Trajectory Fusion for Multiple Camera Tracking

Ariel Amato[1], Murad Al Haj[1], Mikhail Mozerov[1], and Jordi Gonzàlez[2]

[1] Computer Vision Center and Department d'Informàtica. Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain aamato@cvc.uab.es

[2] Institut de Robòtica i Informàtica Industrial(UPC-CSIC),Edifici U Parc Tecnològic de Barcelona. 08028, Spain.

Summary. In this paper we present a robust and efficient method to overcome the negative effects of occlusion in the tracking process of multiple agents. The proposed approach is based on the matching of multiple trajectories from multiple views using spatial and temporal information. These trajectories are represented as consecutive points of a joint ground plane in the world coordinate system that belong to the same tracked agent. We introduce an integral distance between compared trajectories, which allows us to avoid mismatches, due to the possible measurement outliers in one frame. The proposed method can also be considered as an interpolation algorithm of a disconnected trajectory during the time of occlusion. This technique solves one of the most difficult problems of occlusion handling, which is a matching of two unconnected parts of the same trajectory.

1 Introduction

Tracking objects with multiple views is an active area of research that finds application in automated surveillance, video archival-retrieval, and human-computer interaction. This work presents a method based on the matching of multiple trajectories of the same agent which are obtained from multiple views. This approach allows to resolve the temporal discontinuity of a trajectory, which occur due to the occlusion. Consequently, trajectories obtained from different cameras are fused in the same ground plane. Occlusion handling is a complex task that has been thoroughly investigated [1], which can be classified in four areas: region based tracking [2], active contour-based tracking [3], feature- based tracking [4], and model-based tracking[5]. Existing approaches include Extended Kalman Filters (EKF) was used for occlusion analysis [6]. A relationship between camera views using homography is considered in [7]. To resolve mutual occlusion the best view is utilized in [8]. A similar method that based on a fusion of information from every camera into a 3-D geometric coordinate system is presented in the work [9]. Most of such techniques need high computationally cost, and the accuracy depends on the different object conditions. In contrast, our approach needs just a few ground plane points of tracked agents in each frame. Another advantage of the presented method is the strong mathematical conclusion of the tracking process, which easily allows implementing the proposed model to the arbitrary scene. Our experiments with a real data show that it is possible to track the trajectories very effectively even in the presence of occlusion using only the ground plane points of the agents.

This paper is organized as follows. Section 2 describes the method used for segmentation objects. Section 3 describes how to obtain the overlapping camera view process in a joint ground plane, separated in two step, camera calibration and top view conversion. Section 4 shows the main process in order to obtain time continue trajectory. In section 5, experimental results are shown.

2 System Input Data

Our tracking process involves on stage of segmentation. We choose the method of colour background subtraction that is described in [10]. Once the segmentation is performed, the size and position of the ellipsoid that surrounds the object is obtained. The floor point projection can be calculated using intrinsic and extrinsic parameter of each camera. The result of segmentation is shown in Fig. 1.

(a) (b)

Fig. 1: (a) Original Image (b) Segmentation with foot point

3 Data Fusion

To convert the image planes obtained from multiple camera views into one ground plan map of an inspected scene, a calibration process is needed.

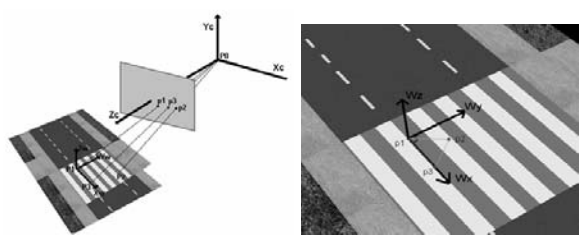
3.1 Camera Calibration

Most traditional camera calibration techniques require specific knowledge about the geometric characteristics of the referenced object such as Direct Linear Transformation (DLT) [11] which solves the perspective matrix linearly; similar methods include Tsais [12] and Zhangs [13]. Plane based methods [14] use the same DLT paradigm, but show more flexibility. The method we use in this paper is based on geometric derivation [15]. We choose this because it uses only distance between three points in the world coordinate system to achieve calibration. Rotation matrix $\mathbf{R}$ and translation vector $\mathbf{T}$ for each camera coordinate system are obtained via the given distance between the vertices of the marker triangle formed by the three points of reference.

(a) (b)

Fig. 2: Two viewpoints of the same scene

The intrinsic parameter like focal radio can be obtained independently; we could assume that the focal radio is known (from the data sheet) or calculated in a simple experimental process with two known points and the distances between these points and the camera. Two different points of views of the same scene with the landmark of three points are represented in Fig. 2. The pin-hole camera model is shown in Fig. 3(a).

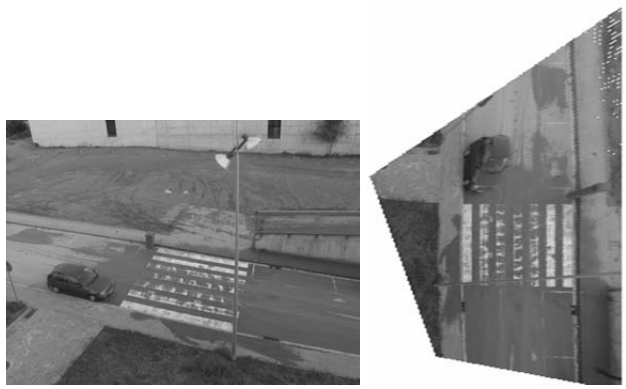


(a) (b)

Fig. 3: (a) Cameral model and ground plane (b) New coordinate systems

3.2 Joint Ground Plane

To match the points from different points of view, a joint coordinate system is needed. Once the segmentation algorithm localizes a pixel of interest (i,j) in the image of one camera (in our case it is a foot point of an agent), this pixel should be projected onto the joint plane in the world coordinate system for matching with another point in a different image obtained from a second camera. This projection process is referred to as the Inverse Perspective Mapping (IPM). Fig. 4 show an example of the IPM in a complete image.



(a) (b)

Fig. 4: Top View Transformation (IPM)

4 Trajectory Tracking

Initially, we have a set of A cameras Cα, and T images $\mathrm{I}_{(t,\alpha)}$ obtained from each camera. The segmentation algorithm plus the joint ground plane transformation provide us with a set of floor points:

$$\mathbf{Q} = \bigcup_{A,T,N(\alpha,t)} \mathbf{p}^t_{\alpha,n} \tag{1}$$

where p is a point of the word coordinate system in the joint ground plane. The number of points in a single image at a certain snapshot time t depends on the camera index α, due to occlusion and non overlapped fields of view of different cameras. So, it is convenient to compliment the set of real points Q by a set of invisible or imaginary points

$$\tilde{\mathbf{Q}} = \bigcup_{A,T,\tilde{N}(\alpha,t)} \tilde{\mathbf{p}}_{\alpha,n}^{t} \tag{2}$$

where the coordinates of point $\tilde{p}$ are calculated using some rules that will be explained bellow. An example of such extension is illustrated in Fig. 5. It is useful to define a set of points that belong to the same time of the cameras snapshot as

$$\mathbf{P}_{m(t)}^{t} = \bigcup_{A} \mathbf{p}_{\alpha,n(m)}^{t} \tag{3}$$

where; $1 \leq m(t) \leq \max_{\alpha} \{N(\alpha,t)\}$ and the number of points in this set always is equal to number of cameras, which means that the set can include also imaginary points.

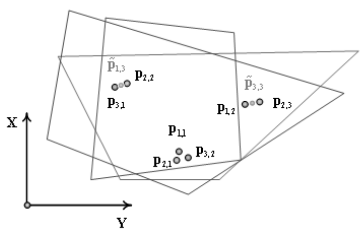


Fig. 5: Overlapping View in a joint plane

Our main goal now is to associate all these points with several tracks.

$$\mathbf{O}_k = \bigcup_{T} \mathbf{P}_k^t \tag{4}$$

Where $1 \leq k \leq K$, and, in principle,$\max_{\alpha,t} \{N(\alpha,t)\} \leq K$, which means that a certain track can include void elements in Eq.(3), in other words, the time domain of a track possibly cannot coincide with the full interval of the analyzed scene performance [1,T]. We introduce distance Δ_k^t between two consecutive sets in the same track as

$$\Delta_k^t = \left\| \mathbf{P}_k^t - \mathbf{P}_k^{t-1} \right\| = \sum_{\alpha \in A} \left\| \mathbf{p}_{\alpha,n(k)}^{t} - \mathbf{p}_{\alpha,n(k)}^{t-1} \right\| \tag{5}$$

And the intrinsic distance of a set $\mathbf{P}_k^t$ itself:

$$D_k^t = \sum_{i<j<A} \left\| \mathbf{p}_{i,n(k)}^{t} - \mathbf{p}_{j,n(k)}^{t} \right\| \tag{6}$$

It is reasonable to admit that for the track of a real agent these distances have to be minimal.So, now we formulate an optimization problem as follows: find the sequences of points of a track , which minimize the sum of distances:

$$\mathbf{O}_k = \arg\min_{\mathbf{P}_k^t} \sum_{t \in T} \left(D_k^t + \Delta_k^t \right) \tag{7}$$

This problem can be solved recurrence growing of each track:

$$\mathbf{O}_k^\tau = \mathbf{O}_k^{\tau-1} \bigcup \mathbf{P}_k^\tau \tag{8}$$

where

$$\mathbf{O}_k^\tau = \bigcup_{t=1}^{\tau} \mathbf{P}_k^t \tag{9}$$

$$\mathbf{P}_k^\tau = \arg\min_{p_{\alpha,n}^\tau} \{ D_k^\tau + \Delta_k^\tau \} \tag{10}$$

The expression of Eq.(10) can be easily solved by direct search of the minimum of the objective function $D_k^\tau + \Delta_k^\tau$;

5 Experimental Results

Our experiments were conducted with two cameras. To illustrate the result of tracking we chose a situation with only two agents. Camera 1 was able to track both agents without occlusion, but camera 2 was unable to do so because one agent occluded the second for a long time. In Fig. 6(a),(b) show two snapshots from camera 1 and 2 respectively. Fig. 6(c) shows the trajectory of both agents, captured from camera 1, in joint ground plane. In Fig. 6(d) the same is shown but from camera 2.

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

Fig. 6: (a) Snapshot camera 1 (b) SnapShot camera 2 (c)Trajectory in joint ground plane camera 1 (d) Trajectory in joint ground plane camera 2

When we applied image matching, the discontinuity is resolved and then our algorithm can track efficiently the trajectory Fig. 7.

We tested our method with more than 5000 frames with occlusion and multiple agents, and we were able to track correctly all the trajectories. Fig. 8

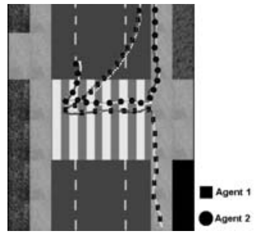
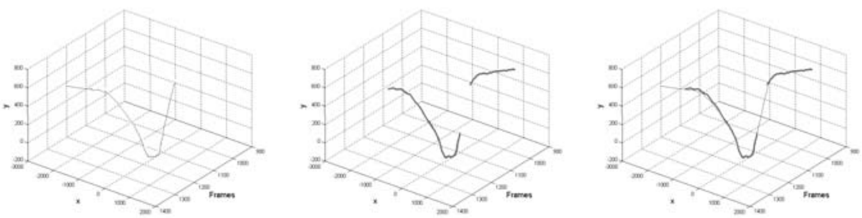


Fig. 7: Overlapping Trajectories camera1 camera 2

(a) and (b) shows one occluded agent trajectory in time-space domain,from camera 1 and camera 2 and Fig. 8 (c) shows the result of trajectory interpolation.



(a) (b) (c)

Fig. 8: (a)Trajectory of agent 1 in (time space) domain of camera 1 (b) Trajectory of agent 1 in (time space) domain of camera 2 (c) Trajectory overlapping Camera 1 and Camera 2 in (time-space) domain.

6 Conclusion

This paper presents a method for resolving occlusion of tracked agents using multiple views. The experimental results show that it is possible to track the trajectories effectively, even in the presence of occlusion, using only the ground plane points of the agents. This method uses strict mathematical model, which makes the tracking process robust and highly efficient.

Acknowledgements

This work is supported by EC grants IST-027110 for the HERMES project and IST-045547 for the VIDI-video project, and by the Spanish MEC under projects TIN2006-14606 and DPI-2004-5414. Jordi Gonzàlez also acknowledges the support of a Juan de la Cierva Postdoctoral fellowship from the Spanish MEC.

References

1. YWeiming Hu, Tieniu Tan, Fellow, IEEE, Liang Wang, and Steve Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviors", *Systems, Man and Cybernetics, Part C: Applications and Reviews, IEEE Transactions,* vol 34, pp. 334–352, 2004.
2. M. Kilger, "A shadow handler in a video-based real-time traffic monitoring system", *IEEE Workshop Applications of Computer Vision Palm Springs, CA,* pp. 11-18, 1992.
3. L. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components", *IEEE Trans. Pattern Recognit.Machine Intell,* vol. 23, pp. 349-361, 2001.
4. R. Polana and R. Nelson, "Low level recognition of human motion", *IEEE Workshop Motion of Non-Rigid and Articulated Objects,* pp. 77-82, 1994.
5. Dae-Sik Jang, Hyung-Il Choi, "Active models for tracking moving objects", *Pattern Recognition ,* vol. 33, no. 7, pp.1135-1146, 2000.
6. Rosales, Sclaroff, "Improved Tracking of Multiple Humans with Trajectory Prediction and Occlusion Modeling", *IEEE Conference on CVPR,Workshop on the Interpretation of Visual Motion, Santa Barbara, CA,* pp. 117–123, 1998.
7. A. Stein, "Tracking from Multiple View Points: Selfcalibration of Space and Time", *In DARPA IU Workshop,* pp. 1037–1042, 1998.
8. A. Utsumi, H. Mori, J. Ohya, and M. Yachida., "Multiple-view-based tracking of multiple humans", *in Proc. Int. Conf. Pattern Recognition,* pp. 197–601, 1998.
9. R. T. Collins, A. J. Lipton, H. Fujiyoshi, and T. Kanade., "Algorithms for cooperative multi-sensor surveillance", *Proceedings of the IEEE,* vol. 89 pp. 1456-1477, 2001.
10. I.Huerta and D. Rowe and M. Mozerov and J. Gonzàlez, "Improving Background Subtraction based on a Casuistry of Colour-Motion Segmentation Problems", *3rd IbPRIA, Gerona, Spain,* vol. 2 pp. 475–482, 2007.
11. Y. I. Abdel-Aziz, H.M. Karara, "Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry", *Proc., of the ASP Symposium on Close-Range Potogrammetry,* pp.1-18, 1971.
12. R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using offthe-shelf TV cameras and lenses", *IEEE J. Robot. Autom.,* vol 3 pp. 323-344, 1987.
13. Z. Zhang, "A flexible new technique for camera calibration", *LIEEE Trans. Pattern Analysis and Machine Intelligence,* vol 22 pp.1330-1334, 2000.
14. L. Lucchese, "Geometric calibration of digital cameras through multiview rectification", *Image Vis. Comput,* pp. 517-539, 2005
15. Huang T., Netravali, A., "Motion and Structure from Feature Correspondences", *Proceedings of the IEEE,* vol 82 pp. 252–268, 1994.