

Automatic Behavior Model Selection by Iterative Learning and Abnormality recognition

Heping Li, Jie Liu , and Shuwu Zhang

High-Tech Innovation Center,
Institute of Automation, Chinese Academy of Sciences, Beijing, China
{hpli, jliu, swzhang}@hitic.ia.ac.cn

Abstract— Automatic behavior recognition is one important task of community security and surveillance system. In this paper, a novel method is proposed for automatic selection of behavior models by iterative learning and abnormality recognition. The method is mainly composed of the following two steps: (1) The models of normal behaviors are automatically selected and trained by combining Dynamic Time Warping based spectral clustering and iterative learning; (2) Maximum A Posteriori adaptation technique is used to estimate the parameters of abnormal behavior models from those of normal behavior models. Compared with the related works in the literature, our method has three advantages: (1) automatic selection of the class number of normal behaviors from large unlabeled video data according to the process of iterative learning, (2) semi-supervised learning of abnormal behavior models, and (3) avoidance of the running risk of over-fitting during learning the Hidden Markov Models of behaviors in case of sparse data. Experiments demonstrate the effectiveness of our proposed method.

Keywords—human motion analysis; behavior modeling; abnormality recognition; Hidden Markov Model

I. INTRODUCTION

Behavior modeling and recognition is one key technology of community security and surveillance system, and it is an active research topic in information analysis and computer vision [1].

Most of the existing works on modeling behaviors require manually labeling like those in [2-9]. For example, Gong and Xiang [4] learned a Dynamically Multi-Linked Hidden Markov Model (DML-HMM). Li and Greenspan [5] built a multi-scale model from time-varying contours. However, manual labeling of behavior patterns is laborious, impractical and error prone [10]. To this problem, Some behavior modeling methods based on semi-supervised/unsupervised learning [10-14] have been proposed. For instance, Xiang and Gong [10] used unsupervised model selection and feature selection to discover natural grouping of behavior patterns. Zhong et al. [11] proposed an unsupervised method to detect anomaly events. This method clustered observed patterns and labeled small clusters as abnormal without the need for modeling behaviors explicitly. Zelnik-Manor and Irani [12] used the multiple temporal scales technology and the normalized-cut approach to automatically cluster the data and then build the statistical behavior model. Zhang et al. [13] gave a semi-supervised method with a two-stage training process: (1) using labeled normal samples to learn one normal behavior model and (2)

obtaining abnormal behavior models from the normal behavior model by combining Maximum A Posteriori (MAP) adaptation and unsupervised method.

During the course of modeling behaviors with semi-supervised/unsupervised method, Hidden Markov Model (HMM) based distance is often used to build affinity matrix for spectral clustering because HMM provides a suitable tool for solving the time-warping problem of behaviors. If every sequence sample is long enough, it is feasible for HMM based distance to calculate affinity matrix, otherwise the running risk of over-fitting is inevitable, which will in turn result in false labeling of samples during spectral clustering. In addition, after obtaining affinity matrix, how to faithfully determine the clustering number is another difficult problem for spectral clustering.

To overcome these difficulties and build an effective model for abnormality recognition, we propose a novel method for automatic selection of behavior models by iterative learning in this paper. Our method has three advantages: (1) automatic selection of the number of normal behaviors from large unlabeled video data, (2) semi-supervised learning of abnormal behavior models, and (3) avoidance of the running risk of over-fitting during learning the HMMs of behaviors in case of sparse data.

The remainder of the paper is organized as follows: Video segmentation and representation are outlined in section II. Section III is a detailed description of normal behavior modeling. Section IV discusses abnormal behavior modeling and recognition. Experimental results are reported in section V, and followed by some conclusions in section VI.

II. VIDEO SEGMENTATION AND REPRESENTATION

We will first segment a continuous video sequence V into N short segments: $V = \{v_1, v_2, \dots, v_i, \dots, v_N\}$. Broadly speaking, there are three types of methods for the video segmentation in the literature, namely, non-activity gaps, fixed time duration with overlapping window, and the points of abrupt change. Here, we adopt the second one with a temporal window of T_{win} frames and a step length of T_{step} frames. For each short segment, the technology of spatiotemporal filtering is adopted to extract behavior features with the following steps:

Step 1: Do spatiotemporal filtering for each frame in the video with the following equation [11]:

$$I_t(x, y, t) = \|I(x, y, t) * G_t * G_{x,y}\|_2,$$

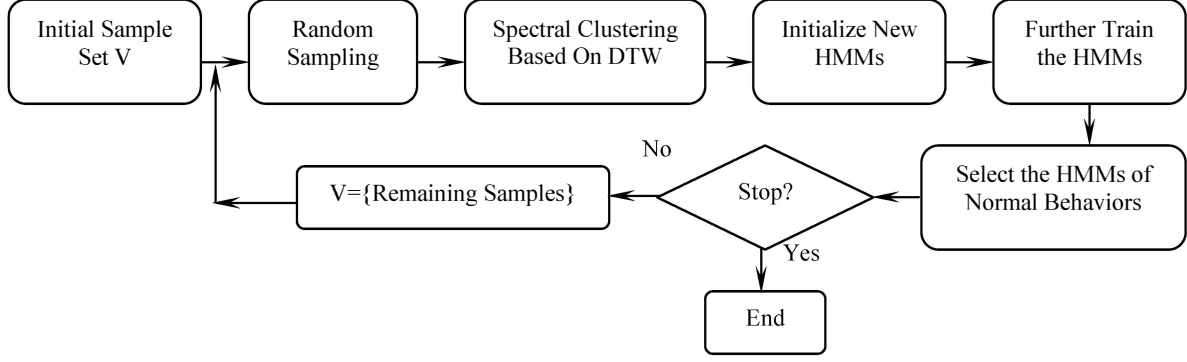


Figure 1. The flowchart of the process of iterative learning.

$$G_t = te^{-(t/\sigma_x)^2}, G_{x,y} = e^{-[(x/\sigma_x)^2 + (y/\sigma_y)^2]},$$

where t is the frame index, (x, y) is pixel point coordinate, $I(x, y, t)$ is the grey value at (x, y, t) ;

Step 2: Binarize the filtered image by thresholding;

Step 3: Equidistantly divide the bounded rectangle of the foreground into $J = U \times F$ non-overlapping sub-blocks and then calculate the normalized value of each sub-block of frame t as follows: $d_{it} = u_i / \max(u_j)$, $(1 \leq i \leq J, 1 \leq j \leq J)$, where u_i is the number of the foreground pixels in the i^{th} sub-block;

Step 4: Represent the feature vector of frame t as:

$$D_t = [d_{t1}, d_{t2}, \dots, d_{tj}, \dots, d_{tJ}].$$

Then, each short segment can be represented as:

$$v_i = \{D_1, D_2, \dots, D_{T_{\text{seg}}}\}.$$

III. NORMAL BEHAVIOR MODELING

Given the data set $V = \{v_1, v_2, \dots, v_i, \dots, v_N\}$, the next step is to automatically discover the natural groupings of these data and build normal behavior models.

We build the affinity matrix by the normalized distance based on Dynamic Time Warping (DTW). DTW is a process of dynamic programming and has a high computational complexity. To alleviate this computational burden, we only select a small sample set randomly to do spectral clustering based on DTW, then decide the number of normal behaviors and build the corresponding models by iterative learning. The process of iterative learning, our most important contribution in this paper, is shown in Figure 1.

In this section, we will first introduce the spectral clustering based on DTW and HMM structure of behavior, then present the detailed descriptions about iterative learning, as well as HMM topology of normal behaviors.

A. Spectral Clustering Based On DTW

Dynamic Time Warping can align two signals and get the warping path. The path can be found very efficiently using dynamic programming. Given two time series v_Q and v_P , the best warping path will minimize the cumulative distance:

$$\gamma(i, j) = d(i, j) + \min\{\gamma(i-1, j), \gamma(i, j-1), \gamma(i-1, j-1)\}, \quad (1)$$

where $1 \leq i \leq m, 1 \leq j \leq n, \gamma(1, 1) = d(1, 1), \gamma(i, 1) = d(i, 1) + \gamma(i-1, 1), \gamma(1, j) = d(1, j) + \gamma(1, j-1), m$ is the length of time series v_Q

and n is the length of time series v_P , $d(i, j)$ is the similarity distance between the feature vector v_Q of frame i and the feature vector v_P of frame j as: $d(i, j) = 1 - v_Q \cdot v_P / (\|v_Q\| * \|v_P\|)$.

Considering that the best warping paths may have different lengths, we use the normalized distance to measure the similarity of two motion signatures:

$$s = \exp[-\gamma(m, n) / G\sigma], \quad (2)$$

where $\gamma(m, n)$ is the minimum cumulative distance, G ($\max(m, n) \leq G \leq m+n-1$) is the length of the warping path, and σ is a constant factor.

For H short segments selected randomly from the data set V , an $H \times H$ affinity matrix $S = [s_{mn}]_{H \times H}$ with $1 \leq m, n \leq H$ is obtained according to the equation (2). Let the eigenvalue of S be e_i ($i \in [1, H]$) with $e_1 \geq e_2 \geq \dots \geq e_H$, and then the contribution rate of square deviation is by: $\mu = \sum_{i=1}^L e_i / \sum_{j=1}^H e_j$ ($L \leq H$).

When $\mu \geq T_\mu$ ($T_\mu = 0.8$), L is selected as the number of clusters. We then use the spectral clustering method, i.e. the normalized-cut approach [15], to cluster H short segments into L classes. From these L classes, the C' classes whose sample numbers are bigger than $T_{\text{sam}} = H/10$ are selected as the main classes of behaviors in this randomly selected small set. If no class satisfies this condition, the class with maximum sample number will be selected, and at this time, $C' = 1$. Denote the sample sets of main classes as W_i ($i = 1, 2, \dots, C'$) and put back the remaining samples from the small set into W^a .

B. HMM Structure of Behavior

Each sample set W_i of the main classes is used to initialize one HMM with e hidden nodes. The output probability density function of each hidden node is a Gaussian Mixture Model (GMM) as: $p(D_t | \theta) = \sum_{k=1}^K \alpha_k p_k(D_t | \mu_k, \Sigma_k)$, where $\theta = \{\alpha_k, \mu_k, \Sigma_k, k = 1, 2, \dots, K\}$ represents the parameter of GMM, including weight α_k , mean value μ_k and covariance matrix Σ_k of every mixture component, and $\sum_{k=1}^K \alpha_k = 1$. We

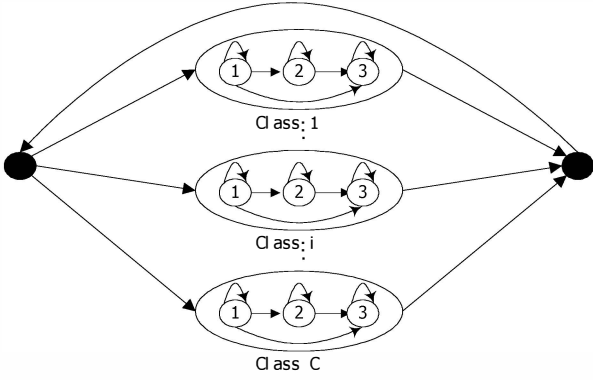


Figure 2. HMM topology of normal behaviors

adopt Bayesian Information Criterion (BIC) [16] to decide the number of mixture components of GMM.

C. Iterative Learning

Unlike the approach [10], we use the spectral clustering based on DTW to select the initial sample sets to initialize the HMMs. This technology can avoid the running risk of over-fitting. In order to automatically determine the true number of normal behaviors and build the corresponding models, the method of iterative learning is adopted with the following steps:

Step 1: Random sampling: Randomly select H samples from the big sample set V to form a new set V^1 , and let A be the set of the remaining samples.

Step 2: Spectral clustering based on DTW: Obtain C' classes of sample sets W_i ($i=1,2,3,\dots,C'$) and the remaining sample set W^a of V^1 by the algorithm in section 3.1.

Step 3: Initialize new HMMs: Use W_i to initialize the corresponding HMM with λ_i^n . And W^a is used to initialize another HMM with λ^a .

Step 4: Train the HMMs again:

(1). Decide the class type of every sample in set A according to the following rule: For a sample v , if $j = \max_i \{P(v | \lambda_i^n) | P(v | \lambda_i^n) > P(v | \lambda^a)\}$, then v belongs to the class with parameters λ_j^n . Assume there are n_i samples belonging to the class with parameters λ_i^n ($i=1,2,\dots,C'$) and n^a samples belonging to the class with parameters λ^a ;

(2). Arrange the samples in set A according to their probability of being the class with parameters λ_i^n (or λ^a) from large to small, then take the first fifth out of A , and put them into the corresponding set W_i (or W^a);

(3). Use all the samples in set W_i (or W^a) to further train the corresponding HMM;

(4). Repeat (1)~(3) until no class whose sample number is bigger than T' exists in set A , then go to Step 5.

Step 5. Select the HMMs of normal behaviors: If the number of samples in set W_i is bigger than the preset threshold num_th , the class with HMM's parameter- λ_i^n is considered as a normal behavior.

Step 6. Stopping rule of iterative learning: If no class is considered as normal behavior in step 5, then stop learning, otherwise go to Step 7.

Step 7. Set $V = \{A\} + \{W^a\}$ and go to Step 1.

During the course of iterative learning, it is reasonably considered that a behavior is normal if it occurs with a high frequency in a large video data set, and otherwise, it will be considered as an unusual behavior. So the samples of normal behaviors have more chance to be randomly selected into set V^1 than those of abnormal behaviors in Step 1. From Step 1 to 4, we get C' classes of behaviors that represent main content of the small sample set V^1 . But it is not always true that the C' classes are all the normal behaviors in a large sample set V as V^1 is only a small fraction of V . It is possible that C' classes are only a subset of all the normal behaviors or the C' classes include several classes of abnormal behaviors. So in step 5, we use the threshold num_th to further set the bar for classes belonging to normal behaviors and then repeat Step 1 to 7 until finding all the normal behaviors.

D. HMM Topology of Normal Behaviors

Having obtained HMMs of the normal behaviors, we design an HMM topology of normal behaviors as shown in Figure 2. It is a topology with $C+2$ (C is the number of all the normal behaviors) nodes where each state is a sub-HMM. The two black nodes are virtual start point and end point, and the transition probability between them is 1. The transition probability between any other two nodes is $1/C$.

IV. ABNORMAL BEHAVIOR MODELING AND RECOGNITION

In this section, based on the established normal behavior models, MAP adaptation is used to build abnormal behavior model from a small labeled set of abnormal behavior data.

A. MAP Adaptation

MAP adaptation is suitable for learning the parameters of models in the case of sparse data and has been widely used in speaker and face verification [17]. During the course of learning the parameters of GMM-based HMM in [13][18], the state-transition probabilities are kept fixed while mean, variance and mixture weights are adapted as follows (More details about MAP adaptation can be found in [13][17][18]):

(1) According to the existing parameters, new statistical values are computed:

$$P(i | D_t) = \alpha_i p_i(D_t | \mu_i, \Sigma_i) / \sum_{k=1}^K \alpha_k p_k(D_t | \mu_k, \Sigma_k), \quad (3)$$

$$\alpha_i^{new} = \sum_{t=1}^T P(i | D_t) / T, \quad (4)$$

$$\mu_i^{new} = \sum_{t=1}^T D_t P(i | D_t) / \sum_{t=1}^T P(i | D_t), \quad (5)$$

$$\Sigma_i^{new} = \sum_{t=1}^T P(i | D_t) (D_t - \mu_i^{new})(D_t - \mu_i^{new})^T / \sum_{t=1}^T P(i | D_t). \quad (6)$$

(2) New parameters are estimated by:

$$\hat{\alpha}_i = \rho \cdot \alpha_i^{new} + (1 - \rho) \cdot \alpha_i^{old}, \quad (7)$$

$$\hat{\mu}_i = \rho \cdot \mu_i^{new} + (1 - \rho) \cdot \mu_i^{old}, \quad (8)$$

$$\hat{\Sigma}_i = \rho \cdot \Sigma_i^{new} + (1 - \rho) \cdot [\Sigma_i^{old} + (\hat{\mu}_i - \mu_i^{old})(\hat{\mu}_i - \mu_i^{old})^T], \quad (9)$$

where ρ ($0 \leq \rho \leq 1$) is the scale factor.

B. Abnormal Behavior Modeling

The samples of abnormal behaviors usually constitute a small part of the whole video sequence V . We use the MAP adaptation technology instead of EM algorithm to estimate the parameters of abnormal behavior models with the following two steps: (1) For a given sample of abnormal behavior, we select one HMM with maximum estimation probability from the C HMMs of normal behaviors; (2) The parameters λ_i'' ($1 \leq i \leq C$) of the selected HMMs are regarded as prior knowledge, and the HMMs of abnormal behaviors are adapted from those of normal behaviors by equations (3)~(9). Denote the parameters of abnormal behavior models as λ_j' ($1 \leq j \leq C'_a$), where C'_a is the number of all the abnormal behaviors.

Having built the models of all abnormal behaviors, an HMM topology of abnormal behaviors is built according to the method in section 3.4. Different from the topology as Figure 2, this topology has $C'_a + 2$ nodes and the transition probability between any two nodes except the two virtual nodes is $1/C'_a$.

C. Abnormality Recognition

Denote the parameter set of the topology for normal behaviors as λ_M and that of the abnormal behaviors as λ_{AM} . For the current video sequence O , if $P(O | \lambda_M) > P(O | \lambda_{AM})$, O belongs to normal behavior, otherwise it belongs to abnormal behavior. Furthermore, if O is a normal behavior and $z = \arg \max_k \{P(O | \lambda_k'')\}$ ($1 \leq k \leq C$), then O belongs to the z^{th} class of normal behaviors, and if O is a abnormal behavior and $q = \arg \max_l \{P(O | \lambda_l')\}$ ($1 \leq l \leq C'_a$), then O belongs to the q^{th} class of abnormal behaviors.

V. EXPERIMENTS

Extensive experiments are carried out to verify the effectiveness of our proposed method. In our experiments, $U=9$ and $F=5$ are used for dividing the bounded rectangle of foreground into 45 sub-blocks. And then the Principal Component Analysis (PCA) is used to reduce the 45-dimensional features to the 8-dimensional ones. We adopt the overlapping window with a temporal window of size $T_{win}=30$ frames and a step length of $T_{step}=10$ frames to segment the long video. The experiments are detailed as below.

A. Data Acquisition

In this paper, we take some video sequences of 24912 frames from Schult's dataset [19] including five types of behaviors: "clap", "wave_two_hands", "walk", "box" and "run" and some video sequences of 13981 frames from Li's

dataset [20] including four types of behaviors: "standup", "kick", "look_around" and "wave_right_hand".

In the two datasets, each video sequence only includes one type of behavior, and so in our experiments, we synthesize long video sequences using the above nine types of behaviors by the following steps: (1) Segment the video of every type of behavior with non-overlapping. For the normal behavior sequence, an integer from 100 to 400 is randomly selected as the length of the segmented clip, and for the abnormal behavior sequence, an integer from 30 to 60 is randomly selected. (2) Array all the clips randomly to form a long video sequence V . According to the above synthesis step, we get the training/testing sample sequences. In our experiments, we adopt 8 groups of training sample sequences, where the normal behaviors include "clap" (5500 frames), "standup" (4500 frames), "wave_two_hands" (3500 frames), "walk" (3000 frames) and "box" (2500 frames), and the abnormal behaviors include "kick", "look_around", "run" and "wave_right_hand". The number of each abnormal behavior in the 8 groups is 60, 150, 350, 500, 700, 1000, 1300, 1500 respectively. The testing sequences include normal behaviors: "clap" (2984 frames), "standup" (2580 frames), "wave_two_hands" (1543 frames), "walk" (1584 frames) and "box" (2081 frames), and each type of abnormal behavior has 500 frames.

B. Spectral Clustering and Parameters Selection

(1) The number of normal behaviors

In this paper, the parameters related with spectral clustering are preset as: $T'=50$, $num_th=200$ and $H=150$. The two parameters T' and num_th can be easily adjusted according to demands. For example, if we think that the behaviors whose sample numbers are smaller than 50 belong to abnormal behaviors, we can set $num_th=50$ and $T'=20$. The magnitude of H is chosen as $1/30 \sim 1/10$ over the total sample number of the whole data set because if H is too small, the reliability of the initialized HMMs' parameters will be decreased, and if it is too big, the computational complexity will be too high.

Figure 3(a)~(f) show the clustering results when only one class with the maximum sample number ($C'=1$) is selected as the main class during the course of iterative learning. Figure 3(a)~(e) are the first five clustering results, where one normal behavior model is built from every clustering result. In order, they are "standup", "clap", "walk", "wave two hands" and "box". Figure 3(f) is the sixth clustering result, and for this clustering, no class of behavior satisfies the condition of thresholding in step 5 in the iterative learning. Figure 3(g)~(i) are the clustering results when C' is determined by the method in section 3.1. In this experiment, we get the accurate models of normal behaviors by repeating Steps 1~7 in section 3.3 with three times. C' is respectively assigned as 3, 3, 4 according to spectral clustering based on DTW. For the first

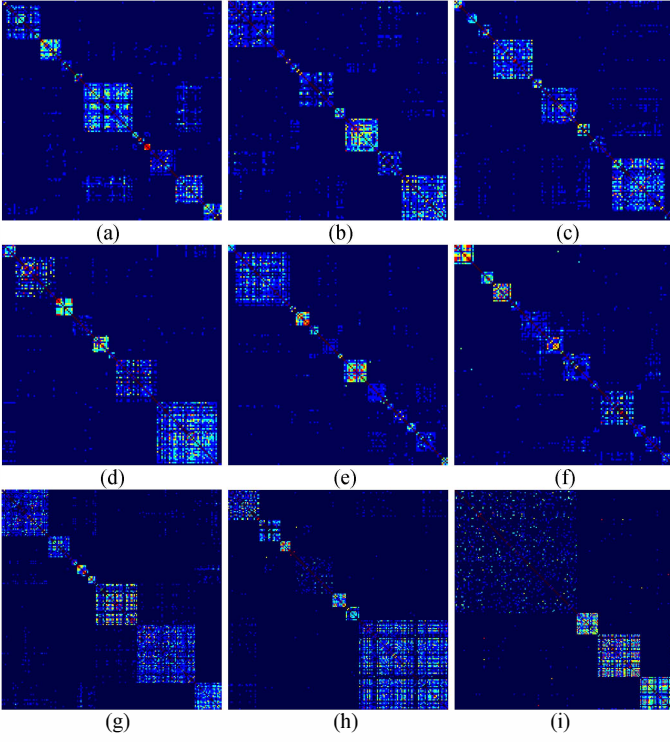


Figure 3. Spectral clustering results

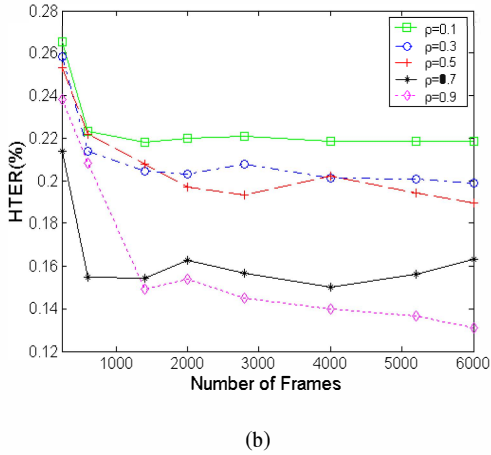
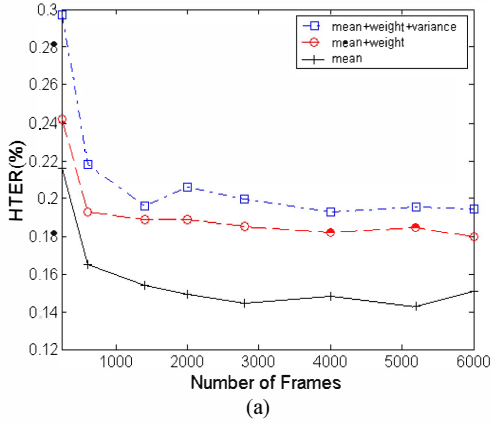


Figure 4. Selection of adaptation parameters.

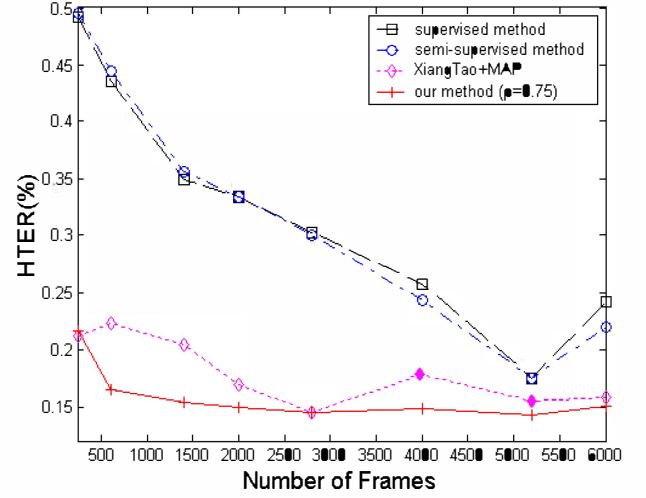


Figure 5. Comparison on abnormality recognition results.

clustering, we get three normal behaviors: “clap”, “standup” and “walk”. There are only two classes of behaviors satisfying the condition of thresholding in step 5 in the second clustering, they are “wave two hands” and “box”. But none of the four classes in the third clustering satisfies the condition of thresholding in step 5. Since the latter is faster, we use the results from the latter in the following experiments.

(2) Adaptation parameters

Figure 4 shows the results of abnormality detection with different parameter combinations and different ρ values. The horizontal axis is the number of frames used in the training set of abnormal behaviors. The vertical axis is the Average Half-Total Error Rate (HTER), where $HTER = (FAR + FRR) / 2$ [13], FAR is false acceptance rate and FRR is false rejection rate.

Figure 4(a) gives three HTER curves when $\rho = 0.75$: (1) “mean+weight+variance” curve, where we adapt all the parameters including mean, weight and variance; (2) “mean+weight” curve, where we only adapt the mean and the weight; (3) “mean” curve, where we only adapt the mean. From this figure, we find when the mean is only adapted, the result of abnormality detection is best. In case of sparse data, it is possible to get worse if a large number of parameters are adapted. Figure 4(b) shows five different HTER curves when we only adapt mean value and vary ρ for 0.1, 0.3, 0.5, 0.7, 0.9. We see that when $0.7 \leq \rho \leq 0.9$, the result of abnormality detection is better. It shows that abnormal behavior model adapted from normal behavior model depends more on the samples of abnormal behaviors even if they are sparse. Based on these results, we only adapt mean value and set $\rho = 0.75$ in the following experiments.

C. Comparison and Analysis

Recognition of normal behaviors---We first compare the performance of our method for normal behavior recognition with the following methods: (1) Supervised method, of which all the HMMs are directly trained by EM with labeled samples; (2) XiangTao’s method [10].

TABLE I. COMPARISON OF RECOGNITION RESULTS ON NORMAL BEHAVIORS

	Supervised method		Xiangtao's method		Our method	
	CRR(%)	FA _c R(%)	CRR(%)	FA _c R(%)	CRR(%)	FA _c R(%)
clapping	73.80	0.80	26.60	14.45	73.80	0.85
stand_up	93.80	5.28	56.70	16.47	90.30	1.82
wave_two_hand	95.40	5.04	71.20	8.03	95.30	5.04
walk	98.30	10.12	49.70	20.92	86.00	3.43
box	97.60	5.87	49.90	9.24	82.10	2.99
Learning time	---		1955 minutes		379 minutes	

Table I shows the comparison of recognition results on normal behaviors. CRR represents Correct Recognition Rate and FA_cR is False Acceptance Rate which is different from FAR (False Alarm Rate). From the results, we can clearly see that our method is better than XiangTao's method, and although the CRR of our method is lower than that of supervised method, the FA_cR of supervised method is higher than that of our method. Moreover, our method is faster.

Abnormality recognition---We compared our semi-supervised method with the following baseline methods: (1) Supervised method, where all the HMMs of normal/abnormal behaviors are directly trained by EM with labeled samples; (2) Semi-supervised method, where all the HMMs of normal behaviors are built by our method and all the HMMs of abnormal behaviors are directly trained by EM with labeled samples; (3) "XiangTao+MAP" method, where all the HMMs of normal behaviors are built by XiangTao's method [10] and all the HMMs of abnormal behaviors are obtained by the MAP adaptation technology introduced in this paper. From Figure 5, we can see that approaches (1) and (2) have bad performance when the number of frames of abnormal behaviors is small. And when the number becomes bigger, the performance becomes better. The performance of approach (3) is better than the above two approaches. And our method has the best performance and stability.

VI. CONCLUSIONS

The proposed method combines DTW and HMM to avoid the running risk of over-fitting during learning the HMMs of behaviors and use the iterative learning technology to automatically select the number of normal behaviors from large unlabeled video data. A large number of experiments including comparisons with related works in the literature demonstrate the effectiveness of our proposed method.

ACKNOWLEDGMENT

This work has been supported by the National Key Technology R&D Program of China under Grant No. 2009BAH48B02, 2009BAH43B04, 2011BAH16B01 and 2011BAH16B02.

REFERENCES

- [1] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognition*, vol. 36, pp.585-601, 2003.
- [2] I. Haritaoglu, D. Harwood, and L. Davis, "W4: real-time surveillance of people and their activities," *IEEE Trans Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 809-830, 2000.
- [3] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257 - 267, March 2001.
- [4] S. Gong and T. Xiang, "Recognition of group activities using dynamic probabilistic networks," In *IEEE Int'l Conf. On Computer Vision*, October 2003, pp. 742-749.
- [5] H. Li and M. Greenspan, "Multi-scale gesture recognition from time-varying contours," In *IEEE Int'l Conf. On Computer Vision*, , October 2005, pp. 236-243.
- [6] I. Laptev and T. Linderberg, "Space-time interest points," In *IEEE Int'l Conf. On Computer IEEE Conf. on Computer Vision and Pattern Recognition, Vision*, October 2003, pp.432-439.
- [7] J. Niebles and F. F. Li, "A hierarchical model of shape and appearance for human action classification," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2007
- [8] K. Mikolajczyk and H. Uemura, "Action recognition with motion appearance vocabulary forest," In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2008
- [9] J. G. Liu, J. B. Luo, and M. Shas, " Recognizing realistic actions from video "in the wild"," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2009.
- [10] T. Xiang and S. G. Gong, "Video behaviour profiling and abnormality detection without manual labeling," In *IEEE Int'l Conf. On Computer Vision*, October 2005, pp. 1238-1245.
- [11] H. Zhong, J. Shi and M. Visontai, "Detecting unusual activity in video," In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2004, pp. 819-826.
- [12] L. Zelnik-Manor and M. Irani, "Event-based analysis of video," In *IEEE Conf. on Computer Vision and Pattern Recognition*, December 2001, pp. 123-130.
- [13] D. Zhang, D. Gatica-Perez, S. Bengio and I. McCowan, "Semi-supervised adapted HMMs for unusual event detection," In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 611-618.
- [14] F. Nater, H. Grabner, and L. V. Gool, " Exploiting single hierarchies for unsupervised human behavior analysis," In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2010
- [15] Y. Weiss, "Segmentation using eigenvectors: a unifying view," In *IEEE Int'l Conf. On Computer Vision*, September 1999, pp. 975 -982
- [16] T. Xiang and S. G. Gong, "Visual learning given sparse data of unknown complexity," In *IEEE Int'l Conf. On Computer Vision*, October 2005, pp. 701-708
- [17] D. A. Reynolds, T. F. Quatieri, and R. B. Dumn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, pp. 19-41, 2000.
- [18] D. Zhang, D. Gatica-Perez, and S. Bengio, "Semi-supervised meeting event recognition with adapted hmms," In *IEEE International Conference on Multimedia Expro, ICME*, 2005.
- [19] C. Schudt, I. Laptev and B. Caputo, "Recognizing human actions: a local SVM approach," *International Conference on Pattern Recognition*, 2004, pp. 32-36
- [20] H. P. Li, Z. Y. Hu, Y. H. Wu, and F. C. Wu, "Behavior modeling and recognition based on space-time image features," *International Conference on Pattern Recognition*, vol. 1, 2006, pp. 243-246.