IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 9, Issue 2, Ver. II (Mar - Apr. 2014), PP 22-25
www.iosrjournals.org

Pedestrian Detection and Tracking through Hierarchical Clustering

[1]Mr. Sathiyanarayana R, [2]Mr.R. Poovendran M.E.,
M.E in Communication Systems (ECE), Final year Adhiyamaan College of Engineering, Hosur, India
Asst Prof, ECE Dept., Adhiyamaan College of Engineering, Hosur, India

Abstract: *Building upon state-of-the-art algorithms for pedestrian detection and multi-object tracking, and inspired by sociological models of human collective behavior, we automatically detect small groups of individuals who are traveling together. These groups are discovered by bottom-up hierarchical clustering using a generalized, symmetric Hausdorff distance defined with respect to pairwise proximity and velocity. We validate our results quantitatively and qualitatively on videos of real-world pedestrian scenes. Where human-coded ground truth is available, we find substantial statistical agreement between our results and the human-perceived small group structure of the crowd. Results from our automated crowd analysis also reveal interesting patterns governing the shape of pedestrian groups. These discoveries complement current research in crowd dynamics, and may provide insights to improve evacuation planning and real-time situation awareness during public disturbances.*

I. Introduction

There has been increasing interest in using surveillance trajectory data for human behavior analysis, ranging from activity recognition based on the motion pattern of a single individual or interactions among a few (e.g., [1]), to analysis of the flow of a large crowd, for example, to discover pathways or monitor for abnormal events (e.g., [2]). Less well studied is the collective behavior of small groups of people in a crowd. In this paper, we build upon state-of-the-art pedestrian detection and tracking techni- ques to discover small groups of people who are traveling together. Determining the group structure of a crowd provides a basis for further midlevel analysis of events involving social interactions of and between groups.

Our main contribution is a hierarchical clustering algorithm that, informed by social psychological models of collective behavior, automatically discovers small groups of individuals traveling together in a low to medium human-perceived small group structure of the crowd. We also qualitatively on three outdoor sequences with different camera elevation angles, target sizes, and crowd densities, to demonstrate our method's tracking and group clustering capabilities across a range of conditions density crowd (Fig. 1). A pairwise distance that combines proximity and velocity cues is extended to form a robust distance between groups of people using a generalized, Agglomeration of clusters is further constrained by an intragroup tightness measure inspired by sociological research into group behavior, enabling the number of groups in the scene to be determined automatically.

We validate our approach extensively on several video sequences taken in public pedestrian areas from elevated viewpoints typical of surveillance camera footage. compare results of our algorithm with consensus ground truth labeled by multiple human coders. We find that there is substantial statistical agreement between our algorithm's estimated groups and the human-perceived small group structure of the crowd. We also qualitatively evaluate our method on three outdoor sequences with different camera elevation angles, target sizes, and crowd densities, to demonstrate our method's tracking and group clustering capabilities across a range of conditions.

Analyzing the group structure of crowds has important practical applications. Current models of evacuation treat all people as separate agents making independent deci- sions. These "particle flow" models tend to underestimate the time it takes for people to leave an area because groups of individuals who are together try to leave together, limiting the speed of the group to that of its slowest member. A small group behavior model also suggests new strategies for police intervention during public distur- bances and the identification of group behavior the group structure of crowds has important practical applications.

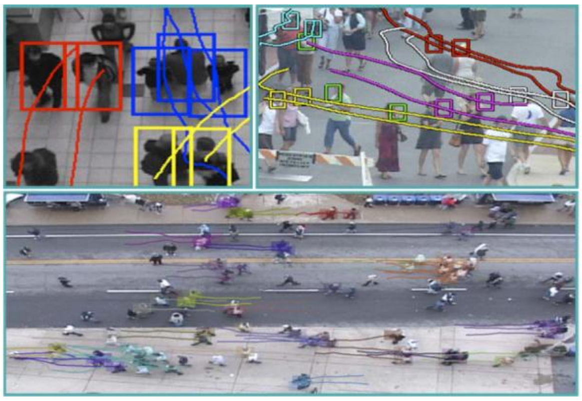

Fig. 1. Small groups are prevalent in pedestrian scenes. Our algorithm detects groups of people traveling together via hierarchical clustering on trajectories automatically extracted from video of crowds under various conditions. empirical data on real crowds faster and more thoroughly than previously possible.

II. Background And Related Work

This section explains why the composition of a crowd is important for modeling social behavior and reviews related computer vision work on crowd scene analysis.

Collective behavior and small groups. Collective behavior is the generic term for the often extraordinary and dramatic actions of groups and of individuals in groups Models of collective behavior tend to be bimodal. At one extreme are models that consider the entire crowd as one entity. Scholars have assumed that crowds transform individuals so that the resulting collective begins to exhibit a homo- geneous “group mind” that is highly emotional and irrational. At the other extreme are models treating everyone as independent members acting to maximize their own utility. For example, crowd behavior has been simulated by considering people as particles making local decisions based on the principle of least effort

As with most dichotomies, the truth is likely to lie in the middle. One hypothesis is that crowds are composed primarily of small groups, defined as a “collection of individuals who have relations to one another that make them interdependent to some significant degreeDespite being intuitively reasonable, there has been surprisingly little work to validate this hypothesis. Johnson argues that most crowds consist of small groups rather than isolated individuals.A n unpublished study by McPhail found that 89 percent of people attending an event came with at least one other person, 52 percent with at least 2 others, 32 percent with at least 3 others, and that 94 percent of those coming with someone left with the people they came with.

Behavior analysis. Behavior recognition involving inter- preting sequences of actions of one person or interactions of two or three are commonly built upon Hidden Markov Models or Dynamic Bayes Networks. These approaches are typically limited to a small, known number of individuals due to the combinatorics involved in the coupled interpretation of multiple time series. There is recent evidence that more efficient recognition of group activities is possible by using a model of the group activity process to guide interpretation of the actions of individual members..

III. Detecting And Tracking Individuals

There is no shortage of explanations for crowd behavior, but there is a shortage of explanations supported by empirical sociological research. The few empirical studies that have analyzed video data of people in public spaces have required hundreds of person hours to hand code just minutes of film, greatly limiting the amount and type of video that can be quantitatively analyzed. The use of automated computer vision methods therefore could repre- sent a substantial methodological improvement. However, generating a reliable set of trajectories for people in crowded public spaces is a nontrivial task due to frequent occlusions and the presence of nearby confusers. In this section, we describe an approach for pedestrian detection and tracking that is capable of producing reasonable trajectories in crowded scenes containing closely spaced people. Cluster- ing these trajectories to hypothesize small pedestrian groups is presented in Section 4.

We combine a pedestrian detector, a particle filter tracker, and a multi-object data association algorithm to extract long- term trajectories of people passing through the scene. The detector is run frequently (at least once per second), and therefore, in addition to any new individuals entering the scene, people already being tracked are detected multiple times. For each detection, a particle filter tracker is instantiated to track that person through the next few seconds of video, yielding a short-term trajectory, or tracklet. The goal at this stage is to generate a set of overlapping tracklets for each person. For example, if detection is run every 20 frames and a particle filter tracks each detection through the next 80 frames, at any one moment in time roughly four temporally overlapping trajectory fragments will be measuring the location of any given person in any given frame. A second phase of trajectory-to-fragment data association is then run to link and merge these multiple fragments into single, longer trajectories. Below, we describe our detection and tracking approaches in more detail.

3.1 Detection

We employ two different detection strategies. For videos captured from high elevation/wide angle views where people are small, we tackle pedestrian detection as a "covering" problem. Individual pedestrians are detected by using Reversible Jump Markov Chain Monte Carlo (RJMCMC) to find a set of overlapping rectangles that best explain or "cover" the foreground pixels in a binary segmentation generated by adaptive background subtrac- tion. This method is similar to that of and is capable of extracting overlapping individuals in crowds up to moderate density.

For higher resolution videos, pedestrian detection is performed in each frame using a combination of motion and contour (edge gradient) information, using a set of stationary regions of background clutter, to avoid finding false positives in those areas. Sample detection results from both methods are shown in Fig. 2.

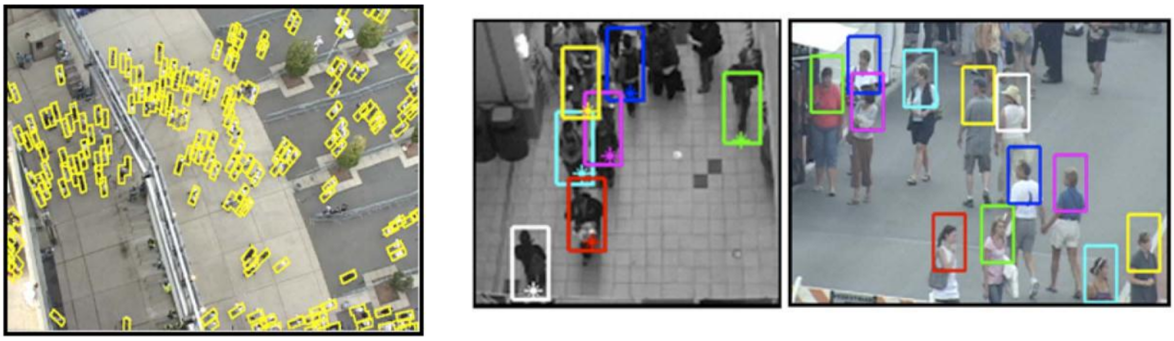

Fig. 2. Left: Sample detections in low-resolution video by estimating a rectangular covering using RJMCMC. Right: Sample detections in higher resolution video using an HoG detector for body (left) and head-and-shoulders (right).

3.2 Tracking

For each detected pedestrian, a Sampling Importance Resampling (SIR) particle filter [63] is instantiated as a short-term tracker to track the detected person for the next few seconds of video. The state space is 4D ðx; y; u; vÞ, where ðx; yÞ is the hypothesized image location of the object centroid and ðu; vÞ is the interframe velocity. We use constant velocity motion prediction with a Gaussian noise model. Roughly 50 particles are propagated for each target. The likelihood measure for determining particle weights for resampling can vary depending on resolution and quality of the video, e.g., normalized correlation of grayscale intensity templates, or Earth Mover's Distance (EMD) on marginal R, B, G color histograms. Since we reinitialize tracking frequently, the short-term tracker does not need to consider appearance model updates.

Sets of tracklets extracted in overlapping sliding windows of time are combined into longer trajectories by recursively merging each new set of tracklets into an evolving set of trajectories, one window at a time, in a single forward scan. Given a set of existing long-term pedestrian trajectories and a new set of tracklets from the next sliding window, we match up trajectories to tracklets through a process of data association. Specifically, if there are N trajectories and M new tracklets, we form an N M affinity table where each element contains a score rating the affinity of one trajectory with one tracklet that overlaps it in time. The affinity measure is a combination of geometric and appear- ance terms: a measure of "continuity" computed by the average distance between corresponding locations in the area of temporal overlap and appearance similarity of the targets. We also augment the affinity table with one row and column of "slack variables" to take into account that a new tracklet may not correspond to any existing trajectory (trajectory birth), or that a trajectory may not have been corroborated by any tracklet (which eventually leads to trajectory death).

To find the best assignment of trajectories to tracklets from the affinity table, we solve the corresponding Linear Assignment Problem (LAP) using the Hungarian algorithm [64]. Matched trajectory-tracklet pairs in the LAP solution are merged to extend the trajectory. Tracklets that have no matching trajectory are used to start new trajectories. Trajectories for which there is no matching tracklet have their "health" decremented. When a trajectory's healthdrops to zero, it is terminated. Trajectories that still exist at the end of this stage become the new trajectory set for another round of data association with tracklets in the next sliding window, and so on. The result of this forward scan procedure is the merging of multiple overlapping tracklets into a set of longer individual trajectories. An actual merge between two contiguous trajectories is a simple average of spatial locations. When two noncontiguous trajectories are merged, the locations in the gap between the two are computed by linear interpolation.

IV. Identifying Small Groups

In this section, we present a clustering approach that hypothesizes small groups traveling together using the notion of group "entitativity" [65], defined in terms of criteria from Gestalt psychology: common fate (same or interrelated outcomes), similarity (in appearance or beha- viors), proximity, and pregnance (patterning). Given a set of automatically extracted pedestrian trajectories, we identify potential groups within a sliding time window using hierarchical clustering based on robust measures computed from the noisy trajectories.

Our automatic grouping algorithm is inspired by McPhail and Wohlstein [57], who present the only objective measure we know of in the social science literature to determine

which people are traveling together through the scene. In [57], group membership is determined via a cascaded set of three tests: 1) Any two people who are within 7 feet of each

other and not separated by another individual are consid- ered to be contiguous and pass on to the next test; 2) any two contiguous people whose speeds are the same to within 0.5 feet per second are judged to have the same speed and pass on to the next test; and 3) any two contiguous people traveling at the same speed whose directions of motion are the same to within 3 degrees are judged to have the same direction. A group-expand procedure is also defined to test whether a new individual should be added to an existing group. Note that in [57], these tests are applied by human observers who analyze frames of video offline.

4.1 Measurements

Consider the trajectory of a person in the scene as a set of tuples ðs; v; tÞ, where s is the position vector of the tracked person's centroid (projected into the ground plane using a homography estimated offline) and v is the velocity vector at frame t. Let be the temporal overlap of the trajectories between persons i and j within a temporal window sequences alone are not adequate to model behavioral where is the mean configuration, P is the matrix of correlations between group members.

Statistical shape modeling. In order to study correlated patterns, we use a statistical shape analysis method [71] to analyze the spatial position of all group members jointly and estimate the typical group formations of walking pedestrians, which we refer to as group configurations. A group configuration S at a particular time consists of a point set of member K dominant eigenvectors associated with the K largest eigenvalues of the covariance matrix, and b is a vector of K model parameters.

Principal component analysis is applied to the covariance matrix H ¼ $\hat{S}\hat{S}^{0}$ to study the joint variation in the group configuration samples. We model each configuration.

References

[1] A. Hoogs and A.G.A. Perera, "Video Activity Recognition in the Real World," Proc. Nat'l Conf. Artificial Intelligence, pp. 1551-1554, 2008.

[2] X. Wang, K. Tieu, and E. Grimson, "Learning Semantic Scene Models by Trajectory Analysis," Proc. European Conf. Computer Vision, pp. 111-123, 2006.

[3] R.W. Brown, "Mass Phenomena," Handbook of Social Psychology, G. Lindzey, ed., vol. 2, pp. 833-876, Addison Wesley, 1954."Crowd Dynamics," PhD thesis, Univ. of Warwick, 2000.

[5] D. Cartwright and A. Zander, Group Dynamics: Research and Theory, third ed. Harper, 1968.

[6] N.R. Johnson, "Panic at the Who Concert Stampede: An Empirical Assessment," Social Problems, vol. 34, pp. 362-373, 1987.

[7] A. Aveni, "The Not-So-Lonely Crowd: Friendship Groups in Collective Behavior," Sociometry, vol. 49, pp. 96-99, 1977.

[8] C. McPhail, "Withs across the Life Course of Temporary Sport Gatherings," unpublished manuscript, Univ. of Illinois, 2003.

[9] T. Moeslund, A. Hilton, and V. Kruger, "A Survey of Advances in Vision-Based Human Motion Capture and Analysis," Computer Vision and Image Understanding, vol. 103, nos. 2/3, pp. 90-126, Nov.