Tracing and Recognizing Persons in Visual Surveillance System using Modified overlap Tracker

NEELIMA K.S

***Abstract*—The supreme contest to examine characters from a monocular video scene is to trace targets within the occlusion circumstances. In this work, we present a scheme to involuntarily trace and tally people in an inspection system. First, an active environment calculation unit is employed to model light change and then to choose ordinary objects from a stationary pictures. To recognize foreground objects as characters, positions and sizes of front regions are treated as decision features. Likewise, the performance to trace persons is enhanced by using the modified overlap tracker, which investigates the centered distance between adjacent objects to help on target tracing in occlusion states of integration and separation. On the experiments of tracing and counting public in three video sequences, the consequences show that the proposed scheme can improve the averaged recognition ratio about 10% as compared to the previous work.**

Here tracked persons are identified and recognized by comparing with external database which was already created .In this work, noisy and blurry videos are also taken as input and persons are easily tracked from these distorted videos.

***Index Terms*— Intelligent Surveillance System, People Tracking, People counting, people recognizing, Overlap Tracker, Occlusion**

1. INTRODUCTION

Illustration-based target tracing is difficult to necessarily observe object actions in video sequences. To view and find object actions from a monocular scene, object occlusions frequently include detection errors due to objects in packed areas. In this work, an object tracingsystem is proposed to overcome the occlusion effects and then to increase the correctness of counting characters in a visual surveillance system.

Up to now, usual effort has applied computer-visualization skills to notice movements and realize actions of characters in a static camera. In Marcenaro*et al.*'s study, for relieving the result of active occlusion, alinear Kalman filter is used to execute tag tracing by matching outline features. Lien *et al*. taught a multi-mode process to increase accuracy and efficiency for tracingmany targets in a crowed scene. Six modes for target tracing are defined with heuristic consideration, and the people count is finally done by model tilling. In, an object tracing framework is planned by separating the foreground regions into many parts in which color features are extracted for object matching. In addition, Haritaoglu*etal*. recommended selecting the contours of persons tocategorize combine and divide states for crowded situations in outside visual surveillance method. For gettingperfect counting of group people, Fehr *et al*. compared the people counting results from using the extended Kalman filter in combination with unlike background segmentation techniques. Their research reportedbetter people counting results achieved from using the method of layering foreground recognition. In the contest of target tracing method, the Kalman filter was used for guessing trajectories of people flow in the successive frames. However, the aboveexplained schemes track the targetswith features of color, shape and contour, which depend on efficient object segmentation, sensitive to light differences. In the condition of high crowded densities, it is also necessary to carefully notice the tracing states of unite and split.

This work presents character tracing from a static video scene to count the number of characters and recognize. Anactive background subtraction is first implemented to sense whether a target existing or not, and then states of merged and split targets are derived to prevent inaccurate people counting at occlusion situations. To do so, a modified overlie tracker is used to finish the character label and then to reach the goal of tracing characters. In addition, the centered distance between adjacent objects is further analyzed to attain fairly good people tracing and counting results in successive frames.

2. PROPOSED SYSTEM ARCHITECTURE

The proposed system block diagram of people tracing and counting is shown in Fig. 1. Anactive background subtraction module is first used to division moving objects from each captured video frame. In order to overcome light changes, aactive threshold value associated with finding regions of interest from the distinguished image is iteratively calculated according to the allocations of background and foreground pixels in each frame. After obtaining theforeground regions, four states including new, leaving, merged and split are allotted to the detected moving objects according to their appearances in the present frame. In particular, targets recognized as conditions of unite and split further pass through backward tracing for relieving the occlusion results by examining the centered distances among objects in the previous frame. To conclude, targets in four states are tagged to give the outcomes of people tracing and counting.

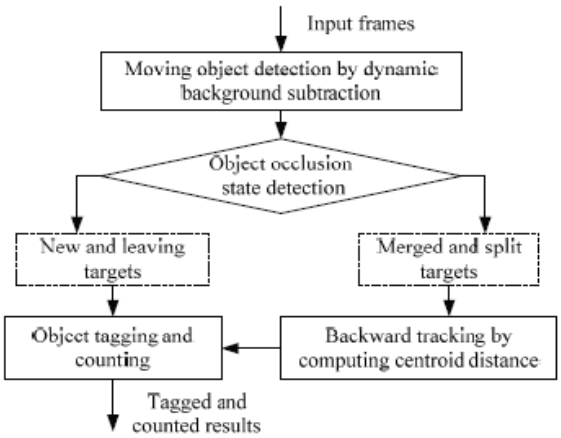

Fig. 1 Block diagram of the proposed people tracing and countingsystem.

3. MOVING OBJECT DETECTION AND RECOGNITION

In the background removing stage, a differential image is obtained by subtracting a background image from the current one, and then the foreground areas are recognized by thresholding the differential image. Therefore, two variables are corresponded to light variations in surroundings: one is to build the background model dynamically and the other one is to select asuitable threshold for obtaining foreground objects. Here, the background model and threshold value are both adaptively decided according to frame contents.

3.1. Building Background Model

To adaptively build the background model, theconsistency of pixel gray-level values of consequent framesis explored. Assume $F_m(x, y)$ denote the pixel gray-level value on (*x*,*y*) of the *m*-th frame, and $B_m(x, y)$ present the corresponding background pixel gray-level value calculated from previous frames. Hence, each background pixel can be updated by the following,

$$B_m(x,y)=\frac{1}{m}\sum_{i=0}^{m-1}F_i(x,y) \quad (1)$$

In Eq. (1), *m* is the index of present frame and also indicates the gathered frames for background pixel averaging. In our research, for primarily building background model, the starting 100 frames in a video sequence are used. After that, the background image is obtained by taking the mean values of the pixels and their associated background ones. When *m* being large, pixels engaged by moving objects can besmoothed to approach true ones on background model.

3.2 Thresholding Differential Image

A distinguished image can be generated by subtracting the binary background image from the current frame. The threshold value for finding the foreground areas need to be appropriately determined with considering the stochastic difference of frame contents. For doing this, the threshold for each distinguished frame is iteratively derived with regard to the distributions of background and foreground regions [7]. The procedure for estimating threshold is illustrated as below.

Step1: An initial threshold is set by averaging the pixelvalues of the differential image and then utilized for segmenting an image into foreground and background regions. To observe the distributions of these two regions, the means of pixels belonging to the background and foregroundregions are separately calculated and denoted as μ_B and μ_O

$$\mu_B = \frac{\sum_{(i,j)\in background} F(i,j)}{\#background\_pixels} \quad (2)$$

$$\mu_O = \frac{\sum_{(i,j)\in object} F(i,j)}{\#object\_pixels} \quad (3)$$

Step2: A temporary value *T* is computed by $T=(\mu_B+\mu_C)/2$

Step3: The updated *T* is used for thresholding the differentiated image.

Step4: Steps 1 to 3 are iterated till that is close to μ_B and μ_O additionally; the foreground image is instantaneously obtained.

After obtaining the binary foreground image,morphological operations are further employed to eliminatenoisy pieces and to fix broken contours of regions

4. PEOPLE TRACING AND RECOGNIZING

In a visual surveillance system, what we point is to get spatial locations of objects and to monitor their trajectoriesalong with time slots. Here, by considering the centereddistances between objects, a customizedoverlap tracker isdeveloped for preventing incorrect tracing people in the occluded condition.

Denoising the frames:

Although the recent advances in the sparse representations of images have achieved outstanding denosing results, removing real, structured noise in digital video frames remains a challenging problem. We show the utility of reliable motion estimation to establish temporal correspondence across frames in order to achieve high-quality video denoising. In this paper, we propose an adaptive video denosing framework that integrates robust optical flow into a non-local means (NLM) framework with noise level estimation. The spatial regularization in optical flow is the key to ensure temporal coherence in removing structured noise. Furthermore, we introduce approximate K-nearest neighbor matching to significantly reduce the complexity of classical NLM methods. Experimental results show that our system is comparable with the state of the art in removing Noise, and significantly outperforms the state of the art in removing real, structured noise.

4.1. Modified Overlap Tracker

After having the segmented outputs from backgroundsubtraction, regions of interest are recognized from theforeground image based on the completeness of regioncontours. In addition, finding character regions from theinterested ones is then considered under the physicalconstraints of people, including the dynamic range ofpixel gray-level values and the outlines of regions. As objectsconforming to the constraints, ellipses having the

reduced areas that can cover the regions are buildingto obtain theirrelated position parameters, ellipse radiuses, centeredand distances.In the modified overlap tracker, as depicted in Fig. 2,four tracing states containing new target, leaving target,merged target and split target are used to realize characters in the current frame. New target means an objectentering a video scene; on the contrary, leaving target describes an object out of a video scene. Particularly, for merged and split target states, the touch event of objects is detected in adjacent frames. The decision of target merging and splitting in the current frame also considers the target states by backward tracing the objects that show in the previous frame. The tracker finally allocates tags to individual objects by means of positional continuity preserved by the tagged characters in the previous frame.

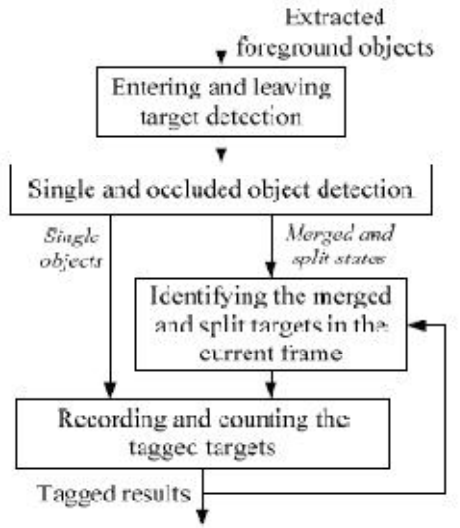

Fig. 2 Processing steps for tracing and tagging targets

(a)

(b)

(c)

Fig. 3 (a) Frame with noise (b) Noise removed frame (c) background separation and tracing people

4.2. Computing Centroid Distances of Objects

When compared to the previous overlap tracker, weadditionally modified the tracker to make backward tracing the marked targets on the earlier frame. The centered distances among targets of merged and split states are searched in the current and previous frames. As discussed at subsection 4.1, the long and short radiuses of each ellipse representing an object are averaged to get a dynamic radius by $s = (w+h)/2$ where w and h denote the long and short radiuses of an ellipse, respectively. A target with the merged state in the current frame is identified by analyzing the centered distance of two neighboring ellipses to be smaller than the sum of their dynamic radiuses in the previous frame. On the other hand, to determine split targets, it is to conform to that, thesum of dynamic radiuses owned by two neighboring ellipses being larger than their centered distance in the current frame.

After studying the target states of the present frame, the tracker ultimately allocates tags to individual objects. Likewise, the tag of each target will be noted and referenced for target tracing in the next frame. When calculating the number of targets, states of targets are also considered to help on people counting. Particularly, for a merged target, the number of people is counted as two to provide consistency of people counting.

5. EXPERIMENTAL RESULTS

The proposed scheme is tested on three video clips from the public testing data sets of PET 2009 and 2011. Two clips are shot in a bus stop at different view angles, and the third one at the outdoor scene on the street. Figure 3 shows the background subtracted images from the experimental clips. Results from Fig. 3 expose that the dynamic moving object detection module can effectively overcome the disturbance from light differences to obtain dependent separated objects.

To get the people counting results, Figs. 4, 5 and 6 show the total number of characters counted in each frame by the proposed scheme, Yilmaz*et al's* scheme in and the ground truth. By taking a closer look at Fig. 4, due to characters moving parallel in the video scene from S3-T7-A, good counting results are achieved by the proposed scheme. Particularly, the correct judgment on entering and leaving states yields reliable people counting and tracing on the

500th to 700th frames. On the other hand, Fig. 5 shows the people counting results on the same scene from S7-T6-B, which is shot in a different view angle from S3-T7-A. Although occlusions of target merge and split are frequently occurred in the 250th to 350th frames, the proposed scheme still tracks the characters well. However, because of the characters far away from the camera, the segmented objects are too small to result in incorrect people counting in the 400th to 600th frames due to over merging the characters. In contrast, Yilmaz*et al.'s* scheme using the Kalman filter for people tracing can have better prediction on trajectories of people far from a camera. In Fig. 6, the counting results from S2-L1-V7 are depicted frame by frame. In the video clip of S2-L1-V7, complicated situations of people entering and leaving the video scene often occur. Consequently, many occlusions take place from frequent object merging and splitting. However, the proposed scheme has the benefit of tagging the characters by appropriately using the centered distances of objects in backward tracing. As compared to the results from the Yilmaz*et al.'s scheme*, better performance on tracing people can be achieved by using the proposed scheme.

For evaluating the tracing performance, accuracies of people counting from both schemes are further compared by calculating the detection ratio and Root-Mean Squared error (RMS) in average. The detection ratio and RMS are calculated in each frame by the following formulations,

$$D(i) = \frac{T(i) - C(i)}{T(i)}, \quad (4)$$

$$E_{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(T(i) - C(i))^2}. \quad (5)$$

Here, $D(i)$ denotes the detection ratio in the i-th frame, and $T(i)$ and $C(i)$ represent the numbers of characters from theground truth and the counted results, respectively, and E_{rms} represents the averaged RMS from counting people in video clips. The averaged detection ratios and errors associated with these three video clips are listed in Table 1. From Table 1, the proposed scheme has higher detection ratios and smaller errors on character counting in S3-T7-A and S2- L1-V7 than the Yilmaz*et al.'s* scheme. When comparing the detection ratios in S7-T6-B, the proposed scheme yields little lower accuracy than the Yilmaz*et al.'s* scheme. This is because many small objects are extracted from a long-distance view. The proposed scheme may not tag the characters adequately. However, in average, the proposed scheme outperforms about 10% detection accuracy than the Yilmaz*et al.'s* scheme.

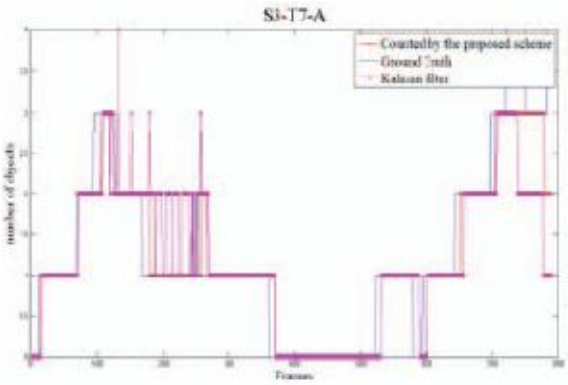

Fig. 4 Number of characters counted in S3-T7-A by the proposedscheme, Yilmaz*et al'*sscheme and the ground truth.

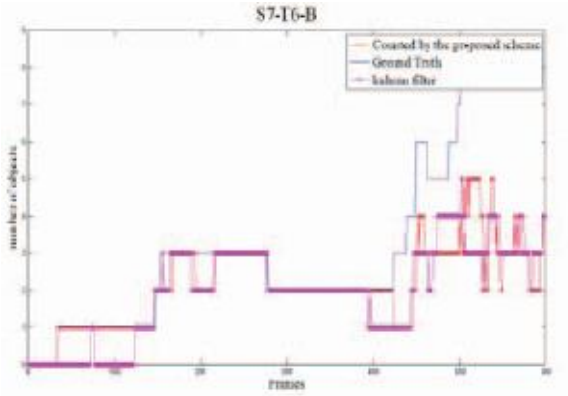

Fig. 5 Number of characters counted in S7-T6-B by the proposedscheme, Yilmaz*et al'*sscheme and the ground truth.

6. CONCLUSION

In this work, we propose a scheme to automatically track, count and recognize people from a stationary video scene in a surveillance system. Foreground regions are segmented by a dynamic background subtraction module for modeling light variations in an environment. Then, objects are recognized as characters with considering the positions and sizes of the obtained foreground regions. For tracking characters, a modified overlap tracker is developed and used to achieve an improvement on tracking characters in occlusion circumstances of target merging and splitting by means of evaluating the centered distances between the objects. Our experimental results demonstrate that the proposed scheme is superior to the conventional work about 10% increase of the detection ratio.

Table 1. Comparisons of the averaged detection ratios and RMSfrom three video clips.

Results / Sequences	Detection ratios		E_{rms}	
	Proposed scheme	Scheme in [6]	Proposed scheme	Scheme in [6]
S3-T7-A	0.84	0.74	0.47	0.66
S7-T6-B	0.65	0.59	2.14	1.19
S2-L1-V7	0.79	0.59	0.93	2.21
Averaged	0.76	0.67	1.14	1.35

REFERENCES

[1] L. Marcenaro etc., "Multiple object tracking under heavy occlusions by using Kalman filters based on shape matching," *Procc. Of IEEE ICIP*, vol. 3, pp. 341-344, 2002.

[2] C.-C. Lien, Y.-L.Huang and C.-C.Han, "People counting using multi-mode multi-target tracking scheme," *Procc ofIEEE IIH-MSP*, pp. 1018-1021, 2009.

[3] S. Khan and M. Shah, "Tracking people in presence of occlusion," *Proc. of Asian Conf. on Computer Vision*, pp.1132-1137, 2000.

[4] I. Haritaoglu, D. Harwood and L. S. Davis, "W4real-time surveillance of people and their activities," *IEEE Trans. onPattern Analysis and Machine Intelligence*, vol. 22, pp. 809- 830, 2000.

[5] Duc Fehr etc., "Counting people in groups," *Proceedings of AVSS*, pp.152~157, 2009.

[6] A.Yilmaz, O. Javed and M. Shah, "Object tracking: a survey," *ACM Computing Surveys*, vol. 38, no. 4, article 13, Dec. 2006.

[7] V. Faber, "Clustering and the continuous k-means algorithm," *Los Alamos Science*, pp. 138-144, 1994.

[8] V. Van der Tuin, *Computer-aided Security Surveillance Design of the Quo Vadis Object Tracker*, Master's thesis, Faculty ofElectrical Engineering, Mathematics and Computer Science,University of Twente, 2009.

[9] PETS 2009 dataset website *http://www.pets2009.net/*.

[10] PETS 2011 dataset website*http://www.pets2011.net/*.