ABNORMAL EVENT DETECTION BASED ON TRAJECTORY CLUSTERING BY 2-DEPTH GREEDY SEARCH

Fan Jiang, Ying Wu, and Aggelos K. Katsaggelos

Northwestern University, EECS Dept
2145 Sheridan Rd, Evanston, IL 60208 USA

ABSTRACT

Clustering-based approaches for abnormal video event detection have been proven to be effective in the recent literature. Based on the framework proposed in our previous work [1], we have developed in this paper a new strategy for unsupervised trajectory clustering. More specifically, an information-based trajectory dissimilarity measure is proposed, based on the Bayesian information criterion (BIC). In order to minimize BIC, the agglomerative hierarchical clustering is applied using a 2-depth greedy search process. This strategy achieves better clustering results compared to the traditional 1-depth greedy search. The increased computational complexity is addressed with several bounds on the trajectory dissimilarity.

***Index Terms*—** Video surveillance, event detection, unsupervised clustering.

1. INTRODUCTION

Many surveillance applications require analysis of the events taking place in video streams recorded in specific situations, in order to find suspicious or abnormal actions, which might present a threat and should be signaled to a human operator. For example, in a traffic monitoring system, it is useful to detect vehicles that make U-turns or brake suddenly, pedestrians trespassing the street, and other abnormal traffic behavior or violations.

It is hard to describe the abnormality of an event in absolute terms and ahead of time. Alternatively, many recent works [1–4] base their abnormal events detection on an unsupervised clustering approach. Our previous work [1] falls in this category. It includes two phases: a training and a testing phase, as illustrated in Fig. 1. First, object trajectories $\{s\}$ are extracted from historical video data. A video event is represented by a sequence of object features of any one trajectory, such as x and y coordinates of the object center at each frame. The historical events or trajectories may contain both normal and abnormal ones and they are unlabeled. Unsupervised clustering is then performed which ends up with several clusters. Those clusters containing large number of trajectories (*i.e.*, more than the average number) are regarded as normal event clusters and a hidden Markov model (HMM) is used to characterize each normal cluster, denoted by $\{\theta\}$. In the testing phase, given any query trajectory q extracted from the unseen video, a decision is made on whether it is abnormal or normal, based on the likelihood of observing q given models $\{\theta\}$.

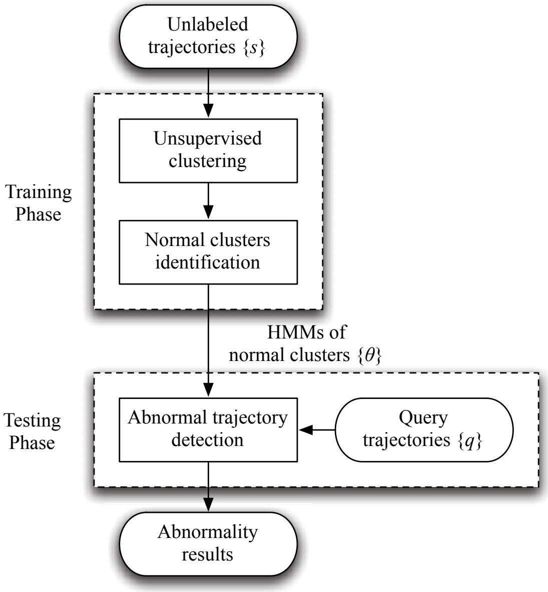


Fig. 1. Clustering-based abnormal trajectory detection

Based on this framework, we have developed in this paper a novel strategy for unsupervised clustering. The proposed approach differs from current works [1–4] in two aspects. First, an information-based trajectory dissimilarity measure is introduced and the Bayesian information criterion (BIC) is used to automatically determine the number of clusters. Second, as we use the agglomerative hierarchical clustering algorithm, the traditional nearest-neighbor-merged-first strategy (1-depth greedy search) is improved by considering the optimal two merging steps each time (2-depth greedy search). The 2-depth search strategy achieves better results but increases the computational complexity dramatically. This is further handled by excluding many possible merging groups with several theoretical bounds on the trajectory dissimilarity.

This paper is organized as follows. Sec. 2 describes the dissimilarity measure used for trajectory clustering. The 2-depth greedy search strategy for clustering is presented in Sec. 3. Experimental results are shown in Sec. 4, and we conclude the paper in Sec. 5.

2. TRAJECTORY DISSIMILARITY MEASURE

In order to perform clustering, we define a dissimilarity measure between trajectories based on the Bayesian information criterion (BIC) which is a statistical criterion for model selection. This measure imposes a trade-off between model quality and model complexity and was previously used in the area of speech recognition [5, 6]. Generally, BIC of an estimated model for N observations is given by

$$BIC = -\log L + \frac{1}{2} K \log N, \tag{1}$$

where L is the likelihood for the estimated model and K is the number of parameters needed to describe the model. Given any two estimated models, the model with the lower value of BIC should be chosen.

For our trajectories dataset, initially we can model each trajectory (actually the feature sequence) by an HMM, *i.e.*, N HMMs are trained for N trajectories respectively. For this modeling,

$$BIC(i, j, \cdots) = -\sum_{n=1}^{N} \log L_n + \frac{1}{2} N K_0 \log N,$$

where L_n denotes the likelihood of trajectory n generated by the HMM trained by itself θ_n, $P(n \mid \theta_n)$, and K_0 is the number of parameters for each HMM. Then, if trajectory i and j are merged and modeled together by one HMM, the BIC for the new modeling becomes

$$BIC(ij, \cdots) = -\sum_{\substack{n=1 \\ n \neq i,j}}^{N} \log L_n - \log L_{ij} + \frac{1}{2}(N-1) K_0 \log N.$$

The difference of BIC for the two models is equal to

$$\begin{aligned} d(i,j) &= BIC(ij, \cdots) - BIC(i, j, \cdots) \\ &= \log L_i + \log L_j - \log L_{ij} - \frac{1}{2} K_0 \log N. \end{aligned} \tag{2}$$

It is clear that if $d(i,j) < 0$, this merge is favored, while if $d(i,j) > 0$, this merge is not favored. Actually, the smaller $d(i,j)$ is, the greater the tendency for i and j to merge, thus the more similar i and j should be. Therefore, we define $d(i,j)$ as our dissimilarity measure of trajectories. This dissimilarity measurement can be easily extended to groups of samples, because i and j can indicate two groups of trajectories, with L referring to the likelihood of all trajectories in one group.

3. 2-DEPTH GREEDY SEARCH

The standard agglomerative hierarchical clustering that was used in our previous work [1] is a greedy algorithm. It merges at each step the two most similar trajectory groups, resulting in the greatest decrease of BIC. However, this does not guarantee the largest decrease of BIC after two or more merging steps. For example, assume that within the whole dataset, the dissimilarities of trajectories i, j and k, l are the smallest ones, *i.e.*, $d(i,j) < d(k,l) < \cdots$. According to the 1-depth merging rule, i and j are merged in the first step while k and l are merged in the second step. However, it is possible that the decrease of BIC by merging another three trajectories m, n and o is greater than the decrease of BIC with merging i, j and k, l, *i.e.*, $d(m,n,o) < d(i,j) + d(k,l)$. Thus, the performance of the algorithm can be improved by considering a 2-depth search. Specifically, each time we take two merging steps that cause BIC to decrease the most. The hierarchical clustering process continues until no merging can further decrease BIC.

Obviously this 2-depth search is still a greedy approach, though it is better than the 1-depth search. The reason that we do not consider D-depth $(D > 2)$ search is that the computation load becomes extremely heavy if D is large. A 2-depth merging may either merge two different group pairs or merge three groups, depending on which has the smallest BIC. Therefore, an exhaustive 2-depth search requires training HMMs for all pairs and triplets of groups, which requires a large number of computations, increasing with the number of training trajectories. However, there are wasteful computations for the triplets, given that all pairs have already been computed. For example, if all $d(i,j)$, $d(j,k)$, and $d(i,k)$ are large, $d(i,j,k)$ also tends to be large, thus its computation can be possibly excluded. Basically, the triplets i, j, k can be classified into 15 categories based on their nearest neighbors, as illustrated in Fig. 2. The arrows in Fig. 2 are pointing to the nearest neighbors, defined as the one with smallest negative dissimilarity. For example, referring to Fig. 2(b), trajectory group i has the smallest dissimilarity to group l. If the smallest dissimilarity is positive, no arrow is shown, *e.g.*, i, j, k in Fig. 2(a). Based on these 15 categories, we can establish certain exclusions for fast rejection of merging any three groups.

We take category (f) as an example. Merging of i, j, k together can be rejected if $BIC(ijk, l, \cdots) > BIC(il, jk, \cdots)$. A sufficient condition for this to be satisfied is that

$$d(i,j) - d(j,k) + d(i,k) - d(i,l) > \frac{1}{2} K_0 \log N. \tag{3}$$

The proof is given below. First of all, it is reasonable to assume that the trajectories are better represented (with a larger likelihood) by the model trained on themselves, than by the model trained on other samples. Thus we have

$$L_{ij} > L_{ijk}^{ij}, \qquad L_{jk} > L_{ijk}^{jk}, \qquad L_{ik} > L_{ijk}^{ik}, \tag{4}$$

where $L_{ij} = P(ij \,|\, \theta_{ij})$ and $L_{ijk}^{ij} = P(ij \,|\, \theta_{ijk})$ (similarly for the other combinations). In addition, our HMM training assumes that each trajectory is probabilistically independent of others, thus

$$\begin{aligned}(L_{ijk})^2 &= (L_{ijk}^i \cdot L_{ijk}^j \cdot L_{ijk}^k)^2 \\ &= (L_{ijk}^i L_{ijk}^j) \cdot (L_{ijk}^j L_{ijk}^k) \cdot (L_{ijk}^k L_{ijk}^i) \\ &= L_{ijk}^{ij} \cdot L_{ijk}^{jk} \cdot L_{ijk}^{ik} \end{aligned} \tag{5}$$

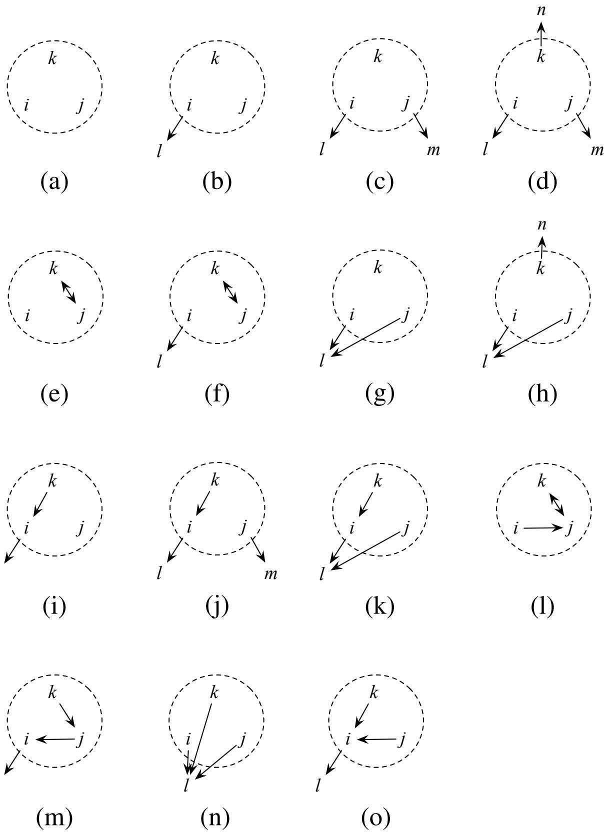
Based on Eqs. 4 and 5, it can be deduced that

$$\begin{aligned} & d(i,j) - d(j,k) + d(i,k) - 2d(i,l) > \frac{1}{2}K_0 \log N \\ \Longrightarrow\; & (\log L_i + \log L_j - \log L_{ij} - \frac{1}{2}K_0 \log N) \\ & - (\log L_j + \log L_k - \log L_{jk} - \frac{1}{2}K_0 \log N) \\ & + (\log L_i + \log L_k - \log L_{ik} - \frac{1}{2}K_0 \log N) \\ & - 2(\log L_i + \log L_l - \log L_{il} - \frac{1}{2}K_0 \log N) \\ & \qquad > \frac{1}{2}K_0 \log N \\ \Longrightarrow\; & -\log L_{ij} - \log L_{jk} - \log L_{ik} \\ & \qquad + 2\log L_{jk} + 2\log L_{il} - 2\log L_l > 0 \\ \Longrightarrow\; & -\log L_{ijk}^{ij} - \log L_{ijk}^{jk} - \log L_{ijk}^{ik} \\ & \qquad + 2\log L_{jk} + 2\log L_{il} - 2\log L_l > 0 \\ \Longrightarrow\; & -\log (L_{ijk})^2 + 2\log L_{jk} + 2\log L_{il} - 2\log L_l > 0 \\ \Longrightarrow\; & -\log L_{ijk} - \log L_l + \log L_{il} + \log L_{jk} > 0 \\ \Longrightarrow\; & BIC(ijk, l, \cdots) - BIC(il, jk, \cdots) > 0 \\ \Longrightarrow\; & BIC(ijk, l, \cdots) > BIC(il, jk, \cdots) \end{aligned}$$

Similar exclusion conditions have been derived for categories (a)-(m) in Fig. 2 (details can be found in [7]). We do not consider exclusions for categories (n) and (o), because for these categories merging of i, j, k can not be easily rejected as they are all neighbors. In our 2-depth searching process, quickly checking the exclusion conditions for any triplet that belongs to categories (a)-(m) reduces the computations dramatically.

4. EXPERIMENTAL RESULTS

We have made experiments on a 5-hour-long surveillance video sequence. In this video scene, people/bicycles/vehicles (i.e., objects) may move along their normal paths most of the time, while some objects may not follow their normal paths (i.e., abnormality). We performed background subtraction and target tracking to obtain the motion trajectories of the targets in the scene. Our event database includes 1000 trajectories in total, with 898 normal ones and 102 abnormal ones. Fig. 3a and Fig. 3b show examples of a normal trajectory (solid line) and an abnormal one (dashed line), respectively. All experiments were done on a leave-one-out basis, *i.e.*, each time with 900 randomly chosen trajectories from the database for the modeling of normal events, and 100 for the testing of abnormality detection.



(a) (b) (c) (d)
(e) (f) (g) (h)
(i) (j) (k) (l)
(m) (n) (o)

Fig. 2. 15 categories of any three trajectory groups according to different nearest neighbors

First, we calculated the BIC at several different clustering levels, *i.e.*, with different number of clusters in the hierarchical clustering. The BIC values shown in Fig. 4 keep on decreasing as the number of clusters decreases, until clustering terminated when the number of clusters reached 22. However, we kept on merging the trajectory groups with smallest dissimilarity until all the trajectories were clustered into one group. It is observed that for a number of clusters smaller than 22 BIC increases. This justifies that our method terminated at a right number of clusters with a minimum BIC.

In addition, both 1-depth search and 2-depth search strategies were performed in the clustering approach. The 2-depth search strategy achieves an average false rejection rate of 7% and an average false alarm rate of 10%, about 6% drop from the 1-depth search strategy.

Finally, a computational comparison was made for the

(a) (b)

Fig. 3. Examples of normal and abnomal trajectories

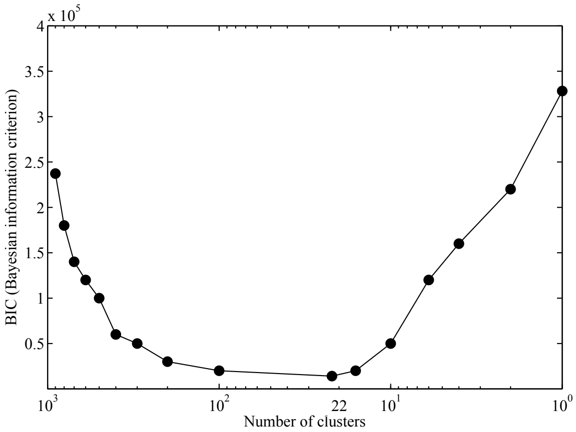


Fig. 4. BIC for different number of clusters

two clustering strategies. The number of HMMs trained were recorded for both strategies at different clustering levels, *i.e.*, with different number of clusters, in the hierarchical process. As shown in Fig. 5, the computational load for the 2-depth greedy search is really high, but with exclusions implemented, it drops off to a comparable level with the 1-depth greedy search. This decrease in computation becomes quite meaningful when the size of the database is very large.

5. CONCLUSION

Using unsupervised clustering to find natural groups of object trajectories in surveillance video provides a practical approach to detect video abnormality. In this paper we have proposed a novel strategy for trajectory clustering. Particularly, an information-based trajectory dissimilarity measure is used as a clustering criterion. A 2-depth greedy search strategy is applied in the hierarchical clustering, with a goal of searching for the clusters that minimize BIC. This strategy can achieve better results than the traditional 1-depth greedy search, but it requires many more computations. We have derived some easy-to-check exclusion conditions to substantially reduce the searching space.

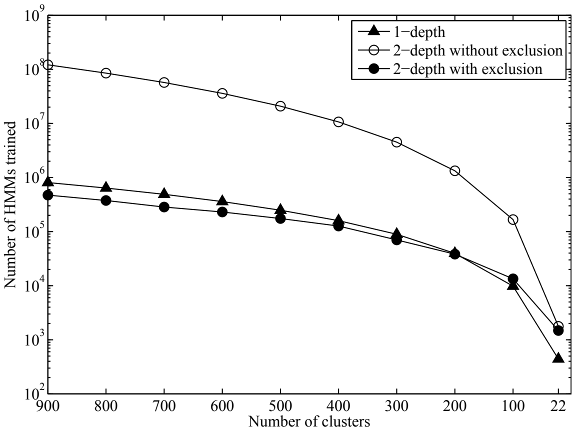


Fig. 5. Computational comparison

6. REFERENCES

[1] F. Jiang, Y. Wu, and A. K. Katsaggelos, “Abnormal event detection from surveillance video by dynamic hierarchical clustering,” in *Proc. IEEE Int’l Conf. on Image Processing*, Sept. 2007.

[2] H. Zhong, J. Shi, and M. Visontai, “Detecting unusual activity in video,” in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, June 2004, vol. 2, pp. 819–826.

[3] T. Xiang and S. Gong, “Video behaviour profiling and abnormality detection without manual labelling,” in *Proc. IEEE Int’l Conf. on Computer Vision*, Oct. 2005, vol. 2, pp. 1238–1245.

[4] H. Zhou and D. Kimber, “Unusual event detection via multi-camera video mining,” in *Proc. IEEE Int’l Conf. on Pattern Recognition*, 2006, vol. 3, pp. 1161–1166.

[5] S. S. Chen and P. S. Gopalakrishnan, “Clustering via the bayesian information criterion with applications in speech recognition,” in *Proc. IEEE Int’l Conf. on Acoustics, Speech, and Signal Processing*, May 1998, vol. 2, pp. 645–648.

[6] J. Ajmera and C. Wooters, “A robust speaker clustering algorithm,” in *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding*, Nov. 2003, pp. 411–416.

[7] F. Jiang, Y. Wu, and A. K. Katsaggelos, “Dynamic hierarchical clustering method for trajectory-based abnormal video event detection,” *IEEE Trans. Image Process.*, submitted.