Crowd Target Positioning under Multiple Cameras Based on Block Correspondence

Qiuyu Zhu[1], Sai Yuan[1], Bo Chen[1], Guowei Wang [1], Jianzhong Xu[1], and Lijun Zhang[2]

[1] School of Communication & Information Engineering, Shanghai University, Shanghai, China
Zhuqiuyu@staff.shu.edu.cn, yuansai200888@qq.com
[2] Shanghai Advanced Research Institute, Chinese Academy of Sciences, China
zhanglj@sari.ac.cn

Abstract. In the research of crowd analysis in a multi-camera environment, the key problem is how to get target correspondence between cameras. Two main popular methods are epipolar geometric constraint and homography matrix constraint. For large view-angle and wide baseline, these two methods exist obvious disadvantages and have a low performance. The paper utilizes a new correspondence algorithm based-on the constraint of line-of-sight for the crowd positioning. Since the target area is discrete, the paper proposes to use blocking policy: dividing the target regions into blocks with certain size. The approach may provide appropriate redundancy information for each object and decrease the risk of objects missing which is caused by large view-angle and wide baseline between different perspective images. The experimental results show that the method has a high accuracy and a lower computational complexity.

Keywords: multiple cameras, constraint of line-of-sight, target positioning, blocks correspondence.

1 Introduction

With rapid development of economic, there are more and more skyscraper, underground constructions, and large commercial entertainment. The requirements for the function of intelligent video analysis platform are improving. For researchers, it is one of the principal subjects of concern about how to effectively detect crowd and to predict the crowd behavior. Current researches on intelligent video surveillance technology mainly focus on the fusion of multiple cameras, camera calibration, target detection, target positioning, target tracking, activity recognition etc.

The paper accomplishes crowd detection and positioning in the condition with multi-camera collaborative environments. It is a fundamental work for various advanced processing such as behavior analysis, behavior recognition, as well as advanced video processing and application. Under multi-camera setup, the key problem is how to realize targets correspondence between cameras. Two main popular methods are epipolar geometric constraint and homography matrix constraint. The former needs to segment foreground region and extract key-points of the targets firstly, and then, depending on epipolar geometric constraint, the correspondent object and

N. Streitz and P. Markopoulos (Eds.): DAPI 2014, LNCS 8530, pp. 509–518, 2014.

point in the other image can be found. But, in the actual dense crowd of video scene, it is too hard to segment single object from crowd accurately. Meanwhile, larger angle of view between cameras leads to larger difference between images and the matching task based on appearance would also be difficult to achieve. So, object matching method based on epipolar geometric constraint cannot be effectively applied to the crowd situation.

For the latter, it needs the matching points are projected from reality points belong to one plane which usually is ground plane and thus we ought to find feet in the image. Owing to high occlusion in crowds, it is impossible to find each foot. Afterwards, researchers utilize homography matrix constraint among multi-plane in different height [2]. This method needs accurately foreground segmentation and the familiar appearance of multi-view object in large baseline. These requirements are difficult to meet in dense crowd.

As Figure 3.1, there are many targets in the scene. It's hard to get a single individual in the crowd. Considering the cameras are generally set up higher than human, even in the crowded situation, the human head is still visible, so the paper regards head region as interesting region. The paper utilizes line-of-sight constraint between cameras to realize the positioning of objet. After detecting the interesting region, target area is discrete, the paper proposes to use of blocking policy: dividing the target region into blocks with certain size. The approach may provide appropriate redundancy information for each object and decrease the risk of objects missing which is caused from large view-angle and long baseline between different perspective images.

2 Related Work

In order to reduce the influence of occlusion in crowd, many researchers have analyzed the crowd under multi-camera. Eshel and Moses [1] associate several views data to detect head in crowd scene and obtain height information of targets from segmented head. The correspondence of head takes use of plane's homography. The paper [2] put forward to use homography existing among different height planes to detect and trace object. Based on their former work, W. Gee [3] et al proposed a statistics model of crowd structure which could manipulate the subordination in crowd and is not discrete for the space position of the object. Castle [4] et al proposed to utilize triangulation method that based on SIFT key-points extracted from key frame to recognize, reconstruct, and localize. They also extend to raise frame rate through FAST key-points and parallel tracking technology. Mazzonand Cavallaro [5] proposed to conduct re-recognize using SFM model for the object that are occluded in the process of trace. Huadong Ma [6] et al, proposed an algorithm that based on PMHL detector and multi-view fusion. This algorithm could detect static object rapidly and robustly. And authors make use of synergy between multiple cameras to improve the detection rate.

3 Target Positioning Based on Block Correspondence

Multi-camera positioning of crowd scene shows in Figure 3.1. If there are too many objects to be localized in the scene, occlusion would be worse. It is too hard to get complete single individual. As the cameras are generally set up higher than human, even in the crowded situation, the entire head contour is still visible. So we consider head as the whole body to detect as well as further vision processing.

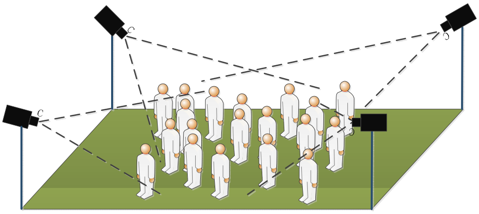


Fig. 1. Crowd scene under multi-camera

3.1 Moving Detection

Frame difference and background difference are two main methods suitable for motion detection in the static background scene. The former is simple and clear, and possess a stable robustness. However, there are holes in the moving object after detecting process. The latter could get entire object contour, but it is a hard work to do. Common methods are averaging method, Gaussian background modeling method, and so on. The paper also utilizes Gaussian background modeling method. In this step, the mathematical average of six images at different time is the background image. In the later step of processing, we join the morphology filter for binary image in order to get a full target area. Due to the influence of frame difference and noise, there may be non-moving region in foreground of the binary image. So we process this image with morphology filter to get better moving region.

After getting moving foreground, the paper detects head region by detecting skin color region and hair area. In the head region, the obvious and tractable regions are mainly skin color region and hair area. So, under the premise of no target missing, the paper put skin region and hair region as the target areas in order to improve the correct rate of head detection in dense crowd. Usually, target areas are some merged connected regions came from different individuals. Therefore, we need a target corresponding algorithm to separate each individual.

3.2 Object Matching Based On Line-Of-Sight Distance

Object matching based on line-of-sight distance is shown as Fig. 2. Here, the line-of-sight is the space line through camera optical center and object point in the world coordinate, for example L1, L2 in Figure 3.2. C1 and C2 stand for cameras, P1, stands for targets in the scene.

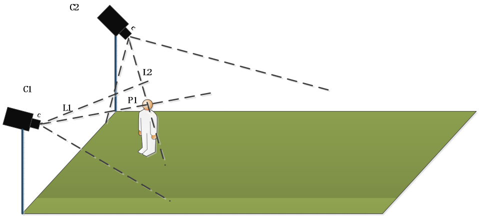


Fig. 2. Line-of-sight constraint

As we know, the pixel (u, v) in the image coordinate and its counterpart in the world coordinate (Xw, Yw, Zw) has a relationship as follows

$$Z\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K\begin{bmatrix} R & T \end{bmatrix}\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix}\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{1}$$

where:

K-is a 3×3 Matrix of camera intrinsic parameters;

R-is a 3×3 unit rotation matrix;

T-is a3×1 translation vector, together with rotation matrix they are called external parameters;

M-is a 3×4 camera matrix.

In the case of the known camera matrix P, since the matrix P is a 3 × 4 matrix, then the linear algebraic theorems about linear equations can know right of formula (3.1), that is a known spatial points coordinates, the equation is well posed equations, then you can easily access a variety of corresponding points by solving linear equations method of pixel coordinates; But in the case knowing pixel coordinates, since P is not invertible matrix, the equations are underdetermined equations, then only get a simplified equations by solving the equations, the geometric concepts of equations at this time is represented a spatial line which through optical center of the camera, the image coordinates of a spatial point, linear spatial points. If the spatial point is visible to the two cameras, then two lines are certain. Two intersect lines determine the location of the target spatial position. So the method is able to obtain the spatial location of the target, to achieve the purpose of positioning.

If pixel p1 in image I1 and pixel p2 in image I2 are matched, L1 and L2 must intersect at point P.

$$\begin{cases} L1: \dfrac{X - x2}{x1 - x2} = \dfrac{Y - y2}{y1 - y2} = \dfrac{Z - z2}{z1 - z2} = m \\ L2: \dfrac{X - x4}{x3 - x4} = \dfrac{Y - y4}{y3 - y4} = \dfrac{Z - z4}{z3 - z4} = n \end{cases} \tag{2}$$

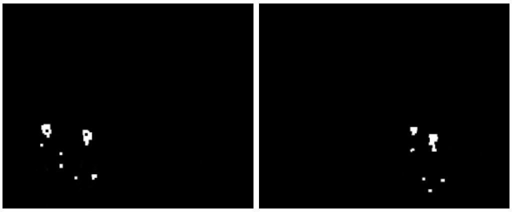
For correctly matched targets, we take the center point, and calculate the corresponding equation of the line-of-sight.

With common perpendicular line and pedals solution, we can get a series points of target P1 $\{P_{11}\ \ P_{12}\ \ ...\ \ P_{1n}\ \ ,n=2*camerasnum\}$. Due to the center of the corresponding blocks are not correspond to the same target, line-of-light equations do not always intersect. In most cases will be non-coplanar lines, but the common perpendicular of the line, which is the shortest distance between non-coplanar lines, will be included in the destination point for the sphere, spherical region within a certain distance radius.

3.3 Blocks Correspondence

Actual detected skin and hair region mostly is not completed, and broke, and discrete, thus the above positioning algorithm should do some modifications: first, each target area breaks into blocks, and in the process of line-of-sight correspondence we use the center of blocks take place of the whole block; addition, an image may contain several targets, the paper utilizes relaxation labeling algorithm to calculate the optimal solution corresponding multi-block problem.

An example of blocking is shown as Figure 3, where image size is 768*576 and the block size is 20*20. The object correspondence will be based on all blocks.



a. blocks of I1 image b. blocks of I2 image

Fig. 3. Results of blocking

If the image I1 corresponding to the camera C1 has m1 blocks and image I2 corresponding to camera C2 has m2blocks. Thus, for the m1 blocks, we use following matrix:

$$X=\begin{bmatrix} x_{1,1} & \dots & x_{1,N} \\ \dots & & \dots \\ x_{m1,1} & \dots & x_{m1,N} \end{bmatrix} \tag{3}$$

Each row of the above matrix represents the characteristic of each correspondence block. In detail, $\left[x_{i,1}\ \ ...\ \ x_{i,N-2}\right]$ represents the parameters of line equation of sight through the block, and $\left[x_{i,N-1}\ \ \ \ x_{i,N}\right]$ is the block color information. Similarly, for the m2 blocks of I2, there is corresponding matrix:

$$Y = \begin{bmatrix} y_{1,1} & \dots & y_{1,N} \\ \dots & & \dots \\ y_{m2,1} & \dots & y_{m2,N} \end{bmatrix} \tag{4}$$

Thus, the problem of multi-block correspondence is transformed into the corresponding problem between rows of X and rows of Y.

Under the previous definition, the paper expresses the correspondence problem as the model of constrained minimization problem:

$$\begin{aligned} P^* = \arg \quad & \min_P \quad J(X,Y,P) \\ s.t. \quad & P \in \rho_P(m1+1, m2+1) \end{aligned} \tag{5}$$

where J is the objective function, ρ_P is a permutation matrix.

Optimal solution P^* is a partial permutation matrix, and the arrangement of its internal element 0-1 is exactly the situation between the two correctly corresponding image block, the value of its elements are the relation of row of Y and the corresponding row of X: for each input, when the block X_i (i-th row of X) and block Y_j (j-th row of Y) are correctly correspondence, $P_{i,j}$ is 1; on the contrary, $P_{i,j}$ is 0.

In order to make the corresponding algorithm robust to outliers, P will be constructed as a (m1 +1) × (m2 +1) matrix, the additional rows and columns present target blocks which have not correspondence: there is no candidate row of Y which correspondence to X_i, the matrix P in the row i before m_2 elements are 0, $P_{i,\, m2+1}$ is 1; on the contrary, if there is no block corresponding to the block Y_j, then the matrix P, before the first element m1 in the row j are 0, $P_{i,\, m1+1}$ is 1.

The constraint matrix P as follows:

$$\begin{cases} P_{i,j} \in \{0,1\}, \forall i \le m1+1, \forall j \le m2+1 \\ \sum_{i=1}^{m1+1} P_{i,j} \le 1, \forall j \le m2+1 \\ \sum_{j=1}^{m2+1} P_{i,j} \le 1, \forall i \le m1+1 \end{cases} \tag{6}$$

Since each element of matrix P represents a correspondence between all the blocks, so that the model could not only achieve the correct correspondence between the blocks, but also eliminate false correspondence between the target block. Thus the correspondence problem between multi-target block is transformed into linear programming model to strike the global optimal solution of the problem, global optimal solution is to get all the possible correspondence between the actual correspondences with the most consistent situation.

In this paper, taking into account the characteristic of multi-objective corresponding process in solving the global optimal solution, the paper takes relaxation labeling (Relaxation Labeling) algorithm [7] [8] to optimize the correspondence to improve the accuracy and agility of the process. Relaxation labeling is a recognition method of using label to descript pattern. The whole process is similar to the human reasoning process, which uses a variety of constraints and gradually narrows the search scope, eventually obtaining correct results. In the application situation of this paper, the I1 image of each block corresponds to a straight line as the target space, each block in I2 image corresponding spatial line as a marker. When processing starts, correctly correspondence can't be achieved, since the attribute object is blurred. Relaxations marked the formal relationships between the target and the constraints of the system and gradually decrease ambiguity. First, an initial target correspondence condition is given. Then through constant iteration and gradually update the corresponding relationship, we can finally obtain an accurate representation of the target correspondence.

Once obtain the 3D position of each blocks, we project space points onto the ground plane. And use ISODATA clustering to categorize these spaces and to calculate a space region of the target. Thus target positioning is achieved.

4 Experimental Results

In our experiment, we captured two camera's images whose size are 768*576. After skin and hair region extraction and 20*20 blocking, there are 17 blocks in I1 image, and 18 blocks in I2 image, respectively. Through the calibration of internal and external camera parameters, 17 blocks in C1 correspond to 17 space straight lines, 18 blocks in C2 correspond to 18 space straight lines. After each calculation of lines, we got a 17 × 18 matrix of CC. In the matrix, each row represents spatial distance of a certain space line of C1 and each space line of C2. During the relaxation labeling iteration, the minimum value of each line is selected as the starting corresponding states.

In order to reduce the amount of calculation in the correspondence process, the paper sets a threshold distance Trb to do some preprocessing for matrix CC. In the case of exceeding the threshold Trb for each elements, the larger distance between two straight lines means larger space, this element's correspondence is deleted; If the number of elements which are less than Trb for each row or column is more than 15% of the total elements of a row or column(in the paper is 3), the minimum 15% elements are selected as candidate correspondences; otherwise, all these elements are used as an candidate elements, which are added to the process of relaxation iteration.

The paper sets Trb = 40 as the preprocessing threshold of matrix CC, the result of correspondence between each block of C1 and C2 and the obtained target three-dimensional information are shown as Table 1.

After projecting block spatial point onto the ground plane, we utilize ISODATA clustering algorithm to classify these spatial points and calculate the target region of space. As the result, these points are divided into two categories: G1{(2022.47 , 390.698), (2020.92 , 395.729), (2022.04 , 404.396)}, G2{(1729.82 , 825.062), (1741.47 , 826.806), (1735.62 , 832.622) , (1749.25 , 816.067)} as Fig. 4.

Table 1. Results of blocks positioning

ID	C1 blocks	C2 blocks	Distance mm	Spatial positioning mm
1	124, 346	430, 350	24.7812	2022.47 , 390.698 , 1715.35
2	124, 350	430, 254	21.6338	2020.92 , 395.729 , 1711.79
3	128, 368	435, 366	9.80092	2022.04 , 404.396 , 1706.48
4	238, 362	492, 372	14.1256	1729.82 , 825.062 , 1757.61
5	246, 368	494, 376	8.44677	1741.47 , 826.806 , 1770.3
6	274, 376	490, 384	8.4012	1735.62 , 832.622 , 1714.46
7	238, 394	492, 410	6.88672	1749.25 , 816.067 , 1733.53
8	206, 492	514, 498	19.4985	1741.98 , 675.042 , 795.833
9	260, 486	518, 498	10.3048	1702.55 , 978.119 , 853.541

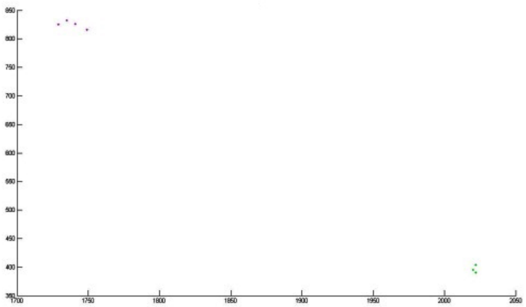


Fig. 4. Projected location

As each block includes the hands skin regions and the head skin regions, the horizontal plane of the spatial distance difference between hands and head should be taken into account in the classification of these regions. Finally, the target positioning in the scene shown in the figure 5.

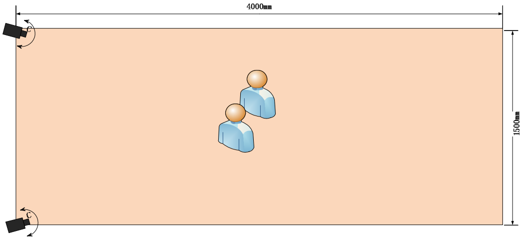


Fig. 5. Targets positioning in the scene

In the preprocessing step of matrix CC optimization, The corresponding results of C1 and C2 blocks are influenced greatly by the threshold value Trb, which is determined by experiment in this paper. With the reduction of the threshold the correspondence performance is reduced, this is because the restrictions are too stringent, resulting in the increasing of false matching, eventually leading to the decreasing of correct correspondence rate. Conversely, when the threshold value is increased, the result of correspondences will increase, including the false correspondence. Although the threshold value can be increased, due to the increasing of error correspondence rate, the overall correct correspondence rate didn't improve.

5 Conclusion

In the paper, we propose a new method for multi-camera target correspondence between different views. The method based on the idea of the line-of-sight constraints to obtain positioning of the target. The paper divides regions of interest into multiple blocks, providing appropriate redundancy information for each object and decreasing the risk of objects missing between different perspective images which are caused by large view-angle and wide baseline. When dealing with multi-objective correspondence problem, we constructed a multi-objective corresponding model; the target is transformed into the corresponding minimum objective function to solve a linear programming problem. The experimental results of the paper show that, the method has high accuracy corresponding based on multi-target line-of-sight constraints, the theory is more intuitive thinking, and algorithm is relatively simple. In order to further improve the measurement accuracy, in the next step, we plan to adopt more cameras and add more constraints in relaxation labeling processing, and then apply it in more crowded situation.

Acknowledgement. This work was supported by the Development Foundation of Shanghai Municipal Commission of Science and Technology (13dz1202404).

References

1. Eshel, R., Moses, Y.: Homography based multiple camera detection and tracking of people in a dense crowd. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 24-26, pp. 1–8 (2008)
2. Eshel, R., Moses, Y.: Tracking in a Dense Crowd Using Multiple Cameras. International Journal of Computer Vision 88(1), 129–143 (2010)
3. Ge, W., Collins, R.T.: Crowd detection with a multi-view sampler. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 324–337. Springer, Heidelberg (2010)
4. Castle, R.O., Murray, D.W.: Key-frame-based recognition and localization during video-rate parallel tracking and mapping. Image and Vision Computing 29(8), 524–532 (2011)
5. Mazzon, R., Cavallaro, A.: Multi-camera tracking using a Multi-Goal Social Force Model. Neuro-Computing 100, 41–50 (2013)

6. Ma, H., Zeng, C., Ling, C.: A Reliable People Counting System via Multiple Cameras. ACM Transactions on Intelligent Systems and Technology 3(2), 1–22 (2012)
7. Lloyd, S.A.: An optimization approach to relaxation labelling algorithms. Image and Vision Computing 1(2), 85–92 (1983)
8. Kittlera, J., Illingworth, J.: Relaxation labelling algorithms — A review, Image and vision computing. Image and Vision Computing 3(4), 206–210 (1985)
9. Rosenfeld, A., Hummel, R.A., Zucker, S.W.: Scene labeling by relaxation operations. IEEE Trans. SMC. 6(6), 420–433 (1976)
10. Cui, L., Dongqing, F.: Improved image segmentation algorithm based on K-means clustering. Journal of Zhengzhou University (Natural Science) 43(1), 109–113 (2011)