ABNORMAL EVENT DETECTION FROM SURVEILLANCE VIDEO BY DYNAMIC HIERARCHICAL CLUSTERING

Fan Jiang, Ying Wu, Aggelos K. Katsaggelos

Department of Electrical Engineering and Computer Science, Northwestern University
2145 Sheridan Rd, Evanston, IL 60208, USA
{fji295, yingwu, aggk}@eecs.northwestern.edu

ABSTRACT

The clustering-based approach for detecting abnormalities in surveillance video requires the appropriate definition of similarity between events. The HMM-based similarity defined previously falls short in handling the overfitting problem. We propose in this paper a multi-sample-based similarity measure, where HMM training and distance measuring are based on multiple samples. These multiple training data are acquired by a novel dynamic hierarchical clustering (DHC) method. By iteratively reclassifying and retraining the data groups at different clustering levels, the initial training and clustering errors due to overfitting will be sequentially corrected in later steps. Experimental results on real surveillance video show an improvement of the proposed method over a baseline method that uses single-sample-based similarity measure and spectral clustering.

Index Terms— Surveillance, event detection, clustering

1. INTRODUCTION

The development of complex video surveillance and traffic monitoring systems has captured recently the interest of both research and industrial communities due to the growing availability of inexpensive sensors and processors, and the increasing safety and security concerns. The goal of an intelligent surveillance system is to automatically process video streams, continuously recorded in specific situations for several days (even weeks), in order to characterize the actions taking place and to infer whether they present a threat that should be signaled to a human operator. Examples of abnormal events in traffic surveillance videos include pedestrians trespassing the street, vehicles that make U-turns or brake suddenly or pull over the road.

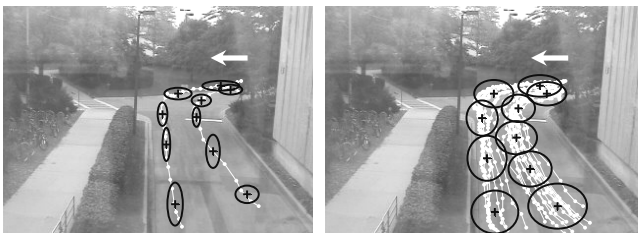
In real videos, the suspicious events are rare, difficult to describe, hard to predict and can be subtle. However, based on the assumption that an abnormal event is associated with the distinctness of the activity (e.g., a running person where everybody walks is interpreted as abnormal as well as a walking person where the rest run) and a normal event indicates the commonality (e.g., a path that most people walk on), some researchers [1-6] define events as either clusters of parameter space components (normal events) or outliers (abnormal events). In order to perform this clustering-based approach, a similarity measure between two events, probably with different time lengths, needs to be specified. Some recent results [1-4] define the distance of two HMM-represented sequences based on the likelihood of observing one sequence given the HMM trained from another sequence. To be exact, the larger their likelihood of being generated from each other's model will be, the more similar these two sequences are. However, this cross likelihood measurement has the problem of model overfitting due to data shortage, as the HMM is trained on only one sample. Therefore data clustering using this single-sample-based similarity is quite unreliable, especially for the popular spectral clustering algorithm [2, 4-6], which is extremely sensitive to the construction of the similarity matrix (whose eigenvalues are utilized).

In this paper we propose a multi-sample-based similarity measure to suppress the overfitting problem, where HMM representation is based on several similar samples. The acquisition of these multiple training data is by hierarchically clustering and iteratively retraining the whole dataset, which is summarized as dynamic hierarchical clustering (DHC) algorithm. This algorithm can dynamically correct initial overfitting errors as the numbers of training samples increase (i.e. data clusters become larger). In addition, it is not sensitive to the absolute values of similarity, because simple comparison operation instead of eigenvalue decomposition is needed in the proposed approach.

2. CLUSTERING-BASED APPROACH FOR ABNORMAL EVENT DETECTION

2.1. HMM representation of video events

In many existing work on surveillance video analysis [2, 4, 7, 8], video events are represented as object trajectories or time evolutions of certain frame features, which can be further modeled by HMM. For example, Fig. 1(a) shows two trajectories of pedestrians (white lines) extracted from a



(a) (b)

Fig. 1. HMM modeling of object trajectories.

surveillance video monitoring a road crossing. A 5-state HMM with Gaussian emission probability is fitted to the 2-D trajectory feature vector $\{(x_1, y_1), (x_2, y_2), \ldots (x_T, y_T)\}$, where $\{x, y\}$ denotes the coordinate of object center at every frame and T is the length of the trajectory. The black ellipses and crosses in Fig. 1(a) show the means and variances of every state.

2.2. Modeling of normal events

The clustering-based approach detects abnormal events by first modeling normal events. After training data that are acquired from the history videos are represented /parameterized by HMMs as described in Sec. 2.1, unsupervised clustering is performed on them based on a certain similarity measure (will be described later in Sec. 3). The clustering process ends up with a few data groups. Those groups containing large number of samples (e.g., more than the average number) are chosen as normal pattern groups, and then HMMs are learned for every normal group. These HMMs, denoted by $\{m_k\}$ (k = 1, 2, …), are models of normal events.

2.3. Detection of abnormal events

Based on the models of normal groups, detection of abnormal events can be performed to new video data. Specifically, given an unseen object trajectory i, the likelihood of observing i given any HMM of normal events m_k is denoted by $L(i \mid m_k)$. If the maximum likelihood is less than a threshold, i.e.,

$$\max_k \{L(i \mid m_k)\} < Th_A, \qquad (1)$$

where Th_A is a threshold, this query trajectory i is detected as abnormal.

3. CLUSTERING ALGORITHM

3.1. Multi-sample-based similarity measure

In some recent work [2, 4], the distance d_{ij} between two events/trajectories i and j, modeled by two HMMs m_i and m_j respectively, is defined as:

$$d_{ij} = L(i \mid m_i) + L(j \mid m_j) - L(i \mid m_j) - L(j \mid m_i), \qquad (2)$$

where $L(i \mid m_j)$ denotes the log-likelihood of trajectory i utilizing the model m_j for trajectory j, normalized by trajectory length T, that is,

$$L(i \mid m_j) = \frac{1}{T_i} \log P(i \mid m_j) \cdot \qquad (3)$$

If the trajectories i and j are different, their likelihood of being generated by each other's model, $L(i \mid m_j)$ and $L(j \mid m_i)$, will be smaller than the likelihood of being generated by itself's model, $L(i \mid m_i)$ and $L(j \mid m_j)$, thus the distance d_{ij} will be large. If the two trajectories are similar, the difference between the cross likelihood and likelihood of self modeling will be small, thus the distance is small. The extreme case is that distance of two identical trajectories will be qual to zero.

However, this HMM-based distance measure has the problem of overfitting with trajectory data extracted from real videos. Note that the variances of the fitted Gaussian distributions indicated by black ellipses in Fig. 1(a) are very small. This is because HMM is trained on only one sample thus it fits the data too closely. This overfitted model will generate very different parameters for similar trajectories in the same direction (e.g., the two trajectories in Fig. 1(a)). As a result, the distance defined in Eq. 2 becomes too large to group similar trajectories into one cluster. One solution to this problem is to use more similar data to train a model that allows for larger variation as illustrated in Fig. 1(b). In terms of this multi-sample-based modeling, the distance between two *groups* of trajectories (groups i and j) can be defined similarly to Eq. 2, except for a modification of the likelihood term. That is, we propose the following definition

$$L(i \mid m_j) = \frac{1}{N_i} \sum_r \frac{1}{T_r} \log P(i_r \mid m_j), \qquad (4)$$

where i_r denotes the r-th trajectory in group i (with its length equal to T_r) and N_i is the number of trajectories in group i. In other words, $L(i \mid m_j)$ is refined as the average of the likelihood of all trajectories in group i, generated by the HMM trained on group j.

The multi-sample-based distance measure is more reliable than the one based on a single sample. For example, the distance between the two trajectories in Fig. 1(a) calculated by Eqs. 2 and 3 is equal to 263.72, while the distance between the two groups containing 20 trajectories each in Fig. 1(b) calculated by Eqs. 2 and 4 is equal to 22.16. As the trajectories shown in Figs. 2(a)(b) are all on the same road and in the same direction, thus they need to be clustered into one group. This can be accomplished much easier with a smaller distance calculated using Eq. 4.

3.2. Dynamic hierarchical clustering (DHC)

HMM modeling based on multiple samples provides a better representation of the trajectory data. However, this is a "chicken-and-egg" problem. On one hand, models are

0). Initialization: each trajectory in the dataset forms a group and is fitted with a HMM. There are N groups and N HMMs;

1). Distance measurements: calculate distances $\{d_{ij}\}$ between two groups i and j in the dataset by Eqs. 2 and 4;

2). Merging: the two groups i and j with smallest d_{ij} are merged into one if the following criterion is satisfied

$$\frac{L(i \mid m_i) \cdot L(i \mid m_j)}{L(i \bigcup j \mid m_{i \bigcup j})} < 1 \qquad \text{(hypothesis testing)}$$

where $L(i \mid m_i)$ and $L(j \mid m_j)$ are likelihood of group i and j generated by HMMs trained on the two groups respectively, $L(i \bigcup j \mid m_{i \bigcup j})$ is the likelihood of samples of both groups generated by HMM trained on all these samples, as defined in Eq. 4; otherwise no groups can be merged and the clustering process ends;

3). Reclassifying: m_i and m_j are replaced by $m_{i \bigcup j}$; then based on the N-1 HMMs, all the data are classified into N-1 groups by the maximum likelihood (ML) criterion;

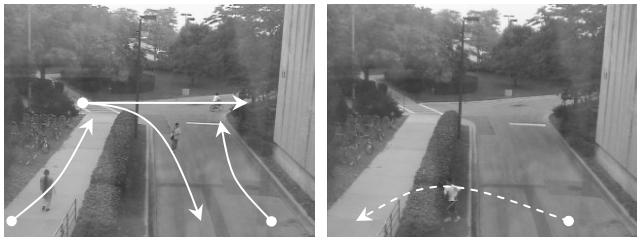
4). Retraining: the N-1 HMMs are retrained based on the updated N-1 data groups;

5). $N = N$ -1; go back to step 1).

Fig. 2. Proposed dynamic hierarchical clustering algorithm.

acquired by training on samples in one group; while on the other hand, groups are acquired by model-based clustering. The common approach to solve such an interlocked problem is to use an iterative approach. For instance, the EM algorithm is an iterative way to solve the embedded problem of data segmentation and model parameters estimation. To allow for an iterative solution, trajectory clustering can not be accomplished in one-step but in a hierarchical fashion instead. In fact, our proposed dynamic hierarchical clustering (DHC) algorithm is based on classic hierarchical clustering [9], incorporated with an iteration process of data reclassifying and model retraining, as described in Fig. 2.

At the beginning of this clustering algorithm (step 0), data samples are possibly overfitted as each HMM is trained on just one trajectory. However, when samples are clustered into larger groups, the number of training data increases as retraining is performed on groups of samples instead of on a single sample at step 4. Therefore, the overfitted HMMs at the beginning can be sequentially refined/updated. Meanwhile, the first few samples that are probably clustered incorrectly due to overfitting will be gradually corrected at step 3 of reclassifying during the iteration process. In other words, the proposed DHC algorithm has the ability of self-adjustment in both model training and data clustering. Another advantage of this algorithm is that it is not sensitive to the absolute similarity/distance values, as at step 2 only the comparison of distance is required to find two group candidates for merging, compared to the complex



(a) (b)

Fig. 3. Trajectory database.

eigenvalue decomposition used in spectral clustering [2, 4]. In addition, testing is used at step 2 to automatically decide at which level the clustering process stops.

4. EXPERIMENTS

The trajectory data used in our experiment is extracted from a 5-hour-long surveillance video, monitoring pedestrians walking at the intersection of four roads. Normal trajectories include paths with people frequently walking on, such as the ones shown in Fig. 3(a). Abnormal trajectories are those rare ones when somebody does not follow the usual routes, as shown for example in Fig. 3(b). Our database includes 1102 trajectories in total, with 1004 normal ones and 98 abnormal ones. All the experiments are done on a leave-one-out basis, each time with 992 (90%) randomly chosen trajectories from the whole database for the modeling of normal events, and 110 (10%) for abnormality detection testing. And the average performance of 10 times' testing is shown below.

The problem of abnormal event detection is a two-class classification problem (normal events vs. abnormal events), with two types of errors, i.e., the false alarm (FA) error, when the method passes a normal trajectory, and the false rejection (FR) error, when the method rejects an abnormal trajectory. Accordingly, the false alarm rate (FAR) and the false rejection rate (FRR) are adopted to evaluate the performance. We also use the half-total error rate (HTER), which combines FAR and FRR into a single measure, that is, HTER = (FAR + FRR) / 2.

Fig. 4 shows the performance of the abnormal events detection system at different levels of dynamic hierarchical clustering, i.e., using a different number of iteration. In order to show the whole trends, the two groups with the smallest distance are always merged, without performing the hypothesis testing at step 2 of Fig. 2. As the performance changes very slowly at the beginning of the iteration (groups are too small), we only show the results after number of iterations is greater than 800. We observe that FAR always decreases while FRR continuously increases as the number of iterations increases. As groups become larger, less normal trajectories are falsely accepted as abnormal ones, while some abnormal trajectories may be incorrectly clustered into normal groups. For comparison we have implemented a baseline method, spectral clustering

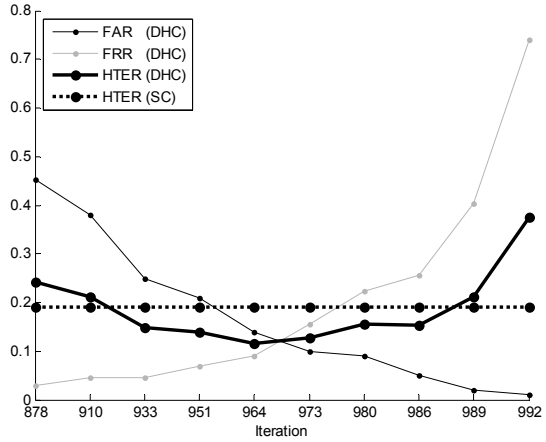


Fig. 4. Results at different numbers of the iteration.

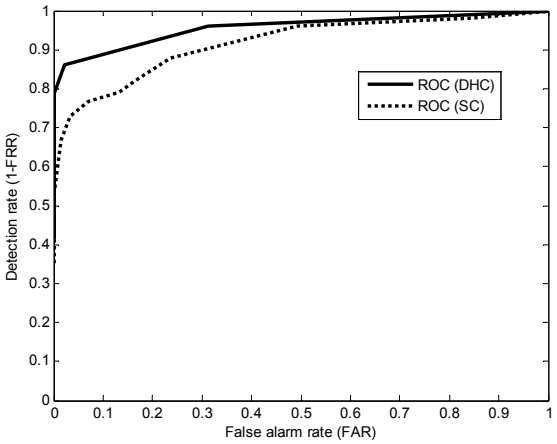


Fig. 5. Results at the termination point of the iteration.

(SC) using single-trajectory-based similarity measure [2], with HTER results illustrated by the dashed line in Fig. 4. It is noticed that within a "good range" of iterations, the HTER of the proposed method is lower than that of the baseline method.

If hypothesis testing is used to terminate the iteration process, the clustering stops at iteration 964 (when 992 – 964 = 28 groups are left) with HTER = 11% (8% drop from baseline HTER = 19%). At the termination point of the iteration, ROC curves are plotted in Fig. 5 for both proposed and baseline methods by varying the abnormality threshold Th_A in Eq. 1. There is a clear improvement of the performance for our proposed method.

5. CONCLUSION

The HMM representation of object trajectories enables the measure of similarity between video events by cross likelihood but suffers from the overfitting problem due to data shortage. We proposed in this paper a novel dynamic hierarchical clustering (DHC) approach, where the HMMs are trained on multiple samples and the initial clustering errors caused by overfit are corrected in the iterative process. The proposed method is not sensitive to the absolute similarity values and calculates the number of clusters automatically. These advantages are also demonstrated experimentally over a baseline algorithm.

6. REFERENCES

[1] J. Ajmera and C. Wooters, "A Robust Speaker Clustering Algorithm," in IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 411-416, December 2003.

[2] F. Porikli and T. Haga, "Event Detection by Eigenvector Decomposition Using Object and Frame Features," in IEEE Conference on Computer Vision and Pattern Recognition Workshop, pp. 114-114, June 2004.

[3] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan, "Semi-supervised Adapted HMMs for Unusual Event Detection," in IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 611-618, June 2005.

[4] T. Xiang and S. Gong, "Video Behaviour Profiling and Abnormality Detection without Manual Labelling," in IEEE International Conference on Computer Vision, vol. 2, pp. 1238-1245, October 2005.

[5] L. Zelnik-Manor and M. Irani, "Event-Based Analysis of Video," in IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 123-130, 2001.

[6] H. Zhong, J. Shi, and M. Visontai, "Detecting Unusual Activity in Video," in IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 819-826, July 2004.

[7] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi, "Traffic Monitoring and Accident Detection at Intersections," in IEEE Transactions on Intelligent Transportation Systems, vol. 1, pp. 108-118, June 2000.

[8] G. Medioni, I. Cohen, F. Bremond, S. Hongeng, and R. Nevatia, "Event Detection and Analysis from Video Streams," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 873-889, August 2001.

[9] R. Duda, P. Hart, and D. Stork, "Pattern Classification," by John Wiley & Sons, Inc. pp. 550-556, 2001.