PARTIAL MOTION TRAJECTORY GROUPING THROUGH ROOTED ARBORESCENCE

*Fan CHEN**

School of Information Science
Japan Adv. Insti. of Sci. and Tech.
Nomi, Ishikawa, Japan

Christophe De Vleeschouwer

Ecole Polytechnique de Louvain
Université catholique de Louvain
Louvain-La-Neuve, Belgium

ABSTRACT

A novel method is presented for tracking multiple deformable objects in a multiview scenario. We avoid applying fixed constraints of object sizes for object isolation, but guide the splitting/merging of objects by grouping partial motion trajectories extracted from the occupancy status of objects in the homography ground-plane. We treat this clustering task as solving a rooted arborescence problem, and investigate the performance of the proposed method by experimental results.

***Index Terms*—** Multi-View Tracking, Trajectory Analysis, Rooted Arborescence Problem

1. INTRODUCTION

Considering object tracking in the multi-view environment could efficiently overcome traditional hurdles such as reflection, occlusion and shadow in the single view case, by fusing the clues from multiple view-angles to resolve the interpretation ambiguities of each camera view. Many previous methods have been proposed in the literature. A top-down method proposed in [1] considers a grid over the ground-plane as the solution space.The likelihood of people occurrence in each position is validated by projecting a fixed-size bounding box back into all camera views, where back-projection and validation processes are computationally expensive, especially when a dense grid is used. In a bottom-down method, objects are detected by only observing the occupancy mask on the ground plane, which is computed once by accumulating foreground occupancy masks from all camera views, and is thus more computationally efficient [2][3].

[2] explained a simple solution to locate objects on the accumulated occupancy mask, where they find all good matchs of a fixed-size guassian pattern by using a greedy searching to visit each local maximum of the occupancy mask, and then use a fixed threshold to filter the most reliable results. Obviously, this method cannot deal with deformable moving objects well, due to both its constraint on object size (which is insufficient in handling variable sizes of accumulated patterns in the ground-plane) and its fixed-value thresholding (which is insufficient in handling variable vertical depth of objects).

*Supported by the MEXT Grant-in-Aid for Young Scientists (B) No.23700110.

Considering the continuity of object movement,temporal information is used to link deformable objects in [3], where they built a temporal Markov Random Field from the accumulated occupancy mask, and minimized an energy function defined on the spatial/temporal occupancy correlations between neighboring ground-plane positions. Due to the heavy computational cost, only a small temporal window (15 Frames) could be considered, which hence cannot exploit the splitting/merging information of multiple objects sufficiently. This is especially a problem when the information used to discriminate the targets is only sporadically available (e.g. the numbers printed on sport players' jerseys).

In the present paper, we present a novel method for tracking objects on the ground-plane occupancy mask, which has three major contributions: 1) Instead of using the whole occupancy mask as in [3], we consider a set of partial motion trajectories extracted from local features of the accumulated occupancy mask, which significantly reduces the computational complexity, and enables direct consideration of long-range temporal/spatial relationships between objects; 2) In contrast to those methods [4][5] that analyze trajectories extracted from a single camera view, we consider the trajectories extracted from the occupancy mask on the ground-plane, where objects are mutually unobstructed in general; 3) We propose an efficient way for grouping trajectories by solving it as a rooted arborescence problem.

We will explain the proposed method with more details in Section 2, and investigate the performance of our method with experimental results in Section 3. We then conclude this paper with Section 4.

2. OBJECT TRACKING BASED ON PARTIAL MOTION TRAJECTORY GROUPING

We aim at exploiting the temporal continuity of the accumulated patterns of moving objects, by extracting and grouping partial motion trajectories from their local features.

In Fig.1, we summarize the overall framework of the proposed tracking approach, which includes four major steps:

1) **Foreground/Background Separation**, which extracts foreground occupancy mask from each camera view;

2) **Multi-layered Accumulation of Occupancy Mask**, which accumulates the forground occupancy masks on the

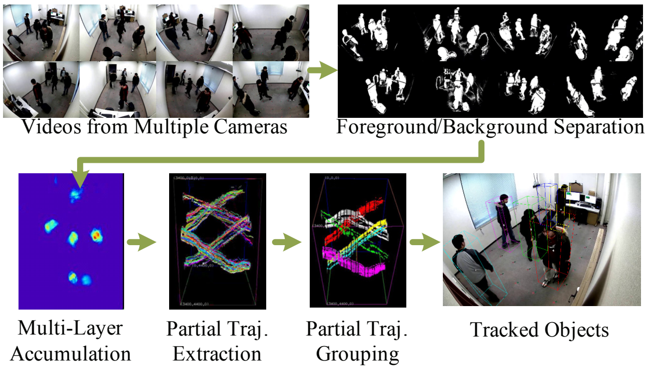


Fig. 1. Overall framework of the proposed tracking algorithm based on grouping partial motion trajectories.

homography ground-plane, where layerwise thresholding is adopted to reduce the ”ghost” effects;

3) **Extraction of Partial Motion Trajectories**, where a simple tracker is used to extract continuous partial motion trajectories of objects on the accumulated occupancy mask;

4) **Partial Motion Trajectory Grouping**, which groups extracted partial motion trajectories into several clusters, where each cluster represents one moving object.

In the later part of this section, we will explain these four steps with more details.

2.1. Foreground/Background Separation

Based on the conventional GMM background remover[6], we extract the foreground occupancy mask from each individual camera view, where some sample results are shown in Fig.1.

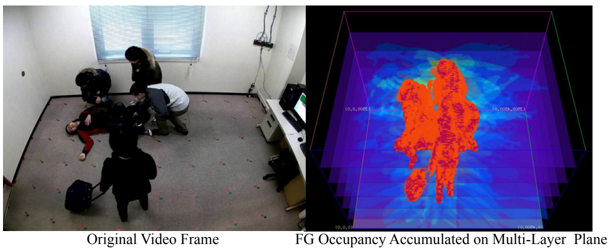


Fig. 2. The occupancy mask of objects in the ground plane by accumulating multiple parallel homography planes.

2.2. Multi-layered Accumulation of Foreground Occupancy Mask on Homography Ground Plane

The accumulated occupancy mask represents the probability map of an object occupying a ground position, which can be obtained by directly summing up the object density in the vertial direction, as shown in Fig.2.

Although an efficient way has been proposed in [2] for speeding up this naive accumulation, we still need to deal with the ”ghost” effect, which is due to the ambiguity of the homographic geometry. As shown in the left of Fig.3, the ”ghost” effect could blur the boundary between different objects severely, which hinders accurate detection of local maximal response as used in [2]. In naive accumulation, it is thus difficult to discern real objects with low heights from the accumulated ghosts existing in more layers.

Occupancy masks generated by real objects have higher values than those from ghosts in each individual homography plane, as shown in the middle of Fig.3, due to their higher consistency in multiple camera views. Hence, we perform layer-wise thresholding before accumulation of multiple layers, so as to surpress the ”ghosts”, which not only refines the boundary between different objects but also enhances the profiles of low-sized objects in the accumulated occupancy mask, e.g. the suitcase and the people lying on the ground (Fig.3).

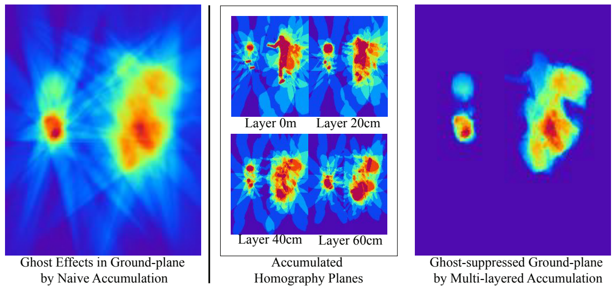


Fig. 3. We perform layer-wise thresholding to surpress the ghost effect and obtain a neat occupancy mask.

2.3. Extraction of Partial Block Motion Trajectories

Owing to the temporal continuity of object movement, we are able to track deformable objects by including the temporal information, represented here by many partial motion trajectories extracted from local block features of different objects.

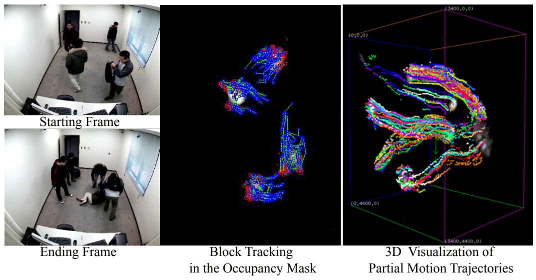


Fig. 4. Partial block motion trajectories extracted by the simple block tracker from the starting frame to the ending frame.

We use a very simple tracker for tracing block features of objects, which has three major processing phases: tracking seed spawning, local searching, and template updating. In Fig.4, we show the tracked block features over a short video segment and their 3D visualization. Although three objects are moving together when one person is falling down, we could still isolate these three objects, based on the historic behavior of their partial block trajectories, which is our major motivation of using temporal information.

2.4. Partial Block Motion Trajectory Grouping by Solving A Rooted Arborescence Problem

Given the extracted partial motion trajectories on the homography groundplane, we consider to group them into different clusters, where each cluster represents one moving object.

Formally, assume that we have N partial trajectories, where the i-th partial trajectory is denoted by τ_i. They are extracted from L objects, and each object is represented by a set of partial trajectories, i.e., $\mathbf{\Theta} = \{\mathbf{\Theta}_l | l = 1...L\}$. We infer the optimal clustering of those trajectories, by maximizing the following posterior probability, i.e.,

$$\{\mathbf{\Theta}_l^*\} = \arg\max_{\{\mathbf{\Theta}_l\}} \log P(\{\mathbf{\Theta}_l\}|\{\tau_i\}). \tag{1}$$

Regarding partial trajectories as observed data, and assuming the prior of $\mathbf{\Theta}$ to be uniformly distributed, we know from Bayes' theorem that it is equivalent to

$$\{\mathbf{\Theta}_l^*\} = \arg\max_{\{\mathbf{\Theta}_l\}} \log P(\{\tau_i\}|\{\mathbf{\Theta}_l\}). \tag{2}$$

We let each moving object mutually independent, and assume 1^{st} order Markov dependency, i.e., each partial trajectory τ_i only depends on one neighbouring trajectory,

$$\begin{aligned} P(\{\tau_i\}|\{\mathbf{\Theta}_l\}) &= \prod_{l=1}^{L} P(\{\tau_i\}|\mathbf{\Theta}_l, \tau_i \in \mathbf{\Theta}_l) \\ &\approx \prod_{l=1}^{L} \prod_{\tau_i \in \mathbf{\Theta}_l} P(\tau_i | \mathcal{P}_{\mathbf{\Theta}_l}(\tau_i)). \end{aligned} \tag{3}$$

Here we use $\mathcal{P}_{\mathbf{\Theta}_l}(\tau_i)$ to denote the predecessor trajectory of τ_i. For each cluster, there is only one trajectory that has no predecessor, denoted by τ_{l0} for Cluster l. We make them depend on a virtual root trajectory τ^{R}, and derive

$$P(\{\tau_i\}|\{\mathbf{\Theta}_l\}) = \prod_{l} \prod_{\substack{\tau_i \in \mathbf{\Theta}_l \\ \tau_j = \mathcal{P}_{\mathbf{\Theta}_l}(\tau_i)}} P(\tau_i|\tau_j) P(\tau_{l0}|\tau^{\mathrm{R}}) \tag{4}$$

We hence obtain the optimal clustering by

$$\{\mathbf{\Theta}_l^*\} = \arg\min_{\{\mathbf{\Theta}_l\}} \sum_{l} \sum_{\substack{\tau_i \in \mathbf{\Theta}_l \\ \tau_j = \mathcal{P}_{\mathbf{\Theta}_l}(\tau_i)}} \mathcal{D}_{ij} + \mathcal{D}_l^{\mathrm{R}} \tag{5}$$

where we define

$$\mathcal{D}_{ij} = -\log P(\tau_i|\tau_j), \mathcal{D}_l^{\mathrm{R}} = -\log P(\tau_{l0}|\tau^{\mathrm{R}}). \tag{6}$$

as the distance between two trajectories τ_i and τ_j and the distance between τ_{l0} and τ_{R}, respectively. Maximization of the posterior probability is thus solvable by finding the minimum-weight spanning tree of a directed graph rooted at τ^{R}, i.e., by solving the rooted arborescence problem.

We set a fixed value to all $\mathcal{D}_l^{\mathrm{R}}$, which in fact serves as a threshold for merging/spliting neighbouring trajectories into objects. For any pair of partial trajectories, we define $\mathcal{D}_{ij}$ as

- if they are temporally overlapped, we define

$$\mathcal{D}_{ij} = -\log P(\tau_i|\tau_j) = -\log \mathcal{N}(D_{ij}^{AVG}|0, R^2), \tag{7}$$

  where R is the average size of a moving object and D_{ij}^{AVG} is the average L2 distance of the longest overlapped parts between these two trajectories .

- if they are not overlapped, we define

$$\mathcal{D}_{ij} = -\log P(\tau_i|\tau_j) = -\log \mathcal{N}(\Delta X_{ij}|0, \sigma_{ij}^2), \tag{8}$$

  where ΔX_{ij} is defined as the L2 distance between the tail of τ_j and the head of τ_i. Standard variation $\sigma_{ij} = \Delta T_{ij} v$, where ΔT_{ij} is the temporal distance while v is the expected maximal speed of the moving object.

Compared to centroid-based clustering, e.g. K-means, a problem of linkage-based clustering is the "chaining phenomenon" [7], which is alleviated by the following processes:

1) Different from planer linkage, we introduced a virtual root node, which in fact achieves a hierarchical clustering by building a direct bridge between the centroids of two clusters;

2) An edge pruning process is inserted before solving the rooted arborescence problem, so as to avoid grouping trajectories of different objects together. Namely, for 3 trajectories τ_m, τ_n and τ_l, edge $\mathcal{E}_{mn}$ from τ_m to τ_n is removed, when $D_{mn}^{AVG} < R$, $D_{ml}^{AVG} < R$ and $D_{ln}^{MAX} > R^{\mathrm{MAX}}$ are all satisfied. Here, R^{MAX} is the maximum size of objects.

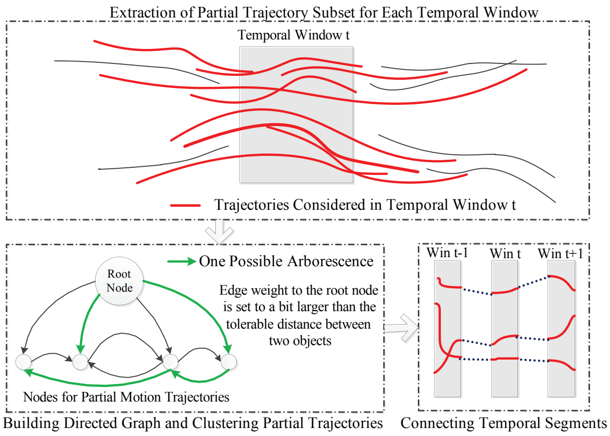


Fig. 5. Workflow of our rooted arborescence based approach for object tracking.

In Fig.5, we show the workflow of our clustering based tracking algorithm. For each time t, we consider all partial trajectories occuring in its local temporal window of size W. We first build a directed graph from these partial trajectories, along with a root node connecting to partial trajectories. Chu-Liu/Edmonds' algorithm (Complexity $O(E + VlogV)$)[8][9] is then applied to find the arborescence solution. Each object is then detected by computing the gravity center of all trajectories that are derived from the same path to the root node. Finally, a simple bipartite graph mapping is performed to link objects between two consecutive temporal windows.

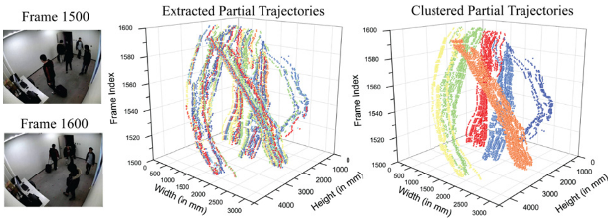
3. EXPERIMENTAL RESULTS

Due to page limitation, we only show in the paper a limited part of our results. We made experiments on some video data collected from eight cameras, where several people were asked to make different interactions during the general walking, such as chatting, exchanging bags/suitcases and falling down, in a small room. Some samples are available in Fig.1.

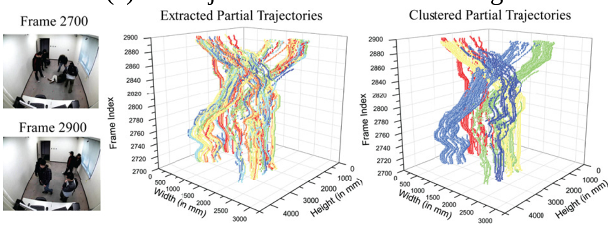
By taking the temporal window size as 200 frames, and setting $\mathcal{D}_l^{\mathrm{R}}$ and R^{MAX} to be equivalent to 45cm and 160cm, we present some results of trajectory grouping in different temporal segments in Fig.6, so as to explore the behavior of the proposed method under different applicative scenarios. Especially, we show results for three different scenarios:

1) In Fig.6(a), it represents the common case when all people are walking independently. It is easy to see that all objects are well separated by the proposed clustering method. Since a low $\mathcal{D}_l^{\mathrm{R}}$ is taken, the suitcase is also separated (in yellow) from its owner (in light cyan), when it is dragged.

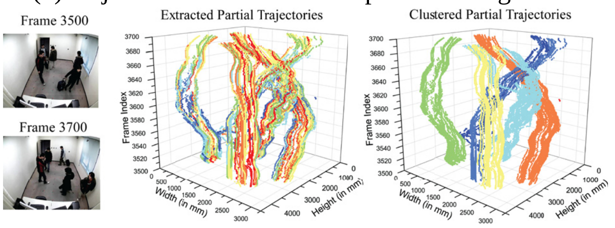
2) In Fig.6(b), it represents the case with object interactions, where one person is falling down and some other people surround him for a while. It is obvious that two persons (in dark blue and yellow) first surround the person (in light cyan), and then these three persons diverge in different directions. Also we notice that the falling person (in light cyan) covers a wider area at first, which reflects his falling status.



(a) All objects for a common walking



(b) Object interaction with one person falling down



(c) Object interaction with bag/suitcase exchange

Fig. 6. Results of trajectory grouping in three different temporal segments.

3) In Fig.6(c), it represents the case with another usual object interaction, where two persons are exchanging bags or suitcases. In Frame 3500,the girl first holds a suitcase. After two times of exchanges, she delivered the suitcase to one person, but received the bag from another person. The exchange of suitcase is clearly visible from the graph, as the suitcase (in light blue) first stands with the girl (in yellow) and then moves along with another man (in darker blue). We also notice that part of the cyan cluster merges with the girl (in yellow), which in fact is caused by the bag exchange. When the bag is carried, it is quite close to its carrier, which makes it difficult to isolate with the current implementation. We leave this detection as one future task, by further exploring the information brought by the partial trajectories.

Owing to its capability to detect deformable objects with various sizes, we found that the proposed method has a higher rate of positive detections ($\sim 85\%$) on the whole 8-min video, compared to that of the method in Ref.[2] ($\sim 75\%$), when the acceptable bias goes higher than 220mm, which is a tolerable bias for general surveillance applications. In the supplemental materials [10], we provide the detailed graph and all related resultant tracking videos, along with those from the method proposed in [2]. As a conclusion, the proposed method is able to detect objects with various accumulation heights (human or suitcase) and also various occupied areas in the ground plane (standing up or falling down), by utilizing the temporal information on the continuity of their movements, which is also our major motivation.

4. CONCLUSION

We proposed a novel method to track deformable moving objects in the multi-camera scenario. We extract partial motion trajectories from the occupancy mask in the ground plane, and then group them into different objects, by solving a rooted arborescence problem. In future, we need to further evaluate the proposed method against other methods, in various scenarios where partial motion trajectories are more noisy. The probability of using partial motion trajectories for scene understanding and abnormal detection will also be discussed.

5. REFERENCES

[1] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera people tracking with a probabilistic occupancy map," *TPAMI*, vol. 30, no. 2, pp. 267–82, 2008.

[2] D. Delannay, N. Danhier, and C. De Vleeschouwer, "Detection and recognition of sports(wo)men from multiple views," in *ICDSC 2009*, Sept. 2009, pp. 1–7.

[3] S. M. Khan and M. Shah, "Tracking multiple occluding people by localizing on multiple scene planes," *TPAMI*, vol. 31, no. 3, pp. 505–519, 2009.

[4] G. Antonini and J. P. Thiran, "Counting pedestrians in video sequences using trajectory clustering," *TCSVT*, vol. 16, no. 8, pp. 1008–1020, 2006.

[5] N. Anjum and A. Cavallaro, "Multifeature object trajectory clustering for video analysis," *TCSVT*, vol. 18, no. 11, pp. 1555–1564, 2008.

[6] N. Martel-Brisson and A. Zaccarin, "Moving cast shadow detection from a gaussian mixture shadow model," in *CVPR'05*, 2005, vol. 2, pp. 643–648.

[7] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Comput. Surv.*, vol. 31, pp. 264–323, 1999.

[8] Y.J. Chu and T.H. Liu, "On the shortest arborescence of a directed graph," *Science Sinica*, vol. 14, pp. 4, 1965.

[9] J. Edmonds, "Optimum branchings," *J. Research of the National Bureau of Standards*, vol. 71, no. B, pp. 7, 1967.

[10] Additional results, "http://www.jaist.ac.jp/project/prime-proj/results-tracking.htm," 2012.