

# Human Activity Recognition for Video Surveillance

Weiyao Lin\*, Ming-Ting Sun\*, Radha Poovandran\*, and Zhengyou Zhang\*\*

\* Department of Electrical Engineering,  
University of Washington, Seattle, WA 98195  
{wylin,mts,rp3}@u.washington.edu

\*\*Microsoft Research (MSR)  
Microsoft Corp., Redmond, WA 98052  
zhang@microsoft.com

**Abstract**—This paper presents a novel approach for automatic recognition of human activities from video sequences. We first group features with high correlations into Category Feature Vectors (CFVs). Each activity is then described by a combination of GMMs (Gaussian Mixture Models) with each GMM representing the distribution of a CFV. We show that this approach offers flexibility to add new events and to deal with the problem of lacking training data for building models for unusual events. For improving the recognition accuracy, a Confident-Frame-based Recognizing algorithm (CFR) is proposed to recognize the human activity, where the video frames which have high confidence for recognition an activity (Confident-Frames) are used as a specialized model for classifying the rest of the video frames. Experimental results show the effectiveness of the proposed approach.

## 1. INTRODUCTION AND RELATED WORK

Video surveillance is of increasing importance to many applications, such as elder-care, home-nursing, and unusual event alarming [1]. The automatic activity recognition plays as the key part for video surveillance. In this paper, besides addressing the issue of human activity recognition, we also address the following two key problems: (1) the recognition of unusual events which lack training data, and (2) the flexibility and accuracy of the algorithms to recognize human activities. They are described in details below.

In many surveillance applications, events of interest may occur rarely. For these unusual events (or abnormal, rare events), it is difficult to collect sufficient training data for supervised learning to develop unusual event models. In this case, many unusual event detection algorithms [8] which require large numbers of training data become unsuitable. Methods for learning from small numbers of examples are needed [5]. Several algorithms have been proposed to address the difficulty of unusual event recognition with lack of training data. Zelnik-Manor et al. [3] and Zhong et al. [4] cluster the divided video clips into different groups based on a similarity measure. The groups with relatively small numbers of video clips are detected as unusual events. However, since unusual events have insufficient training data, clusters for these events may not be representative enough to predict future unusual events. Zhang et al. [5] proposed a method by developing the unusual event model from that of usual events. This method provides a hint on how to deal with the lack-of-training-data problem. However, they obtain all unusual event models by adapting from the general usual-event model, while in reality, the usual events and unusual events can be far different from each other in nature. In this paper, we propose a new approach for recognizing unusual events. We first group features into “Category Feature Vectors” (CFVs). Each activity is then described by a combination of GMMs with each GMM representing the distribution of a CFV. The unusual event model is derived from combining the most relevant CFVs from different usual event models.

Human activity recognition is a challenging task due to the non-rigidness of the human body. Currently, many algorithms have been proposed to recognize human activities. Lv et al. [6] and Ribeiro [7] focus on the selection of suitable feature sets for

different events. Models such as Hidden Markov Model [8], state machine [9], Adaboost [2] are also widely used for activity recognition. However, most of the methods proposed in these works are not flexible for adding new activities. They are trained or constructed to recognize predefined events. If new activities are added, the whole model has to be re-trained or the whole system has to be re-constructed. Other methods [3, 4] tried to use a similarity metric so that different events can be clustered into different groups. This approach has more flexibility for new added events. However, due to the uncertain nature of the activity instances, it's difficult to find a suitable feature set so that all samples of an event are clustered closely around a center. In this paper, we show that our proposed CFV-based structure gives systems the flexibility to handle new added activities. We also proposed a Confident-Frame-based algorithm, which essentially derives an individualized model from the general model to give higher accuracies for recognizing events.

The rest of the paper is organized as follows. Section 2 describes our approach for activity representation. Section 3 discusses our methods for training models for new events and unusual events. Section 4 presents our Confident-Frame-based activity recognition approach. Experimental results are shown in Section 5, and Section 6 concludes the paper.

## 2. ACTIVITY REPRESENTATION

Activities can be described by a combination of feature attributes. For example, a set of human activities (*Inactive*, *Active*, *Walking*, *Running* and *Fighting*) (see the definition in [7]) can be differentiated using the combinations of attributes of two features: *CBS*(*Change of Body Size*) and *Speed*. Each feature can have attributes such as “*High*”, “*Medium*”, and “*Low*”. “*Inactive*”, which represents a static person, can be described as “*Low CBS*” and “*Low Speed*”. “*Active*”, which represents a person making movements but without translations, can be described as “*Medium CBS*” and “*Low Speed*”. “*Walking*”, which represents a person making movements and translations, can be described as “*Medium CBS*” and “*Medium Speed*”. “*Running*”, which is similar to walking but with a larger translation, can be described as “*High CBS*” and “*High Speed*”. “*Fighting*”, which has large movements with small translations, can be described as “*High CBS*” and “*Low Speed*”.

It is efficient to represent the activities by feature attributes as shown above, which can describe and differentiate a large number of activities with a relative small number of features and attributes. However, this approach has low robustness. The misclassification of one feature attribute can easily lead to a completely wrong result (e.g., if *Speed* is misclassified from “*Medium*” to “*Low*”, “*Walking*” will be misclassified into “*Active*”). Furthermore, the extent of “*Medium*” or “*Low*” is also difficult to define.

In this paper, we extend the idea of the above Feature-Attribute (FA) description by representing activities by a combination of GMMs, as shown in Fig. 1. Each GMM

represents the distribution  $p(F_i|A_k)$  of a Category Feature Vector (CFV)  $F_i$  of an activity  $A_k$ .  $F_i$  is defined by

$$F_i = [f_{1i}, f_{2i}, f_{3i}, \dots, f_{mi}]^T \quad (1)$$

where  $f_{1i}, f_{2i}, \dots, f_{mi}$  are features related to the *same* category  $i$ . The likelihood function  $p(F_i(t)|A_k)$  of the observed CFV  $F_i(t)$  for video frame  $t$ , given activity  $A_k$  can be described as

$$p(F_i(t) | A_k) = \sum_{j=1}^{N_i,k} \pi_j^{i,k} N(\mu_j^{i,k}, \sigma_j^{i,k}) \quad (2)$$

where  $\pi_j^{i,k}$  is the weight of Gaussian distribution  $N(\mu_j^{i,k}, \sigma_j^{i,k})$  for CFV  $F_i(t)$  and activity  $A_k$ .

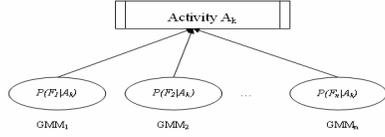


Fig. 1. Activity  $A_k$  is described by a combination of GMMs

Essentially, CFV is the extension of the ‘feature’ in the FA description. In our description, features with high correlations for describing activities are grouped into the same CFV. The feature similarity measure such as K-L distance can be used to cluster and group correlated features into CFVs. The GMM model  $p(F_i(t)|A_k)$  is the extension of the ‘feature attribute’ in the FA description. With the use of the GMM distribution, we will have more robustness in representing and recognizing activities compared with the FA description.

By grouping correlated features into a CFV, the correlations of the features can be captured by the GMM. Also, we can reduce the total number of GMM models, which can facilitate the succeeding classifier which is based on fusing the GMM results. Furthermore, by separating the features into CFVs, it facilitates the handle of new added activities and the training of models for unusual events as will be described in Section 3.

An overall block diagram of our proposed activity recognition approach is shown in Fig. 2. For each video frame, the object features are first extracted and then grouped into different CFVs. A GMM classifier (a Bayesian classifier based on GMM) is used for each CFV. Finally, the proposed CFR-based algorithm combines results from the multiple GMM classifiers and gives the recognition results (as shown in Sec.4).

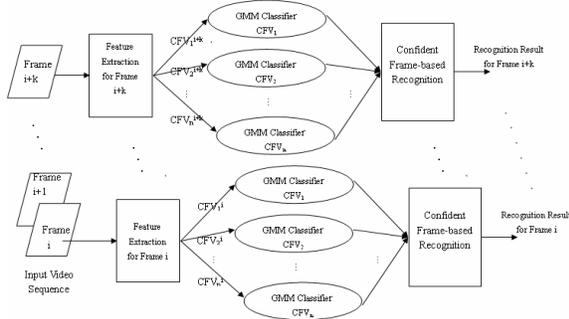


Fig. 2. An overall block diagram for activity recognition.

### 3. TRAIN MODELS FOR NEW & UNUSUAL EVENTS

Due to the uncertain nature of human actions, new or unexpected unusual activities may often need to be added to the system for recognition. In these cases, new models need to be trained for these new activities. This section will show the flexibility of our CFV-based representation for new activities and propose an approach for generating models for unusual events lacking training data.

#### 3.1 Training Models for New Activities

The flexibility of our representation method for new

activities can be demonstrated in the following.

First, with the increasing number of activities, the already defined features may not be enough to differentiate all activities, necessitating the adding of new features. With our CFV-based representation, we only need to define new categories for the added features (i.e., define new CFVs) and train new models for them (i.e., add a new GMM for a new CFV), while keeping other GMMs unchanged. However, for traditional methods [8], the whole set of models has to be re-trained if new features are added.

Second, the models of traditional methods will become increasingly complicated with the addition of new features. However, with the CFV-based description, we only need to increase the number of CFVs while keeping the GMM model for each CFV simple. By using our proposed Confident Frame-based algorithm (see Sec. 4.3) to combine the GMM classifying results of different CFVs, the increasing number of CFVs will not increase the complexity for the recognition process significantly.

#### 3.2 Training Models for Unusual Activities

Since unusual activities rarely occur, we often do not have enough training data to construct the GMM models for these actions. To solve this problem, we observe that people often describe a rare object by combining different parts from familiar objects. For example, people may describe a llama as an animal with the head similar to a camel and the body similar to a donkey. Similarly, with our CFV-based representation of activities, it is possible for us to derive a good initial unusual event model from the CFVs of the known activities. For example, as shown in Fig. 3, we have trained two CFVs  $F_{CBS}$  and  $F_{Speed}$  for recognizing the three usual events: *Active*, *Walking*, and *Running* described in the example in Section 2.  $F_{CBS}$  is the CFV for the category *CBS*, and  $F_{Speed}$  is the CFV for the category *Speed*. Assume *Fighting* is an unusual event we try to recognize but lacking training data. For the *CBS* category, we can reason that the behavior of *Running* is the most similar among all the usual events to that of *Fighting*, therefore the GMM for  $F_{CBS}^{fighting}$  will be adapted from that of  $F_{CBS}^{running}$ . Similarly, for the *Speed* Category, we find that the behavior of *Active* is the most similar to that of *Fighting*, therefore the GMM for  $F_{Speed}^{fighting}$  will be adapted from that of  $F_{Speed}^{active}$ . By this way, we can have a good initial model for *Fighting* even in the case of lacking training data.

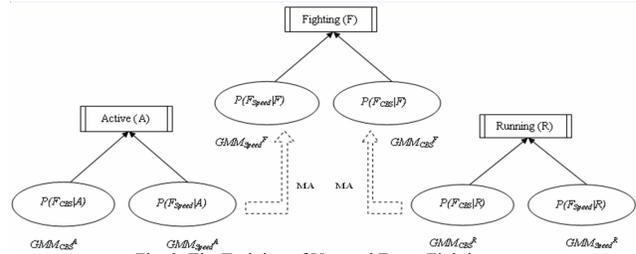


Fig. 3. The Training of Unusual Event *Fighting*.

From above discussions, we summarize our proposed model-training method for unusual activities as follows. For each  $CFV_i$  in category  $i$  of the unusual activity  $A_u$ , find the trained GMM model ( $GMM_i^{A_k}$ ) where the behavior of activity  $A_k$  is most similar to the unusual activity  $A_u$  in this *specific* category. With this, we can obtain an initial CFV-based model for the unusual activity. The initial model can then be further adapted using the limited training data. The MAP-based Adaptation (MA) [5][11] is used to derive the new model  $GMM_i^{A_u}$ .

### 4. CONFIDENT-FRAME-BASED RECOGNITION (CFR)

In Section 3, the activities have been described by a combination of CFV-based GMMs. Therefore, we can construct a GMM classifier  $C_i$  for each CFV  $F_i$  with MAP (Maximum a Posteriori) principle, as shown in Equation (3)

$$\begin{cases} C_i(F_i(t)) = \arg \max_{A_k} p(A_k | F_i(t)) \\ p(A_k | F_i(t)) = \frac{p(F_i(t) | A_k) \cdot p(A_k)}{p(F_i(t))} \end{cases} \quad (3)$$

where  $p(F_i(t)|A_k)$  is the likelihood function for the observed CFV <sub>$i$</sub>  (i.e.  $F_i(t)$ ) at frame  $t$ , given  $A_k$ , calculated by Equation (2).

It is obvious that GMM classifiers for different CFVs are able to differentiate activities with different confidence (e.g.  $C_{CBS}$  is more able to differentiate *Inactive* and *Fighting*, while  $C_{Speed}$  may have more difficulty in doing this), leading to various possible inconsistencies between results from classifiers for different CFVs (due to the misclassification of some classifiers). Thus, it is desirable to fuse the classification results from different classifiers to obtain the final improved result. In this section, we propose a Confident Frame-based Recognition algorithm (CFR) which can recognize the activities more accurately.

#### 4.1 Combining Global Model and Local Model

Due to the uncertain nature of human actions, samples of the same action may become dispersed or clustered into several groups. The ‘global’ model derived from the whole set of training data collected from a large population of individuals with significant variations may not give good results in classifying activities associated with an individual. As shown in Fig. 4, suppose there are two global models:  $A_k$  for activity ‘walking’ and  $A_j$  for activity ‘running’. The cross samples are frame positions in the feature space with each cluster of crosses representing one period of action taken by one person. Due to the non-rigidity of human actions, each individual person’s activity pattern may be ‘far’ from the ‘normal’ patterns of the global model. For example, as shown in Fig. 4, if Person 1 walks (cluster  $W1$ ) faster than normal people while Person 2 walks (cluster  $W2$ ) slower than normal people, then most of the samples in both clusters will be ‘far’ from the center of the ‘global’ model for  $A_k$ . If using the global model to do classification on Cluster  $W1$ , only a few samples (the boldfaced crosses) in  $W1$  can be correctly classified, while the other samples will be mis-classified as  $A_j$ . However, based on the self-correlation of samples within the same period of action, if we use those boldfaced-cross samples to generate a ‘local’ model, it could greatly help the global model.

Therefore, we can outline the ideas of our proposed Confident-Frame-based Recognition algorithm as follows:

1. During an activity  $A_k$ , we use the ‘global’ model to detect frames  $T_k$  which have *high confidence* for recognizing  $A_k$ , instead of trying to match *every* frame directly by the global model. We call  $T_k$  *Confident Frames*, while the rest frames are called *Left Frames* (denoted as  $L_k$ ) as shown in Fig. 5.

2. These *confident frames* will be used to generate a ‘local’ model (or specialized model) for the current period of activity  $A_k$ . The local model and global model will be combined together to classify the Left Frames  $L_k$ .

The detailed description of the proposed CFR algorithm is shown as follows (Section 4.2-4.5).

#### 4.2 Detection of Confident Frames

The detection function of Confident Frame can be described as:

$$t = \begin{cases} T_k & \text{if } \sum_i w_{k,i} \cdot p(A_k | F_i(t)) > th_k \\ L_k & \text{if } \sum_i w_{k,i} \cdot p(A_k | F_i(t)) \leq th_k \end{cases} \quad (4)$$

where  $t$  is the current frame,  $T_k$  is a *Confident Frame* and  $L_k$  is a *Left Frame*.  $P(A_k|F_i(t))$  is calculated by Equation (3).  $w_{k,i}$  is the weight for category  $i$  under action  $A_k$ .  $th_k$  is the threshold for action  $A_k$ .  $w_{k,i}$  and  $th_k$  can be determined by the validation set so that the detection rate is maximized while keeping the false-alarm rate smaller than a threshold.

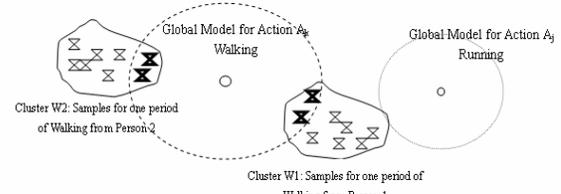


Fig. 4. Global and local models.

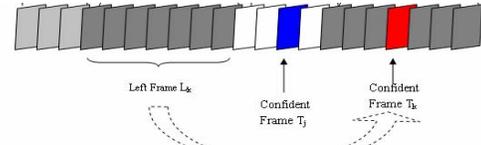


Fig. 5. Confident Frames and Left Frames associated with an activity  $A_k$ .

#### 4.3 Multi-category Dynamic Partial Function

In Qamra and Li’s work [12,13], the Dynamic Partial Function (DPF) was proposed for measuring the dissimilarity of two images ( $X$  and  $Y$ ). In our representation of activities, since features are grouped into CFV according to the categories, we here extend the DPF into a Multi-category DPF, described as:

$$D_w(X, Y) = k_1 \left( \sum_{i \in \bar{\Delta}} w_{i1} \cdot (f_{i1}^X - f_{i1}^Y)^r \right)^{\frac{1}{r}} + \dots + k_n \left( \sum_{i \in \bar{\Delta}} w_{in} \cdot (f_{in}^X - f_{in}^Y)^r \right)^{\frac{1}{r}} \quad (5)$$

where  $F \setminus \bar{\Delta}$  means samples in  $F$  but not in  $\bar{\Delta}$ ,  $k$  are weights for each category  $i$ ,  $\bar{\Delta} = \{ \text{the largest } n \text{ } (f^X - f^Y)^r \text{ s of } [F_1^X - F_1^Y, F_2^X - F_2^Y, \dots, F_n^X - F_n^Y] \}$ , and  $F_i = [f_{i1}, f_{i2}, f_{i3}, \dots, f_{im}]^T$  is the CFV for category  $i$ .

#### 4.4 The CFR process

The CFR recognition process is summarized as follows:

a. For a given video sequence, first detect all the *confident frames* associated with each activity by Equation (4).

b. For each *Left Frame*  $t_L$ , pick the two *most possible* candidate activities for this frame based on Equation (6):

$$\begin{cases} candi_1 = \arg \max_k \sum_i w_{k,i} \cdot p(A_k | F_i(t)) \\ candi_2 = \arg \max_{k, k \neq candi_1} \sum_i w_{k,i} \cdot p(A_k | F_i(t)) \end{cases} \quad (6)$$

c. Select the two confident frames  $T_{candi1}$  and  $T_{candi2}$  corresponding to the two most possible candidate activities which are spatially closest to  $t_L$ . Check the dissimilarity of  $D_w(T_{candi1}, t_L)$  and  $D_w(T_{candi2}, t_L)$  by Equation (5).

d. The candidate that has the smaller  $D_w$  value will be the resulting activity of frame  $t_L$ .

In the above process, we have combined the global model and the local model together for classifying the *left frames*, in order to increase the robustness of using local models. The global model based on the GMMs is first used to select the *two most possible* candidate activities and discard the rest activities, then the local model is used to classify a left frame into one of the two candidate activities.

#### 4.5 Discussion of CFR

Due to the limited space, we only summarize some of the advantages of the CFR method as follows:

a. In case that global model fails to detect during one event period  $P_{A_k}$ , CFR may still be able to detect the event by checking with local models of  $A_k$  outside  $P_{A_k}$  (see Fig. 5).

b. By introducing the dissimilarity checking (DPF) for local model, we take the advantage of using more features which is unsuitable for global models (e.g. location, duration).

c. The CFR algorithm greatly facilitates the fusing process of results from different CFV-based GMMs by only requiring fusing on *Confident Frames*.

## 5. EXPERIMENTAL RESULTS

We perform experiments on PETS'04 database [10] (see Figure 6), and try to recognize five activities: *Inactive(I)*, *Active(A)*, *Walking(W)*, *Running(R)*, and *Fighting(F)*. The total numbers of video frames for each activity are listed in Table 1. In order to exclude the effect of tracking algorithm, we use the ground-truth tracking data (MBB, Minimum Bounding Box [10]). The features derived from the MBBs are classified into two CFVs with  $CFV_{bm} = [f_{avg\_c\_mbb\_size}, f_{avg\_c\_mbb\_width}, f_{avg\_c\_mbb\_height}]$  for category *body movement*, and  $CFV_{bt} = [f_{avg\_speed}, f_{avg\_vector}, f_{mean\_vector}]$  for category *whole body translation* (Note:  $avg\_c\_mbb\_feature$  means the average change of the MBB feature of the object over  $n$  frames and the definition of  $avg\_vector$  and  $mean\_vector$  are the same as [7]). As mentioned in Section 4.5, when using DPF to check with local model similarities, we include 7 more features which are object's MBB center (x,y), MBB width, height, size and width/height ratio, and duration since the object appears. It is a challenging experiment because all features in the experiment are derived from the very simple object's MBB information. We compare five cases as in Table 2.



Fig. 6. PETS'04 Database [10].

Table 1 Number of positive and negative samples for activities

	Positive Samples #	Negative Samples #
<i>Inactive</i>	9077	18642
<i>Active</i>	3397	24322
<i>Walking</i>	14349	13370
<i>Running</i>	490	27229
<i>Fighting</i>	406	27313

Table 2 Five cases in the experiment

I. Use a GMM classifier for $CFV_{bm}$ only. ( $C_{bm}$ in the tables)
II. Use a GMM classifier for $CFV_{bt}$ only. ( $C_{bt}$ in the tables)
III. Use weighted addition of results for the 2 classifiers above. (WA)
IV. Use <i>one</i> GMM classifier for all features in both CFVs (One GMM)
V. Our proposed CFR algorithm. (CFR)

Due to the limited space, we only show results for two settings: (1) 50% Training and 50% Testing, and (2) 75% Training and 25% Testing (For each setting, we perform five independent experiments and average the results). The experiment results are shown in Table 3 (In Table 3, *Miss* is for Miss Detection Rate and *FA* for False Alarm Rate [6]).

The results of  $C_{bt}$  and  $C_{bm}$  show the contributions of each individual CFV. The results of WA and One GMM show the results with and without the CFV structure, respectively. Finally, the results of CFR show that our proposed CFR algorithm which uses the CFV structure and combines the global and local models gives the best recognition performance, besides giving the flexibility and the capability to handle unusual events. For activities such as *Active*, *Running*, and *Fighting* where the GMM classifiers have low detection rates, our CFR algorithm can greatly improve the detection performance by introducing the local model from the *confident frames*.

From Table 3, we can also find that the detection rate for actions such *Running* and *Fighting* are relatively low (although CFR has greatly improved the detection rate). It's because the number of training samples is low (see Table 1). The training

Table 3 Result Comparison (White Columns:50% Training and 50% Testing and Gray Columns: 75% Training and 25% Testing)

		$C_{bt}$	$C_{bm}$	WA	One GMM	CFR	$C_{bt}$	$C_{bm}$	WA	One GMM	CFR
<i>Inactive</i>	Miss	3.5%	4.6%	0.65%	6.6%	0.96%	1.92%	4.18%	1.37%	3.69%	0.49%
	FA	2.8%	3.1%	5.8%	1.4%	1.5%	3.76%	2.20%	4.19%	2.37%	1.88%
<i>Active</i>	Miss	42.2%	52.1%	43.1%	23%	12.2%	40.36%	64.0%	47.0%	23%	9.4%
	FA	2.99%	4.43%	0.47%	2.0%	1.25%	2.83%	3.95%	0.92%	1.95%	0.91%
<i>Walking</i>	Miss	3.99%	7.49%	2.9%	1.2%	2.3%	5.33%	6.12%	1.5%	1.57%	2.01%
	FA	12.0%	19.1%	14.7%	5.7%	5.09%	7.02%	16.3%	11.7%	4.48%	2.26%
<i>Running</i>	Miss	58%	92%	80.1%	50%	25%	50.52%	89.1%	74.1%	44.75%	21.4%
	FA	0.22%	0%	0.04%	0.39%	0.29%	0.44%	0.04%	0.06%	0.29%	0.7%
<i>Fighting</i>	Miss	68.7%	79.2%	71.7%	59.5%	37.3%	56.7%	64.2%	60.7%	47.5%	32.4%
	FA	0.1%	0.13%	0.02%	0.32%	0.22%	0.23%	0.23%	0.12%	0.39%	0.24%

samples are not sufficient to model the whole distribution of these activities. In this case, we use our proposed method in Section 3.2 to deal with the insufficient training data problem, where we adapt both CFVs' GMM models of *Running* from *Walking*, while both CFV GMM models of *Fighting* are adapted from *active*. The recognition results based on our adapted-GMM models are shown in Table 4, which shows the effectiveness of our proposed method in dealing with insufficient training data

Table 4 Our proposed method in dealing with insufficient data (50% training, 50% testing, Gray columns: Directly from Training Data, White columns: Adapted by our proposed method)

		$C_{bt}$ Direct	$C_{bt}$ Adaption	$C_{bm}$ Direct	$C_{bm}$ Adaption	CFR Direct	CFR Adaption
<i>Running</i>	Miss	58%	36.4%	92%	73.1%	25%	15.1%
	FA	0.22%	0.22%	0%	0.27%	0.29%	0.30%
<i>Fighting</i>	Miss	68.7%	62.52%	79.2%	71.12%	37.3%	28.14%
	FA	0.1%	0.12%	0.13%	0.13%	0.22%	0.24%

## 6. CONCLUSION

In this paper, we proposed (1) a flexible framework for representing activities, (2) a method to deal with unusual events which lacks training data, and (3) a CFR algorithm to improve the recognition accuracy. The experiment results demonstrate the effectiveness of our proposed methods.

## ACKNOWLEDGEMENT

This work was supported in part by the following grant: ARO PECASE grant, W911NF-05-1-0491.

## References

- [1] P. Harjo, T. Taipalus, J. Knuutila, J. Wallet and A. Halme, "Needs and Solutions- Home Automation and Service Robots for the Elderly and Disabled," *Int'l Conf. Intelligent Robot, Systems*, pp. 3201-3206, 2005.
- [2] P. Viola, M. Jones and D. Snow, "Detecting Pedestrians Using Patterns of Motion and Appearance," *Int'l Journal of Computer Vision*, vol. 63, pp. 153-161, 2005.
- [3] L. Zelnik-Manor and M. Irani, "Event-based Video Analysis," *Proc. IEEE Conf. CVPR*, vol. 2, pp. 123-130, 2001.
- [4] H. Zhong, J. Shi, and M. Visontai, "Detecting Unusual Activity in Video," *Proc. IEEE Conf. CVPR*, vol. 2, pp. 819-826, 2004.
- [5] D. Zhang, D. Gatica-Perez, S. Bengio and I. McCowan, "Semi-supervised adapted HMMs for unusual event detection," *CVPR*, 2005.
- [6] F. Lv, J. Kang, R. Nevatia, I. Cohen and G. Medioni, "Automatic Tracking and Labeling of Human Activities in a Video Sequence," *Int'l Workshop on Performance Evaluation of Tracking and Surveillance*, 2004.
- [7] P.C. Ribeiro and J. Santos-Victor, "Human Activity Recognition from Video: modeling, feature selection and classification architecture," *Int'l Workshop on Human Activity Recognition and Modeling*, pp. 61-70, 2005.
- [8] T.V. Duong, H.H. Bui, D.Q. Phung and S. Venkatesh, "Activity recognition and abnormality detection with the switching hidden semi-Markov model," *IEEE Conf. CVPR*, vol. 1, pp. 838-845, 2005.
- [9] D. Ayers and M. Shah, "Monitoring Human Behavior from Video Taken in an Office Environment," *Image and Vision Computing*, vol. 19, pp. 833-846, 2001.
- [10] CAVIAR Project, <http://homepages.inf.ac.uk/bf/CAVIAR/>
- [11] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital Signal Processing*, vol. 10, pp. 19-41, 2000.
- [12] A. Qamra, Y. Meng and E.Y. Chang, "Enhanced perceptual distance functions and indexing for image replica recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 2, no. 3, pp. 379-391, 2005.
- [13] B. Li, E. Chang, C.T. Wu, "DPF-a perceptual distance function for image retrieval," *Int'l Conf. Image Processing*, vol. 2, pp. 597-600, 2002.