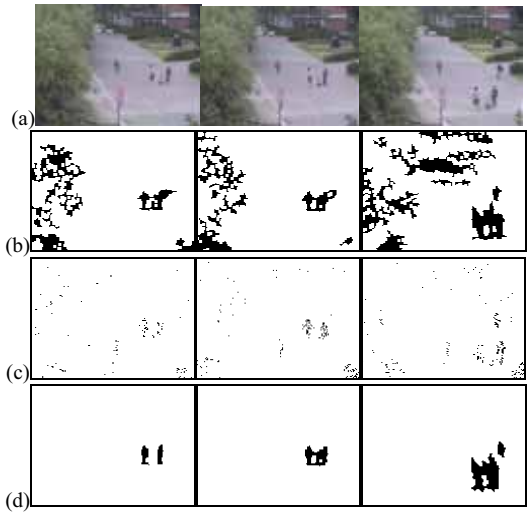


Fig. 6. (a) Outdoor scene. (b) The objects are detected by using pixel-wise temporal probability model. (c) Foreground detection using the dynamic texture model. (d) The extracted objects after the texture noise removing process.

6.2 Multi-mode Target Tacking Scheme

In Fig. 7, the multi-mode target tracking on a crowd outdoor scene is illustrated. The split, merge, and occlusion among the targets occur repeatedly. The merged multiple-target is labeled "M". Meanwhile, some important features, e.g., color, weight, height, ground point position, are record as the tracking measurements for each target. The detailed description of the multi-mode target tracking can be seen in [14].

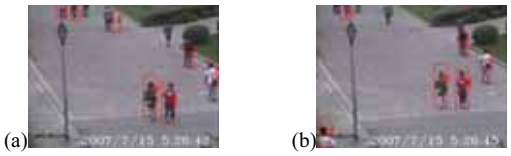


Fig. 7. Multi-mode tracking on an outdoor crowd scene.

6.3 Handoff Scheme

In Fig. 8-(a), the target is walking from the yellow region to the green region shown in Fig. 8-(b). When the target walks across two non-overlapping camera views, the handoff scheme starts to predict the position that the tracked target will appear on the second camera scene shown in Fig. 8-(b). Experimental results show that the prediction accuracy is about 85%.

Fig. 8. (a) The trajectory of the tracked target on the first camera view. (b) The position that the tracked target will appear on the second camera scene is denoted as green block.

7. Conclusion

In this study, the spatial-temporal probability background model, dynamic texture modeling, multi-mode target tracking scheme and handoff scheme are integrated to develop a robust multi-mode target tracking system that can overcome the main shortcomings in the conventional surveillance system. The experimental results show the proposed system can perform the target detection and tracking with the rate above 10 fps.

References

[1] Y. Jo, J. Han, W. Nam, "Object handoff between uncalibrated views without planar ground assumption", *Pattern Recognition Letters,* Vol. 29, 2008, pp. 2099-2108.

[2] A. Elgammal, D. Harwood and L. Davis, "Non-parametric model for background subtraction," *in Proceedings of the 6th European Conference on Computer Vision, 2000,* pp. 751-767.

[3] R. Jain, W. Martin and J. Aggarwal, "Segmentation through the detection of changes due to motion," *Compute Graph Image Process 11,* 1979, pp. 13–34.

[4] Y. Ren, C. S. Chua and Y. K. Ho, "Motion detection with nonstationary background," *Machine Vision and Application,* Vol. 13, No. 5-6, Mar. 2003, pp. 332–343.

[5] C. C. Lien and S. C. Hsu, "The target tracking using the spatial-temporal probability model," *IEEE International Conference on Nonlinear Signal and Image Processing, NSIP 2005,* May 2005, pp. 34-39.

[6] M. Xu, J. Orwell, L. Lowey and D. Thirde, "Architecture and algorithms for tracking football players with multiple cameras" *Image and Signal Processing, IEE Proceedings,* Vol. 152, Issue 2, April 2005, pp. 232-241.

[7] K. Nummiaro, E. K. Meier and L. J. V. Gool, "An adaptive color-based particle filter" *Image Vision Computing,* Vol. 21, Issue. 1, 2002, pp. 99-110.

[8] W. Hu, M. Hu, X. Zhou, Tieniu Tan, "Principal axis-based correspondence between multiple cameras for people tracking" *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 28, No. 4, April 2006, pp. 663-671.

[9] E. Alpaydin, Introduction to Machine Learning. MIT Press, Cambridge 2004.

[10] C.-C. Lien, J.-C. Wang and Y.-M. Jiang, "Multi-mode target tracing on s crowd scene," IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing 2007 (IIH-MSP 2007), Nov. 26-28, Kaohsiung, Taiwan.

[11] T. Ojala, M. Pietika¨inen, and T. Ma¨enpa¨a¨, "Multiresolution Gray Scale and Rotation Invariant Texture Analysis with Local Binary Patterns," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 7, pp. 971-987, July 2002.