An Improved Multi-Target Tracking Algorithm for Pedestrian Counting

Tian Xia, Hong Fan¹, Suping Yu, Liping Zhang and Jiajing Wen

College of Information Science and technology, Donghua University, Information and communication Engineering, Shanghai, China
Email: xiatianmy@163.com, ccfanhong@163.com

Abstract. In view of the traditional Camshift algorithm is easy to lose the target in the case of serious colour interference and occlusion, and the traditional method of people counting is inefficient and has low accuracy. This paper uses three-frame difference method and Gaussian Mixture Model to detect the pedestrians, and then an improved multi-target tracking algorithm is used to implement automatic tracking of multiple pedestrians and draw a motion trajectory. Based on this, the motion direction is judged and pedestrians are counted. In the tracking process, the problem of colour interference and severe occlusion is solved according to the trajectory prediction. The experimental results show that the proposed algorithm not only improves the tracking accuracy, but also realizes the automatic pedestrian counting. It also greatly improves the efficiency of the number of people’s statistics and can accurately count the number of people in the presence of overlap and occlusion.

1. Introduction

With the popularity of video surveillance systems, people statistics have been widely applied to crowd monitoring and intelligent transportation [1]. The number of entrances and exits to shopping malls and supermarkets, as well as the Public transportation facilities, can be used not only to understand the location of shopping malls and supermarkets, it is also reasonable to carry out crowd scheduling [2]. When there are too many people, they can issue warnings to avoid danger and provide better security for our lives. So how to quickly track pedestrians in surveillance videos and perform demographic statistics has also become the focus of research in the field of image processing [3]. Therefore, this paper combines some algorithms to improve on the basis of the existing research.

2. Pedestrian Counting

This paper uses three-frame difference method and Gaussian Mixture Model (GMM) to detect the pedestrians, and then uses the Camshift algorithm to achieve automatic tracking of multiple pedestrians. Then, the tracked pedestrians are objectized one by one, and the object is matched continuously using the minimum centroid distance method, the centroid position information of the moving object is stored respectively, so that the trajectory drawing and the number of pedestrians in each direction of motion are realized.

2.1. Pedestrian Detection

2.1.1. Three-frame Difference Method Combined with Gaussian Mixture Model. The three-frame difference method uses pixel-based time difference between adjacent three frames of the image sequence to extract motion regions in the image by blocking [4]. GMM is a background representation method based on statistical information of pixel samples. The statistical information of a large number
of sample values such as the probability density of pixels in a relatively long time is used to represent the background. Use the statistical difference (such as 3σ principle) to judge the target pixel [5].

In this paper, we use the three-frame difference method and the GMM combined algorithm to process the image sequence [6]. After respectively extracting the moving objects, median filtering and morphological processing were respectively performed on the moving objects extracted by the two algorithms to remove unnecessary noise. These threshold images respectively obtained are then logically ORed to obtain the target image. This combination algorithm can effectively suppress the "hole" and "Ghosting" phenomenon and avoid the disturbance of the environment change to the mixed Gaussian model and obtain a relatively complete moving target.

2.2. Pedestrian Tracking

2.2.1. The CamShift Tracking Algorithm. The Camshift algorithm is a tracking algorithm based on the colour histogram as the target mode. Because the RGB colour space has strong external dependence on illumination, the Camshift algorithm converts the image from the RGB colour space to the HSV colour space. The real-time performance of the algorithm is better, and the target tracking is mainly achieved through the colour information of the moving target in the video. The algorithm first selects an initial search space to determine the size and position of the window, because the approximate location of the moving target is not known, we need to calculate the value of the H channel of each pixel of the window to find the exact location of the target. Therefore, some unnecessary calculations are greatly increased, and the algorithm operation time is increased. Moreover, in the case of complex interference such as colour interference and blockage, the target will be lost [7].

2.2.2. The Improved Tracking Algorithm Based on Trajectory Prediction. The traditional Camshift algorithm can not accurately track pedestrians during the tracking process when the pedestrian is obstructed or the colour interference is serious [7]. To overcome this problem effectively, an improved tracking algorithm based on trajectory prediction is proposed.

The object of the newly emerged pedestrian is objectified, and then the pedestrians detected in each frame of the image and the previous pedestrian object are matched one by one using the minimum centroid distance method to determine whether the size of the search window between the current pedestrian target and the matched pedestrian target has changed greatly.

$$d_1 = \sqrt{[x_{i}(1) - x_{i-1}(j)]^2 + [y_{i}(1) - y_{i-1}(j)]^2} < T, O_i = O_{j} = 1, 2, 3, \ldots, m$$

$$d_2 = \sqrt{[x_{i}(2) - x_{i-1}(j)]^2 + [y_{i}(2) - y_{i-1}(j)]^2} < T, O_i = O_{j} = 1, 2, 3, \ldots, m$$

$$\vdots$$

$$d_k = \sqrt{[x_{i}(k) - x_{i-1}(j)]^2 + [y_{i}(k) - y_{i-1}(j)]^2} < T, O_i = O_{j} = 1, 2, 3, \ldots, m, k = 1, 2, 3, \ldots, n$$

$d_k$ is the minimum centroid distance, $x_{i}(k)$ denotes the abscissa of the k-th centroid position of the i-th frame ($i>1$), and $x_{i-1}(j)$ is the abscissa of the j-th centroid position of the i-1th frame, $y_{i}(k)$ is the ordinate of the k-th centroid position of the i-th frame, and $y_{i-1}(j)$ is the ordinate of the jth centroid position of the i-1th frame. If the minimum centroid distance is smaller than the threshold $T$, it is considered to match the same target, in which case $O_k = O_j$.

$$\text{Current}_{\text{area}} = \text{Matched}_{\text{area}} \times \beta$$

Current$_{\text{area}}$ indicates the current pedestrian target search window, and Matched$_{\text{area}}$ indicates the matched pedestrian target search window. When $0.6 < \beta < 1.5$, it is considered that there is no partial occlusion, otherwise, it is considered that partial occlusion exists. We can use the trajectory
information stored in the first few frames of the pedestrian target object as position prediction to replace the currently detected pedestrian target position and size, then store the information in the pedestrian target object. If there is no major change, the actual position information is stored in the pedestrian target object.

When the pedestrian overlaps and the colour interference is serious in the tracking process, in order to reduce the calculation time, each three-frame image sequence is used to cycle to compare whether the number of centroids stored before and after the three frames is the same. If it is not the same, then determine whether the pedestrian target disappears in the edge region according to the location of the last centroid, if it disappears to the edge, the pedestrian target is considered to disappear and the moving target object is destroyed along with it. On the contrary, if the target is temporarily disappeared, the position information according to the trajectory is predicted and the position information of the target object is continuously updated. In this way, the goal is to solve the problem of large-area occlusion or severe colour interference. Algorithm flow chart shown in Figure 1.

![Diagram](image_url)

**Figure 1.** Improved Camshift algorithm flow chart with occlusion and colour interference
2.3. Pedestrian Tracking and Counting Function
Pedestrians detected in the first frame of the image are objectified one by one, and their centroid position information is stored, and the number of people is counted at this time. Then the pedestrian detected by the next frame image and the previous pedestrian object are matched one by one using the minimum centroid method to determine whether they are the same pedestrian. If they are the same pedestrian, the centroid position information is recorded, otherwise it is determined that there are new pedestrians. Similarly, new pedestrians are objectified to store location information and count the number of people. When judging the direction of movement, the judgments of occlusion and disappearance (As already described in detail above) are performed. If non-occlusion and non-colour interference are observed, compare the x value of the first point and the last point in ld (centroid position information) to determine whether the direction of movement is left or right, and count the number of pedestrians in each direction of movement, complete the statistics and destroy the objects. After the processing of each frame of the image is completed, the trajectory is drawn on the position information vector set recorded by the existing pedestrian object to form a pedestrian trajectory. If the pedestrian disappears, the trajectory disappears as the object is destroyed. The number of pedestrians in the current area is the number of current pedestrian target objects, which is denoted by mo.size ( ).

The algorithm flow chart to draw the trajectory is shown in Figure 2.

Figure 2. Flow chart of pedestrian trajectory algorithm
3. Experimental Results and Analysis

In order to verify the validity of the algorithm, this paper uses VS2013 combined with OpenCV function library to write programs on a computer with Intel(R) Celeron(R), 1.5GHz CPU, 4GB RAM and windows10 operating system.

3.1. The Experiment Results of Pedestrian Detection

In order to verify that the algorithm combined with the three-frame difference method and the GMM will make the detection result better, the video was taken in an experiment in a simple background area of the school. The image size of each frame is 480×360. Figure 3(b) shows the result of pedestrian detection using three-frame difference. It can be seen that there is a "hole" phenomenon. After morphological processing is performed, as shown in Figure 3(c), noise is reduced but "hole" still existed. Figure 3(d) shows the results of pedestrian detection using GMM. It can be seen that there are many noises. At this time, morphological processing is performed to eliminate most of the noise. The result is shown in Figure 3(e). Figure 3(f) shows the result of morphological processing after logical OR of the two results. The result of the two morphological treatments is taken as "or". It is clear from the results that the combination of the two methods not only solves the "hole" phenomenon in which the three-frame difference appears, but also improves the accuracy of the test results due to changes in light and other factors.

3.2. The Camshift Algorithm

Figure 4(a) shows the result of pedestrian tracking and counting when there are two pedestrians in the 84th frame area. Figures 4(b) and (c) show that in the 103th and the 111th frames, when the pedestrians overlap, the traditional Camshift algorithm will lose the target, which will result in counting errors.
3.3. The Improved Tracking Algorithm Based on Trajectory Prediction

Figure 5(b) shows that the improved algorithm can accurately track pedestrians when the pedestrians overlap in the 103rd frame. From Figure 5(c) it can be seen that the overlapping pedestrians can still accurately track and display the trajectory after separation.

3.4. Results of the People Statistics

Figure 6. Results of demographic statistics
In order to verify the accuracy of the demographic statistics, video was taken in front of the school administration building. A total of six people passed through the video, and their movement trajectory and direction of movement were different. In Figure 6(a), 4 pedestrians appear in the 30th frame area, and Number is displayed as 4.

Figure 6(b) shows the experimental results of the 79th frame. The pedestrians detected by the image and the previous pedestrian object are matched one by one using the minimum centroid method, and one pedestrian is added. There are five pedestrians. Number is now displayed as 5.

Figure 6(c) shows a reduction of 1 person in the 99th frame area and the remaining 4 persons in the current area. The total number of people passing through the area is 5, so Number is displayed as 5 and Surplus is displayed as 4. The reduced pedestrian movement track is to the right, so To Right is displayed as 1.

Figure 6(e) shows the remaining 1 pedestrian after the reduction of 4 pedestrians in the 245th frame region. The total number of people passing through this region is 5, so Number is displayed as 5 and Surplus is displayed as 1. One reduced pedestrian trajectory is to the right, and three pedestrians to the left, so To Right is displayed as 1, and To Left is displayed as 3.

Figure 6(f) shows that the total number of 267 frames passing through this area is 6, and the current number is 1, of which 4 pedestrians have their left trajectory and 1 pedestrian has their right trajectory. Number is displayed as 6, Surplus is displayed as 1, To Right is displayed as 1, and To Left is displayed as 4.

4. Conclusion

This article makes use of the object-oriented features of C++ language to objectified the pedestrian object so as to store the corresponding information separately, and the objects do not interfere with each other. The above-mentioned thinking algorithm solves problems such as target occlusion and serious colour interference that cannot be solved in the Camshift tracking process. At the same time, multi-target real-time tracking, number of pedestrian targets, and automatic statistics of movement direction are realized. Of course, this article uses background modeling to realize the idea. Therefore, it is based on the premise of background static to realize functions, and how to improve the applicability and stability of the algorithm in dynamic environment and complex background still needs further study.

5. References

[1] Li H, Zhang E H and Duan J H Population statistics method based on PCA and multivariate statistical regression 2014 Computer Engineering and Applications 50 206 - 9.
[2] Lin F X 2009 Statistics of People in Intelligent Video Surveillance (Hebei University of Technology) pp 1-2.
[3] Xiong W and Zhu Q Y A video-based bus passenger flow counting method 2008 Modern Electronic Technology 95-98.
[4] Lu Z P, Kong D F, Li X L and Wang J W Combination of background difference and three-frame differential motion target detection algorithm 2013 Computer Measurement and Control 21 3315-18.
[5] Tian F Research on video image person counting algorithm based on Gaussian Mixture Model 2017 Computer Applications and Software 34 187-192.
[6] Zhou T X and Zhu M Moving target detection of the video images 2017 Chinese Journal of Liquid Crystals and Displays 32 40 - 47
[7] Wu L J, Kuang L, Deng Q L and Liu H H An improved multi-target tracking algorithm based on combining Camshift and Kalman Filtering 2010 Modern Scientific Instruments.
[8] Cai N, Chen S W, Guo W T and Pan Q Motion target detection based on Gaussian Mixture Model and Wavelet Transform 2011 Chinese Journal of Image and Graphics 162-7.
[9] Lou W J 2012 OpenCV-based moving object direction recognition (Hubei University)
[10] Liu X, Liu H, Qiang Z P and Geng X T Adaptive background modeling based on mixture Gaussian model and frame subtraction. 2008 Journal of Image and Graphics 4 729-735.
[11] Stauffer C and Grimson W. Adaptive background mixture models for real time tracking. 1999 *In Computer Vision and Pattern Recognition* **43** 198-208.