Proceedings Paper:

Stephens, Kyle and Bors, Adrian Gheorghe orcid.org/0000-0001-7838-0021 (2016) Group Activity Recognition on Outdoor Scenes. In: IEEE International Conference on Advanced Video and Signal-based Surveillance (AVSS). IEEE , pp. 59-65.

https://doi.org/10.1109/AVSS.2016.7738071

Group Activity Recognition on Outdoor Scenes

Anonymous AVSS submission for Double Blind Review

Paper ID 90

Abstract

In this research, we propose an automatic group activity recognition approach by modelling the interdependencies of group activity features over time. Unlike simple human activity recognition, the distinguishing characteristics of group activities are often determined by the way how the movement of people are influenced by one another. We propose to model the group interdependences in both motion and location spaces. These spaces are represented in time-space and time-movement spaces using Kernel Density Estimation (KDE). Such representations are then fed into a machine learning classifier. Unlike other approaches to group activity recognition, we do not rely on any long term tracklets or manual annotation of tracks.

1. Introduction

The area of human activity recognition is of interest for a variety of different applications such as video surveillance, human-computer interaction and semantic annotations of multimedia. Despite being a critical part of overall scene understanding, group activity recognition gained a significant interest only recently.

Research in simple human activity recognition was undertaken for several years [2, 18], often by modelling the activities using local features [11, 10] followed by their modelling. Recently, the focus of activity recognition has moved on to more complex problems such as scene understanding and analysis. One of such approaches is to detect abnormalities or uncommon activity events. Examples of such methods include [12], where the motion patterns are modelled using Gaussian Mixture Models (GMMs) of 3D distributions of local space-time gradients. Similarly, GMMs of Markov random fields (GMM-MRF) were used in [16] for abnormal activity detection. Dynamic texture models [13], which considers both appearance and dynamics, have also been considered for abnormal activity detection.

Group activity recognition requires more complex descriptions of the group interaction in the context of a given scenario assumption. Ni *et al.* [17] recognised group activities using localized causalities based on manually initialized tracklets. Lin *et al.* [14] used a heat-map based algorithm for modelling human trajectories when recognising group activities in videos. Chang *et al.* [4] 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.* [9] proposed a framework for analysing collective group activities based on different levels of semantic granularity. Zhang *et al.* [20] proposed an approach using histograms of the different features extracted from the tracklets of moving pedestrians. More recently, Cheng *et al.* [6] modelled group activity as a framework composed of multiple layers and Gaussian processes were used for representing motion trajectories. One dominating issue with the current group of approaches is that they mainly rely on some manual initialization of tracklets. Furthermore, each person in the scene is observed as a single tracklet entity, ignoring the potential discriminant features that could be extracted from more localised motions. Activities containing complex individual human movements cannot be well modelled by such approaches.

In this paper, we propose a automatic group activity approach by modelling the relationships of inter-dependant group movements and locations over time. In our approach, we avoid the use of manual tracklets and instead make use of medium term automatic movement estimation by using streaklines [15]. Distinct moving regions in the scene are segmented in space-time and the moving regions are modelled by their interdependencies by evaluating the differences in relative movement and locations. Kernel Density Estimation (KDE) is utilised to model the changes in the regions interdependencies over time in both time-location and time-motion spaces. Furthermore, the proposed model tracks the stopping of pedestrians by marking the locations when they stop moving. We also propose a scaling method to compensate for the perspective distortion present in video sequences acquired from lowly located cameras of wide view.

The rest of the paper is organised as follows: Section 2 describes the interdependency features used for representing moving regions, and the modelling of such interdependencies in the context of group activity is explained in Section 3. Section 4 describes the modelling of such interdependencies over time and discusses the classification of group activities. Section 5 shows the experimental results and Section 6 draws the conclusions of this research study.

2. Group Activity Modelling

The proposed methodology for group activity recognition has several stages, including extracting streaklines representing medium-time trajectories of movement, using these for modelling group interaction 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, crossing and occluding each other. To alleviate this problem, we propose to use the medium-time movement tracking method of streaklines, proposed in [15]. 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 [15]. This approach provides a smooth and robust representation of the movement flow over several frames. Unlike the approach in [15], we associate each streakline with blocks of pixels by using the marginal median as the streakline estimate. A first degree polynomial is then fit to the streakline in order to obtain a smoother representation. This differs from [19], where the authors use PCA for estimating the principal streak. One issue with the approach from [19] is that it does not consider the motion consistency over several frames. In this research paper we ensure the consistency of the streaklines over several frames. Furthermore, we make the assumption that each compact region of streakflows may contain several distinct movements, which are represented by clusters. Firstly, we begin by segmenting the streakflow field into distinct moving regions using the Expectation-Maximization (EM) algorithm, under the Gaussian Mixture Model (GMM) assumption. The number of clusters and the centres of the Gaussian functions in the EM algorithm are initialised using the modes of the histogram of flow improving the convergence. The space of clustering is defined jointly by both movement and localisation, as given by the streakflows and their locations in the frame, respectively.

We also address the effects of perspective distortions 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 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. 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 appropriately identified and the respective moving region is dropped from the existing movements dictionary considered for the scene. 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.

3. Modelling Interdependent Relationships of Moving Regions

The key characteristics of group activities are often present in the interdependent relationship between the pedestrians/moving objects. In this research study we propose to model the interdependent relationships 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 relative movement between streakflow models in the scene, considering both direction and intensity of movements. This models the inter-

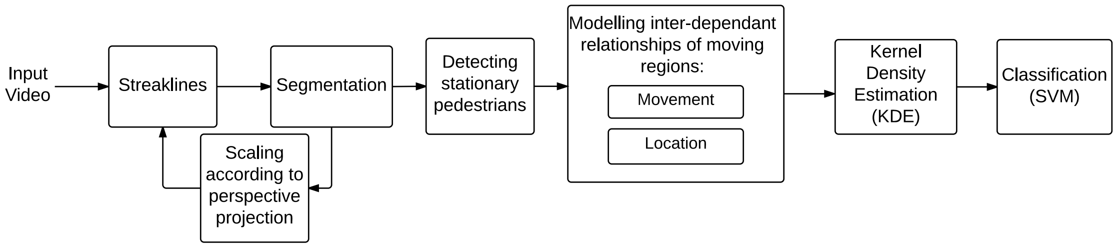


Figure 1. Overview of the proposed group activity recognition approach

dependant relationship of the group movement 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 scaled value within the range $[0, 1]$, representing the difference between two streakflow models, each characterising the relative movement of one region with respect to another. The differences are computed by considering all pairs of moving regions in the scene at a particular time t by using equation (3). The differences 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 all streakflows 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 the moving region. By this model, the mean will approximate the centre of the region, 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] which represents the difference between the two locations. 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 very 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. 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 of location differences, representing all the inter-dependant relationships of location points between time t and $t + n$, is then obtained.

One further issue that arises when computing such differences is that the rate of movement change and rate of location change is not clearly characterised. To overcome this, we consider the background as an additional region for both the streakflow model and the location model. 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 is characterised representing the absolute movement of people in the scene. 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 $B(t)$ (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 pairwise motion and location differences between the moving regions in the scene.

4. Classification of Group Activities

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 inter-dependences 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 (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, helping 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 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 for training a Support Vector Machines (SVM) algorithm.

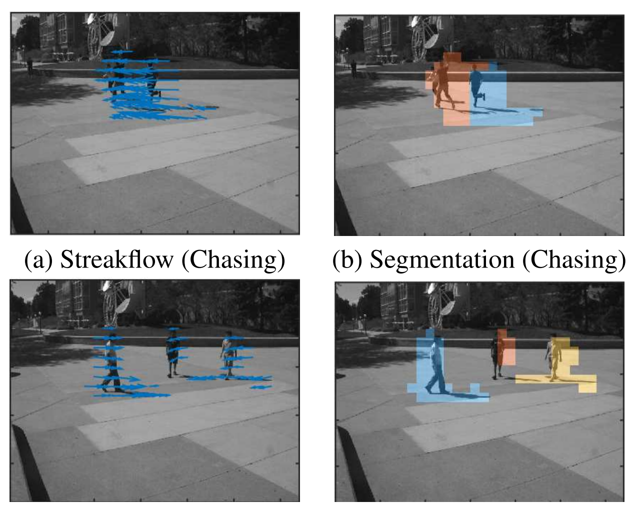
5. Experimental Results

For all experiments, we follow the same recognition routine. To begin, the streakflows are extracted for each set of frames and the moving regions are segmented based on the streakflows in each inter-connected region. Streakflow models and their location models are extracted for the moving regions in each set of frames. The features of the moving regions are then modelled by the inter-dependant differences between all moving regions across a set of frames. The dynamic changes of the features are modelled by the inter-dependant differences between all moving regions in one set of frames and the next set. Then, 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 difference in features along one axis and time along the other. Finally, the densities are used as features to build a classifier and make recognition decisions via a Support Vector Machine (SVM) (with RBF kernel).

5.1. New Collective dataset

The new Collective dataset [8] consists of 6 collective activities: gathering, talking, dismissal, walking together, chasing and queueing. The dataset consists of 32 video sequences, where each video sequence contains multiple examples of each activity. The video sequences are recorded using a hand-held camera, and therefore the perspective distortion is quite strong.

To start, the video sequence is segmented spatio-temporally into blocks of 20×20 pixels by 10 frames, where the streaklines are extracted for each block of 10 frames. The motion filter is applied over each 3 sets of frames. The movement segmentation is applied as in Section 2, and examples of the streakflows and movement segmentation are shown in Figure 2 for the chasing and gather activities. In both cases, the moving regions are well segmented, particularly in the chasing example where the chaser and chasee are segmented separately despite forming one connected region moving in the same direction.



(a) Streakflow (Chasing) (b) Segmentation (Chasing)

(c) Streakflow (Gather) (d) Segmentation (Gather)

Figure 2. Examples of streakflow and segmentation on the new Collective dataset

The next step involves applying the stationary pedestrian detector as in Section 2 where the prior frames p is $p = 25$ and the boundary parameter is set to 15% of the region size. In the collective dataset, the pedestrians transition between different activities, some of which include the pedestrians remaining stationary. An example of the transitioning stationary pedestrians through three activities are shown in Figure 3. At the start, shown in Figure 3 a), the pedestrians are moving towards each other performing the gathering activity. At the end of the gathering activity, the pedestrians have gathered and transition to the talking activity shown in Figure 3 b). The stationary pedestrian detection has successfully recorded the last locations of the individuals as seen in Figure 3 b), despite the individuals having stopped moving. Finally, after a period of time, the individuals begin to move again performing the dispersing activity shown in Figure 3 c). In Figure 3 c), the new moving regions are detected and replace the previously identified stopped regions which are no longer recorded.

Next, the features (streakflow differences, streakflow dynamics, location differences and location dynamics) are computed for each moving region as described in Section 3. The scaling parameters (σ_m and σ_l) for the feature equations from Section 3 are varied and the best parameter values are selected for each feature. The best recognition results are obtained when $\sigma_m = 15$ and $\sigma_l = 450$ for both motion features and location features respectively. The size of the dynamic window for the motion dynamics and location dynamics n is set to $n = 5$.

Following the computation of the streakflow differences, streakflow dynamics, location differences and location dynamics, the data is represented over time using KDE as described in Section 4. The KDE is applied over a fixed grid size using the 2-column feature matrices as input data. In this work, we choose to utilise the bi-variate KDE method proposed in [3] which is based on using linear diffusion processes. The KDE methodology from [3] assumes the kernel to be Gaussian and uses a bandwidth selection method such that the bandwidth parameters are automatically selected depending on the data. The use of KDE over traditional histograms has several key advantages, most notably adaptive smoothing of the data which not only helps with the smoothing of noise but provides smooth transitions of the feature differences over time. Secondly, the automatic bandwidth selection method allows for different granularity of different features to be represented depending on the feature data. Next, we compare the use of the proposed KDE method to conventional histograms using the same fixed grid size of $K \times K$. In this experiment, K is varied and the recognition accuracy is compared between histograms and KDE. The results are shown in Figure 4. In Figure 4, the KDE results shows a notable improvement over their equivalent-sized histograms, demonstrating the effectiveness of KDE over histograms. In our experimental work, there was no improvement in recognition results by using grid sizes larger than $K = 8$. Furthermore, the computational complexity increases significantly when grid sizes larger than $K = 16$ are used. Therefore, in our experiments, we choose $K = 8$. Finally, the KDEs are used as input to the SVM classifier with RBF kernel.

Table 1. Recognition results on the new Collective dataset

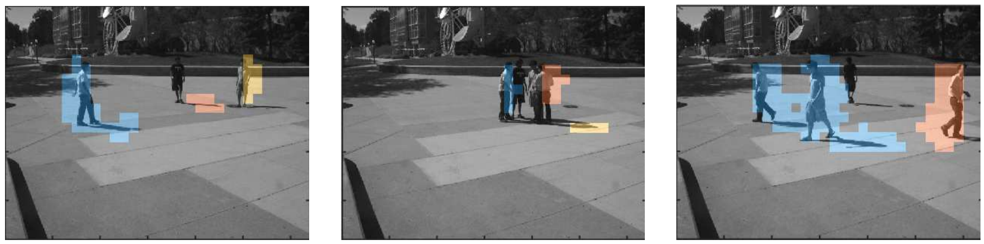
Method	Result (%)
Monte Carlo Tree Search [1]	77.7%
Collective activities [9]	79.2%
MIR [5]	80.3%
Motion differences	75.4%
Motion dynamics	76.8%
Location differences	64.3%
Location dynamics	71.6%
Motion and location differences	76.5%
Motion and location dynamics	78.4%
Combined differences and dynamics	79.7%

To compare with state of the art, we follow the recommended evaluation protocol from [8] and divide the dataset into 3 subsets for 3-fold training and testing. Since the data sequences contain an unknown quantity of activities of an unknown length, we split the sequences during training and testing to short sequences of 60 frames each for evaluation. We compare our results to state of the art using average recognition accuracy across all activity classes. Confusion matrices of the results of our combined features compared to the approach from [5] are shown in Figure 5. One observation of the confusion matrices is that the queuing activity is not well classified in our method. This is due to the stationary pedestrians not moving at all for the duration of the sequence, therefore our stationary detector fails to detect the pedestrian. Considering this, a further observation from Figure 5 is that we achieve an improvement in overall recognition results when the queuing activity is not considered, and also greater consistency in the results across the other activities. Comparison of our recognition results when compared to state of the art are shown in Table 1. Notably, our method is comparative to state of the art and superior when the queuing activity is removed, despite using an automatic method.

5.2. NUS-HGA Dataset

We also evaluate our method on the NUS-HGA dataset [17]. This data set consists of six different group activities collected in five different sessions. We follow the same experimental outline as described above.

To begin, streaklines are extracted for blocks of size 14×14 over 10 consecutive frames. The motion filter described in Section 2 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 2. 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



a) Gathering b) Talking b) Dispersing

Figure 3. Example of pedestrians transitioning through activities in the new Collective dataset.

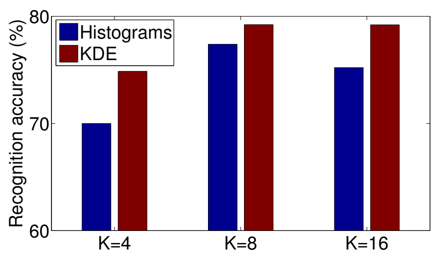
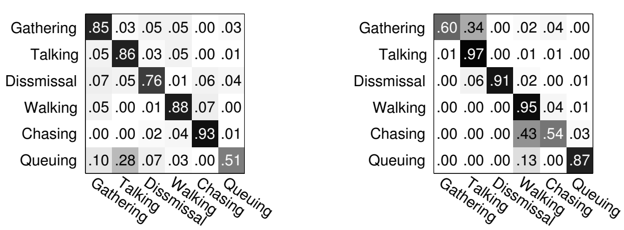


Figure 4. Difference in recognition accuracy between histograms and KDE for 3 different grid sizes.



(a) Ours - 79.7% (b) [5] - 80.3%

Figure 5. Confusion matrices for the recognition results on the new Collective dataset

detector is applied as in Section 2 where the number of prior frames is set to $p = 25$. We define the boundary parameter as 10% of the region size.

The streakflow movement model, streakflow dynamics, location and location dynamics relationship differences are computed as in Section 3, 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 dynamic window from Section 3 is set to $n = 13$. The data is represented by a 2-column matrix over time as described in Section 4. KDE is applied over a fixed grid size using the 2-column feature matrices as input data where $K = 16$.

For classification purposes, the density estimations are sub-sampled and fed to the classifier independently. For the classifier we use SVM with the RBF kernel, and we follow the evaluation protocol described in [17], where the NUS-HGA dataset is split into 5-fold training and testing and the performance is evaluated by average classification accuracy.

Table 2. Recognition results on the NUS-HGA dataset

Method	Result (%)
Localized Causalities [17]	74.2%
Group interaction zone [7]	96.0%
Multiple-layered model [6]	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 2. 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%. In comparison to state-of-the-art methods, we achieve a clear improvement in results of about 2%, while using a fully automated method.

6. Conclusions

In this paper, we proposed 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. We also proposed a scaling method to compensate for the effect of perspective projection in video sequences with perspective distortion. A stationary pedestrian detector is used in order 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 such group movement interactions. Experimental results on a group activity dataset demonstrate the effectiveness of the approach, without relying on any manual annotation of tracks like other methods.

References

[1] M. R. Amer, S. Todorovic, A. Fern, and S.-C. Zhu. Monte carlo tree search for scheduling activity recognition. *Proc. IEEE Int. Conf. on Computer Vision*, pages 1353–1360, 2013.

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

[3] Z. Botev, J. Grotowski, and D. Kroese. Kernel density estimation via diffusion. *Annals of Statistics*, 38(5):2916–2957, 2010.

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

[5] X. Chang, W.-s. Zheng, and J. Zhang. Learning personperson interaction in collective activity recognition. *IEEE Transactions on Image Processing*, pages 1905–1918, 2015.

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

[7] 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*, page 1555007, 2015.

[8] W. Choi and S. Savarese. A unified framework for multi-target tracking and collective activity recognition. In *Proc. European Conference on Computer Vision, vol. LNCS 7575*, pages 215–230, 2012.

[9] W. Choi and S. Savarese. Understanding collective activities of people from videos. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 36(6):1242–1257, 2014.

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

[11] 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*, pages 65–72, 2005.

[12] 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*, pages 1446–1453, 2009.

[13] W. Li, V. Mahadevan, and N. Vasconcelos. Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):18–32, 2014.

[14] 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*, 23(11):1980–1992, 2013.

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

[16] 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,*, pages 343–348, 2014.

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

[18] 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*, pages 1593–1600, 2009.

[19] K. Stephens and A. G. Bors. Observing human activities using movement modelling. In *Proc. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pages 1–6, 2015.

[20] 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*, pages 249–255, 2012.