# A New Heat-Map-based Algorithm for Human Group Activity Recognition

Hang Chu[1], Weiyao Lin[1*], Jianxin Wu[2], Xingtong Zhou[1], Yuanzhe Chen[1], Hongxiang Li[3]

[1] Dept. of Electronic Engineering, Shanghai Jiao Tong University, China (*Corresponding author)

[2] School of Computer Engineering, Nanyang Technological University, Singapore

[3] Department of Electrical and Computer Engineering, University of Louisville, USA

## ABSTRACT

In this paper, a new heat-map-based (HMB) algorithm is proposed for human group activity recognition. The proposed algorithm first models people trajectories as series of "heat sources" and then applies a thermal diffusion process to create a heat map (HM) for representing the group activities. Based on this heat map, a new surface-fitting (SF) method is also proposed for recognizing human group activities. Our proposed HM feature can efficiently keep the temporal motion information of the group activities while the proposed SF method can effectively catch the characteristics of the heat map for activity recognition. Experimental results demonstrate the effectiveness of our proposed algorithm.

## Categories and Subject Descriptors

I.2.10 [**Vision and Scene Understanding**]: Video analysis

## Keywords

Human Group Activity Recognition, Heat Map, Surface Fitting

## 1. INTRODUCTION

Detecting human group activities or human interactions has attracted increasing research interests in many applications such as video surveillance and human-computer interaction [1-6].

Many algorithms have been proposed for recognizing group activities or interactions [1-6]. Zhou et al. [2] detect pair-activities by extracting the causality features from bi-trajectories. Ni et al. [3] further extend the causality features into three types of individuals, pairs and groups. Besides, Chen et al. [5] detect group activities by introducing the connected active segmentations for representing the connectivity among people. Cheng et al. [4] propose the Group Activity Pattern for representing group activities where Gaussian parameters from trajectories are calculated from multiple people. However, most of the existing algorithms extract the overall features from the activities' entire motion information (e.g., the statistical average of the motion trajectory). These features cannot suitably embed activities' temporal motion information (e.g., fail to indicate when and where a person is in the video). Thus, they will have limitations when recognizing more complex group activities. Although other methods [6] incorporate the temporal information with the chain models such as the Hidden Markov Models (HMM), they have the disadvantage of requiring large-scale training data, as will be discussed later.

In another part, handling motion uncertainties is also an important issue in group activity recognition. Since the motions of people vary inherently in group activities, the recognition accuracy may be greatly affected by this uncertain motion nature. Although some methods utilize Gaussian Processes Estimation or filtering to handle this uncertain problem [3, 4], they do not simultaneously consider the issue for reserving the activities' temporal motion information.

Furthermore, the recognition method is a third key issue for recognizing group activities. Popularly-used recognition models include Support Vector Machine (SVM) [2-4], Linear Discriminative Analysis, and HMM [6]. Although these models show good results in many scenarios, their training difficulty and the requirement of the training data scale will increase substantially when the feature vector length becomes large or the group activity becomes complex. Therefore, it is also non-trivial to develop more flexible recognition methods for effectively handling the recognition task.

In this paper, we propose a new heat-map-based (HMB) algorithm for human group activity recognition. The contributions of our work can be summarized as follows.

1) We propose a new heat map (HM) feature to represent group activities. The proposed HM can effectively catch the temporal motion information of the group activities.
2) We propose to introduce a thermal diffusion process to create the heat map. By this way, the motion uncertainty from different people can be efficiently addressed.
3) We also propose a new surface-fitting (SF) method to recognize the group activities. The proposed SF method can effectively catch the characteristics of our heat map feature and perform recognition efficiently.

The remainder of this paper is organized as follows. Section 2 describes the basic ideas of our proposed HM feature and SF method. Section 3 presents the details of our HMB algorithm. The experimental results are shown in Section 4 and Section 5 concludes the paper.

## 2. BASIC IDEAS FOR HM AND SF

As mentioned, given the activities' motion information (i.e., motion trajectory in this paper), directly extracting the overall features will lose the useful temporal information. In order to avoid such information loss, we propose to model the activity trajectory as a series of heat sources. As shown in Figure 1, (a) is the trajectory of one person. In order to transfer the trajectory into heat source series, we first divide the entire video scene into small non-overlapping patches (i.e., the small squares in (b)). If the trajectory goes through a patch, this patch will be defined as one heat source. By this way, a trajectory can be transferred into a series of heat sources, as in Figure 1 (b). Furthermore, in order to further catch the temporal information of the trajectory, we also introduce a decay factor on different heat sources such that the thermal energies of the "older" heat sources (i.e., patches closer to the starting point of the trajectory) are smaller while the "newer" heat sources will have larger thermal energies. By this way, the thermal values of the heat source series can be arranged increasingly according to the direction of the trajectory and the temporal information can be effectively embedded.

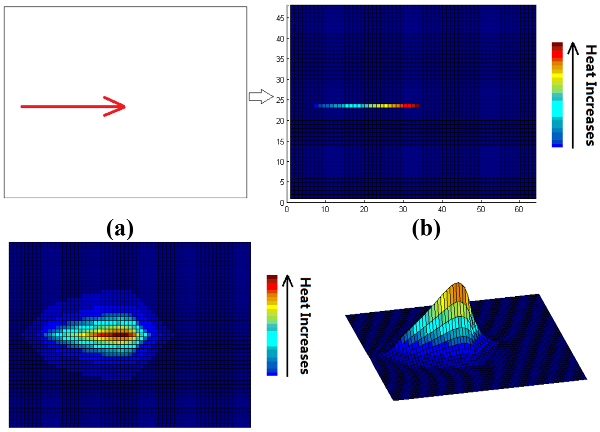


**Figure 1. (a): The activity trajectory; (b) The corresponding heat source series; (c) The heat map (HM) diffused from the heat source series in (b); (d) The HM surface of (c) in 3D.**

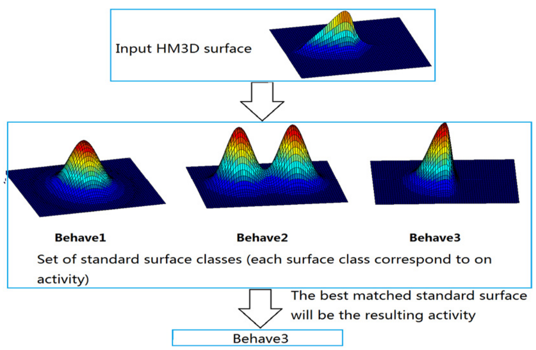


**Figure 2. The process of the surface-fitting (SF) method.**

Furthermore, since people's trajectories may have large variance, directly using the heat source series as features will be greatly affected by this motion fluctuation. Therefore, in order to reduce the motion fluctuation, we further propose to introduce a thermal diffusion process to diffuse the heats from the heat source series to the entire scene. We call this diffusion result as the heat map (HM). Note that although the heat diffusion was introduced in object segmentation in some works [8], the mechanism and utilization of HM in our algorithm is far different from them. And to the best of our knowledge, this is the first work to introduce HM into group activity recognition. Figure 1 (c) and (d) show the HM of the trajectory in Figure 1 (a). More details will be described in the next section.

With our HM feature, we can describe the activities' motion information by 3D surfaces (as in Figure 1 (d)). Since the HM feature includes rich information, the problem then comes to the selection of a suitable method for performing recognition based on this HM feature. In this paper, we further propose a surface-fitting (SF) method for activity recognition. In our SF method, a set of standard surface classes are first identified for representing different activities. Then, the similarities between the surface of the input HM and the standard surface classes are calculated. And finally, the best matched standard surface class will be picked up and its corresponding activity will become the recognized activity for the input HM. The process of our SF method is shown in Figure 2.

With the basic ideas of the HM feature and the SF method described above, we can propose our heat-map-based (HMB) group activity recognition algorithm. It is described in detail in the following section.

## 3. THE HMB ALGORITHM

The framework of our HMB algorithm can be described by Figure 3. In Figure 3, the input group activities' trajectories are first transferred into heat source series, then the thermal diffusion process is performed to create the HM feature for describing the input group activity. Then, the SF method is used for activity recognition. As mentioned, the heat source series transfer, the thermal diffusion, and the SF method are the key parts of our proposed algorithm. Thus, we will focus on describing these three parts in the following.

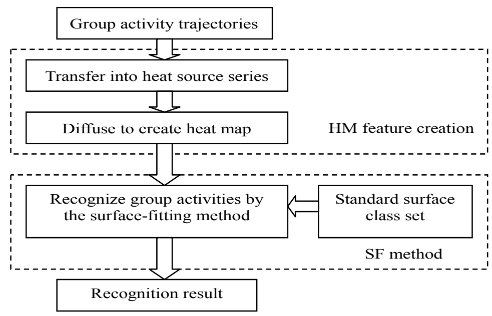


**Figure 3. The process of the HMB algorithm.**

### 3.1 Heat Source Series Transfer

Assume that we have totally $j$ trajectories in the current group activity. The thermal energy $E_i$ of the heat source patch $i$ can be calculated by:

$$E_i = \sum_j \overline{E_{i,j}} \cdot e^{-k_t\left(t_{cur}-t_{id,j}\right)} \tag{1}$$

where $e^{-k_t\left(t_{cur}-t_{id,j}\right)}$ is the time decay term [10], $k_t$ is the temporal decay coefficient, $t_{cur}$ is the current frame number, and $t_{id,j}$ is the frame number when the $j$-th trajectory leaves patch $i$. $\overline{E_{i,j}}$ is the accumulated thermal energy for trajectory $j$ in patch $i$. It can be calculated by Eq. (2).

$$\overline{E_{i,j}} = \int_0^{t_{id,j}-t_{is,j}} C \cdot e^{-k_t t} dt = \frac{C}{k_t}\left(1-e^{-k_t\left(t_{id,j}-t_{is,j}\right)}\right) \tag{2}$$

where $t_{is,j}$ and $t_{id,j}$ are the frame number when the $j$-th trajectory enters and leaves patch $i$, respectively. $k_t$ is the temporal decay coefficient as in Eq. (1), and $C$ is a constant. In the experiments of our paper, $C$ is set to be 1. From Eq. (2), we can see that the accumulated thermal energy is proportional to the stay length of trajectory $j$ at patch $i$. If $j$ stays in $i$ for longer time, more thermal energy will be accumulated in patch $i$. On the other hand, if $j$ does not go through patch $i$, the accumulated thermal energy will be 0 indicating that patch $i$ is not a heat source patch.

### 3.2 Thermal Diffusion

After getting the source heat series by Eq. (1), the thermal diffusion process will be performed over the entire scene to create the HM. The HM value $H_i$ at patch $i$ after diffusion [10] can be calculated by:

$$H_i = \frac{\sum_{l=1}^{N} E_l \cdot e^{-k_p d(i,l)}}{N} \qquad (3)$$

where $E_l$ is the thermal energy of the heat source patch $l$, $N$ is the total number of heat source patches. $k_p$ is the spatial diffusion coefficient, and $d(i, l)$ is the distance between patches $i$ and $l$.

The advantage of the thermal diffusion process can be described by Figure 4. In Figure 4, the left column lists two different trajectory sets for the group activity "gather". Due to the variation of people activity, these two trajectory sets are obviously different from each other. And these differences are exactly transferred to their heat source series (the middle column). However, with the thermal diffusion process, the trajectory differences are suitably "blurred", which makes their HMs (the right column) close to each other. At the same time, the temporal information of the two group activities is still effectively reserved in the HMs.

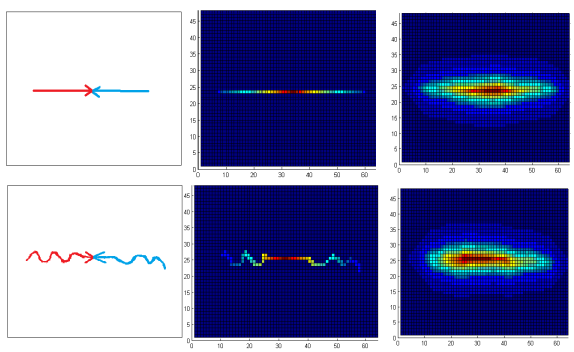

**Figure 4. Left column: two trajectory sets for group activity "Gather"; Middle column: the corresponding heat source series; Right column: the corresponding heat maps.**

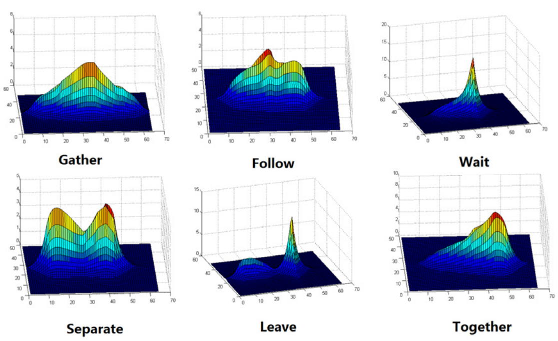


**Figure 5. The HM surfaces for different group activities.**

Furthermore, Figure 5 shows the example HM surfaces for different group activities defined in Table 1. From Figure 5, it is clear that our proposed HM can precisely catch the activities' temporal information and show obviously distinguishable patterns among different activities.

**Table 1. Definitions to different group activities**

| | |
|---|---|
| Gather | Two or more persons gathering to a point. |
| Follow | One or group of person is followed by one person. |
| Wait | One of one group of person is waiting for one person. |
| Separate | Two or more persons are separating from each other. |
| Leave | One person is leaving one or one group of person. |
| Together | Two or more persons are walking together. |

## 3.3 The SF Method

The surface-fitting process can be described by Eq. (4):

$$M = \arg\min_{m} \left( \min_{\boldsymbol{\theta}} \| S_{HM} - S_{SD,m}(\boldsymbol{\theta}) \| \right) \qquad (4)$$

where $M$ is the final recognized activity. $S_{HM}$ is the surface of the input HM, $S_{SD,m}(\boldsymbol{\theta})$ is the standard surface class for activity $m$, $\boldsymbol{\theta}$ is the parameter set for controlling the surface class $S_{SD,m}$. For example, if $S_{SD,m}$ is a Gaussian surface class, $\boldsymbol{\theta}$ will include the mean vector and the covariance matrix of the Gaussian surface. From Eq. (4), we can see that the SF method includes two steps. In the first step, an optimal parameter set $\boldsymbol{\theta}$ is found to make each standard surface class fit with the input HM surface [7]. And then in the second step, the standard surface class that best fits the input HM surface will be selected as the recognized activity.

Generally, the standard surface class can be selected manually or by some training data such as in Eq. (5).

$$S_{SD,m} = \arg\min_{S_{can}} \left( \min_{\theta_1, \theta_2, \ldots \theta_{n_m}} \sum_{i, i \in Tr_m} \| S_{HM,i} - S_{can}(\boldsymbol{\theta}_i) \| \right) \qquad (5)$$

where $S_{can}(\boldsymbol{\theta})$ is the candidate standard surface class, $S_{HM,i}$ is the HM surface of activity $m$ for the $i$-th training sample, $Tr_m$ is the training data set for activity $m$, and $n_m$ is the number of training samples in $Tr_m$.

However, since the process of fitting two 3D surfaces is normally computation intensive and sometimes even unsolvable [7], we further propose a simplified surface fitting (SFF) method. The SFF method is described by Eq. (6).

$$M = \arg\min_{m} \left( \| S_{HM} - \widetilde{S}_{SD,m} \| \right) \qquad (6)$$

where $\widetilde{S}_{SD,m}$ is the approximated standard surface for activity $m$ and it can be achieved by the following two steps:

1) For activity $m$, estimate its corresponding standard trajectories $\boldsymbol{T_m}$ from the input HM surface $S_{HM}$.

2) Create the HM for $\boldsymbol{T_m}$ by Eqs (1)-(3), and the resulting HM surface will be the approximated standard surface for $m$.

There can be various ways to create the estimated standard trajectories $\boldsymbol{T_m}$. In this paper, we create $\boldsymbol{T_m}$ by some pre-set rules. Due to the limited space, we only show one rule for creating $\boldsymbol{T_m}$ for the group activity "follow" in Figure 6. The rules for other activities can be defined in a similar way.

The advantage of using the SFF method can be summarized in the following:

1) Since the approximated standard surface is derived from the input HM surface $S_{HM}$, it is inherently fitted with $S_{HM}$. Thus, the computation-intensive surface fitting step can be skipped.

2) The rules for creating the estimated standard trajectories $\boldsymbol{T_m}$ can be viewed as the pre-embedded information for describing the activity $m$. Thus, it also provides the flexibility for allowing the embedding of different rules for different activities.

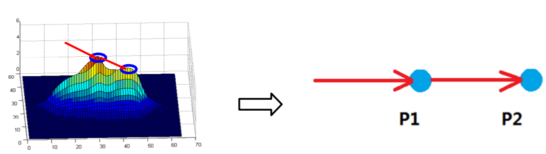


**Figure 6. The rules for creating $T_m$ for "follow": Detect the largest two peaks from $S_{HM}$, and generate two traces in the same direction from the higher peak to the lower peak.**

# 4. EXPERIMENTAL RESULTS

In this section, we show experimental results for our proposed

HMB algorithm. In our experiment, the patch size is set to be 10x10, and $k_t$ and $k_p$ in Eqs (1) and (3) are set to be 0.125 and 0.5, respectively. And the SSF method is used for surface fitting in our experiments. We perform experiments on the BEHAVE dataset [1] where 325 video clips are selected for the six group activities in Table 1 (with about 50 clips for each activity). We compare our HMB algorithm with the other three methods: the WF-SVM algorithm [2], the PGTB algorithm [5] and the GRAD algorithm [6]. Note that since the WF-SVM and the GRAD algorithms require the training process, we split the dataset into 75% training-25% testing parts and perform recognition on the testing part [6]. Four independent experiments are performed and the results are averaged. Furthermore, in order to get rid of the effect of tracking errors, we use the ground-truth trajectories in the experiments. In practice, various tracking algorithms [9] can be utilized to achieve trajectories. Table 2 shows the Miss, False Alarm (FA), and Total Error Rates (TER) [6] for different methods.

**Table 2 Miss, FA, and TER rates for different methods**

| | | HMB | WF-SVM [2] | PGTB [5] | GRAD [6] |
|---|---|---|---|---|---|
| Gather | Miss | **0.8%** | 12.0% | 32.9% | 11.3% |
| | FA | **1.3%** | 0.3% | 4.3% | 1.6% |
| Follow | Miss | **5.8%** | 8.3% | 8.3% | 16.2% |
| | FA | **0.9%** | 2.3% | 34.4% | 1.3% |
| Wait | Miss | **4.5%** | 9.0% | 32.1% | 14.3% |
| | FA | **0.6%** | 2.0% | 0.0% | 1.8% |
| Separate | Miss | **3.3%** | 5.0% | 35.0% | 8.2% |
| | FA | **0.1%** | 0.4% | 7.1% | 1.0 % |
| Leave | Miss | **2.7%** | 4.8% | 44.8% | 9.6% |
| | FA | **1.1%** | 1.4% | 0.0% | 2.5% |
| Together | Miss | **5.1%** | 2.9% | 61.2% | 2.6% |
| | FA | **0.3%** | 1.8% | 0.2% | 0.8% |
| TER | | **3.7%** | 7.0% | 36.7% | 10.9% |

From Table 2, we can see that the PGTB algorithm produces poor results. This is because the PGTB algorithm recognizes activities mainly based on the change of connectivity among people [5]. These simple features cannot precisely describe group activities and will fail to work in various scenarios such as people approaching each other but haven't met. Comparatively, The WF-SVM and the GRAD algorithms have much improved performances by using more distinguishable features. However, due to the complexity and uncertainty of human activities, they still produce unsatisfactory results for some group activities such as "Gather" and "Wait". Compared to these methods, our proposed HMB algorithm has the best performance. This demonstrates that our HM features are able to precisely catch the characteristics of activities and our SSF method can effectively cooperate with the HM features for performing recognition.

In order to further demonstrate our HM features, we perform another experiment for recognizing two complex activities: exchange (i.e. two people first approach each other, stay together for a while and then separate) and return (i.e., two people first separate and then approach to each other later). In Figure 7, (a) shows the trajectories of the two complex activities, (b) shows the values of the major features in the WF-SVM algorithm [2], and (c) shows the HM surfaces. From Figure 7 (b), we can see that the features in the WF-SVM algorithm cannot show much differences between the two complex activities. Compared to (b), our HMs in (c) are obviously more distinguishable. The recognition results for the WF-SVM algorithm and our HMB algorithm are shown in Table 3. The results in Table 3 further demonstrate the effectiveness of our HM features in representing complex group activities.

# 5. CONCLUSION

In this paper, a heat-map-based (HMB) algorithm is proposed for recognizing group activities. We propose to create the heat map for representing the group activities and use the surface fitting method for activity recognition. Experimental results demonstrate the effectiveness of our proposed algorithm.

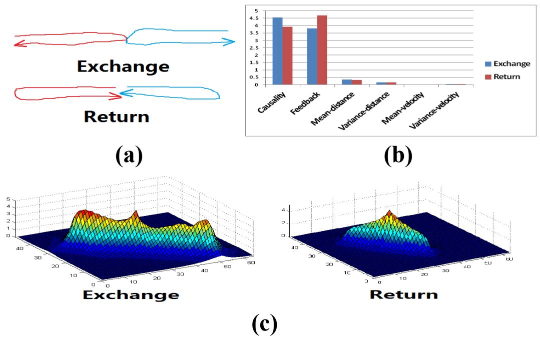


**Figure 7. (a) The trajectories for the two complex activities; (b) The major feature values for the WF-SVM algorithm [2]; (c) The HMs for the two complex activities.**

**Table 3 Miss and TER rates for the complex activities**

| | | HMB | MF-SVM |
|---|---|---|---|
| Exchange | Miss | **6.3%** | 50.0% |
| Return | Miss | **12.5%** | 43.8% |
| TER | | **9.4%** | 46.9% |


# 6. ACKNOWLEDGMENTS

This paper is supported in part by the following grants: National Science Foundation of China grants (61001146, 61025005), Chinese national 973 grants (2010CB7314016), the Open Project Program of the National Laboratory of Pattern Recognition (NLPR), and SMC scholarship of SJTU.


# REFERENCES


[1] BEHAVE data: http://homepages.inf.ed.ac.uk/rbf/BEHAVE.

[2] Y. Zhou, S. Yan, and T. Huang, "Pair-activity classification by bi-trajectory analysis," *IEEE Conf. Computer Vision and Pattern Recogntion (CVPR)*, pp. 1-8, 2008.

[3] B. Ni, S. Yan, and A. Kassim, "Recognizing human group activity with localized causalities," *IEEE Conf. Computer Vision and Pattern Recogntion (CVPR)*, pp. 1470-1477, 2009.

[4] Z. Cheng, L. Qin, Q. Huang, S. Jiang, and Q. Tian, "Group activity recognition by Gaussian process estimation," *Int'l Conf. Pattern Recognition (ICPR)*, pp. 3228-3231, 2010.

[5] Y. Chen, W. Lin, H. Li, H. Luo, Y. Tao, and D. Liu, "A new package-group-transmission-based algorithm for human activity recognition in videos," *IEEE Visual Communications and Image Processing(VCIP)*, pp. 1-4, 2011.

[6] W. Lin, M.-T. Sun, R. Poovendran, and Z. Zhang, "Activity recognition using a combination of category components and local models for video surveillance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 8, pp. 1057-1067, 2008.

[7] R. Strongin and Y. Sergeyev, *Global Optimization with Non-Convex Constraints-Sequential and Parallel Algorithms*, Springer Press, 2000.

[8] Y. Fang, M. Sun, M. Kim, and K. Ramani, "Heat-mapping: a robust approach toward perceptually consistent mesh segmentation," *IEEE Conf. Computer Vision and Pattern Recogntion (CVPR)*, pp. 2145-2152, 2011.

[9] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–574, 2003.

[10] H. Carslaw and J. Jaeger, *Conduction of Heat in Solids*, Oxford University Press, 1986.