PEOPLE COUNTING USING MULTI-MODE MULTI-TARGET TRACKING SCHEME

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

Ya-Lin Huang
Industrial Technology Research Institute, ISTC, Taiwan, ROC

Chin-Chuan Han
Dept. of CSIE, National United University, Taiwan, ROC
cchan@nuu.edu.tw

ABSTRACT

Conventional video surveillance systems often have several shortcomings. First, target detection can't be accurate under the light variation environment. Second, multiple target tracking becomes difficult on a crowd scene. Third, it is difficult to the partition the tracked targets from a merged image blob. Finally, the tracking efficiency and precision are reduced by the inaccurate foreground detection. In this paper, the fusion of temporal and texture background model, multi-mode tracking scheme, color-based difference projection, and ground point detection are proposed to improve the abovementioned problems. In addition, we propose a people counting scheme based on the multi-mode multi-target tracking method on a crowd scene. Experimental results show that the targets on the scene may be detected robustly with the rate above 10 fps and counted with the accuracy above 90%.

Index Terms—Multi-mode multi-target tracking, People counting

1. INTRODUCTION

In the conventional tracking systems, some typical methods are applied to extract the moving objects, e.g., background subtraction [1] and temporal difference analysis [2]. Both kinds of methods are sensitive to the illumination variation and background changing. In the following, the pixel-based temporal probability model [3] is proposed to extract the moving objects. In this paper, we improve the pixel-based temporal statistical model by constructing the spatial-temporal statistical model [4]. By considering the spatial distribution around each pixel (texture distribution) the problem of slight background variation may be overcome.

In addition, conventional target tracking systems are developed on a less crowd scene. However, target tracking on a crowd scene are an important issue in a large open space. Multiple target tracking become difficult on a crowd scene because the split and merge or occlusions among the tracked targets occur frequently. In this study, a bottom-up multi-mode target tracking scheme is proposed to improve the accuracy and efficiency of target tracking on a crowd scene. In most target tracking systems, central point can be influenced easily by the inaccurate foreground detection. Here, the method of principle-axis detection [5] is applied to extract the ground point of each target to server as the reference point in the target tracking algorithms [6].

Early works for locating and counting people were fulfilled by looking for heads in the vertical histograms of image blobs, where the number of peaks was assumed to be the number of heads. Hydra [7] proposed a method based on silhouettes to identify the moving people groups and detect the number of heads to count people in groups. However, heads may be an unreliable cue when they occupy only a few pixels in the image. Based on the multi-mode multi-target tracking scheme [9] the number of people appeared on the crowd scene can be estimated. The system block diagram is shown in Fig. 1.

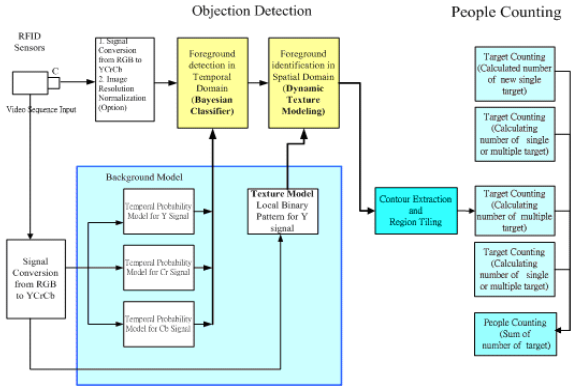


Fig. 1 The block diagram of the multi-mode multi-target tracking system.

2. FOREGROUND DETECTION USING TEMPORAL AND TEXTURE BACKGROUND MODELS

In general, target detection can't be accurate under the light variation environment or clustering background. Especially, the light reflection and back-lighted problems can deteriorate the target detection seriously. Here, we apply a pixel-wise temporal probability background model [4] and voting rule [9] to segment the foreground and background on a light variant or clustering background. To improve the accuracy of foreground detection, the dynamic texture model is used to eliminate the false foregrounds. In general,

978-0-7695-3762-7/09 $26.00 © 2009 IEEE
DOI 10.1109/IIH-MSP.2009.239

the texture model for background can be modeled by using the LBP [8] defined as:

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

Here, we apply the modified LBP to perform the dynamic texture 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:

$$P^{frame(t+1)}(\eta) = \begin{cases} foreground, & if \quad \eta \geq \eta\_th \\ background, & if \quad \eta < \eta\_th \end{cases}, \quad (2)$$

where $P^{frame(t+1)}$ separates the foreground from background according to the bit difference between the captured scene and LBP-based background model. The bit difference η is defined as:

$$\eta = \sum_{p=0}^{8} (LBP_p^{frame(t+1)} \; XOR \; LBP_p^{frame(t)}) \quad (3)$$

where, p is the index of the pixel on the circular chain. Based on the careful observation of foreground detections, the foreground variation rule is then designed as:

If $I(O_c) \in$ *foreground*
  count $R(F^c_{LBP} \mid O_c)$,
  If $p(R(F^c_{LBP} \mid O_c)) > D_{th}$,
    $O_c \in$ True *foreground*,
    update $I(O_c)$,
  else
    $O_c \in$ F*alse foreground*,
    clear $I(O_c)$,
End if

where, O_c denotes the target that is detected by pixel-wise temporal probability model on the current frame c, I denotes the information which about the target size, number, appearance of the target, F^c_{LBP} is the foreground detected by pixel-wise LBP texture model on the current frame c, R represents the candidate detecting region, p denotes the pixel number of the LBP foreground in the region R, and D_{th} is a threshold used for noise filtering. In order to correct the false detection, we propose the update/clear method as follow:

$$I(O_c) = \begin{cases} F(O_c) \cup F^c_{LBP}, & if \; Update_{foreground} \\ Null, & if \; Clear_{foreground} \end{cases} \quad (4)$$

Consequently, we can not only correct the detected regions of objects, but also can remove the false foreground. Fig. 2 illustrates the results of foreground variation. The regions of moving leaves are removed by applying the proposed detection algorithm.

Fig. 2. (a) Original image. (b) Foreground detected with pixel-wise temporal probability model. (c) Foreground detected with modified LBP. (d) Foreground variation using the proposed method.

3. MULTI-MODE MULTI-TARGET TRACKING

Here, a multi-mode multi-targets tracking scheme [9] is applied to overcome the complex target tracking problem on a crowd scene.

3.1. Multi-Mode Multi Targets Tracking Scheme

Based on the careful observation of target tracking on a crowd scene, there are totally six target tracking modes shown in Fig. 3 exist in the complex target tracking situation, which is described as follows.

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

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

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

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

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

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

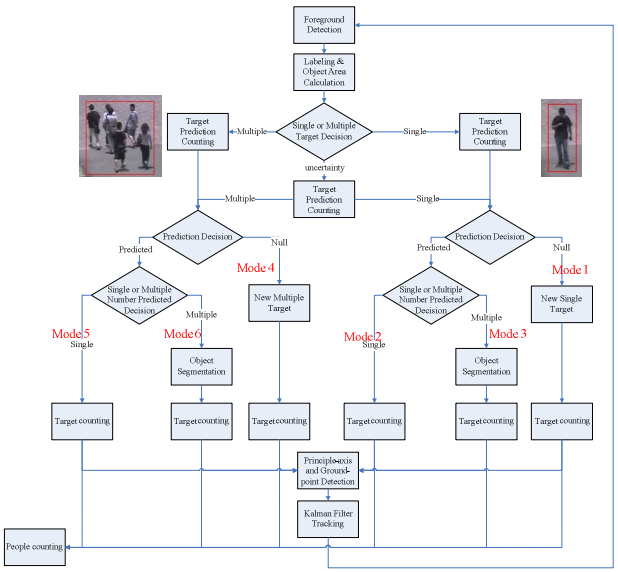


Fig. 3 Flowchart of the multi-mode multi targets tracking scheme.

3.2. Target Segmentation

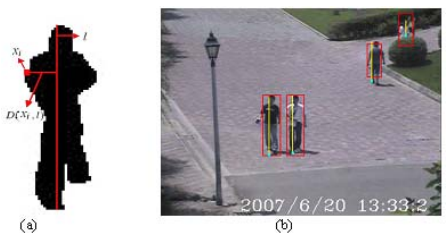
When the tracked targets are slightly occluded it is possible to separate them into individual objects. In general, color features are effective to separate the slightly merged targets. However, to develop robust target segmentation we apply the color-based difference projection method [9] to separate the targets from a merged image blob.

3.3. Principal Axis and Ground-point Detection

We apply the method of least median of squares to find the principal axis for an isolated target. Based on the global shape constraint the pixel on human body is distributed symmetrically about the principal axis. The principal axis is determined by minimizing the median of squared perpendicular distances from the foreground pixels to a vertical axis. Let $D(x_i, l)$ be the perpendicular distance from the i^{th} foreground pixel x_i to the axis l shown in Fig. 4. The principal axis l is estimated by minimizing

$$L = \arg\min_{l} \text{Median}_i \left\{ D(x_i, l)^2 \right\} \tag{5}$$

After detecting the principle axis, the ground point is found from the intersection of the principle axis and the bottom boundary.



(a) (b)

Fig. 4 (a) Principal axis of a tracked subject. (b) Example of the detected principal axis and ground point

4. PEOPLE COUNTING

With the multi-mode multi-target tracking scheme, people counting on a crowd scene can be developed. The contour of people is used to locate human and estimate the number of people in the crowd scene. The proposed people counting method is described in Fig. 5.

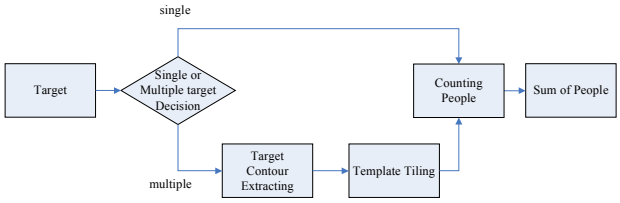


Fig. 5 Flowchart of people counting scheme.

By analyzing the target contour, we can compute how many people inside in the contour. Fig. 6 shows the template tiling process that can estimate the number of people in a merged image blob. First, we use contour extraction to get target contour shown in Fig. 6-(c). After obtaining the target contour, we propose a template tiling method to estimate people number in the target region. Second, we apply a rectangular box with the average human height and width to match object contour from left-up to right-down. Finally, we can estimate how many people in the merged image blob. The template tiling result is shown in Fig. 6-(d).

(a) (b) (c) (d)

Fig. 6 Template tiling. (a) Original image. (b) Detected foreground. (c) Object contour. (d) Template tiling result.

5. EXPERIMENTAL RESULT

Here, the open space in the Chu-Hua University is served as a test bed. First, the proposed spatial-temporal probability model is applied to detect the moving targets. Second, the multi-mode target tracking scheme with the ground point extraction and color-based target segmentation is used to track the targets. Finally, people counting scheme with the contour corner detection and template tiling is used to estimate the number in the crowd scene.

5.1. Object Extraction

Fig. 7 shows an outdoor scene and the detected foreground with the spatial-temporal probability model. By using the spatial-temporal background model the targets can detect accurately even though the illumination changing is serious.

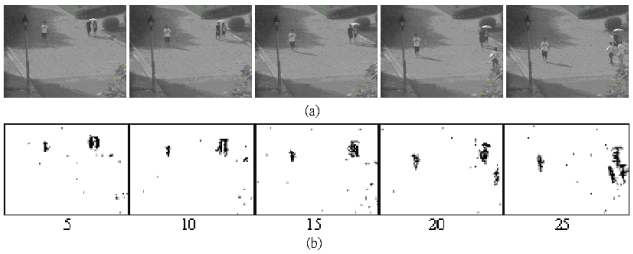


Fig. 7 Moving target detection on an outdoor scene. (a) Outdoor scene. (b) The objects are detected by using spatial-temporal probability model.

5.2. Multi-Mode Target Tracking Scheme

The multi–mode multi targets detection on a crowd outdoor scene is shown in Fig. 8 and each target will be tracked with the mode according to the situation of target occlusion.

Fig. 8 Multi-mode tracking on an outdoor crowd scene.

5.3. People Counting Scheme

Here, the people counting over a large area using the templates tiling is illustrated. Fig. 9-(a) shows a people counting result from outdoor scene. The estimated numbers of people are shown in the left-up corner of images in Fig. 9-(a). Fig. 9-(b) shows the multiple object partition using templates tiling respectively.

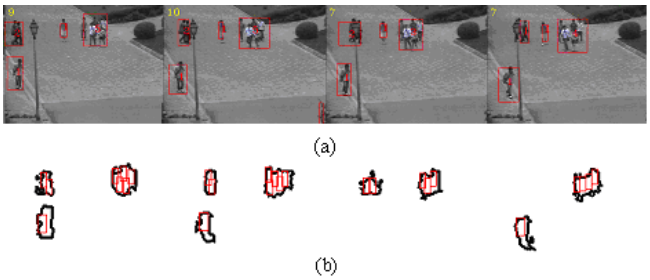


Fig. 9 People counting on an outdoor crowd scene. (a) The estimated numbers of people are shown in the left-up corners. (b) The results of template tiling for people counting.

6. CONCLUSION

In this paper, the spatial-temporal probability background model and texture background model are fused to detect the foregrounds robustly even thought light changes seriously. Furthermore, the multi-mode multi-target tracking scheme can track the targets on the crowd scene with proper mode transitions. By applying the template tiling method we can estimate the number of people in the crowd scene. Experimental results show that the targets on the scene may be tracked with the correct tracking modes and with tracking rate above 10 fps. The number of people also can be estimated correctly

7. REFERENCES

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

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

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

[4] 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*, pp. 34-39, 2005.

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

[6] D. Salmond, "Target tracking: introduction and Kalman tracking filters," *IEEE Target Tracking: Algorithms and Applications,* Vol. 2, pp. 1/1-1/16, 2001.

[7] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4：Who? When? Where? What? A real-time system for detecting and tracking people," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 222-227, April 1998.

[8] 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.

[9] 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.