Formation period matters: Towards socially consistent group detection via dense subgraph seeking

Yanhao Zhang[⋆], Lei Qin[†], Shengping Zhang[⋆], Hongxun Yao[⋆], Qingming Huang[⋆†]
[⋆]Harbin Institute of Technology, Harbin, 150001, China
[†]Inst. of Comput. Tech., Chinese Academy of Sciences, Beijing, 100190, China
{yhzhang,h.yao,s.zhang}@hit.edu.cn {lqin,qmhuang}@jdl.ac.cn

ABSTRACT

Group detection becomes an important task in crowd behavior surveillance. However, most existing methods ignore the formation persistency characteristics, which predict unreliable interactions when the crowd is realistic and complex. To address this issue, we propose a novel graph-based method to declare that the formation period really matters for detecting social groups in crowd. First, we develop a socially motivated representation by modeling the formation period probability in a Bayesian manner, which results in social and temporal consistency for group member interactions. A graph is then established using individuals as nodes and formation periods as edge weights to reflect pedestrian relationships. In this way, seeking of socially consistent groups is converted into an optimization problem which seeks dense subgraphs with maximum formation likelihood within the graph structure. We employ graph shift optimization to detect groups by finding all the dense subgraphs due to its robust performance. In the experimental results on public datasets, our proposed method clearly outperforms other related state-of-the-art methods.

Categories and Subject Descriptors

H.3.1 [**Information Systems Applications**]: Content Analysis and Indexing methods

General Terms

Algorithms, Social factors, Experimentation.

Keywords

Social group, Formation period, Dense subgraph seeking

1. INTRODUCTION

Crowd activity analysis plays a significant role in video surveillance fields with diverse applications including public safety management and activity understanding for predicting crimes. Understanding complex interactions in crowded scenes just by treating each pedestrian individually is unrealistic, due to the inherent social characteristics of human behavior. Typically, interactions involve in small subsets of a crowd, at neither an individual level nor in a holistic level. Many existing works on detecting groups tend to find people getting together based on the persistence of their proximity and direction of motion [8, 3, 7, 10, 2, 1, 12, 11], which are measurable from the extracted trajectories. Pellegrini et al. [8] jointly predict trajectories and estimate group memberships by employing a Conditional Random Field (CRF), which is modeled as latent variables over a short time window. Yamaguchi et al. [10] formulate the problem of predicting groups as minimization of an energy function which encodes physical condition, personal motivation and pairwise interactions features. Ge et al. [3] present an agglomerative approach to associate trajectories, which hierarchically merges clusters by evaluating an inter-group closeness measure. Chang et al. [1] propose a soft segmentation process to partition the crowd by constructing a weighted graph, where the edges represent the probability of individuals belonging to the same group.

However, existing works are lack of sociological motivation in the choice of features. Furthermore, they tend to understand the groups without considering the social consistency of formation behavior. To give an intuitive interpretation, social formation among small gathering of people sharing a common goal would indicate interaction behavior at group level. It is preserved by a dynamic process of trajectories, which can be inferred more from the group formation and the temporal evolution among their relationships. To address these issues, we underline formation characteristics inspired by social studies to declare that formation period really matters for detecting the social groups. We summarize our contributions as:

1) We present a socially consistent group detection method by exploiting the temporal formation representation for the social interactions. To our best knowledge, this is the first effort towards improving social consistency by formation period modeling.

2) Considering social cues of the groups, our proposed formation period is capable of capturing group structures and compositions of social interactions, which makes it robust and greatly invariant to represent the group behaviors.

3) We propose to construct a graph between the individuals by using formation period probability. We seek the dense subgraphs as the desired social groups within the graph structure that reflects socially consistent formation period in the crowd.

ICMR'15, June 23–26, 2015, Shanghai, China.

DOI: http://dx.doi.org/10.1145/2671188.2749305.

2. SOCIAL GROUP ANALYSIS

In this section, we analyze the social features by exploring the social formation priors in Bayesian formulation. We present a formation period representation which is competent for social group detection, especially when the distance and angle cues show inadequate capabilities for indicating social interactions.

2.1 Formation probability modeling

Given a frame $t \in \{1, \cdots, T\}$ from a crowd sequence, each tracked subject $i \in \{1, \cdots, N\}$ corresponds to one local trajectory $H_i \in \mathcal{H}$ of a person. H_i consists of T tuples $h_i^t = \{s_i^t, v_i^t, \alpha_i^t\}^{t=1 \dots T}$, where s_i and v_i denote the position and velocity vectors of the subject i with the estimated orientation α_i at time t. Let $\mathcal{F}$ denote the set of formation within the interacting individuals, $\mathcal{G}$ be the set of all possible social groups along the interacting time $\mathcal{T}$, where each element $g \in \mathcal{G}$ is a binary vector indicating the group relationship among the subjects i and j. Our goal is to find an optimal group clusters $\mathcal{G}^*$ from $\mathcal{G}$ which best fits a given interaction time and formation set $\mathcal{T}$ and $\mathcal{F}$. To this end, we model the probability of the group indicator $\mathcal{G}$ based on $\mathcal{F}$ and $\mathcal{T}$ by considering both the spatial formation of interaction and the temporal overlap of period in a Bayesian manner as

$$p(\mathcal{G}|\mathcal{F},\mathcal{T}) \propto \underbrace{p_s(\mathcal{F}|\mathcal{G})}_{Formation} \underbrace{p_t(\mathcal{F}|\mathcal{G},\mathcal{T})}_{Period} \underbrace{p(\mathcal{G}|\mathcal{F})}_{Prior} \quad (1)$$

where $p_s(\mathcal{F}|\mathcal{G})$ measures the spatial consistency between the assignment of group $\mathcal{G}$ and the given formation set $\mathcal{F}$. $p_t(\mathcal{F}|\mathcal{G},\mathcal{T})$ measures the "period" properties of $\mathcal{G}$, which indicates the temporally pairwise interactions. Traditionally, this item can be presented as $p_t(\mathcal{T}|\mathcal{G})$, giving the likelihood that the happening of social group is based on the time of getting together. Now, the interacting time $\mathcal{T}$ is not treated equally and is instead valued by the social formation set $\mathcal{F}$. Therefore, the new likelihood $p_t(\mathcal{F}|\mathcal{G},\mathcal{T})$ can be interpreted as the probability of social formation along the interacting time $\mathcal{T}$. The last term $p(\mathcal{G}|\mathcal{F})$ is the prior probability of the formation for the groups that we regard equally for all candidates. When we maximize the probability of Eqn. 1, obviously, we expect to obtain the social groups which have the spatially consistent formation with the social formation encoded in the formation set $\mathcal{F}$ and simultaneously satisfy the pairwise interacting period.

2.2 Socially consistent feature

Formally, social aggregation theories [4] project the pedestrian trajectories onto the space which provides us useful group concepts of sustaining mutual activities. The nonlinearity of the social interaction is correlated with a quantization of their mutual distance into personal, social and public space as F-formation. It is considered as a specific instance of a focused encounter as all the participants co-operate to sustain a continued period of information exchange. The idea behind this is that spatial and orientational behaviors are created by people who sustain the shared interaction space between them. To this aim, we define the proximity based probability $\phi_{ij}^{prox}(h_i, h_j|g)$ and orientation based probability $\psi_{ij}^{ori}(h_i, h_j|g)$ to give the detailed semantics of the formation. An illustration of formation period is shown in Fig. 1. The spatial formation can be measured

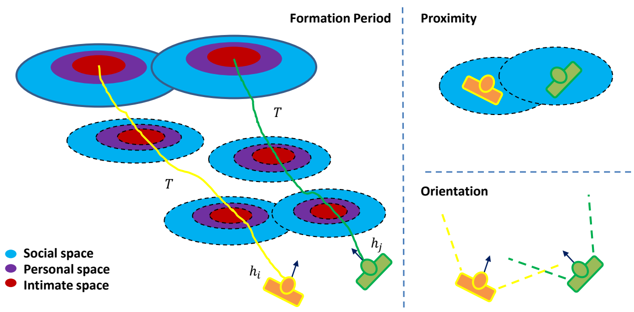


Figure 1: Formation period considers the formation of interaction in terms of both proximity and orientation on the temporal correlation.

by computing the position and angle with respect to local trajectories set $\mathcal{H}$, which is similar as the trajectories and grouping associating process [8] as a log likelihood,

$$log(p_s(\mathcal{F}|\mathcal{G})) = \sum \phi_{ij}^{prox}(h_i, h_j|g) + \sum \psi_{ij}^{ori}(h_i, h_j|g) \quad (2)$$

The proximity probability is defined as

$$\sum \phi_{ij}^{prox}(h_i, h_j|g) = \sum \log[p_{mov}(s_i|s_j\alpha_j, g_{ij} = 1)$$

$$+p_{stop}(d(s_i, s_j)|g_{ij} = 1)] - \sum \log(p(d(s_i, s_j)|g_{ij} = 0) \quad (3)$$

where $d(s_i, s_j)$ is the Euclidean distance between two positions. Corresponding to moving and stopping people in the same group, two modes of probabilities p_{mov} and p_{stop} are estimated by building a normalized histogram of the angles and speeds in the training set [8]. From the statistics, it is easy to distinguish two modes. For pedestrians in different groups, position is enough to tell apart the interaction in the public space. In terms of orientation probability, it is defined as

$$\sum \psi_{ij}^{ori}(h_i,h_j|g) = \sum p(\alpha_i,\alpha_j|g_{ij}=1) - \sum p(\alpha_i,\alpha_j|g_{ij}=0) \quad (4)$$

where the probabilities are based on whether the pedestrians have the same orientational range or not. The focused orientation can help to identify the shared space between all the formation members. This is relevant that pedestrian should be oriented towards an F-formation [2]. As before, all the estimated terms are computed with a smoothed histogram approach overall a temporal window.

Under the objective of obtaining the interaction period, we measure the temporal overlap of the trajectories between the person i and j within a temporal interaction $\mathcal{T}$. We denote the distance measurement of the social formation for each pair h_i and h_j directly as the likelihood

$$log(p_t(\mathcal{F}|\mathcal{G},\mathcal{T})) = \sum_t p_t(\mathcal{F}|g,\mathcal{T}) \propto \sum_t \delta_t(i,j) \quad (5)$$

where $\delta_t(i,j) = 1$ indicates $h_{i,j}$ are in the same group if $\|s_i^t - s_j^t\| \le r_s$ and $\|v_i^t - v_j^t\| \le r_v$, otherwise $\delta_t(i,j) = 0$. This period probability $p_t(\mathcal{F}|\mathcal{G},\mathcal{T})$ is the occurrences between person i and j within $\mathcal{T}$ that the spatial and velocity differences are below the thresholds r_s and r_v. This strategy makes grouping people walking close to each other with similar velocities capable of interacting for a long period of time, which leads to temporal consistency to get stable groups over time and meanwhile guarantees its social formation.

3. FORMATION PERIOD SUBGRAPHS

In this section, we formulate the formation probability modeled in Sec. 2.1 into the graph shift framework, where the problem of maximizing the probability of Eqn. 1 over the possible groups $\mathcal{G}$ is converted to the form of maximizing the density of subgraphs problem. We extend the idea of finding dense subgraph in the social formation period domain to detect local social groups.

3.1 Consistency Graph Construction

A dense subgraph is a strongly connected subset of vertexes in a weighted undirected complete graph. The undirected graph is constructed by using the individuals as the graph nodes and the formation period probabilities as the edge weights. In this way, the valid formation period that is socially consistent would form a dense subgraph and could be robustly detected by the method of graph shift. Given all the trajectories extracted from the individuals, $H(h_1, ..., h_n)$ with the individual number n, we conduct formation period correlation in the product space $Q = |H| \times |H|$ and obtain $n \times n$ pairs. In order to obtain a continuous and more general measure, we substitute the original quantization with an exponential function and the proximal similarity of any trajectory pair h_i and h_j is computed as follows,

$$f_1(h_i, h_j) = \frac{1}{T}\sum_t \delta_t(i,j) \times e^{p_s(\mathcal{F}|\mathcal{G})} \tag{6}$$

Thus, a T-frame video sequence can be represented by a trajectory graph with 3-tuple $G = (V, E, w)$, in which V is a set of vertices, $E \subseteq V \times V$ is a set of edges, w is the edge weights assigning to E. We next try to specify the detailed description based on the basic graphs. The adjacency matrix representation $A(i, j)$ for the graph G is as follows:

$$A(i,j) = \begin{cases} 0 & i = j \\ f_1(h_i, h_j) & i \neq j \end{cases} \tag{7}$$

We can calculate the formation period similarity by Eqn. 7. The function of the similarity value can be denoted as Eqn. 6 which encodes the evaluation of temporal consistence and social formation correlation.

3.2 Detection by dense subgraph seeking

Obviously, the graph G is symmetric and nonnegative. Suppose a social group has r trajectory-pairs. It corresponds to a dense subgraph D of G with r vertices, which is a weighted counterpart of maximal clique. If we represent a subgraph by a vector x, $x \in \Delta^n$, where $\Delta^n = \{x \in R^n : x \geq 0 \,\&\, |x|_1 = 1\}$. x_i, the ith component of x, denotes the probability of this subgraph containing the vertex i. The modes of a graph G are approximately local maximizers of the graph density $g(x) = x^T Ax$. We can derive that $g(x) = \sum_i x_i(Ax)_i$. Given a vector x, the corresponding subgraph is $G(x)$. If x^* is a local maximizer of $g(x)$, the $G(x^*)$ is a dense subgraph which represents the social group. Therefore we need to calculate all local maximizers of this standard quadratic optimization function:

$$\begin{cases} maximize & g(x) = x^T Ax \\ subject\,to & x \in \Delta^n \end{cases} \tag{8}$$

Table 1: Datasets descriptions.

Dataset	Length	Individuals	Groups	Density	Type
eth	8m 40s	386	81	low	few interactions
hotel	7m 23s	287	55	low	complex interactions
stu003	3m 40s	417	112	high	more complex behaviours

It can be solved by the replicator dynamics with neighborhood expansion in graph shift optimization [6],

$$x_i(t+1) = x_i(t)\frac{(Ax(t))_i}{x(t)^T Ax(t)}, i = 1, \dots, n \tag{9}$$

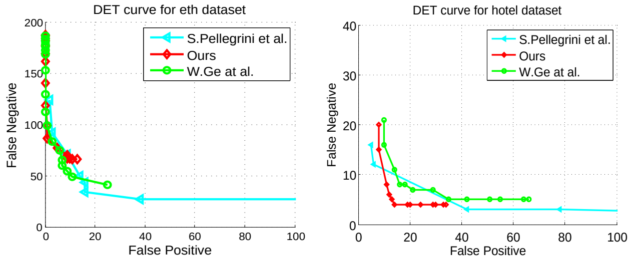
where $x_i(t)$ is a mode of subgraph obtained at the ith iteration. Note that the vertexes (which correspond to the candidate formation period of individuals) contained in the dense subgraph x^* are the most likely to be the valid social groups, and the value of the non-zero bin x^* now indicates the probability that a candidate formation period belongs to the socially consistent group. We conclude the merits of detecting the groups in the crowd via the consistency graph model as follows: 1) Formation periods naturally capture the social inspired knowledge, which robustly reject most of the invalid trajectory interactions, hence are effective in leveraging the influence of formation periods. Meanwhile, the formation period probability enables us to measure the group formation in a more robust and accurate way than traditional similarity of trajectories. 2) The group detection is converted to dense subgraph seeking, which is reasonable because the dense subgraph is a strongly connected subset of vertexes in a graph. According to the definition of the weights on graph edges, the correspondence between the clusters from the same similar connection has the most graph density than that from different connections or noises. 3) Our approach could handle the complex group formation situation during a temporal overlap. When a group is forming social interaction, there exist dense subgraphs corresponding to these candidate formation periods. The problem of finding dense subgraphs is easy to be solved by the graph shift algorithm. So we can obtain high robustness and accuracy of validly social formation group.

4. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of our method, we conduct group detection experiments on three public datasets including the eth and hotel sequences from the BIWI Walking Pedestrians dataset [7] as well as the stu003 sequence from Crowds-By-Examples (CBE) dataset [5].

Datasets. BIWI dataset records two low crowded scenes outside a university and a bus station, while the CBE dataset records a high density crowd outside a university. Table 1 shows the detailed information of these datasets. All sequences come with a group annotation and the trajectory annotations are completed by marking the position of each individual every a few frames. These sequences are particularly challenging ranging from low to high density and various level interactions with complex condition due to low image resolution, perspective changes and cast shadows. We split the eth and hotel sequences as training set (about 2000 frames) to train the statistics features and parameters for formation period and the rest frames for testing set.

Evaluation Protocol and Analysis. To evaluate the performance of the socially consistent formation period, we compare our method with the state-of-the-art methods. We provide two evaluation approaches to justify the performance of the compared methods:



(a) (b)

Figure 2: DET curves for eth (a) and hotel (b) datasets compared with others.

(1) DET curves for pairwise links. We compare with the interaction time based approaches [8, 3] on the relatively sparse and simple BIWI eth and hotel dataset. All three methods produce similarly good results as groups are isolated and easy to tell as shown in Fig. 2. In each approach, the detected groups are compared with the ground truth. The DET curves are generated by varying the parameters of length of interaction to report false positive (when the pair is assigned to the same group which is not in the same group in the ground truth) and false negative (when the method fails to predict the pair as the same group). This clearly demonstrates the effectiveness of exploiting socially consistent interaction in the task of pairwise group relation prediction. Note that the distance and velocity based measure (Pellegrini et al. [8]) is not strong enough to distinguish group relations and produces many false links. Ge et al. [3] performs as well as ours except false positives increase when using longer trajectories. Our proposed formation period achieves the best in this less complex scene. This can be attributed to the fact that we not only measure proximity but also consider the orientational formation of the pedestrians.

(2) The precision and recall of the social group detection. The stu003 sequence is much more crowded and complex, which can be used to test the robustness for complex social interaction. To gain intuition into robustness and effectiveness of our method, we compare our results with the state-of-the-art methods [8, 3, 10, 9]. We run every model on the dataset for evaluating the impact on performances in terms of precision, recall and F-measure as shown in Table 2. Notice that the social interaction based approaches [10, 9] bridge the gaps between the social interaction and the pairwise link in finding groups, among which our method gets a significant improvement than others. This is due to sociological aspect from formation period, which is able to better generalize to describe social interaction and group relations. In addition, the subgraph seeking process enables us to measure the formation period probability in a more robust and accurate way.

Parameter tuning and Complexity. The experiments are conducted by varying the respective formation period threshold r_s and r_v, which indicate the difference between the position and velocity. These thresholds justified by the precision and recall comparison for stu003 are shown in Fig. 3, where r_s and r_v are ranging from (1,3) and (0.1,0.9), respectively. Low values bring noisy formation time which leads to large false negative detection. In general, when the value of r_s is higher than 2.5, or the r_v is smaller than 0.5, the performance will be affected. From the parameter analysis, we choose $r_s = 2$ and $r_v = 0.5$ as reasonable compromise between precision and responsiveness. Our formation period

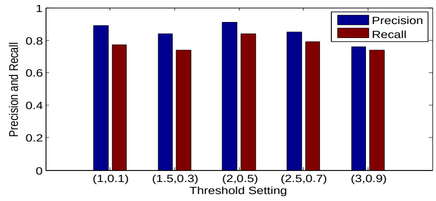


Figure 3: Precision and Recall for stu003 by varying formation period threshold r_s and r_v.

Table 2: Social group detection on stu003 dataset.

stu003 dataset	Precision	Recall	F-measure
Ours	**91.4%**	**83.4%**	**87.2%**
S.Pellegrini et al. [8]	46.0%	82.0%	58.9%
K.Yamaguchi et al. [10]	80.5%	77.0%	80.8%
W.Ge et al. [3]	51.8%	81.7%	63.4%
J.Sochman et al. [9]	85.4%	61.4%	71.3%

can not only obtain outstanding performances when used as the feature representation but also show to be more robust and turned out to be fast with the computational complexity being $O(n \times m)$ and $O(n)$ for computing $A(i, j)$.

5. CONCLUSION

In this paper, we have presented a socially consistent group detection method that represents formation period of group members in terms of social consistency. The formation and temporal information are integrated and encoded into the established graph in a Bayesian manner. Based on a graph relationship constructed with formation period probability features, the seeking of socially consistent group is casted as a dense subgraph seeking problem over the graph structure. Experimental results on three public datasets have confirmed the effectiveness and robustness of the proposed method. For future work, we will explore more promising descriptors for crowds in terms of social consistency.

6. ACKNOWLEDGMENTS

This work was supported in part by National Basic Research Program of China(973 Program): 2012CB316400 and 2015CB351802, in part by National Natural Science Foundation of China: 61133003, 61332016, 61300111, 61390510.

7. REFERENCES

[1] M.-C. Chang et al. Probabilistic group-level motion analysis and scenario recognition. In *ICCV*, 2011.

[2] M. Cristani and Bazzani. Social interaction discovery by statistical analysis of f-formations. In *BMVC*, 2011.

[3] W. Ge, R. Collins, and B. Ruback. Automatically detecting the small group structure of a crowd. In *WACV*, 2009.

[4] E. T. Hall. *The hidden dimension*. Anchor Books New York, 1969.

[5] A. Lerner, Y. Chrysanthou, and D. Lischinski. Crowds by example. In *EUROGRAPHICS*, 2007.

[6] H. Liu and S. Yan. Robust graph mode seeking by graph shift. In *ICML*, 2010.

[7] S. Pellegrini et al. You'll never walk alone: Modeling social behavior for multi-target tracking. *ICCV*, 2009.

[8] S. Pellegrini et al. Improving data association by joint modeling of pedestrian trajectories and groupings. In *ECCV*. 2010.

[9] J. Sochman et al. Who knows who-inverting the social force model for finding groups. In *ICCV Workshops*, 2011.

[10] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg. Who are you with and where are you going? In *CVPR*, 2011.

[11] Y. Zhang et al. Social attribute-aware force model: Exploiting richness of interaction for abnormal crowd detection. *IEEE TCSVT*, 2014.

[12] Y. Zhang et al. Beyond particle flow: Bag of trajectory graphs for dense crowd event recognition. In *IEEE ICIP*, 2013.