Proceedings Paper:

Stephens, Kyle and Bors, Adrian Gheorghe orcid.org/0000-0001-7838-0021 (2017) Human Group Activity Recognition based on Modelling Moving Regions Interdependencies. In: Proc. of International Conference on Pattern Recognition (ICPR). IEEE , pp. 2116-2121.

https://doi.org/10.1109/ICPR.2016.7899948

Human Group Activity Recognition based on Modelling Moving Regions Interdependencies

Kyle Stephens and Adrian G. Bors
Department of Computer Science, University of York, York YO10 5GH, UK
E-mail: adrian.bors@york.ac.uk

Abstract—**In this research study, we model the interdependency of actions performed by people in a group in order to identify their activity. Unlike single human activity recognition, in interacting groups the local movement activity is usually influenced by the other persons in the group. We propose a model to describe the discriminative characteristics of group activity by considering the relations between motion flows and the locations of moving regions. The inputs of the proposed model are jointly represented in time-space and time-movement spaces. These spaces are modelled using Kernel Density Estimation (KDE) which is then fed into a machine learning classifier. Unlike in other group-based human activity recognition algorithms, the proposed methodology is automatic and does not rely on any pedestrian detection or on the manual annotation of tracks.**

Index Terms—**Group Activity Identification, Motion Segmentation, Streaklines.**

I. Introduction

Several algorithms have been proposed for human activity recognition by considering individual actions. This research area has a significant importance for video surveillance, human-computer interaction, semantic annotations of multimedia, retrieval of video data, among many other applications. Meanwhile, group activity classification has attracted interest only very recently, despite being essential in defining the real intention and the context of human activities. Most of the human activity recognition methods begin by modelling low level local features from video sequences, for example using the Dollar gradient cuboids [1] or histograms of gradients (HOG) [2]. In other approaches, Baktashmotlagh *et al.* [3] applied non-linear stationary subspace analysis to activity recognition while Ryoo and Aggarwal [4] introduced a method named spatio-temporal relationship match.

More recently, the main focus of human activity has moved on from simple human activities to those that are more complex, where the main objective is scene analysis rather than determining the activities of a single individual. One group of approaches is to detect abnormalities or uncommon activity events. The method from [5] modelled the motion patterns using Gaussian Mixture Models (GMMs) of 3D distributions of local space-time gradients. Similarly, GMMs of Markov random fields (GMM-MRF) was used in [6] for abnormal activity detection. Dynamic texture models [7], which considers both appearance and dynamics, have also been considered for abnormal activity detection. An observational system, in which new activities are identified in the scene, based on a significant Kullback-Leibler divergence from a dictionary of activities pre-learnt during the training stage, was proposed in [8], [9]. In comparison to human activity recognition, group activity recognition requires more complex descriptions of the people's interaction in the group. Ni *et al.* [10] recognizes group activities using manually initialized tracklets. Lin *et al.* [11] used a heat-map based algorithm for modelling human trajectories when recognising group activities in videos. Chang *et al.* [12] used a probabilistic approach to group human activity by forming various probabilities depending on the tracks between individuals using a multi-camera system. Choi *et al.* [13] proposed a framework for analysing collective group activities based on different levels of semantic granularity. Zhang *et al.* [14] addressed the problem of group event recognition by computing histograms of different features extracted from the tracklets, representing localized movement in the video. Similarly, Cheng *et al.* [15] modelled group activity as a framework composed of multiple layers and Gaussian defined processes were used for representing motion trajectories. One common issue with all these methods is that they rely on either the training of a pedestrian detector for each scene, or on the manual annotation of tracklets.

In this research study we propose an automatic method for group activity recognition by modelling the inter-dependant relationships between features over time. Unlike other methods, we do not rely on any manual initialisation of tracklets and instead make use of medium term tracking as provided by streaklines [16]. Compact moving regions are then segmented. The interdependency between moving regions is represented by evaluating the relative movement and location of each moving region with respect to all the others. Kernel Density Estimation (KDE) is used to model both time-location and time-motion spaces, resulting in representing the dynamics of such interactions. Moreover, the model keeps track of stationary pedestrians by marking the locations where they stop moving and considers these locations in modelling their following movements. We also propose a scaling procedure in order to compensate for the effect of perspective projection in video sequences acquired by lowly located cameras of wide view and compensate in the group activity model for such effects. Section II describes the features used for representing moving regions, while how their inter-dependencies are modelled in the context of group activity is explained in Section III. Section IV describes the classification of group activities. Section V shows the experimental results and Section VI draws the conclusions of this research study.

II. Group Activity Modelling

The proposed methodology for group activity recognition has several stages, including extracting streaklines, representing medium-time trajectories of movement, identifying moving regions and their dynamics, using these for modelling group interactions, and then finally classifying the sequences into group activities using Support Vector Machines (SVM). A block diagram of the proposed method for recognising group activities is shown in Figure 1.

The first processing stage consists of movement estimation. One issue that arises from using traditional optical flow is the difficulty in capturing unsteady movement in scenes with multiple pedestrians interacting and crossing each other. To alleviate this problem, we propose the use of a medium-time movement tracking method such as the streaklines proposed in [16] which was used in [8], [9] as well. Streaklines correspond to tracking fluid particles that have passed through a particular location in the past and its modelling is based on the Lagrangian framework for fluid dynamics [16]. The streakflows represent the fluid like flow in a scene, enabling the filling of spatial gaps. Unlike in [16], where streaklines are computed for each pixel, we associate each streakline with blocks of pixels of a fixed size by computing the marginal median as the streakline estimate for each block of pixels. Following this, we fit a first degree polynomial to each streakline in order to obtain a smoother representation. This differs from [8], where the principal direction of movement was obtained from applying PCA on the vectors forming each streakline. One issue with the approach from [8] is that it does not consider the motion consistency over several frames. In the approach from this study the consistency of the streaklines is enforced over several frames.

We make the assumption that each compact region of streakflows may contain several individual movements, which can be represented by clusters. Firstly, we begin by segmenting the streakflow field into distinct moving regions. The Expectation-Maximization (EM) algorithm, under the Gaussian Mixture Model (GMM) modelling assumption, is used for segmenting and modelling each inter-connected region. The number of clusters and the centres of the Gaussian functions in the EM algorithm are initialised using the modes of the histogram of streakline flow in order to improve the convergence. Moreover, in this study we also address the perspective distortion effects by using a two-step approach to movement segmentation. Such effects are evident in the case of video sequences acquired with wide-angle lens cameras which are located at low heights. In the first step, the segmentation is performed in order to estimate the height of the moving objects, which is used to derive a scaling factor. In the second step, the segmentation is repeated considering this scaling factor, applied appropriately to the estimated movement, according to the location of its corresponding moving region in the scene. A moving region i is scaled as follows:

$$s_i = \tfrac{1}{2h_m}(h_i + \frac{\sum_{j=1}^{n} h_j}{n}) \tag{1}$$

Where h_i is the height identified for each moving region in the first step, $j = 1, \ldots, n$ are the segmented moving regions, h_m is the predetermined overall mean height of all moving regions and s_i is the scaling factor for moving region i. This is repeated for all compact moving regions which are identified in the scene. The motion $\mathbf{M}_i$ of region i is then scaled by a factor s_i:

$$\mathbf{M}_i^{'} = s_i \mathbf{M}_i. \tag{2}$$

Each moving region is therefore represented by a GMM defined by its characteristic parameters representing its movement and location in the scene. Another issue that is addressed in this research study is the modelling of people who become stationary after they have moved through the scene. Under the optical flow detection and motion model such people would not be accounted for. To overcome this situation, we propose to identify when and where people stop moving in the scene. If no movement is present in a particular region where motion was previously detected, during p consecutive frames, this indicates a stationary region that has previously moved. Such stationary regions are characterised by their location and by zero motion. Any movements of a person present near the edge of the scene that subsequently moves out of the scene is identified and the respective moving region is no longer considered. Finally, when movement occurs within a bounding box of the stopped pedestrian, the region is deemed to be no longer stationary and the new emerging moving region in the area is activated in the existing group activity model.

III. Modelling Interdependent Relationships of Moving Regions

The key characteristics of group activities are often present in the interdependent relationship between the pedestrians and moving objects. In this research study we propose to model the interdependent relationship between the features of each pair of moving regions detected in the scene. In this section, we describe how we model four distinct features for representing group activities: streakflows, streakflow dynamics, locations and location dynamics.

To begin, we model the interdependent relationship by evaluating the differences between streakflow models in the scene for each pair of movingregions. This models the interdependant relationship of the movement of the group at a particular time instance. We compute the differences between streakflows, $\mathcal{A}_{I(t)}$ and $\mathcal{A}_{J(t)}$ for two moving regions $I(t)$ and $J(t)$ at time t by:

$$M(I(t), J(t)) = \mathrm{e}^{-\frac{D_{SKL}(\mathcal{A}_{I(t)}||\mathcal{A}_{J(t)})}{\sigma_m}} \tag{3}$$

where σ_m is a scaling factor for movement differences and $D_{SKL}(\mathcal{A}_{I(t)}||\mathcal{A}_{J(t)})$ is the symmetrised KL divergence between the streakline distribution of moving regions $I(t)$ and $J(t)$ at time t. This results in a value within the range $[0, 1]$ which models the difference between two streakflow models, each characterising the movement of one region in the scene, associated to a moving person. For example, individuals moving in completely different directions will have

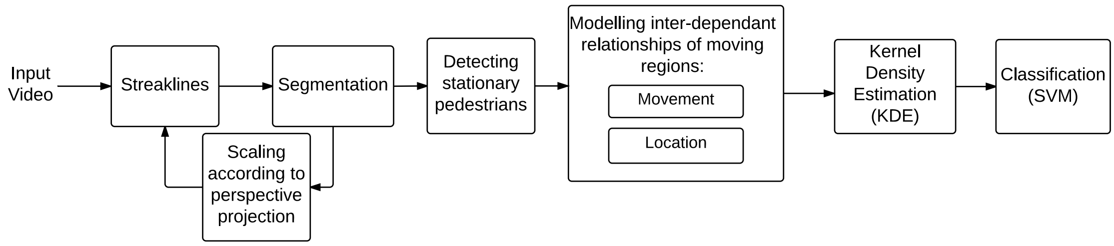


Fig. 1. Overview of the proposed group activity recognition approach

$M(I(t), J(t)) = 0$, whilst individuals moving in the same direction and at the same speed will have $M(I(t), J(t)) = 1$. The differences are computed by considering all pairs of moving regions in the scene at a particular time t by using equation (3). These are then concatenated to form a vector representing the inter-dependant group relationship of the streakflows at a particular time t.

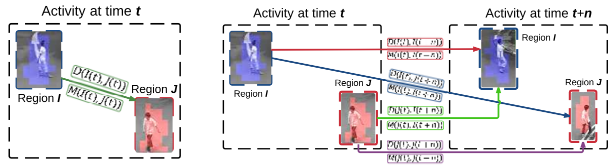
We also model the dynamic changes of differences between moving regions over subsequent frames by computing the differences between all streakflow models at time t and those identified at time $t + n$. These are computed as in equation (3), except that the models are now across subsequent sets of frames instead of at the same time instance. A vector of streakflow differences representing all the inter-dependant relationships of streakflow models between the time instances t and $t + n$ is then formed.

The distributions of relative locations for the people from the scene, both moving or stationary, is modelled similarly by considering differences between the GMM representing the spatial-location of their corresponding moving region. The means will approximate the centres of moving regions, whilst the variance will provide some characteristics of the size and shape of the region. Similarly to the streakflows, the differences between such location GMMs are then computed. Given two location GMMs $\mathbf{C}_{I(t)}$ and $\mathbf{C}_{J(t)}$ for moving regions $I(t)$ and $J(t)$ at time t, the differences between their locations can be computed by:

$$D(I(t), J(t)) = \mathrm{e}^{-\frac{D_{SKL}(\mathcal{C}_{I(t)}||\mathcal{C}_{J(t)})}{\sigma_l}} \tag{4}$$

where σ_l represents the characteristic scale parameter for locations. Similarly to the streakflow model, this provides a value in the range [0,1] representing the spatial relationship between the two moving regions. For example, individuals characterised by moving regions $I(t)$ and $J(t)$ at time t, located far apart, will have $D(I(t), J(t)) = 0$, whilst individuals located close together will have $D(I(t), J(t)) = 1$. A vector, representing all the inter-relationships of locations for the group activity at time t, is then formed.

Similarly to the streakflow model, the dynamics of the locations over time is computed as well. The dynamic changes of differences over subsequent frames are computed by the differences between all location points at time t and all location points at time $t + n$ using equation (4). A vector representing the moving regions location differences, representing all the inter-dependant relationships of location points between time t and $t + n$, is then obtained. These movement models are illustrated in Figure 2.



a) Base model b) Dynamic model

Fig. 2. Modelling the inter-dependencies of moving regions in both space and time.

One further issue that arises when computing such differences is that the rate of movement change and the rate of location change are not clearly characterised. For example, when using the dynamics in both movement and locations alone, the dynamics between walking and running activities may appear quite similar. In order to avoid this situation we consider the background as an additional region for both the streakflow and the location models. In the former case, the background object is defined as the GMM model comprising of all the motion in the scene that does not belong to a moving region (often zero motion if the camera is stationary). In the latter case, the location object is defined as the GMM representing the centre of the scene. By adding the background model, the change in both motion and location relative to the background represents the absolute movement in the scene. In the case of camera movement, such a model would account for this. Given a streakflow background model $\mathcal{A}_{B(t)}$, at time t the difference between the streakflow model $\mathcal{A}_{I(t)}$, for moving region $I(t)$, at time t, and the background $B(t)$ is computed as:

$$M(I(t), B(t)) = \mathrm{e}^{-\frac{D_{SKL}(\mathcal{A}_{I(t)}||\mathcal{A}_{B(t)})}{\sigma_m}} \tag{5}$$

Similarly, given the centre point $\mathbf{C}_{B(t)}$ defined as the location of background model $B(t)$ (the centre of the scene) at time t and the location model $\mathbf{C}_{I(t)}$ for moving region $I(t)$ at time t, the difference is computed as:

$$D(I(t), B(t)) = \mathrm{e}^{-\frac{D_{SKL}(\mathcal{C}_{I(t)}||\mathcal{C}_{B(t)})}{\sigma_l}} \tag{6}$$

Such differences are then computed between every region in the scene and the background model $B(t)$. Finally, the vector of differences in both cases are concatenated with the vector representing the other pairwise movement and location differences, corresponding to the pairs of moving regions.

IV. Group Activity Classification

To model the change in feature relationship over the whole sequence, we propose to use bi-variate Kernel Density Estimation (KDE). KDE would provide smoothing on the dynamics of feature changes over time increasing the robustness of the group activity model. We form two column matrices where the motion and location interdependences for each pair of moving regions are represented along the first column and their corresponding time instances are located in the second column. This matrix representation is used for each feature representing streakflow, streakflow dynamics, locations and location dynamics, separately. The bi-variate kernel density estimation is applied over a fixed grid size of $K \times K$, given the normalized matrix data.

By using a fixed grid size, video sequences of different lengths will be normalized in length. This helps to normalise the difference in speeds at which the activities are performed. The grid size is a important parameter in the density estimation as a too small grid would result in over-smoothed feature data and consequently important characteristics in the relationship features may be lost. If the grid size is too large, then the data will appear too sparse and would not model well the underlying pattern of the data. The kernel for density estimation is assumed to be Gaussian. The bandwidth parameters of the bi-variate Gaussian kernel are used to help control the smoothing effects of the kernel density estimator.

The densities computed over the fixed grid are used as the defining feature vector representation for the group activity. Such densities are computed independently for each dimension, representing the relationships of the moving regions in the movement, movement dynamics, location and location dynamics, respectively. Finally, the feature vectors representing each activities are used to train a Support Vector Machine (SVM).

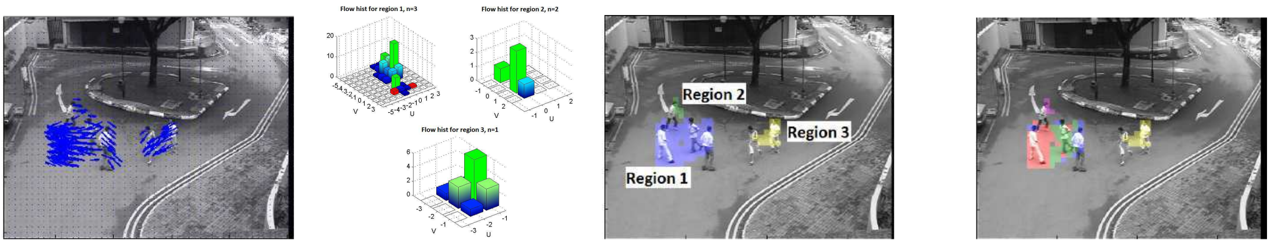
V. Experimental Results

For all experiments, we follow the same recognition routine. Firstly, the streakflows are extracted for each set of frames as in [16] and the moving regions are segmented based on the streakflows aiming to obtain compact inter-connected regions. Streakflows and their location are calculated for the moving regions in each set of frames. The features of the moving regions are then modelled by the differences between all pairs moving regions across the given set of frames. The dynamic changes of the features are modelled by the differences between all moving regions in one set of frames and the following set. Finally, the vector of differences for each set are used to form a two column matrix with differences along the first column and the time instance along the second column. KDE is applied on a fixed grid size using the data from the feature matrix. The features are then represented by their density estimation obtained from applying the KDE with the difference in movement and location features along one axis for the same timing, while the differences between such features at two different time instances are located along the other. This procedure is repeated for the dynamic model. Finally, the densities are used as features to build a classifier and the recognition decisions are taken by a Support Vector Machine (SVM) with RBF kernel.

Unlike in human activity recognition, the number of group activity datasets are quite limited, and in this study we present results on the NUS-HGA dataset [10]. This data set consists of six different group activities collected in five different sessions, each session representing the actions of various actors taking place in a road area located between office buildings. In total there are 6 group activities with 476 video sequences in total. To begin, streaklines are extracted for blocks of size 14×14 over 10 consecutive frames. The motion filter described in Section II is placed over each set of 5 frames, where motion must be present in 3 out of 5 image frames. The motion is segmented as described in Section II and each moving region is represented by its streakflow Gaussian Mixture Model (GMM) and its location GMM. Figure 3 shows an example of the estimated streakflows, motion histograms, and the moving region segmentation for the fight activity from the NUS-HGA dataset. In this particular activity, movement is very intense and very chaotic. In Figure 3b the solid green bars correspond to peaks of the histogram, while the solid red bars are entries with the height below 15% of the maximum bar height which are removed. The moving regions are well segmented and the small regions obtained in region 1 of Figure 3d help characterise the smaller atomic events performed in the group, for example pushing or kicking which usually happens during the fighting activity.

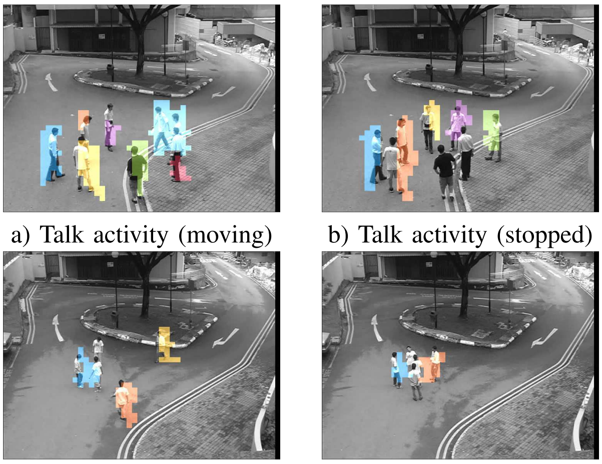
Following the initial movement segmentation, the motion in each moving region is scaled according to the height of the region using equation (2). The segmentation is then performed for the second time using the scaled motion. Following the second movement segmentation step, the stationary pedestrian detector is applied as in Section II where the number of prior frames is set to $p = 25$. We define the boundary parameter from Section II as 10% of the region size. Two examples of detecting stationary pedestrians are shown in Figure 4 for the talking and gathering activities. In Figures 4a and 4c the pedestrians are still moving and therefore their corresponding moving regions are properly detected. In Figure 4b and 4d the individuals have stopped but their stationary regions are properly detected by the stationary pedestrian detector procedure.

The streakflow movement model, streakflow dynamics, location and location dynamics relationship differences are computed as in Section III, considering the scaling parameters $\sigma_m = 15$, $\sigma_l = 550$ for motion and location differences respectively, and $\sigma_m = 17.5$, $\sigma_l = 650$ for the motion and location dynamics. The size of the number of frames, considered for the dynamic window from Section III, is set to $n = 13$. The data is represented by a 2-column matrix over time as described in Section IV. KDE is applied over a fixed grid size using the 2-column feature matrices as input data. In this study, we use the bivariate KDE method proposed in [17] which is based on using linear diffusion processes. The KDE methodology from [17] assumes the kernel to be Gaussian and uses a bandwidth selection method such that the bandwidth



a) Streakflows b) Histograms of flow c) Inter-connected regions d) Moving regions after segmentation

Fig. 3. Example of streakflows, histograms of flow and the moving regions before and after segmentation on a fight sequence from the NUS-HGA dataset. In b) "n" refers to the number of histogram peaks.


a) Talk activity (moving) b) Talk activity (stopped)

c) Gather activity (moving) d) Gather activity (stopped)

Fig. 4. Identifying when pedestrians stop during the video frames showing gathering and talking activities from the NUS-HGA dataset.

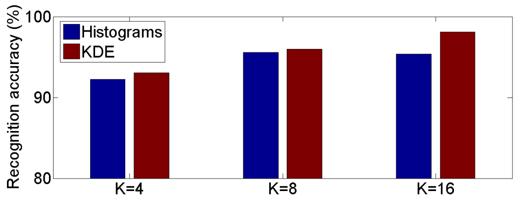
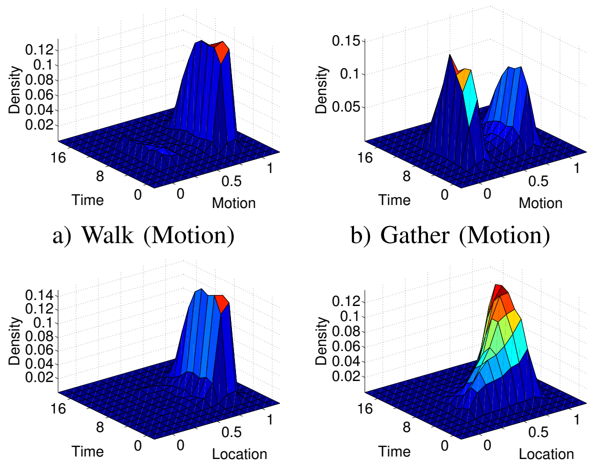


Fig. 5. Recognition results as K is varied when using KDE and histograms.

parameters are automatically selected depending on the data. The bivariate kernel density estimation is computed over a fixed grid size of $K \times K$. In our experiments, we examine the difference in recognition results as K is varied for KDE, when compared to histograms of the same size. Figure 5 shows the difference in recognition results between the histograms and KDE for grid sizes of 4, 8 and 16. In all three cases, a notable improvement can be seen when the KDE is used. We use the value $K = 16$, because the results do not improve further when increasing K, despite a higher computational complexity of the required processing. Representations of the PDFs are shown in Figure 6 for both motion and location. The walking motion shown in Figure 6a has a difference value close to 1 for the entire sequence, this implies that the motion is all quite similar, which is expected of the walking in group activity. The gathering motion shown in Figure 6b displays a variety of difference values, which is expected as some individuals are gathering coming from different direction. The walking activity location differences shown in Figure 6c are all close to 1. This implies that the individuals are tightly grouped, which is expected in the walk group activity. The gather activity location differences shown in Figure 6d display clear transitions between locations far apart to locations close together towards the end of this activity. This is expected, as the gathering activity involves individuals coming from a distance towards gathering in a small group at the end of the activity.



a) Walk (Motion) b) Gather (Motion)

c) Walk (Location) d) Gather (Location)

Fig. 6. KDEs for the motion and location differences of activities from the NUS-HGA dataset.

For classification purposes, the density estimations are subsampled and fed to the classifier independently. The results are then combined to form a discriminant model as the motion and location features are often complimentary. For the classifier we use SVM with the RBF kernel, considering the parameters $C = 2.8284$ and $\gamma = 0.0019531$. For all experiments, we follow the evaluation protocol described in [10], where the NUS-HGA dataset is split into 5-fold training and testing and the performance is evaluated by average classification accuracy.

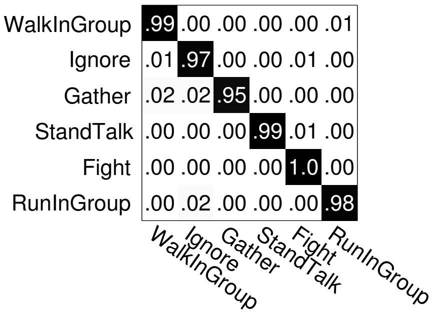


Fig. 7. Confusion matrix showing the recognition results when the combination of all four features are used is 98%

TABLE I
RECOGNITION RESULTS ON THE NUS-HGA DATASET

Method	Result (%)
Localized Causalities [10]	74.2%
Group interaction zone [18]	96.0%
Multiple-layered model [15]	96.2%
Motion differences	86.2%
Location differences	87.1%
Motion dynamics	91.6%
Location dynamics	92.6%
Motion and location differences	94.5%
Motion and location dynamics	97.1%
Combined differences and dynamics	98.0%

A comparison of the results when compared to the state-of-the-art in group activity recognition is shown in Table I. The location features provide a better recognition result than the motion features while the results for the dynamics models for motion and location emphasise their importance for group activity recognition. The combination of all features provides the best overall result of 98%. We should remark that the group interaction zone method from [18] does not evaluate the results using the 5-fold training and testing as suggested in [10], therefore slightly different results are expected from their method. In comparison to the state-of-the-art methods, we achieve a clear improvement in results of about 2%, while using a fully automated method.

VI. CONCLUSION

In this paper, we present an automatic approach for group activity recognition. We propose a model to describe the discriminative characteristics of group activity by considering the relations between motion flows and locations of moving regions in the scene as well as their dynamics in time. We also propose a scaling method to compensate for the effect of perspective projection in video sequences taken by cameras with wide angles located at low height. Moreover, we propose a stationary pedestrian detector to keep track of stationary pedestrians by marking the locations where they stop moving. Kernel Density Estimation (KDE) is used to model both time-location and time-motion spaces for representing such interactions. Experimental results show the effectiveness of the approach, without relying on any manual annotation of tracks like in other approaches.

ACKNOWLEDGMENT

This research work was supported by DSTL grant DSTLX1000074616 ?Human Activity Recognition.?

REFERENCES

[1] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," *Proc. IEEE Int. Work. on Visual Surveillance and Performance*, pp. 65–72, 2005.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.

[3] M. Baktashmotlagh, M. Harandi, and A. Bigdeli, "Non-linear stationary subspace analysis with application to video classification," *Proc. Int. Conf. on Machine Learning*, pp. 450–458, 2013.

[4] M. S. Ryoo and J. K. Aggarwal, "Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities," *International Conference on Computer Vision*, pp. 1593–1600, 2009.

[5] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2009, pp. 1446–1453.

[6] H. Nallaivarothayan, C. Fookes, S. Denman, and S. Sridharan, "An mrf based abnormal event detection approach using motion and appearance features," in *Proc. IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance,*, 2014, pp. 343–348.

[7] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 18–32, 2014.

[8] K. Stephens and A. G. Bors, "Observing human activities using movement modelling," in *Proc. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, 2015, pp. 44:1–6.

[9] ——, "Grouping multi-vector streaklines for human activity identification," in *Proc. IEEE Workshop on Image, Video and Multidimensional Signal Processing*, 2016.

[10] B. Ni, S. Yan, and A. Kassim, "Recognizing human group activities with localized causalities," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1470–1477, 2009.

[11] W. Lin, H. Chu, J. Wu, B. Sheng, and Z. Chen, "A heat-map-based algorithm for recognizing group activities in videos," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 23, no. 11, pp. 1980–1992, 2013.

[12] M. Chang and W. Ge, "Probabilistic group-level motion analysis and scenario recognition," *Proc. Int. Conf. on Computer Vision*, pp. 747–754, 2011.

[13] W. Choi and S. Savarese, "Understanding collective activities of people from videos," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 36, no. 6, pp. 1242–1257, 2014.

[14] Y. Zhang, W. Ge, M. C. Chang, and X. Liu, "Group context learning for event recognition," in *Proc. IEEE Work. on Applications of Computer Vision*, 2012, pp. 249–255.

[15] Z. Cheng, L. Qin, Q. Huang, S. Yan, and Q. Tian, "Recognizing human group action by layered model with multiple cues," *Neurocomputing*, vol. 136, pp. 124–135, 2014.

[16] R. Mehran, B. Moore, and M. Shah, "A streakline representation of flow in crowded scenes," *Proc. European Conference on Computer Vision, vol. LNCS 6313*, pp. 439–452, 2010.

[17] Z. Botev, J. Grotowski, and D. Kroese, "Kernel density estimation via diffusion," *Annals of Statistics*, vol. 38, no. 5, pp. 2916–2957, 2010.

[18] N.-G. Cho, Y.-J. Kim, U. Park, J.-S. Park, and S.-W. Lee, "Group activity recognition with group interaction zone based on relative distance between human objects," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 29, no. 5, pp. # 1 555 007, 1–15, 2015.