Small Group Human Activity Recognition

BMVC 2011 Submission # 573

Abstract

Small human group activity recognition has attracted much attention in recent years, since human activities often represent as small groups in public surveillance systems. Comparing to single human activity recognition or crowd analysis, small human group activity recognition is much more challenging due to mutual occlusions between different people, the varying group size, and inter or intra group interactions. In this paper, we propose a novel structural feature set to represent group behavior as well as a probabilistic framework for group activity learning and recognition. We first apply a robust multiple targets tracking algorithm to track each individual in the entire image region. Small groups are then clustered based on the output positions of the tracker. After that, we introduce a set of social network analysis based structural features to describe the dynamic behavior of small group people in each frame. A Maximum Gaussian Process Dynamical Model(MGPDM) is then employed to learn the temporal activity of small group people overtime. After training, the testing group activity will be identified as the action with the highest conditional probability with respect to each trained activity model. Our experimental results indicate that the proposed features and behavior model can successfully capture both the spatial and temporal dynamics of group people behavior, and correctly identify different small human group activities.

1 Introduction

Human behavior recognition has been studied for decades in the computer vision filed, as it can be applied in many surveillance systems. Most human action recognition research focus on single human action identification under controlled environment[8, 13, 20]. As the importance for public safety increases, much more attention are needed for recognizing interactions between people. Group activity, especially small human group (around ten people) activity recognition has become an essential issue in human action recognition.

As shown in Figure 1, most public safety scenarios consist of small group activities. However, relatively fewer research has been done on this topic, due to the difficulties of describing varying number of participants and the mutual occlusion between people. In contrast with single person action recognition, small groups contain much richer inter-person interactions among group members. Compared to crowd analysis [11], in which each person can be regarded as a point in a flow, small groups contain much detail information about each individual in the group. Small human group activity recognition need to bridge the local description of single human and global description for crowd analysis, as well as addressing both the spatial dynamics (varying group size) and temporal dynamics (varying clip length). Recently, Ni et.al [12] introduced three types of localized causalities for human group activities with different number of people, and their experiment results showed that

Figure 1: Group activities examples from BEHAVE data set [1], such as In Group, Group Split, Group Fight, Chasing

intro-person feature could be used to classify group actions. However, as different group activities were described by feature vectors with varying length, specific classifiers were trained for different feature vectors. Chang et.al [2] proposed a bottom-up method to form a group and calculated the similarity of different groups. Ge et.al [5] also developed a hierarchical clustering algorithm for small group detection in a crowded scene. Guimera et.al [6] proposed a collaboration network structure to determine the team performance, and the experiment result indicated that team assembly mechanism could be used for predicting and describing the group dynamics. All the aforementioned methods have not address the group action recognition with different time durations.

Inspired by recent works [14, 17] on single human motion modeling by Gaussian Process Dynamic Models, we propose a novel structural feature set to represent group activities as well as a probabilistic framework for small group activity learning and recognition. Our framework consists of four stages, as shown in Figure 2. First, we apply a robust mean-shift [3] based tracker to track each individual in a small group sequentially. Second, the output coordinates of each tracker will be clustered and allocated to different small groups. Based on social network feature description, we extracted the structural features from each video clip in the third stage from each video clip. Those feature vectors contain global structure of each group as well as local motion description of each group member, and they all have same size regardless the different number of people inside each group.

In the last stage, the feature vectors from each frame will form a feature matrix for each video clip. A Maximum Gaussian Process Dynamical Model(MGPDM) is trained to model different group behaviors respectively. The group activity matrix will be projected to a low dimensional latent space and get a compact representation. A posterior conditional probability is compute with each trained model to identify different group behaviors. We validate our framework on two publicly available data set: BEHAVE data set [1] and IDIAP data set [16].

Our main contributions are listed as follows: First of all, we proposed a social network analysis based structural feature set to represent the dynamic of small group people. The structural feature characterizes both the global distribution of a group as well as local motion of each individual. In addition, this feature set can keep a fixed length while handling vary group size and group location, which is very important for recognition. Secondly, we established a probabilistic framework (MGPDM) for human behavior classification, which extended the GPDM [17] to address the group action classification. For the coming new activity feature sequence, the conditional probability with respect to each pre-trained GPDM is computed, and the one with the highest probability is selected as the testing group activity. As there is no length constraint for input training and testing sequence, this GPDM based recognition framework can address recognition of video clips with different lengths. The difference between our proposed model with GPDM in [14, 17] is our model introduce the

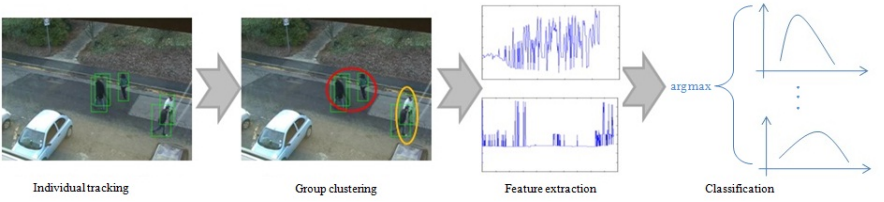

Figure 2: Overview of group activity recognition framework

conditional property of GPDM for classification, while GPDM in [14, 17] are mostly used for single human motion reconstruction.

The rest of this paper is organized as follows: Section 2 will describe the social network based feature for action representation. Section 3 will review the Gaussian Process Dynamic Model and introduce the Maximum GPDM for action recognition. Section 4 presents the overall framework for small group human action recognition. All the experimental results and comparison are shown in the Section 5. Finally, conclusion and future work are summarized in the Section 6.

2 Social Network Analysis Based Feature

As discussed in the previous section, feature extraction plays an essential role in the small group action recognition. Most features used for human action recognition fall into two big categories: general descriptive features and tracking based motion features. General low level feature includes motion vector, optical flow, 3D SIFT [15](Scale Invariant Feature transform) or STIP [9](Spatial Temporal Interest Points), which are directly computed on the entire image region. Therefore general descriptive features are good for single person action classification in the controlled environment. Tracking based features require tracking each human target in the video sequences, then extracting corresponding feature around the target's position. Motion vector, color histogram and other appearance models are widely used tracking based features for action recognition. As the development of multiple camera system, robust and accurate tracking of multiple human is not difficult. Beyond the aforementioned features, recently many middle level features are proposed for human group behavior analysis, as group human behavior involves interactions among different members, and requires the feature set can capture local detail information as well as global structure description. Middle level features characterizes the global properties of low level features rather local description. Ni et.al proposed a middle level feature set for the group structure information above general low level feature, which has been developed for small group human action recognition[12]. Recently Fei.Y et.al also proposed a middle-level representation for human activity recognition [21].

2.1 Social Network Analysis

In [4, 19] social network analysis has emerged as an interdisciplinary technique in modern sociology, information science and economics. Social network analysis(SNA) is originally designed to model the social structure of individuals and relationships among people

in real world societies. As a popular methodology in business or management consulting, it maps the social individuals or "actors" as *nodes* and relationships between them as *ties* into a graphic based network. To measure and understand the network and participants, we can analysis the locations and roles of nodes in their connection flows. Result graph based network structure analysis can be extended to many similar applications. To describe the network structure, there are three most used features in the social network analysis,closeness, centrality, and betweeness. Closeness measures the degree of an individual is near all other individuals in a network (directly or indirectly). It reflects the ability to access information through the "grapevine" of network members. Thus, closeness is the inverse of the sum of the shortest distances between each individual and every other person in the network.Centrality measures gives a rough indication of the social power of a node based on how well they "connect" the network. Betweeness measures the extent to which a node lies between other nodes in the network. This measure takes into account the connectivity of the node's neighbors, giving a higher value for nodes which bridge clusters. The measure reflects the number of people who a person is connecting indirectly through their direct links

2.2 Structural Feature Set

In this paper, inspired by the social network analysis, we extract several structural features to capture the dynamic properties of a small group structure. We believe that the dynamic structure and its theoretical framework can help us to model the group scenario in the real world. To our best knowledge, this is the first time that social network analysis based feature is used to model group behavior in the surveillance videos. Similar to the original definitions of betweenness, Closeness, and Centrality[7, 18] in social network analysis, we define several group structure features which are derived from SNA with modification in the group activity recognition.

1. *group center:* Suppose there are n people in a group, the group center $m = \left(\frac{1}{n}\sum_{i=0}^{n} x_i, \frac{1}{n}\sum_{i=0}^{n} y_i\right)$ is defined as the mass center of the group.

2. *motion histogram:* Motion vector is defined as the position difference of each individual between two consecutive frames. For each person in a group, we can calculate the orientation and magnitude of the motion vector. Suppose there are n people in a group, then we have $\mathbf{M}_t = \{m_i\}_t (i = 1, \ldots, n)$, then the magnitude of m_i is accumulated into orientation histograms and normalized at each direction, as shown in the Figure 3. The length of each arrow is corresponding to the sum of the vector magnitude near that direction. As the orientation has been divided to 8 bins, the motion histogram is a 8-dimension vector for each group in each frame.

3. *closeness histogram:* Closeness describes how close an individual is near to all the other nodes, directly or indirectly in a network. In our experiment, closeness vector is defined as the directional vector between every two different people. Suppose there are n people in a group, then we have $\mathbf{C}_t = \{c_i\}_t (i = 1, \ldots, n)$. Similar to motion histogram, the magnitude of c_i is accumulated into 8-bin orientation histograms and normalized at each direction. Motion histogram is also a 8-dimension vector for each group in each frame.

4. *centrality histogram:* Centrality was originally used for describing the overall network structure based on each node's location in a network. In this paper centrality vector

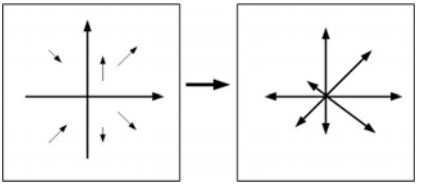

Figure 3: Motion histogram, the motion vector of each group member is shown in the left, and the length of each arrow in the right is corresponding to the sum of the vector magnitude along that direction.

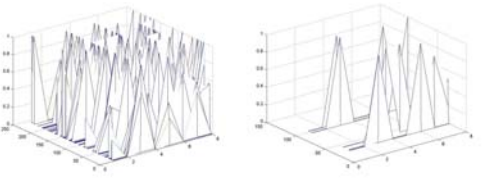

Figure 4: The 3D illustration of motion histogram of In Group (left) and Group Fighting (right) from two video clips

is defined as the directional vector which from the position of each person toward the group mass center. Suppose there are n people in a group, then we have $\mathbf{Ce}_t = \{ce_i\}_t (i = 1, \ldots, n)$. Similar as motion histogram, after being accumulated into 8-bin orientation histograms and normalized, the centrality histogram is a 8-dimension vector for each group in each frame.

As described above, for each frame, a 26 dimensional vector is extracted, including group center, motion histogram, closeness histogram and centrality histogram. Suppose the length of a group activity(total frame number) is m, then the size of the feature matrix is $26 \times m$.

3 Gaussian Process Dynamical Model

The assumption of this paper is that in the normal situation, the motion distribution of a human group is prone to have a Gaussian distribution. If we treat centrality feature in the Figure 4 as a Gaussian process, then the centrality histogram at each frame is a sampling of this process. Different group activities can be seen as a set of Gaussian processes with different means and covariance matrices, thus Gaussian process can be used to model the dynamics in the temporal dimension. In order to describe the dynamic property of the group behavior, here we adopt Gaussian Process Dynamical Model(GPDM) [17] to represent different group activities.

3.1 Gaussian Process Dynamical Model

Gaussian Process Dynamical Model was derived from Gaussian Process Latent Variable Model(GPLVM) [10], which provided a probabilistic mapping from high-dimensional observation data to low-dimensional latent space and represented the joint distribution of observation data. To address the sequential data with GPLVM, J.Wang et.al [17]introduced GPDM, which augmented the GPLVM by adding first-order Markov dynamic in the latent space. Consider a basic discrete model with first order Markov dynamics in equations below:

$$x_t = f(x_{t-1}, U) + \eta_{x,t} \quad (1)$$

$$z_t = h(x_t, V) + \varsigma_{z,t} \quad (2)$$

where x_t is the latent variable and z_t is the observation variable at time t. $\eta_{x,t}$ and $\varsigma_{z,t}$ are zero-mean, isotropic, white Gaussian distributed noise for the latent and observation spaces respectively.

The GPDM model $\{\Lambda\}$ can be derived as in the equation (3) based on the Gaussian priors, first order Markov dynamics and latent space mapping.

$$\begin{aligned}\Lambda &= p(X,Z,\bar{\alpha},\bar{\beta},\Omega)\\ &= p(Z_t|X_t,\bar{\beta},\Omega)p(X|\bar{\alpha})p(\bar{\alpha})p(\bar{\beta})p(\Omega)\end{aligned} \tag{3}$$

$p(X|\bar{\alpha})$ can be reprented as:

$$p(X|\bar{\alpha}) = \frac{p(x_1)}{\sqrt{(2\pi)^{\frac{D(N-1)}{2}}|K_X|^{\frac{N}{2}}}} exp\left(-\frac{1}{2}(K_X^{-1}X_{out}X_{out}^T)\right), \tag{4}$$

where $X_{out} = [x_2,...,x_N]^T$ are considered as testing data, and $\bar{\alpha}$ is a vector of kernel parameters. We assume $p(x_1)$ also has a Gaussian prior. K_X is the $(N-1)\times(N-1)$ kernel matrix constructed from $[x_1,...,x_{N-1}]$, and a linear kernel is used here. $p(Z_t|X_t,\beta,\Omega)$ in the equation (**??**) represents a non-linear projection from the latent space X to the observation space Z,

$$p(Z_t|X_t,\beta,\Omega) = \frac{|\Omega|^N}{\sqrt{(2\pi^{ND}|K_Z|^D)}} exp(-\frac{1}{2}tr(K_Z^{-1}Z\Omega^2Z^T)) \tag{5}$$

where Ω is a scale parameter, N is the length of observation sequences Z,D is the data dimension of Z, K_Z is the *RBF* kernel function.

3.2 Maximum GPDM

Since GPDM is a generative model to represent the time sequence data, it has widely applied in modeling a particular motion or motion reconstruction. In addition, as different motion type may form different distribution in the latent space in tje GPDM, it can be extended for motion classification. Leonid et.al [14] proposed a tracking and classification of human motion with Gaussian process annealed particle filter, and each motion type was pre-trained by GPDM. In the process of tracking, the tracked results were projected to the latent space and the Frechet distance is computed for comparing the similarity of two different distribution in the latent space. This classification method is good for a small number of motions types, since the computational cost will increase dramatically as the number of motion type increase.

Here we proposed Maximum GPDM for classification based on GPDM. The proposed model can select the GPDM with the highest posterior probability among a set of trained GPDMs. Given a trained GPDM, $\Lambda = \{Z^T, X^T, \bar{\alpha}, \bar{\beta}, \Omega\}$, where Z^T is the training observation data, X^T is the corresponding latent variable sets, $\bar{\alpha}$ and $\bar{\beta}$ are hyperparameters vectors, and Ω is a scale parameter. Following the derivation in [17], the conditional probability of a new observation $Z^{(*)}$ can be defined in the equation (6).

$$\begin{aligned}&p(Z^{(*)},X^{(*)}|\Lambda)\\ &\propto\ p(Z,Z^{(*)}|X,X^{(*)},\bar{\beta},\Omega)p(X,X^{(*)}|\bar{\alpha})\end{aligned} \tag{6}$$

Suppose the length of Z and $Z^{(*)}$ is N and M, then the kernel size of $\{Z,Z^{(*)}\}$ is $(N+M)\times(N+M)$. To reduce the computational cost, we define two kernel matrices: $Q_{i,j} = k_Z(x,x^{(*)})$,

and $R_{i,j} = k_Z(x^{(*)}, x^{(*)})$, then we can derive $p(Z^{(*)}|X^{(*)}, \Lambda)$ as:

$$p(Z^{(*)}, X^{(*)}|\Lambda) = \frac{|\Omega|^M}{\sqrt{(2\pi)^{MD}|K_{Z^{(*)}}|^D}} exp\left(-\frac{1}{2}tr(K_{Z^{(*)}}^{-1} P_Z \Omega\Omega^2 P_Z^T)\right) \quad (7)$$

where $P_Z = Z^{(*)} - Q^T K_Z^{-1} Z$ and $K_{Z^{(*)}} = R - Q^T K_Z^{-1} Q$.

When a new observation $Z^{(*)}$ comes, the posterior probability p_i with respect to each GPDM Λ_i will be computed according to equation (7), and the one with the highest probability will be selected as the final classification result. The maximum GPDM model classified each motion type through a probabilistic approach which takes into account both the observation and latent space, rather than just comparing the shape of distribution in the latent space [14]. In addition, this model can be extend for motion classification with a large number of motion types. It should be noted that, the length of new observation can be different with the size of training data.

4 Proposed Framework for Behavior Classification

As shown in Figure 2, our small group activity recognition framework consists of four stages: adaptive mean-shift tracking, small group clustering, group feature extraction and group activities recognition.

4.1 Adaptive Mean-shift Tracking

One of the important factor for small group human activities analysis is the accuracy and robustness of tracking each individual in the group. As the development of multiple camera systems, the accurate tracking of each individual can be well addressed. In this paper we apply adaptive mean-shift tracking[3] on the two data sets.

Compared to general mean-shift tracking, on-line feature selection is applied during the adaptive mean-shift tracking. In [3], the feature consisted of linear combination of pixel valves at R, G, B channels: $F \equiv \omega_1 R + \omega_2 G + \omega_3 B$, where $\omega_i \in [-2, -1, 0, 1, 2], i = 1, \ldots, 3$. By pruning all redundant coefficients of ω_i, the feature set was cut down to 49. Linear discriminative analysis (LDA) was then used to determine the most descriptive feature for target tracking.

In order to reduce the computational complexity during tracking, we just update the feature set every 50 frames instead of updating the feature set at each frame. In addition, we extend the single mean-shift tracking algorithm for multiple targets tracking. As the cameras were fixed in these two data sets, a simple motion detector is applied to detect each new person coming into scene. Once a person comes in the scene, a new tracker will be allocated and track that person overtime. Since our focus of this paper is not reliable multiple targets tracking, we just reinitialize each target manually if the tracking algorithm fails for some reason.

4.2 Small Group Clustering

After obtaining all the positions of each target, a group clustering algorithm [2] will be applied to locate small groups. We first calculate the closeness of each person and use the

Minimum Span Tree (MST) clustering to obtain the distribution of each group. After that, we follow the hierarchical clustering method described in [2] to locate the mass center of each small group.

4.3 Small Group Activity Recognition

The small group activity recognition can be divided to two phases: group activity training and group activity classification. In the training stage, for each small group activity $\{A_i, i = 1,\ldots,n\}$, a GPDM $\{\Lambda_i, i = 1,\ldots,n\}$ will be trained. Suppose we have k samples of a group activity A_i, the length of each each sample is m, then we have k feature matrices of size $26 \times m$. To learn a specific GPDM for A_i, we will first compute the mean value $\bar{Z}$ of k feature matrices, and utilize the mean value for training.

GPDM is applied to learn the specific trajectories of a group activity. The probability density function of latent variable X and the observation variable $\bar{Z}$ are defined by the equation (3). The basic procedure of the Gaussian Process Dynamical Model training is described as below:

1. *Creating GPDM:* GPDM $\Lambda = \{\bar{Z}^T, X^T, \bar{\alpha}, \bar{\beta}, \Omega\}$ is created on the basis of the trajectory training data sets, i.e. extracted structural feature, where $\bar{Z}^T$ is the training observation data, X^T is the corresponding latent variable sets, $\bar{\alpha}$ and $\bar{\beta}$ are hyperparameters.

2. *Jointly initializing the model parameters:* The latent variable sets and parameters $\{X^T, \bar{\alpha}, \bar{\beta}\}$ are obtained by minimizing the negative log-posterior function $-lnp(X^T, \bar{\alpha}, \bar{\beta}, \Omega|\bar{Z}^T)$ of the unknown parameters $\{X^T, \bar{\alpha}, \bar{\beta}, \Omega\}$ with scaled conjugate gradient (SCG) on the training datasets.

3. *Train GPDM for each group activity:* For each group activity $\{A_i, i=1,\ldots,n\}$, repeat the procedure 1 and 2, and create a corresponding GPDM: $\{\Lambda_i, i = 1,\ldots,n\}$.

After training, we have a set of GPDMs: $\{\Lambda_i, i = 1,\ldots,n\}$ for the human group activities. When a new human group activity $Z^\star$ coming in, we will compute the conditional probability with respect to each trained GPDM, and select the one with maximum conditional probability as the classification result.

1. *Calculate the conditional probability with each trained GPDM:* For each trained GPDM $\{\Lambda_i\}$, compute $X_i^{(*)}$ by using the learned parameters: $\{\bar{\alpha}_i, \bar{\beta}_i\}$. This can be obtained by minimizing the negative log-posterior function $-lnp(X^T, \bar{\alpha}_i, \bar{\beta}_i, \Omega|Z^{(*)})$ with scaled conjugate gradient (SCG) on the training datasets. After that, we can calculate the conditional probability $P\left(Z_i^{(*)}, X_i^{(*)}|\Lambda_i\right)$ by the equation (6).

2. *Select the GPDM with the highest conditional probability:* The new group activity can be determined by the following equation:

$$\underset{i=1,\ldots,n}{\operatorname{argmax}} P\left(Z_i^{(*)}, X_i^{(*)}|\Lambda_i\right) \tag{8}$$

As we discussed in the previous section, the length of new observation can be different with the size of training data, which means that the number of frames in test clips can be different with training clips. Therefore our trained model can address the dynamics in the temporal

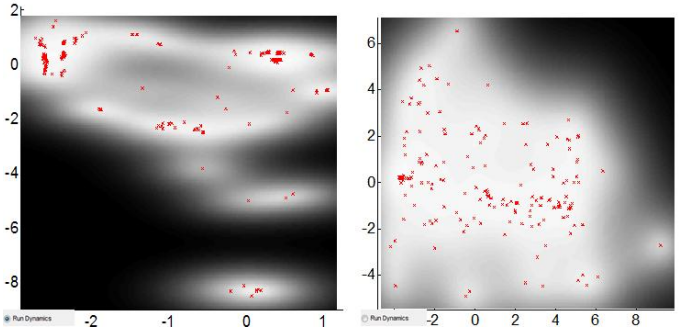

Figure 5: Visualization of trained GPDMs, the left one the InGroup, and the right one is Group Fight

Table 1: Classification results of our method

	IG	WT	F	S
Our method	94.3%	92.1%	95.1%	93.1%

dimension. As the duration of an activity may change under different situation, it is important that the classifier can handle the testing sequences with varying lengths.

5 Experimental Results

There are not many publicly available data set for group activity recognition. We have evaluated our framework on two popular group activity data sets. The first one is the recently released BEHAVE data set [1], which contains the ground truth for each group activity. The second data set is IDIAP [16] data set, which was originally captured for multiple targets tracking.

5.1 Results on BEHAVE data set

The BEHVAE data set consists $76,800$ frames in total. This video data set is recorded at 26 frames per second and has a resolution of 640×480. Different activities include: InGroup, Approach, WalkTogether, Split, Ignore, Following, Chase , Fight, RunTogether, and Meet. There are 174 samples of different group activities in this dataset. As our focus is the small group activity analysis, we select 118 samples from all the group activities data set, excluding these samples with less than three people in the scene. The selected group activities include InGroup (IG), WalkingTogether(WT), Split(S) and Fight(F) as our group activities for classification. For each activity, we divide the samples to ten-fold, with nine-fold for training and one fold for testing, the classification result is shown in the Table 1. Two of learned GPDMs are shown in the Figure 5. Each point in the latent space is corresponding to a feature vector in a single frame. The distribution of InGroup activity is prone to have some local clusters in the latent space, while the distribution of GroupFight activity is similar to a random distribution.

We also compare our results with the best recognition classification results in [1]. As in

Table 2: Comparison of classification results

HMM based method[1]	Our method
93.67%	93.12%

Figure 6: Sampling frames of InGroup, WalkTogether and Split

[1], the training and testing data is divided to 50/50, the comparison results in the Table 2 indicates the competitive performance of our proposed framework. It should be noted that, the recognition rate is the average rate for all the activities. For the HMM based method in [1], the time window size is 100, which means that their proposed method required at least 200 frames to recognize each action type. While our framework can can handle small group action recognition regardless of time durations through the probabilistic approach.

5.2 Results on IDIAP data set

IDIAP data set is firstly used in [16] for multiple targets tracking. The data set contains 37182 frames in total. We manually select 46 clips with different lengths for human group activity recognition. As there is no Fight activity in the IDIAP data set, so we just evaluate three other activities: InGroup, WalkTogether, and Split. To validate the robustness of our framework, we directly apply the trained GPDMs in the BEHVAE data set for activity recognition on the IDIAP data set, and the overall average classification rate is 90.3%. The experiment results indicate that our proposed framework is robust to identify human group activities under different scenariors. Some of the sampling frames from IDIAP data set are shown in the Figure 6.

6 Conclusion and Discussion

In this paper, we propose a novel structural feature to describe the small group activity. Based on the structural feature, we also propose a Conditional Gaussian Process Dynamic Model for group activity recognition. The proposed structural feature can be adapted to many other applications, since its dynamic characteristics can be used to describe different features. The framework can also be used for abnormal group activity detection in surveillance systems. In addition, this framework can be used for soccer sport analysis or team performance evaluation.In the future, we will continue to validate our proposed framework on other group activity data sets with more difficult group activities.

Require reliable multiple targets tracking is one of the limitation of our framework. Recent progress on the multiple detection and tracking ensure this task is becoming much easier. Moreover, we just utilize the rough central point as our location for group behavior analysis, which indicate our framework does not require accurate tracking of each person in the group. The topology of the group is much more important for group behavior analysis.

We have demonstrated impressive results on the group human activity recognition by constructing middle level features only on position cues. More experiments will be conducted by adding other low level features to build middle level features, such as human body part motion. We believe middle level features beyond more descriptive low-level features will benefit complex human group activity recognition.

In addition, besides small group activities, our framework can also used to address the interaction among individual with small groups as well as the interaction between multiple small groups, we will test the performance of our framework in the future.

References

[1] S. Blunsden and R. B. Fisher. The behave video dataset: ground truthed video for multi-person. *Annals of the BMVA*, 4:1–12, 2010.

[2] Ming-Ching Chang, N. Krahnstoever, S. Lim, and Ting Yu. Group level activity recognition in crowded environments across multiple cameras. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*.

[3] Robert T. Collins, Yanxi Liu, and Marius Leordeanu. Online selection of discriminative tracking features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27 (10):1631 –1643, 2005. ISSN 0162-8828.

[4] Linton Freeman. The development of social network analysis. *Empirical Press*, 2006.

[5] Weina Ge, Robert T. Collins, and Barry Ruback. Automatically detecting the small group structure of a crowd. In *Workshop on Applications of Computer Vision, WACV 2009*.

[6] Roger Guimera, Brian Uzzi, Jarrett Spiro, and Luis A. Nunes Amaral. Team assembly mechanisms determine collaboration network structure and team performance. In *Science*, volume 308, pages 697 – 702, 2005. doi: 10.1126/science.1106340.

[7] Valdis Krebs. The social life of routers. In *Internet Protocol Journal*, pages 14–25, dec. 2000.

[8] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. In *Proceedings of Conference on Computer Vision and Pattern Recognition, CVPR 2008.*

[9] Ivan Laptev and Tony Lindeberg. Space-time interest points. In *IEEE 6th International Conference on Computer Vision, ICCV 2003*.

[10] Neil Lawrence. Probabilistic non-linear principal component analysis with gaussian process latent variable models. *Journal of Machine Learning Research*, 6:1783–1816, 2005. ISSN 1532-4435.

[11] Ramin Mehran, Alexis Oyama, and Mubarak Shah. Abnormal crowd behavior detection using social force model. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*.

[12] Bingbing Ni, Shuicheng Yan, and A. Kassim. Recognizing human group activities with localized causalities. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*.

[13] Juan Carlos Niebles, Hongcheng Wang, and Li Fei-fei. Unsupervised learning of human action categories using spatial-temporal words. In *Proceedings of British Machine Vision Conference, BMVC 2006*.

[14] L. Raskin, M. Rudzsky, and E. Rivlin. Tracking and classifying of human motions with gaussian process annealed particle filter. In *ACCV07*.

[15] Paul. Scovanner, Saad. Ali, and Mubrark Shah. A 3-dimensional sift descriptor and its application to action recognition. 2007.

[16] Kevin Smith, Daniel Gatica-perez, and Jean-marc Odobez. Using particles to track varying numbers of interacting people. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2005*.

[17] Jack M. Wang, David J. Fleet, and Aaron Hertzmann. Gaussian process dynamic models for human motion. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 30:283–298, 2008.

[18] Stanley Wasserman and Katherine Faust. *Social Networks Analysis: Methods and Applications*. Cambridge: Cambridge University Press., 1994.

[19] Duncan J. Watts and Duncan H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.

[20] Shu-Fai Wong and Roberto Cipolla. Extracting spatiotemporal interest points using global information. In *IEEE 11th International Conference on Computer Vision, ICCV 2007*.

[21] Fei Yuan, Veronique Prinet, and Junsong Yuan. Middle-level representation for human activities recognition:the role of spatio-temporal relationships. In *3rd Workshop on Human Motion Understanding, Modeling, Capture and Animationn*, September 2010.