International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:11, No:12, 2017

Motion-Based Detection and Tracking of Multiple Pedestrians

A. Harras, A. Tsuji, K. Terada

***Abstract*—**Tracking of moving people has gained a matter of great importance due to rapid technological advancements in the field of computer vision. The objective of this study is to design a motion based detection and tracking multiple walking pedestrians randomly in different directions. In our proposed method, Gaussian mixture model (GMM) is used to determine moving persons in image sequences. It reacts to changes that take place in the scene like different illumination; moving objects start and stop often, etc. Background noise in the scene is eliminated through applying morphological operations and the motions of tracked people which is determined by using the Kalman filter. The Kalman filter is applied to predict the tracked location in each frame and to determine the likelihood of each detection. We used a benchmark data set for the evaluation based on a side wall stationary camera. The actual scenes from the data set are taken on a street including up to eight people in front of the camera in different two scenes, the duration is 53 and 35 seconds, respectively. In the case of walking pedestrians in close proximity, the proposed method has achieved the detection ratio of 87%, and the tracking ratio is 77 % successfully. When they are deferred from each other, the detection ratio is increased to 90% and the tracking ratio is also increased to 79%.

***Keywords*—**Automatic detection, tracking, pedestrians.

I. Introduction

COMPUTER vision is a multifaceted field which deals with how we could use computers to highly understand digital images or videos [1]. That tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions [2]. To achieve these tasks, we capture objects' images and focus on obtaining their attributes, and then apply filters and employ different techniques for image processing to detect and track them in their environment. Detecting and counting moving objects are very important area in the field of computer vision. It is used in many applications to count large numbers of objects that move around with all directions in buildings, on roads, and railway stations, etc. The objective of our study is to detect and track moving multiple pedestrians.

II. Literature Review

Attempts to automatically detect and count of pedestrians using motion based tracking have been proposed by Terada et al. They implemented a method where a stereo camera hung from the ceiling of a gate and the optical axis of the camera is set up so that the passing people can be observed from just overhead. In this system, if there are a crowd of people in the gate, then the data of the passing people are not overlapped on the obtained images. In addition, by using the stereo camera, the human region and road region on the obtained images are able to be segmented accurately. They had some experimental results obtained by using a simple experimental system to verify the effectiveness of the proposed method [3]. Elmarhomy proposed a method to automatically count passersby by recording images using virtual measurement lines using an image sequence obtained from a USB camera placed in a side-view position. His approach included many challenges: (1) two passersby walking in close proximity to each other, at the same time, and in the same direction; (2) two passersby moving simultaneously in opposite directions; (3) a passerby moving in a line followed by another, or more, in quick succession. In his study, the human regions treated using the passerby segmentation process are based on a lookup table and labeling where the number of people passed by is indicated by the shape of the human region as appeared in space- time images. The system uses different color spaces to perform the template matching, and automatically selects the optimal matching and accurately counts passersby. Human images are extracted and tracked using background subtraction and time-space images. Moreover, a relation between passerby speed and the human-pixel area is used to distinguish one or two passersby [4]. Irani and Anandan discussed on the previous approaches to the problem of detecting moving objects that can be broadly divided into two classes: 2D algorithms when the scene is convergent to a flat surface in the case where the camera is only rotating and zooming, and 3D algorithms when the scene has significant variations and the camera is moving. They proposed a unified approach to handling moving object detection in both 2D and 3D scenes with a strategy to gracefully bridge the gap between those two extremes. Their approach is based on a stratification of the moving object detection into scenarios which gradually increase in their complexity. They presented a set of techniques that match the above stratification. These techniques progressively increase in their complexity, ranging from 2D techniques to more complex 3D techniques. Moreover, the computations required for the solution to the problem at one complexity level become the initial processing step for the solution at the next complexity level. They illustrate these techniques using examples from real image sequences [5].

A. Harras, A. Tsuji, and K. Terada are with the Faculty of Science and Technology, Graduate School of Advanced Technology and Science, Tokushima University, Japan (e-mail: rahmyharras@gmail.com, a-tsuji@is.tokushima-u.ac.jp, terada@is.tokushima-u.ac.jp).

Fig. 1 (a) and (b) Two different shots of an unorganized scene for a crowded area

Chen et al. argued that traditional methods; such as frame difference and optical flow may not be able to deal with the problem encountered during the process of moving object detection in an intelligent visual surveillance system, where a scenario with complex background is sure to appear. In such scenarios, they used a modified algorithm to do the background modeling work. They used edge detection to get an edge difference image to enhance the ability of resistance illumination variation then multi-block temporal-analyzing Local Binary Pattern (LBP) algorithm was applied to do the segmentation. In the end, a connected component is used to locate the object. They also produced a hardware platform, the core of which consists of the DSP and FPGA platforms and the high-precision intelligent holder [6]. There are more challenging attempts in the regarding of detecting and tracking moving objects, where objects and the detecting camera are both moving. Thompson and Pong, in their study to examine moving object detection based on optical flow, concluded that in realistic situations, detection using visual information alone is quite difficult, particularly when the camera may also be moving. They highlighted that when additional information about camera motion and/or scene structure is available, it could greatly simplify the problem. They presented two general classes of techniques; the first is based upon the motion epipolar constraint—translational motion produces a flow field radially expanding from a "focus of expansion" (FOE). Epipolar methods depend on knowing at least partial information about camera translation and/or rotation. The second is based on comparison of observed optical flow with other information about depth, for example from stereo vision [7]. In the same line, DeGol and Nam have proposed an approach to motion detection in scenes captured from a camera onboard an aerial vehicle. They argued that it is a challenge to detect slow motion in an aerial video because it is difficult to differentiate object motion from camera motion. An unsupervised learning approach has been adopted in which they require a grouping step to define slow object motion. The grouping is done by building a graph of edges connecting dense feature key points. Then, they use camera motion constraints over a window of adjacent frames to compute a weight for each edge and automatically prune away dissimilar edges. This leaves groupings of similarly moving feature points in the space, which have been clustered and differentiated as moving objects and background. They provide qualitative and quantitative results that demonstrate the effectiveness of their detection approach [8].

III. Proposed Method

The proposed method has two main processes. The first process is to detect moving pedestrians. The second process is to track the passerby in all frames and count them. These processes are performed through three preprocessing operations. The process starts with the object detection and tracking, then noise elimination and motion detection are applied to the detected area. The three operations, object detection and tracking, noise elimination, and motion tracking are discussed in detail.

A. Objects Detection and Tracking

For detecting location of a passerby, background subtraction is applied in order to detect moving pedestrians within video frames. We use a GMM, which is vital to determine moving objects from a sequence of video frames. It is based on a probabilistic model that assumes that all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters [9]. By using the GMM, frame pixels are removed from the required video to obtain the desired results. It reacts to various changes like illumination, starting and stopping of moving objects

B. Noise Elimination

There exists a noise in the acquired frames which are required to be eliminated. Morphological operations are applied to eliminate the noise. These operations rely on the relative ordering of pixel values, not on their numerical values, and therefore, are especially suited to the processing of binary images when applied to grey-scale images which their light transfer functions are unknown. Their absolute pixel values are no or minor interest.

Morphological techniques probe the image with a small shape or template called a structuring element. The structuring element is positioned at all possible locations in the image and compared with the corresponding neighborhood of pixels. Some operations test whether the element "fits" within the neighborhood while others test whether it "hits" or intersects the neighborhood [10].

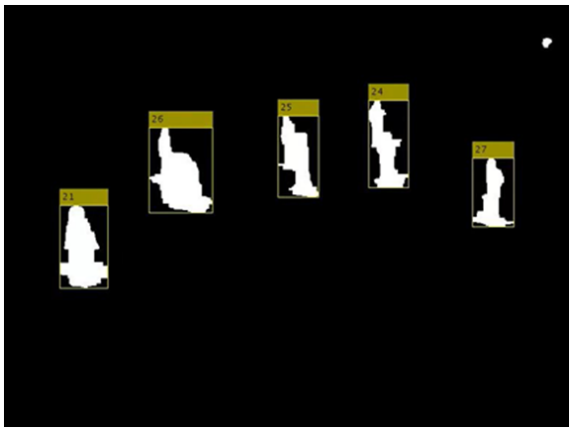

Fig. 2 Binary image after applying the noise elimination

C. Motion Detection

The motion of each track is estimated by a Kalman filter which is used to predict the tracked location in each frame and determine the likelihood of the detection being assigned to each track, i.e. the assigned tracks update is based on their corresponding detections. Some tracks may remain unassigned depending on the situation. Each of those tracks keeps a count of the number of consecutive frames which it remained unassigned. If the count exceeds a specified threshold, it is assumed that the passerby left the field of view and thus deleted.

IV. Experiments

We used the MOT Challenge data set which is a benchmark data set to evaluate our proposed method. The experimental environment is as the following.

1. MOT Challenge data set [11]
2. Acquired video sequences from the data set.
3. A GUI based program implementing the proposed method to display and analyze information

In the experiments, the acquired video clips were shot on a street. Up to eight people evolve in front of the camera for around 35 seconds in the 1st scene and 53 seconds in the second one. The frame rate is 30 fps. A number of frames has been extracted from the video clips and saved to the sequence of image files. They have been sent to the PC in order to be processed and analyzed. We used MATLAB to implement our proposed method. In the experiments, the pedestrians were able to move freely in a fairly random manner and would not necessarily perform regular walking within the detection area. In Addition, their motion was unimpeded by obstacles or by the crowding effects of other pedestrians. Furthermore, we do not care about the effect of weather on lighting conditions, or on causing pedestrians to hurry when the weather is bad.

V. Experimental Results

Detecting and tracking objects undergoes the following consecutive processes. 1. Read the file to load the image into the memory. 2. Conversion to the bitmap file which objects separated from the background. 3. Fill the holes to recover inner pixels that are part of the objects and does not recognize. 4. Edge smoothing for the outline of objects. 5. Filtering by size to remove small chips. 6. Labeling and counting blocks of the pixels. These processes implemented in the MATLAB functions, which is able to access through the command prompt in the workspace. The video sequences show the results of detection and tracking in two experiments. Examples of the experiments scenes are shown in Figs. 1 and 2. The two scenes we used for the evaluation is as the following.

1. In the first experiment, we evaluated our algorithm on the video sequence MOT16-9 from MOTChallenge data set. It shows a pedestrian street scene filmed from a low angle.
2. In the second experiment, we evaluated our algorithm on the video sequence PETS09-S2L1 from MOTChallenge data set. It shows an unorganized scene of pedestrians walking in different directions

One of the binary image corresponds to the second experiment is shown in Fig. 2. Table I shows the experimental results for the evaluation of detection and tracking ratio in the scene 1 and 2. The results showed that the detection ratio is 87% and the tracking is 77 % in scene 1 while the detection ratio is 90% and the tracking is 79% correctly in scene 2. The success cases are those in which pedestrians are moving far from each other and the scene does not contain any reflections or shadows as shown in Figs. 3 (a) and (b). On the contrary, failure cases happened when pedestrians are moving close to each other as shown in Fig. 3 (c). As the result in scene 1, the system detects 4 people successfully in shot (a) and detects 4 people as one object because they are near from each other in shot (b). In scene 2 the system detect 7 people successfully in shot (a) and (b), and detect two people as one object or doesn't detect one person in shot (c). It also sometimes miscounts the number of pedestrians because it detects some shapes as objects while in the fact that they are reflections and shadows.

TABLE I
Experimental Results of Detection and Tracking Ratio

	No. of frames	Detection		Tracking	
		Success	Failed	Success	Failed
Scene 1	1845	**87%**	13%	**77%**	23%
Scene 2	1454	**90%**	10%	**79%**	21%

VI. Conclusion

We presented the implementation for detecting moving pedestrians and their motion-based tracking in an image sequence. The results showed that the system works efficiently with no fatal error, where in case of pedestrians walk in close proximity, the proposed system has achieved the detection ratio of 87% and the tracking ratio has achieved 77 % correctly. When they deferred from each other, the detection ratio has increased to 90% and the tracking ratio to 79%. In the future work, we improve the system by applying other techniques to enhance multiple pedestrians tracking and define their motion direction as well. The system is also required to be evaluated in irregular conditions such as fog, heavy rain, day and night time.

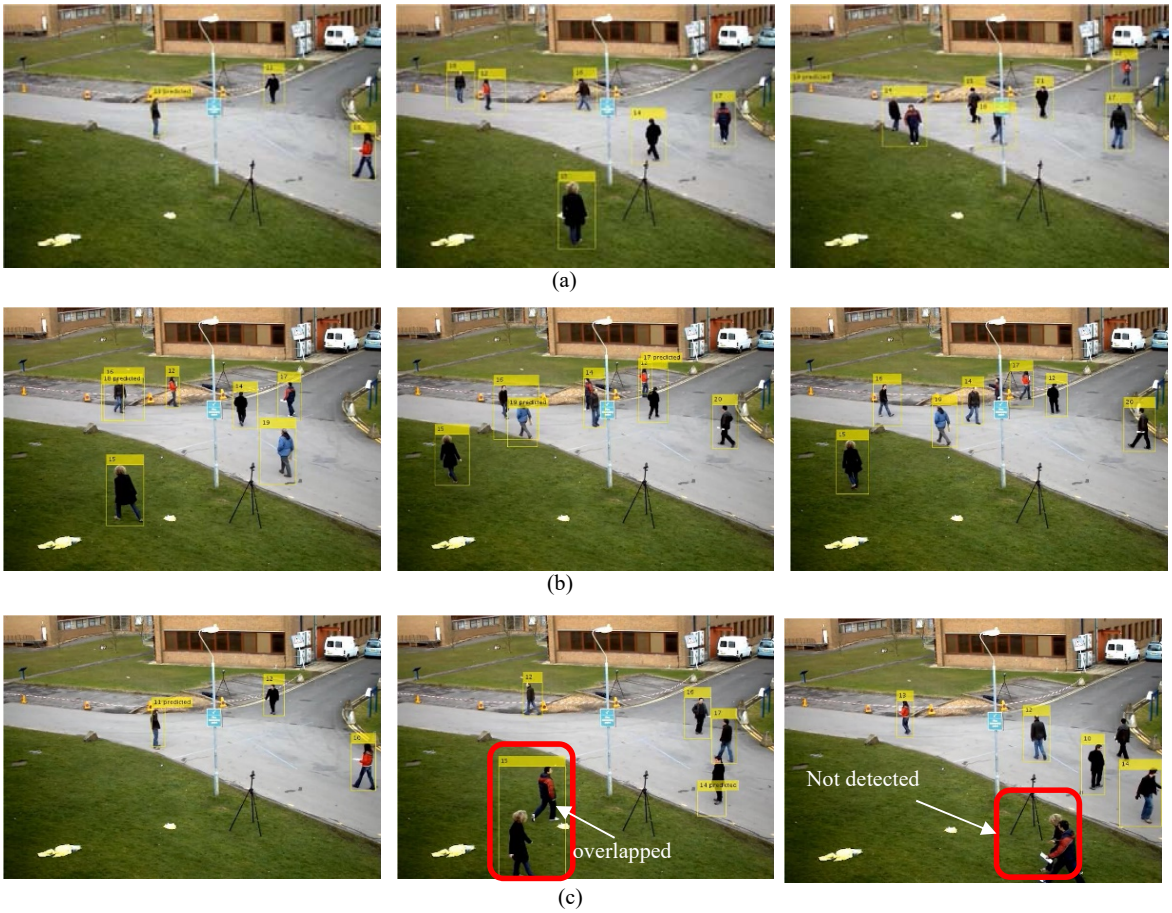

Fig. 3 Experimental results of the Scene2 on the unorganized scene of pedestrians walking in different directions. (a) and (b) Correct detection and tracking. (c) Failure cases happened when pedestrians are moving close to each other

REFERENCES

[1] Dana H. Ballard; Christopher M. Brown (1982). Computer Vision. Prentice Hall. ISBN 0-13-165316-4.
[2] Reinhard Klette (2014). Concise Computer Vision. Springer. ISBN 978-1-4471-6320-6
[3] K. Terada, D. Yoshida, S. Oe and J. Yamaguchi." A counting method of the number of passing people using a stereo camera", IEEE Proc. of Industrial Electronics Conf., Vol. 3, pp.1318-1323, 1999.", International conference on image processing, pp. 338-342,1999
[4] Elmarhomy, A., A Method for Real Time Counting Passersby utilizing Space-time Imagery, 2014.
[5] Irani, M. and Anandan, P., "A Unified Approach to Moving Object Detection in 2D and 3D Scenes," IEEE Trans Pattern Analysis and Machine Intelligence, Vol 20(6), June 1998, pp. 577-589.
[6] S. Chen, T. Xu, D. Li, J. Zhang, and S. Jiang. Moving object detection using scanning camera on a high-precision intelligent holder. MDPI Sensors 2016, October 2016.
[7] Thompson, W. B. & Pong, TC. Int J Comput Vision (1990) 4: 39. doi:10.1007/BF00137442.
[8] Oseph DeGol; Myra Nam, A clustering approach for detectingmoving objects captured by a moving aerial camera, IEEE Xplore 2014, ISBN: 978-1-4799-2893-4.
[9] Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.
[10] Nick Efford. Digital Image Processing: A Practical Introduction Using JavaTM. Pearson Education, 2000.
[11] MOT Challenge 2015: Towards a Benchmark for Multi-Target Tracking. Laura Leal-Taixé, Anton Milan, Ian Reid, Stefan Roth, Konrad Schindler.