

Online Pedestrian Group Walking Event Detection Using Spectral Analysis of Motion Similarity Graph

Vahid Bastani, Damian Campo, Lucio Marcenaro and Carlo Regazzoni
University of Genoa, DITEN

Via all'Opera Pia, 11A - 16145 Genova (GE)

{vahid.bastani, damian.campo}@ginevra.dibe.unige.it, {lucio.marcenaro, carlo.regazzoni}@unige.it

Abstract

A method for online identification of group of moving objects in the video is proposed in this paper. This method at each frame identifies group of tracked objects with similar local instantaneous motion pattern using spectral clustering on motion similarity graph. Then, the output of the algorithm is used to detect the event of more than two object moving together as required by PETS2015 challenge. The performance of the algorithm is evaluated on the PETS2015 dataset.

1. Introduction

In the video surveillance applications, video analysis and scene understanding usually involve object detection, tracking and behavior recognition [8]. A particularly important task in these applications is crowd analysis, which has been a field of great interest in computer vision and cognitive science. The crowd phenomenon has been identified as a topic of great interest in a large number of applications such as crowd management, public space design, visual surveillance and intelligent environments [17].

Detection of events has gained lots of attention in video surveillance domain, it is the case of research related to the identification of lying pose recognition [15, 13], detection of people running [4, 19], crowd safety [7, 16], etc. In videos that include civil interactions, modeling the social behaviors of people plays an important role in describing the individual and group behaviors on crowded scenes [18].

This work focuses on the understanding of events in video sequences where there are more than one person involved and groups of people walking together are conformed. Group event detection is an important application in automatic video surveillance for understanding situations that involve bunches of people doing a common activity. Identification of pedestrian groups is a key topic in crowd monitoring for the automatic recognition of anomalous be-

haviors that can threaten the safety of a place. For defining pedestrian groups, it is necessary to define the notion of interaction among people.

The work described in this paper studies the clustering of pedestrians into groups base on their local motion features. The output can be used for detection of events when there are pedestrian groups of more than two people in the scene. Bounding boxes represent the observation of pedestrians on each frame. The position and dimensions of bounding boxes and the speed of change of these features is tracked using Kalman filters for each object. The filtered state density of Kalman filter is used to form motion similarity graph that represent at each frame how close are the motions of pairs of pedestrians. Then, using spectral clustering techniques the group of people are identified as connected components of the graph. The proposed algorithm is tested on six videos from a dataset provided by PETS 2015 and compared with ground-truth.

2. Proposed Method

The task of detecting objects moving together in the video can be viewed as determining whether they move in a same local trajectory pattern. A simple yet effective way for modeling trajectories is through flow functions [11, 10, 2, 6]. In these approaches a trajectory pattern is characterized by a flow function that basically shows the speed field at each point in the environment for that pattern. Given a set of flow function corresponding to a trajectory classes, it is possible to determine the trajectory pattern class of moving objects. However, in group walking event detection task there is no flow function available beforehand. In this case the problem is to understand if the motion of two contemporary objects belong to a same flow function or not.

In this section a method is introduced for online clustering of moving pedestrians based on their motion pattern. the clustering output then is used to detect the event where more than two people walking together. To this end, first

it is introduced how space-dependent flow can be estimated online for each moving object which is used for measuring the similarity of instantaneous motion pattern between objects. Then a graph will be made to represent the situation at each frame, where nodes are objects and edges are weighted by the similarity rate of motion patterns between two pair of objects. Finally, spectral graph analysis techniques are used for clustering object into groups with similar motion pattern.

2.1. Object Detection and Tracking

The input to the proposed algorithm at each frame is a set of bounding boxes of detected pedestrians in the scene. It is also assumed that each detection is identified using appropriate data association technique. Although this is a strong assumption, recent results on pedestrian detection [3] and multi-object tracking [1] have shown its feasibility. Let $Y_t = \{\mathbf{y}_t^1, \dots, \mathbf{y}_t^{N_t}\}$ be the set of observations extracted from frame t , where $\mathbf{y}_t^i = [x, y, w, h]^T$ is the observation vector of object i consisting of coordinates (x, y) and dimension (w, h) of its bounding box and N_t is the number of objects present in frame t . The flow of coordinates and dimension of bonding boxes have also great importance when we want to analyze the motion of an object. Therefore, it is reasonable to estimate coordinates and dimension together with their flow using the sequence of observations. Kalman filter can be incorporated for this task where the state of the filter is defined as $\mathbf{x}_t^i = [x, y, w, h, \dot{x}, \dot{y}, \dot{w}, \dot{h}]^T$ consisting of coordinates and dimension of object bounding box and respective flows.

Kalman filter at each frame provides the posterior filtered state density as a Gaussian distribution

$$\rho_t^i := p(\mathbf{x}_t^i | \mathbf{y}_{0:t}^i) = \mathcal{N}(\hat{\mathbf{x}}_{t|t}^i, \hat{P}_{t|t}^i) \quad (1)$$

where $\hat{\mathbf{x}}_{t|t}^i$ and $\hat{P}_{t|t}^i$ are the filtered state and its covariance matrix up to time t . Note that the state of Kalman filter in this case consists of position and flow of all four components. Each instance of the state vector is one sample from trajectory pattern flow function. For this reason we use the posterior state of Kalman filter to compute the similarity of the motion of objects at each frame.

2.2. Motion Pattern Similarity Measure

It is possible to use the euclidean distance between the estimated state vector of each object to measure their similarity. However, since the output of Kalman filter has the form of probability distribution, it is more effective to use distance metrics specific for probability distribution. Kullback-Leibler (KL) divergence is a measure of difference between probability distributions denoted by $D_{KL}(P||Q)$. It is available in closed form for Gaussian distributions [5]. However, KL is not a symmetric measure

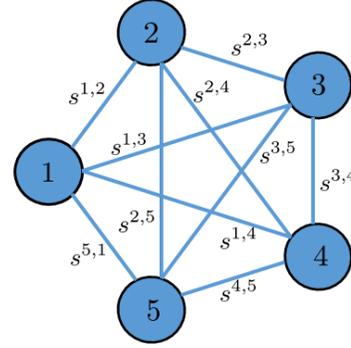


Figure 1: Example of similarity graph for 5 moving objects.

since $D_{KL}(P||Q) \neq D_{KL}(Q||P)$. Thus, here the symmetrized version of KL divergence is used to measure differences of two state distribution of moving objects:

$$d_t^{i,j} := D_{KL}(\rho_t^i || \rho_t^j) / 2 + D_{KL}(\rho_t^j || \rho_t^i) / 2. \quad (2)$$

$d_t^{i,j}$ is a positive value that increases as the difference between ρ_t^i and ρ_t^j becomes larger. A normalized similarity between two state vector distribution then is calculated as

$$s_t^{i,j} = e^{-d_t^{i,j} / k_t^{i,j}}. \quad (3)$$

$k_t^{i,j}$ is a scaling factor whose motivation is to compensate the effect of distance of objects from camera. Note that as objects move away from camera their speed and respective distance get lower in the image plane. Thus a global measure of similarity should be equalized as far as possible in order to give the same score for the same situation whether it is close or far from camera. The scaling factor is defined as

$$k_t^{i,j} = a(\sqrt{w_t^i h_t^i} + \sqrt{w_t^j h_t^j}) / 2 + b. \quad (4)$$

Which is a linear function of the mean of the square root of the area of two objects. Here the area of the objects is used to understand the relative distance of them from camera. a and b both are positive so that the scaling factor increases as the objects become close to camera. In the similarity measure (3), the scaling factor compensates the larger relative distance of objects when they are close to camera.

$s_t^{i,j}$ is one when ρ_t^i and ρ_t^j are completely similar and monotonically decreases toward zero as they become more different. The pairwise similarity scores then are used to form similarity graph such as the one shown in the Fig.1, which is an undirected graph whose nodes represent moving objects and edges are weighted according their motion similarity. The graph later will be used to understand relations between objects and detection of the event when more than two people walking together.

2.3. Group Walking Event Detection

The spectral clustering algorithm [9] is a simple yet effective algorithm for clustering data sets that can be represented using similarity graphs. It often outperforms many conventional algorithms. In the proposed algorithm we use spectral clustering on the generated motion similarity graph in order to find groups of objects moving in a same way. The adjacency matrix for a graph such as Fig.1 is the matrix $W_t = [s_t^{i,j}]_{i,j=1,\dots,N}$ [14]. The main quantity for spectral clustering is graph Laplacian matrix which is defined as

$$L_t = D_t - W_t. \quad (5)$$

where D_t is the graph degree matrix defined as a diagonal matrix whose i th diagonal element is the i th node degree $m_t^i = \sum_{j=1}^n s_t^{i,j}$. The Laplacian matrix can be used to find connected components of the graph. In the application of this paper the connected components of the graph represent groups of people walking together.

Let $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ be the eigenvalues of Laplacian matrix L_t and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ be corresponding eigenvectors, e.g. $L_t \mathbf{v}_i = \lambda_i \mathbf{v}_i$ for $i = 1, \dots, n$. If the number of connected components $m \leq n$ of the graph is known, the connected components and their corresponding nodes can be found using spectral clustering algorithm [9] as follows:

- Let U be the matrix whose columns are the first m eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_m$.
- For $i = 1, \dots, n$, let \mathbf{u}_i be the vector corresponding to i th row of U .
- Cluster vectors $\{\mathbf{u}_i\}_{i=1,\dots,n}$ into m clusters using k -means algorithm into clusters C_1, \dots, C_m .
- Return cluster indicator variables $z_t^i = \{j | \mathbf{u}_i \in C_j\}$ for $i = 1, \dots, n$.

The event of a group of people walking is then triggered when the number of members in at least one connected component (cluster) is greater than or equal to three.

The number of connected components (the number of groups of people in scene) however is not known in this problem. A particular way to estimate number of connected components is eigngap heuristic [14]. Here m is chosen such that all eigenvalues $\lambda_1, \dots, \lambda_m$ are small and λ_{m+1} is relatively large. This is shown in the example of Fig.2 where three first three eigenvalues are close to zero and there is a big gap from third to fourth eigenvalues, which shows that there are three connected components in the corresponding graph. Thus, to find m using eigngap heuristic the following procedure is done:

- Calculate $\gamma_i = \lambda_{i+1} - \lambda_i$ for $i = 1, \dots, n - 1$
- Find $m = \min i$ such that $\gamma_i \geq \frac{0.8}{n} \sum_{j=1}^{n-1} \gamma_j$

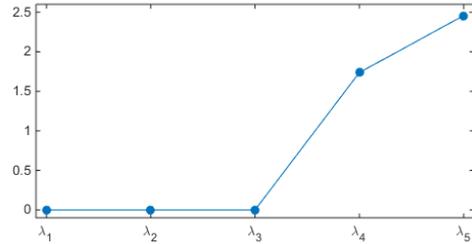


Figure 2: Example of Laplacian matrix eigenvalues for a graph with 5 nodes and 3 connected components. From motion similarity graph of frame 47 of sequence N1_ARENA-Gp_ENV_RGB_3.

The above procedure searches for the first gap in the sequence of eigenvalues. Since the first eigenvalue is always zero [14], the first gap shows the number of connected components in the graph.

3. Experimental results

The PETS2015 dataset¹ is used to evaluate the proposed algorithm. The data sets consist of three different camera views of two situations P5 and ARENA, in which a group of walking people appears in the scene. In ARENA sequences a group of three people appear in an environment in which there are some other individual pedestrians. This group eventually splits into three separate pedestrians. In P5 sequences, a group of six people are shown which eventually splits into one individual, a group of two and a group of three people while walking.

The detection of pedestrians done manually using ViperGT tool². The bounding boxes of objects then are feed sequentially to the algorithm. The id of objects are also passed to algorithm in order to bypass data association problem. For evaluation purpose a ground-truth of group indexes of objects in each frame of video is made by human observer. The ground-truth group (cluster) indicator variables denoted by g_t^i is compared with the algorithm output z_t^i . Since the problem is formulated here as a clustering problem, it is possible to use clustering performance measures such as Adjacent Mutual Information (AMI) [12] to evaluate the algorithm. Given the ground-truth of the cluster assignments this metric compares the result of clustering by returning values in range [0 1]. AMI Values close to zero means two assignments are highly independent and AMI values close to one indicates match between two indexes. The AMI score is calculated at each frame t by comparing $\{g_t^i\}_{i=1}^n$ with $\{z_t^i\}_{i=1}^n$. The mean AMI value for every frame of each sequence then is reported here.

¹<http://pets2015.net/>

²<http://viper-toolkit.sourceforge.net/>

The measurement noise covariance matrix in Kalman filter is set to $10 \times \mathbf{I}_4$, and the process noise covariance matrix is set to

$$\begin{bmatrix} 10 \times \mathbf{I}_4 & \mathbf{0} \\ \mathbf{0} & 2 \times \mathbf{I}_4 \end{bmatrix}.$$

The algorithm is evaluated for different choices of parameters a and b of (4) to see how its behavior changes with different parameter setting. Parameter a is chosen from set $\{2, 4, 6, 8, 10\}$ and parameter b is chosen from set $\{10, 100, 1000\}$. Fig 3 shows snapshots of the output of the algorithm where objects that are in same group have same color bounding boxes. The quantitative results are depicted in Fig. 4 where for each sequence the value of mean AMI is plotted versus the value of parameter a for different value parameter b . As can be observed from Fig. 4 the performance depends on the situation, camera view and parameters setting. However, the performances in any case is reasonably good since the value of mean AMI is close to one. The best average AMI score over all sequences is 0.8566, which is achieved for parameter set $a = 8$ and $b = 10$.

4. Conclusions

In this paper a method for online clustering of walking people is proposed for detecting of the event when more than two people are walking together. The method is based on measuring the similarity of the motion patterns of pairs of moving objects in the scene to form a motion similarity graph. The graph then is used to cluster objects based on spectral clustering algorithm. Experimental results show that the proposed method is able to identify separate walking groups efficiently. This method however is instantaneous and does not take into account the history of objects. An improvement can be achieved by applying probabilistic filtering like Hidden Markov Model (HMM) on the output group indexes to eliminate sporadic joining and splitting of groups.

References

- [1] S.-H. Bae and K.-J. Yoon. Robust online multiobject tracking with data association and track management. *Image Processing, IEEE Transactions on*, 23(7):2820–2833, July 2014.
- [2] V. Bastani, L. Marcenaro, and C. Regazzoni. A Particle Filter Based Sequential Trajectory Classifier for Behaviour Analysis In Video Surveillance. In *2015 IEEE International Conference on Image Processing ICIP*, Quebec City, 2015.
- [3] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(4):743–761, April 2012.
- [4] S. Fouche, M. Lalonde, and L. Gagnon. A system for airport surveillance: detection of people running, abandoned objects, and pointing gestures. In *SPIE 8056, Visual Information Processing XX*, page 8, June 2011.

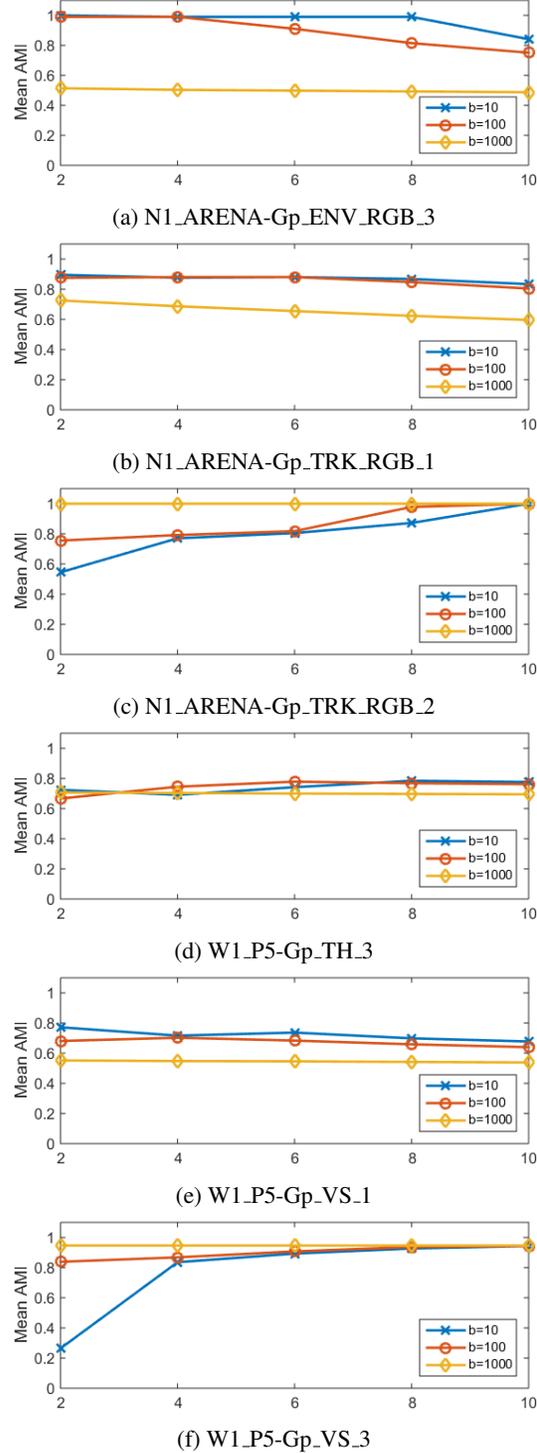


Figure 4: Mean AMI value versus different choices of parameters a (horizontal axis) and b for six sequences from P5 and ARENA dataset of PETS2015.

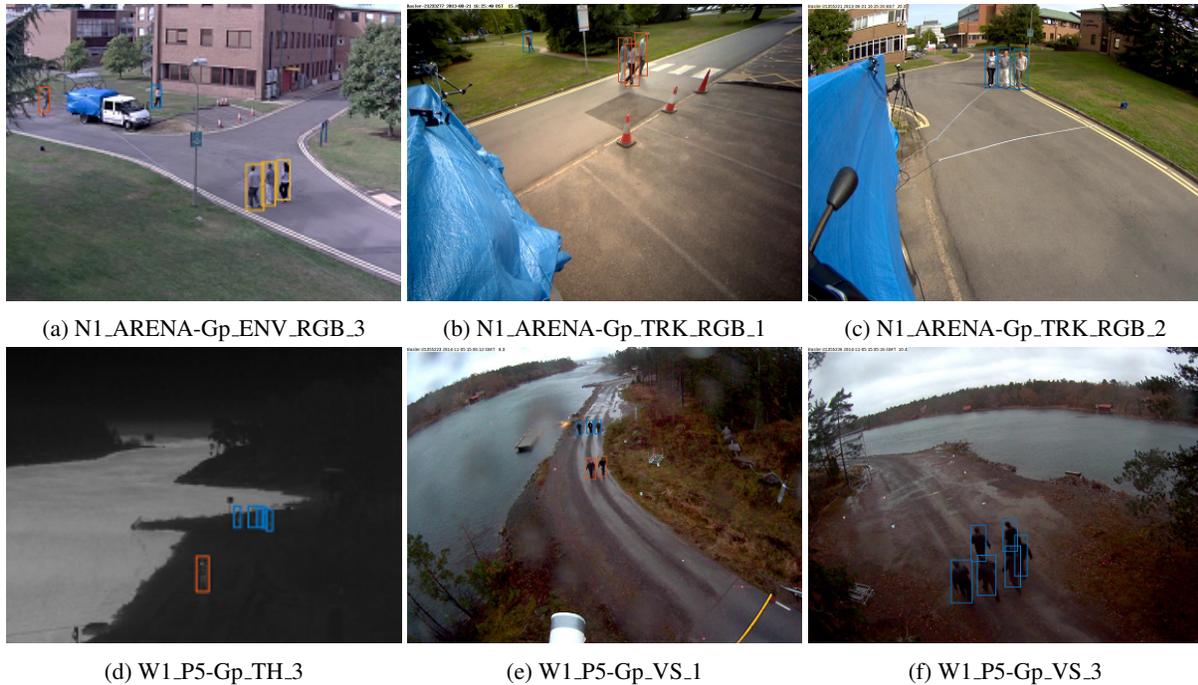


Figure 3: Snapshots of the output of the algorithm from P5 and ARENA sequences. Same color boxes represent same group.

- [5] J. Hershey and P. Olsen. Approximating the kullback leibler divergence between gaussian mixture models. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 4, pages IV-317-IV-320, April 2007.
- [6] K. Kim, D. Lee, and I. Essa. Gaussian process regression flow for analysis of motion trajectories. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1164-1171, Nov 2011.
- [7] R. Knecke and V. Schneider. Basigo: Safety of large scale events crowd flow modeling of ingress and egress scenarios. *Transportation Research Procedia*, 2(0):501 - 506, 2014. The Conference on Pedestrian and Evacuation Dynamics 2014 (PED 2014), 22-24 October 2014, Delft, The Netherlands.
- [8] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, and S. Yan. Crowded scene analysis: A survey. *Circuits and Systems for Video Technology, IEEE Transactions on*, 25(3):367-386, March 2015.
- [9] J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888-905, 2000.
- [10] J. Nascimento, M. Figueiredo, and J. Marques. Activity recognition using a mixture of vector fields. *Image Processing, IEEE Transactions on*, 22(5):1712-1725, May 2013.
- [11] J. C. Nascimento, M. Figueiredo, and J. S. Marques. Trajectory classification using switched dynamical hidden Markov models. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 19(5):1338-48, May 2010.
- [12] N. X. Vinh, J. Epps, and J. Bailey. Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance. *The Journal of Machine Learning Research*, 11:2837-2854, Mar. 2010.
- [13] M. Volkhardt, F. Schneemann, and H.-M. Gross. Fallen person detection for mobile robots using 3d depth data. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 3573-3578, Oct 2013.
- [14] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395-416, 2007.
- [15] S. Wang, S. Zabir, and B. Leibe. Lying pose recognition for elderly fall detection. In *Proceedings of Robotics: Science and Systems*, Los Angeles, CA, USA, June 2011.
- [16] H. Yin, D. Li, and X. Zheng. An energy based method to measure the crowd safety. *Transportation Research Procedia*, 2(0):691 - 696, 2014. The Conference on Pedestrian and Evacuation Dynamics 2014 (PED 2014), 22-24 October 2014, Delft, The Netherlands.
- [17] B. Zhan, D. Monekosso, P. Remagnino, S. Velastin, and L.-Q. Xu. Crowd analysis: a survey. *Machine Vision and Applications*, 19(5-6):345-357, 2008.
- [18] Y. Zhang, L. Qin, H. Yao, and Q. Huang. Abnormal crowd behavior detection based on social attribute-aware force model. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 2689-2692, Sept 2012.
- [19] Y.-Y. Zhu, Y.-Y. Zhu, W. Zhen-Kun, W.-S. Chen, and Q. Huang. Detection and recognition of abnormal running behavior in surveillance video. *Mathematical Problems in Engineering*, 2012:14, 2012.