

Abnormal Behavior Recognition Using Self-Adaptive Hidden Markov Models

Jun Yin and Yan Meng

Department of Electrical and Computer Engineering,
Stevens Institute of Technology, Hoboken, NJ 07030, USA
{jyin, yan.meng}@stevens.edu

Abstract. A self-adaptive Hidden Markov Model (SA-HMM) based framework is proposed for behavior recognition in this paper. In this model, if an unknown sequence cannot be classified into any trained HMMs, a new HMM will be generated and trained, where online training is applied on SA-HMMs to dynamically generate the high-level description of behaviors. The SA-HMMs based framework consists of training and classification stages. During the training stage, the state transition and output probabilities of HMMs can be optimized through the Gaussian Mixture Models (GMMs) so the generated symbols can match the observed image features within a specific behavior class. On the classification stage, the probability with which a particular HMM generates the test symbol sequence will be calculated, which is proportional to the likelihood.

1 Introduction

Recognizing human behaviors in a video stream is critical in many applications, such as video surveillance, video indexing, video annotation, and video summarization. Behavior recognition is difficult since the mapping between the video signal data and event concepts is not always one-to-one. Among various behavior recognition applications, the automatic abnormal behavior/event detection has recently attracted attention in computer vision and multimodal processing under different names, such as abnormal, unusual, or rare events [1], [2], [3]. This is a challenging problem since events of interest occur much less frequently than normal behaviors and occur unpredictably, such as alarm generation in surveillance systems, and extractive summarization of raw video events.

Due to the difficulty for supervised learning, more methods have been proposed recently for unsupervised learning of abnormal behavior models [6], [4], [3], [5]. Some approaches [4], [6] conduct clustering on observed patterns and label those forming small clusters as being abnormal. Another approach proposed in [3] defines the abnormal behaviors as those patterns which cannot be fit into any normal pattern from a database of spatial-temporal patches using only normal behavior. The approach proposed in [6] cannot be applied to online abnormal behavior detection since it cannot handle previously unseen behavior patterns. Most recently, Xiang and Gong [5] proposed an online video behavior

profiling framework for anomaly detection, where a Dynamic Bayesian Network (DBN) is proposed to model each behavior pattern, and a runtime accumulative anomaly measure is introduced to detect abnormal behaviors based on an online Likelihood Ratio Test (LRT) method.

Although it is unrealistic to obtain a large training data set for abnormal behaviors, it is conversely possible to do so for normal behaviors, allowing the creation of a well-estimated model of normal behaviors. In this paper, in order to overcome the scarcity of training material for abnormal behaviors, a self-adaptive Hidden Markov Models (SA-HMMs) based framework is proposed. This is an on-line learning method which can dynamically generate a number of abnormal behavior models in an unsupervised manner. This SA-HMMs based method is able to learn from current data and generates new models, which differs from previous work on abnormality detection through a number of training data.

On the other hand, in the traditional HMMs, only key features are generally of interest. However, for behavior recognition, focusing on key postures are not enough due to a large number of transition postures in human motion. Since the emission distribution of the HMMs is difficult to be evaluated accurately, a single model solution may be insufficient. Therefore, to reduce the influence of transition postures, Gaussian Mixture Models (GMM) are developed in this paper to represent the emission distribution, which allows the model to be more effective at capturing the variability in behavior patterns.

Note that our proposed SA-HMMs based framework is a fully unsupervised learning method, where manual data labeling can be avoided in both feature extraction and classification of behavior patterns. Manual labeling of behavior patterns is tedious and sometimes even infeasible given a large amount of the surveillance video data to be processed. It is worth to mention that the proposed SA-HMMs based framework is a general one which can be applied to any type of scenarios. In particular, the proposed approach is able to be self-adaptive, and the models will become much stronger with the consecutive behavior patterns input.

2 The SA-HMMs Approach

A video sequence V is ideally considered as a kind of behavior including N temporally ordered patterns $V = \{v_1, \dots, v_n, \dots, v_N\}$. For example, the behavior of “going to bed” contains several elementary patterns such as walking, sitting and lying. Moreover, each pattern v_n consisting of X_n image frames is represented as $v_n = [I_{n1}, \dots, I_{nx}, \dots, I_{nXn}]$, where I_{nx} is the x th image frame. Before dealing with the video, we need to effectively process all of frames in order to obtain useful information.

Our system aims at detecting abnormal behaviors, which consists of two major steps. The first step matches the input postures with the templates trained for normal behaviors, and the second step identifies sequences of the discrete postures through a trained HMM and decides whether the behaviors are abnormal. The block diagram of the proposed system is shown in Fig. 1. When an unknown

sequence arrives, firstly, the sequence is processed through the templates matching. Since our approach is based on the silhouette of the postures, we employ the Hausdorff distance to measure the similarity between the test frame and templates. The smaller the distance, the higher the similarity is. Then, we will work with HMMs, where the ‘observations’ of HMMs are drawn from a Gaussian mixture model. The SA-HMMs system can recognize the similar behavior pattern in the recognition process, which includes learning phase and recognizing phase. In the learning phase, the similarity between a new sequence and existent models will be calculated through HMMs. Basically, the decided pattern depends on the threshold for all HMMs. In other words, if the new likelihood dramatically exceeds the threshold, the new sequence will be clustered into a new behavior. In this sense, a new HMM will be generated. The major issue is how to identify appropriate behavior patterns that can enable both behavior recognition and generation. One single sample is obviously insufficient to provide a good estimate of the model parameters. To overcome the lack of training samples, we propose an online learning model, where every test sequence can be considered as a training sample, and the model will be updated after the input sequence.

In this paper, a state space approach is used to define each static posture as a state, where these states are associated with certain probabilities. Each behavior sequence can be mapped into a sequence of states. According to the similarity of the silhouette shapes, the human behavior sequence can be classified into several groups of similar postures. Each group is treated as a model state, and a human behavior can be described by a sequence of model states. In other words, in order to recognize human behaviors, we first recognize postures frame by frame, and the motion sequence can be represented by the recognition results of each frame.

Since it is easy to obtain a well-estimated model for normal behaviors, we start with one state for a normal behavior. A set of parameters Θ^* of the normal behavior HMM is learned by maximizing the likelihood of observation sequences $X = \{X_1, X_2, \dots, X_n\}$ as follows:

$$\Theta^* = \arg \max_{\Theta} \prod_{i=1}^n P(X_i|\Theta) \quad (1)$$

The probability density function of each HMM state is assumed to be a Gaussian Mixture Model (GMM).

When a new HMM is generated, a set of parameters can be evaluated through the Baum-Welch algorithm [7], [8], which is a generalized expectation-maximization algorithm. Baum-Welch algorithm can estimate maximum likelihood and posterior mode for the parameters (i.e. transition and emission probabilities) of an HMM, given emissions as the training data. Then, a Viterbi algorithm [9] is applied to calculate the likelihood, which is a dynamic programming algorithm for finding the most likely sequence of hidden states.

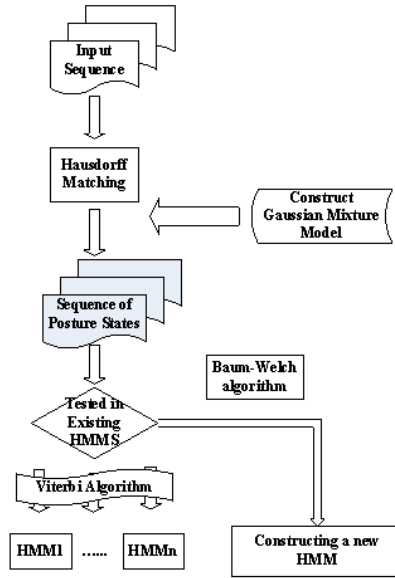


Fig. 1. The block diagram of the SA-HMMs based system

2.1 Shape Matching with Hausdorff Distance

The use of variants of the Hausdorff distance [10] has recently become popular for image matching applications, which is defined as the maximum distance of a set of points to the nearest point in the other set. Generally, Hausdorff distance from set A to set B is a maximum function, which is defined as

$$d_{AB} = \max_{a \in A} \{ \min_{b \in B} \{ l(a, b) \} \} \tag{2}$$

where a and b are points of sets A and B, respectively, and $l(a, b)$ is a metric between these points. Since we are interested in using the Hausdorff distance to identify similarity of a testing frame with a template frame, the distance vector can be defined as $D_i = \{d_{i1}, d_{i2}, \dots, d_{in}\}$, where i denotes the i th frame of the sequence, and n denotes the number of templates pre-stored in the system.

2.2 Hidden Markov Models

A Hidden Markov Model (HMM) [11] is a statistical model in which the system being modeled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters from the observed parameters. The appealing feature of HMM is that no priori assumptions are needed about the statistical distribution of the data to be analyzed. The HMM consists of a finite set of states, each of which is associated with a probability distribution. Transitions among the states are governed by a set of probabilities called

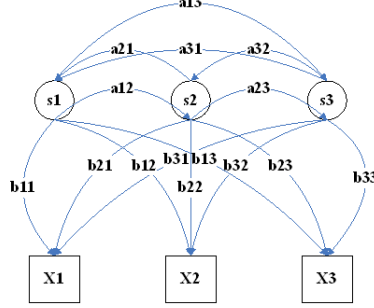


Fig. 2. Probabilistic parameters of a hidden Markov model. S represents states, X represents possible observations, a represents state transition probabilities, and b represents output probabilities.

transition probabilities. In a particular state, an outcome or observation can be generated, according to the associated probability distribution. The probabilistic parameters of a hidden Markov model are shown in Fig. 2. Here, we apply HMM for abnormal behavior recognition.

More specifically, in our system, the parameters of a HMM can be represented as $\Theta = \{\pi, S, A, B\}$, where each parameter is defined as follows:

1. States: $S = \{S_1, S_2, \dots, S_N\}$, where N is number of states. State of HMM at time t is denoted as q_t . In our system, the states are defined as postures, such as walking, sitting, falling, and so on.
2. State transition probability distribution: $A = \{a_{ij}\}$, where a_{ij} denotes the transit probability from state S_i to S_j , which is defined as:

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i), \quad 1 \leq i, j \leq N \quad (3)$$

3. Observation symbol probability distribution: $B = \{b_j(X_t)\}$, where $b_j(X_t)$ defines the probability of observing X at state S_j at time t .

$$b_j(X_t) = P(X_t | q_t = S_j), \quad 1 \leq j \leq N \quad (4)$$

4. Initial state distribution: $\pi = \{\pi_i\}$, where π_i represents the probability of a HMM being at state S_i at time $t = 1$.

$$\pi_i = P(q_1 = S_i), \quad 1 \leq i \leq N. \quad (5)$$

2.3 Gaussian Mixture Model

Observations are emitted on transitions in HMMs and can also be emitted from states. As defined in the above section, $B = \{b_j(X_t)\}$ is a set of emission probabilities, where $b_j(X_t)$ is the probability of observing X on state S_j at time t . In

order to transform the observed Hausdorff distance D into output elements, a Gaussian Mixture Model (GMM) [12] is applied to construct emission functions in our approach.

GMM is an effective tool for data modeling and pattern classification, which is a type of density model comprising a number of component functions, usually Gaussian. GMM assumes that the data under modeling is generated via a probability density distribution which is a weighted sum of a set of Gaussian probability density functions. Due to the flexibility of GMM, it has been successfully applied to numerous applications of data modeling and pattern classification [12]. The single Gaussian function is defined as:

$$f(d; \mu, \Sigma) = 1/\sqrt{(2\pi)^{dim}|\Sigma|} \cdot \exp[-1/2 \cdot (d - \mu)^T \Sigma^{-1} (d - \mu)] \quad (6)$$

where μ is the mean value, Σ is the covariance matrix, and dim denotes the dimension.

The distribution of a random variable $D \in \mathbb{R}^{dim}$ is a mixture of k Gaussians if:

$$f(D = d|\theta) = \sum_{j=1}^k \omega_j \cdot 1/\sqrt{(2\pi)^{dim}|\Sigma_j|} \cdot \exp[-1/2 \cdot (d - \mu)^T \Sigma_j^{-1} (d - \mu)] \quad (7)$$

where the parameters of GMM is defined as $\theta = \{\omega_j, \mu_j, \Sigma_j\}_{j=1}^k$. ω_j is the weights for each Gaussian distribution, and it is constrained by $\sum_{j=1}^k \omega_j = 1$ and $\omega_j > 0, j = 1, \dots, k$. $\mu_j \in \mathbb{R}^{dim}$ is a mean vector and Σ_j is a $dim \times dim$ positive definite covariance matrix. The dimension of μ and Σ is the same with D , the Hausdorff Distance vector.

In our system, by using an Expectation Maximization (EM) method, an optimal set of parameters for GMMs can be identified in an iterative manner. By generating such a Gaussian mixture model for classification, the influence of transition postures can be reduced significantly, which leads to more robustness in recognition.

2.4 Viterbi Algorithm

The Viterbi algorithm is a dynamic programming algorithm for searching the most likely sequence of hidden state, called the Viterbi path, which results in a sequence of observed events in the context of hidden Markov models. The idea of the Viterbi algorithm is to find the most probable path for each intermediate state, and finally for the terminating state in the trellis. At each time n , only the most likely path leading to each state s_i survives. A reasonable optimality criterion consists of choosing the state sequence (or path) that has the maximum likelihood with respect to a given model. This sequence can be determined recursively via the Viterbi algorithm, which is called the state dynamic programming.

This algorithm makes use of two variables:

1. $\delta_n(i)$ is the highest likelihood of a single path among all the paths ending in state S_i at time n , which is defined as:

$$\delta_n(i) = \max_{q_1, q_2, \dots, q_{n-1}} p(q_1, q_2, \dots, q_{n-1}, q_n = s_i, x_1, x_2, \dots, x_n | \Theta) \quad (8)$$

2. $\psi_n(i)$ allows to keep tracking the “best path” ending in state S_i at time n , which is defined as:

$$\psi_n(i) = \arg \max_{q_1, q_2, \dots, q_{n-1}} p(q_1, q_2, \dots, q_{n-1}, q_n = s_i, x_1, x_2, \dots, x_n | \Theta) \quad (9)$$

In our system, $\delta_n(i)$ determines the most possible route to next posture, and $\psi_n(i)$ remembers how to get there. This is done by considering all of the products of transition probabilities with the maximum probabilities derived from the previous step. The largest product is remembered, together with the one that provoked it.

3 Experimental Results

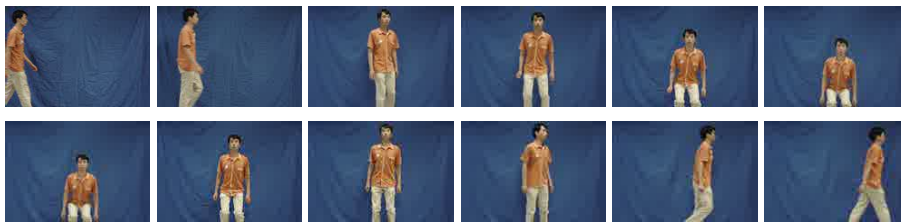
To evaluate the performance of the proposed SA-HMM based framework for abnormal behavior recognition, we capture 35 unlabeled video sequences including both normal and abnormal behaviors, which are used for both testing and



(a)



(b)



(c)

Fig. 3. Examples of normal behaviors patterns. (a) walking; (b) walking-sitting; (c) walking-sitting-walking.

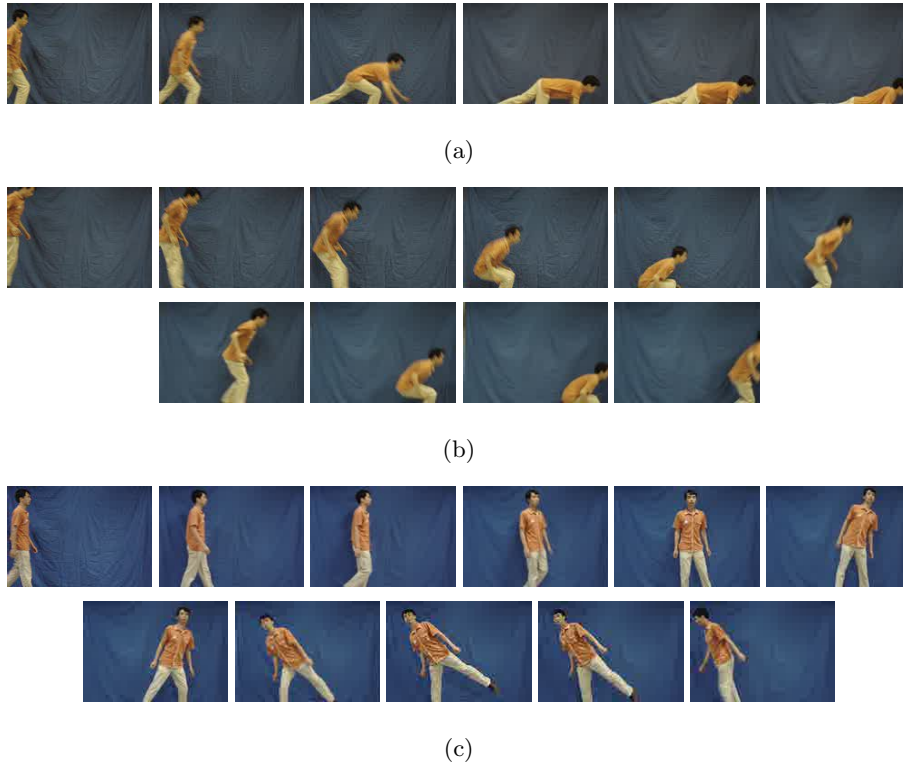


Fig. 4. Examples of abnormal behavior patterns. (a) falling down; (b) jumping; (c) shaking.



Fig. 5. Templates of behavior patterns

on-line training purposes. A set of images representing normal and abnormal human behaviors are shown in Fig. 3 and Fig. 4, respectively. The interested regions are obtained by the background subtraction with fixed color difference threshold. We consider a normal routine consisting of three human activities: (1) walking, (2) walking-sitting, and (3) walking-sitting-walking. The abnormal routine follows the rule of weird actions other than “walking” and “sitting”. Fig. 4 shows three abnormal behavior patterns used in our experiments.

We perform three Gaussian probability density functions on each template image, and consider each template as one state in HMMs. In our experiments, templates consist of six images, as shown Fig. 5, representing “walking”, “standing”, “sitting”, “falling down”, “jumping” and “shaking”, respectively. All of

the estimated Gaussian Mixture Models (GMMs) corresponding to each state are presented as:

$$\begin{aligned}\mu_i &= (\mu_{i1}, \mu_{i2}, \dots, \mu_{ik}) \\ \Sigma_i &= (\Sigma_{i1}, \Sigma_{i2}, \dots, \Sigma_{ik})\end{aligned}$$

where i is the index of GMMs, and $k=6$ denotes the number of states in HMMs.

The HMMs adopted in the experiment are left-to-right type, and each HMM is used to represent one class of behaviors. Suppose that the system has recognized that the current sequence doesn't belong to any of the existing HMMs. Then, a new HMM will be constructed for this sequence, which can be used for training purpose. In this manner, the database can be updated by appending new examples. Initially, the system has no HMM model. After all of 35 random sequences have been trained through our framework, four HMMs have been generated, including one normal classification and three abnormal classifications. The results are listed in Table 1. It can be seen from Table 1 that in 35 sequences, only two normal and three abnormal sequences are incorrectly identified. As expected, our experiments show that our framework can successfully deals with this scenario without any database or priori information.

Table 1. Recognition Results

Behaviors	N	N_r
Normal	20	18
Abnormal(Falling down)	5	5
Abnormal(Jumping)	5	5
Abnormal(Shaking)	5	2

N: Number of input behaviors

N_r : Number of correctly recognized behaviors.

4 Conclusions and Future Work

This paper proposed a SA-HMMs based framework for abnormal behavior detection. Initially, no prior knowledge of possible combinations of valid behaviors in the past and no prior knowledge of what kind of abnormal behaviors may occur in the scene are given. The proposed framework has an inherent flexibility to allow the model to be automatically updated from the testing data, and the online abnormal behavior detection in video sequences can be conducted. In addition, the novel use of the Gaussian Mixture Models in modeling emission probabilities has solved the problem that the emission function is difficult to be evaluated as unknown distribution. Furthermore, the method reduces the influence of many trivial transition states significantly. The experimental results demonstrate the effectiveness of the model for online abnormal behavior detection with good recognition rates.

However, the SA-HMMs based system currently has some limitations. For example, a good shape matching algorithm is required for the system performance. The computational cost of Hausdorff distance is quite high. We will investigate these issues in the future work.

References

1. Chan, M.T., Hoogs, A., Schmiederer, J., Perterson, M.: Detecting rare events in video using semantic primitives with HMM. In: Proc. of IEEE Conf. on ICPR (August 2004)
2. Stauffer, C., Eric, W., Grimson, L.: Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, archive 22 (August 2000)
3. Zhong, H., Shi, J., Visontai, M.: Detecting unusual activity in video. In: Proc. of IEEE Conf. on Vision and Pattern Recognition (June 2004)
4. Boiman, O., Irani, M.: Detecting Irregularities in Images and in Video. In: Proc. 10th IEEE Int'l Conf. Computer Vision, pp. 462–469 (2005)
5. Xiang, T., Gong, S.: Video Behavior Profiling for Anomaly Detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 30(5), 893–908 (2008)
6. Xiang, T., Gong, S.: Video Behavior Profiling and Abnormality Detection without Manual Labeling. In: Proc. 10th IEEE Int'l Conf. Computer Vision, pp. 1238–1245 (2005)
7. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 39(B), 1–38 (1977)
8. Baum, L.E., Petrie, T., Soules, G., Weiss, N.: A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.* 41(1), 164–171 (1970)
9. Viterbi, A.J.: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory* 13, 260–269 (1967)
10. Olson, C.F.: A Probabilistic Formulation for Hausdorff Matching. In: Proc. of IEEE Conf. on Vision and Pattern Recognition (CVPR 1998), Santa Barbara, CA, pp. 150–156 (1998)
11. Rabiner, L.R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE* 77(2), 257–286 (1989)
12. Batu, T., Guha, S., Kannan, S.: Inferring Mixtures of Markov Chains. In: Shawe-Taylor, J., Singer, Y. (eds.) COLT 2004. LNCS, vol. 3120, pp. 186–199. Springer, Heidelberg (2004)