V. Conclusion

A novel subspace analysis method is reported in this correspondence for feature extraction and dimensionality reduction based on FL distance. The algorithm brings better generalization ability. Moreover, it iteratively updates the basis of the subspace—which makes it suitable for dealing with sequentially coming data and online learning. We assume that: there are at least three training points per class. Two training points determine a feature line onto which the third points can be projected.

Acknowledgment

The authors would like to thank the associate editor and all anonymous reviewers for their constructive comments on the first two versions of this paper.

References

[1] C. Blake and C. Merz, UCI Repository of Machine Learning Databases [Online]. Available: http://archive.ics.uci.edu/ml/datasets.html 2008
[2] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1716–1725, Jul. 2007.
[3] J. T. Chien and C. C. Wu, "Discriminant waveletfaces and nearest feature classifiers for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1644–1649, Dec. 2002.
[4] Y. Fu and T. S. Huang, "Image classification using correlation tensor analysis," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 226–234, Feb. 2008.
[5] Y. Fu, S. Yan, and T. S. Huang, "Classification and feature extraction by simplexization," *IEEE Trans. Inf. Forens. Sec.*, vol. 3, no. 1, pp. 91–100, Jan. 2008.
[6] Y. Fu, S. Yan, and T. S. Huang, "Discriminant simplex analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2008, pp. 3333–3336.
[7] A. Hasan and M. Hasan, "Constrained gradient descent and line search for solving optimization problem with elliptic constraints," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2003, vol. 2, pp. 793–796.
[8] X. He, D. Cai, and J. Han, "Learning a maximum margin subspace for image retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 2, pp. 189–201, Feb. 2008.
[9] H. Hu and J. Yang, "A direct LDA algorithm for high-dimensional data—With application to face recognition," *Pattern Recogni.*, vol. 34, pp. 2067–2070, 2001.
[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[11] S. Z. Li, K. L. Chan, and C. Wang, "Performance evaluation of the nearest feature line method in image classification and retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1335–1339, Nov. 2000.
[12] S. Z. Li and J. Lu, "Face recognition using the nearest feature line method," *IEEE Trans. Neural Netw.*, vol. 10, no. 2, pp. 439–443, Feb. 1999.
[13] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.
[14] Y. Pang, Y. Yuan, and X. Li, "Generalised nearest feature line for subspace learning," *Electron. Lett.*, vol. 43, no. 20, pp. 1079–1080, 2007.
[15] P. J. Phillips, H. Moon, S. A. Rivzi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
[16] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[17] D. Tao, X. Tang, and X. Li, "Which components are important for interactive image searching?," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 3–11, Jan. 2008.
[18] H. Wang and N. Ahuja, "Rank-r approximation of tensors using image-as-matrix representation," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2005, vol. 2, pp. 346–353.
[19] D. Xu, S. Yan, D. Tao, S. Lin, and H. Zhang, "Marginal Fisher analysis and its variants for human gait recognition and content- based image retrieval," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2811–2821, Nov. 2007.
[20] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H. Zhang, "Human gait recognition with matrix representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 896–903, Jul. 2006.
[21] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 212–220, Jan. 2007.
[22] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding with high-order tensor data," *IEEE Trans. Systems, Man, Cybern. B*, vol. 38, no. 2, pp. 342–352, Feb. 2008.
[23] Y. J. Zheng, J. Y. Yang, J. Yang, X. J. Wu, and Z. Jin, "Nearest neighbour line nonparametric discriminant analysis for feature extraction," *Electron. Lett.*, vol. 42, no. 12, pp. 679–680, 2006.
[24] Z. Zhou, S. Z. Li, and K. L. Chan, "A theoretical justification of nearest feature line method," in *Proc. Int. Conf. Pattern Recognition*, 2000, vol. 2, pp. 759–762.
[25] J. Zhou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2617–2628, Oct. 2007.

A Dynamic Hierarchical Clustering Method for Trajectory-Based Unusual Video Event Detection

Fan Jiang, Ying Wu, *Senior Member, IEEE*, and Aggelos K. Katsaggelos, *Fellow, IEEE*

***Abstract*—The proposed unusual video event detection method is based on unsupervised clustering of object trajectories, which are modeled by hidden Markov models (HMM). The novelty of the method includes a dynamic hierarchical process incorporated in the trajectory clustering algorithm to prevent model overfitting and a 2-depth greedy search strategy for efficient clustering.**

***Index Terms*—Event detection, unsupervised clustering, video surveillance.**

I. Introduction and Related Work

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. Typically, the video camera is fixed and the site being monitored is mainly static. Object trajectories are extracted from the video and the video events can be represented by time sequence of the various features of the objects. In many cases, no *a priori* knowledge is given for patterns of unusual video events. Thus, we aim to analyze all the trajectories extracted from existing videos, and differentiate unusual trajectories from normal ones automatically. Based on this analysis, we also need to detect unusual events in additional videos.

Manuscript received October 29, 2007; revised November 25, 2008. Current version published March 13, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Dan Schonfeld.

The authors are with the Electrical Engineering and Computer Science Department, Northwestern University, Evanston, IL 60208 USA (e-mail: fji295@eecs.northwestern.edu; yingwu@eecs.northwestern.edu; aggk@eecs.northwestern.edu).

Digital Object Identifier 10.1109/TIP.2008.2012070

To address this problem, a clustering-based approach has been investigated in the literature [1]–[5]. This approach is based on the fact that a normal event is associated with the commonality of the behavior and an unusual event indicates its distinctness. For instance, people running represents an unusual event if most people in the crowd are walking, and a car moving against traffic also represents an unusual event. Clearly, what characterizes normality is the high recurrence of some similar events. Typically, there are only a few such normal patterns in a specific surveillance scenario. Therefore, unsupervised clustering can be performed on all video events, so that those events clustered into dominant (e.g., large) groups can be identified as normal, while those that cannot be explained by the dominant groups (e.g., distant from all cluster centers) are defined as unusual.

Specifically, in order to perform clustering on video events, we need to measure the similarity of two events. Usually, a trajectory can be modeled as a dynamic process with different outputs at different times. We can measure the distance between trajectories based on the similarity between their corresponding models. Commonly used models are HMM and DBN [2], [3], [5]–[10]. Similarity between HMMs can be measured using the cross likelihood ratio (CLR) [11]. As in [2], [3], and [5], let i and j be two feature sequences, modeled by two HMMs θ_i and θ_j, respectively. The dissimilarity between i and j is defined as

$$\begin{aligned} d(i,j) &= \log \frac{L_i}{L_i^j} + \log \frac{L_j}{L_j^i} \\ &= \log L_i + \log L_j - \log L_i^j - \log L_j^i \end{aligned} \tag{1}$$

where L_i and L_j denote the likelihoods of i and j being generated by their own models, i.e., $P(i \mid \theta_i)$ and $P(j \mid \theta_j)$, respectively, while L_i^j and L_j^i denote the cross likelihoods, $P(j \mid \theta_i)$ and $P(i \mid \theta_j)$, respectively. Note that the superscript of L denotes the trajectory that the likelihood is referring to, and the subscript of L denotes the HMM training trajectory. When these two are the same (for self-likelihood), the superscript is omitted. This rule is always followed in the correspondence.

Alternatively, a dissimilarity measure between time series based on the Bayesian information criterion (BIC) was used in the area of speech recognition [12]. BIC is a statistical criterion for model selection, which is defined as

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

where L is the likelihood for the estimated model, N is the number of observations, and K is the number of model parameters. Given any two estimated models, the model with the smaller value of BIC is the one to be chosen. If every trajectory in the database is modeled by an HMM with the same structure, thus with equal number of parameters K_0, and N HMMs are trained for N trajectories, for this modeling we have

$$\mathrm{BIC}(i,j,\ldots) = -\sum_{n=1}^{N} \log L_n + \frac{1}{2} N K_0 \log N \tag{3}$$

where $(i,j,\ldots)$ denotes any two trajectories i,j and all the other trajectories. Then, if trajectories i and j are merged and modeled together by one HMM, the BIC for the new modeling becomes

$$\begin{aligned} \mathrm{BIC}(ij,\ldots) = -\sum_{\substack{n=1 \\ n \neq i,j}}^{N} \log L_n - \log L_{ij} \\ + \frac{1}{2}(N-1) K_0 \log N \end{aligned} \tag{4}$$

where $(ij,\ldots)$ denotes the trajectory group (containing the merged trajectories i and j) and all the other trajectories. The difference of BIC values for the two models is equal to

$$\begin{aligned} d(i,j) &= \mathrm{BIC}(ij,\ldots) - \mathrm{BIC}(i,j,\ldots) \\ &= \log L_i + \log L_j - \log L_{ij} - \frac{1}{2} K_0 \log N. \end{aligned} \tag{5}$$

If $d(i,j) < 0$, this merge is favored because it results in a better model. The smaller $d(i,j)$ is, the greater the tendency for i and j to merge. Thus, $d(i,j)$ measures the dissimilarity of two trajectories. This measure can be easily extended to groups of trajectories, because i and j can indicate two groups of trajectories, with L referring to the likelihood of all trajectories in one group.

This type of model-based measurement automatically takes into consideration both the spatial and temporal nature of the trajectories. However, it is not robust because models tend to overfit with few training samples. For the BIC-based dissimilarity measure, if i and j denote two single trajectories, their trained HMMs can be quite inaccurate, resulting in an unreliable $d(i,j)$.

Based on any defined event similarity, many unsupervised clustering algorithms can be applied, including different versions of the spectral clustering algorithm [1]–[4], [13], [14], hierarchical clustering [5], [12], [15], [16], and sequential grouping method [6], [17]. After that, the dominant (normal) groups can be identified by intercluster similarity [1], [2], group size [18], or the associated eigenvalues in spectral clustering [3], [13], [14].

In this correspondence, we propose a novel dynamic hierarchical clustering (DHC) method for unusual event detection from surveillance video. The proposed approach follows the common steps of the clustering-based approach as described above. First, HMMs are used to characterize object trajectories. The overfitting problem is suppressed by a dynamic process of reclassifying and retraining the trajectory groups at different clustering levels, referred to as dynamic hierarchical clustering (DHC) algorithm. Second, the BIC-based dissimilarity measure is used for event clustering. We have designed a 2-depth search strategy that works better than the typical 1-depth greedy search in minimizing BIC. To handle the computational complexity of the 2-depth search, we have derived exclusion conditions for a large number of searching paths, i.e., early pruning of branches. Finally, a probabilistic framework is used to identify dominant groups from the clustering results. Instead of checking for large size groups, normal clusters can be determined as those with high prior probabilities.

The correspondence is organized as follows. Section II describes in detail the trajectory clustering algorithm. Section III explains how to identify normal clusters from all clusters and use normal models to detect unusual events in additional videos. Experimental results are presented in Section IV, and we conclude the paper in Section V.

II. Trajectory Clustering

A. Dynamic Hierarchical Clustering (DHC)

In our system, each trajectory is represented by a feature sequence $\{(x_1,y_1),(x_2,y_2),\ldots,(x_T,y_T)\}$, where (x,y) denotes the 2-D coordinates of the object center at every frame and T is the length of the trajectory. Following [12], HMMs with Gaussian emission probability are used to characterize the trajectories, and agglomerative hierarchical clustering can be performed using BIC-based dissimilarity measure. In other words, the two trajectories or trajectory groups i,j with smallest $d(i,j)$ as defined in (5) are continuously merged until there is no $d(i,j) < 0$.

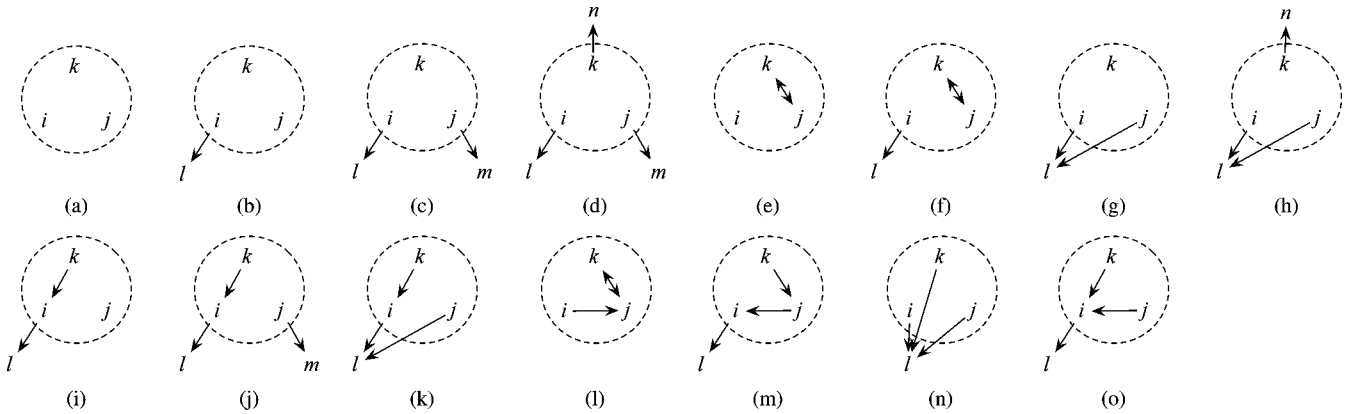

Fig. 2. Fifteen categories of any three trajectory groups according to different nearest neighbors.

0) *Initialization:* each trajectory in the dataset forms a group and is fitted with an HMM. There are N groups with N HMMs;
1) *Dissimilarity Measurement:* calculate the dissimilarity $d(i,j)$ for every two groups i and j in the dataset by the measure in (5);
2) *Merging:* the two groups $\hat{i}$ and $\hat{j}$ with smallest $d(\hat{i},\hat{j})(d(\hat{i},\hat{j}) < 0)$ are merged; if there is no $d(i,j) < 0$, the clustering terminates;
3) *Reclassification:* a new HMM $\theta_{\hat{i}\hat{j}}$ is trained, replacing $\theta_{\hat{i}}$ and $\theta_{\hat{j}}$; then based on the $(N-1)$ HMMs, all trajectories are reclassified into $(N-1)$ groups by the maximum likelihood (ML) criterion;
4) *Retraining:* $(N-1)$ HMMs are retrained based on the updated $(N-1)$ data groups, respectively;
5) *Update:* $N = N-1$; go back to step 1).

Fig. 1. Dynamic hierarchical clustering (DHC).

However, a model-based similarity measure as mentioned above has the overfitting problem given few training samples. This is true for the first several steps of hierarchical clustering: when clusters contain only a few trajectories (starting from only one trajectory), the trained HMMs tend to be overfitted and the dissimilarity measures based on them are quite unreliable, thus resulting in clustering errors. These errors will propagate to future clustering steps. To address this problem, we can update the clustering results at each merging step. In other words, once a new HMM is trained after trajectory merging, all the trajectories in the database are reclassified. Possibly, some incorrectly clustered trajectories at previous steps are associated to the new HMM. All the HMMs are then retrained based on the updated trajectory clusters. This is a typical data reclassification and model retraining process, which is used in many iterative algorithms such as the expectation-maximization algorithm. This updating also enables error correction at later clustering steps, as clusters have gathered more samples and the trained HMMs are more reliable. We refer to this process as dynamic hierarchical clustering (DHC), which is illustrated in Fig. 1. It differs from typical agglomerative hierarchical clustering in that it incorporates into each clustering step the data reclassification process (step 3) and the model retraining process (step 4).

B. 2-Depth Search Strategy and Branch Pruning

Clustering using a BIC-based similarity measure can be regarded as a searching problem, i.e., searching all the clustering possibilities for the minimum BIC. However, the algorithm listed in Fig. 1 is an 1-depth greedy search, as it always merges at every step only two trajectory groups that decrease the BIC the most. A possible improvement of the algorithm is considering a 2-depth search, i.e., each time we take two merging steps that cause BIC to decrease the most. A 2-depth merge may either merge two different group pairs or merge three groups, depending on which has the smallest BIC. However, the problem is that an exhaustive 2-depth search requires training HMMs for all pairs and triplets of trajectory groups. This can be unaffordable if the size of the trajectory dataset is large. In the following, we will establish certain exclusion rules for the fast rejection of merging certain group triplets.

Basically, any triplet i,j,k can be classified into one of the 15 categories based on their "nearest neighbors", as illustrated in Fig. 2. The nearest neighbor of one trajectory (group) is defined as the one (group) with smallest negative dissimilarity to it, which is shown by the arrow in Fig. 2. For example, the arrow in Fig. 2(b) means that the trajectory group i has the smallest dissimilarity to group l, and the double arrow in Fig. 2(e) denotes that trajectory groups k and j have their smallest dissimilarities to each other. If the smallest dissimilarity is positive, no arrow is shown, e.g., i,j,k in Fig. 2(a). For the first 13 categories, there exist simple exclusions of merging i,j,k as one group, according to the BIC minimization criterion, as shown in Table I. It is straightforward to check these sufficient conditions, as they are only based on pairwise dissimilarities.

We take category (f) as an example. Merging i,j,k together can be rejected if

$$\mathrm{BIC}(ijk,l,\ldots) > \mathrm{BIC}(il,jk,\ldots). \tag{6}$$

Substituting the definition of BIC into (6) results in

$$-\log L_{ijk} - \log L_l + \log L_{il} + \log L_{jk} > 0. \tag{7}$$

Multiplied by 2, it becomes

$$-\log(L_{ijk})^2 + 2\log L_{jk} + 2\log L_{il} - 2\log L_l > 0. \tag{8}$$

Our HMM training assumes that each trajectory is probabilistically independent from others, thus

$$\begin{aligned}(L_{ijk})^2 &= \left(L^i_{ijk}\cdot L^j_{ijk}\cdot L^k_{ijk}\right)^2 \\ &= \left(L^i_{ijk}L^j_{ijk}\right)\cdot\left(L^j_{ijk}L^k_{ijk}\right)\cdot\left(L^k_{ijk}L^i_{ijk}\right) \\ &= L^{ij}_{ijk}\cdot L^{jk}_{ijk}\cdot L^{ik}_{ijk}.\end{aligned} \tag{9}$$

TABLE I
EXCLUSION CONDITIONS FOR MERGING i, j, k IN CATEGORIES (A)–(M)

Category	Exclusions	Sufficient conditions
(a)	$BIC(ijk,\cdots) > BIC(i,j,k,\cdots)$	$d(i,j)+d(j,k)+d(i,k) > \frac{1}{2}K_0 \log N$
(b)	$BIC(ijk,l,\cdots) > BIC(il,j,k,\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l) > \frac{1}{2}K_0 \log N$
(c)	$BIC(ijk,l,m,\cdots) > BIC(il,jm,k,\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l)-2d(j,m) > \frac{1}{2}K_0 \log N$
(d)	$BIC(ijk,l,m,n,\cdots) > BIC(il,jm,kn,\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l)-2d(j,m)-2d(k,n) > \frac{1}{2}K_0 \log N$
(e)	$BIC(ijk,\cdots) > BIC(i,jk,\cdots)$	$d(i,j)-d(j,k)+d(i,k) > \frac{1}{2}K_0 \log N$
(f)	$BIC(ijk,l,\cdots) > BIC(il,jk,\cdots)$	$d(i,j)-d(j,k)+d(i,k)-2d(i,l) > \frac{1}{2}K_0 \log N$
(g)	$BIC(ijk,l,\cdots) > BIC(il,j,k,\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l) > \frac{1}{2}K_0 \log N$
	or $BIC(ijk,l,\cdots) > BIC(i,jl,k,\cdots)$	or $d(i,j)+d(j,k)+d(i,k)-2d(j,l) > \frac{1}{2}K_0 \log N$
(h)	$BIC(ijk,l,n,\cdots) > BIC(il,j,kn,\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l)-2d(k,n) > \frac{1}{2}K_0 \log N$
	or $BIC(ijk,l,n,\cdots) > BIC(i,jl,kn,\cdots)$	or $d(i,j)+d(j,k)+d(i,k)-2d(j,l)-2d(k,n) > \frac{1}{2}K_0 \log N$
(i)	$BIC(ijk,l,\cdots) > BIC(il,j,k,\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l) > \frac{1}{2}K_0 \log N$
	or $BIC(ijk,\cdots) > BIC(ik,j,\cdots)$	or $d(i,j)+d(j,k)-d(i,k) > \frac{1}{2}K_0 \log N$
(j)	$BIC(ijk,l,m,\cdots) > BIC(il,jm\cdots)$	$d(i,j)+d(j,k)+d(i,k)-2d(i,l)-2d(j,m) > \frac{1}{2}K_0 \log N$
	or $BIC(ijk,m,\cdots) > BIC(ik,jm,\cdots)$	or $d(i,j)+d(j,k)-d(i,k)-2d(j,m) > \frac{1}{2}K_0 \log N$
(k)	$BIC(ijk,l,\cdots) > BIC(ik,jl,\cdots)$	$d(i,j)+d(j,k)-d(i,k)-2d(i,l) > \frac{1}{2}K_0 \log N$
(l)	$BIC(ijk,\cdots) > BIC(i,jk,\cdots)$	$d(i,j)-d(j,k)+d(i,k) > \frac{1}{2}K_0 \log N$
(m)	$BIC(ijk,l,\cdots) > BIC(il,jk,\cdots)$	$d(i,j)-d(j,k)+d(i,k)-2d(i,l) > \frac{1}{2}K_0 \log N$

Then (8) can be expanded as

$$-\log L_{ijk}^{ij} - \log L_{ijk}^{jk} - \log L_{ijk}^{ik} + 2\log L_{jk} + 2\log L_{il} - 2\log L_l > 0. \quad (10)$$

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

$$L_{ij} > L_{ijk}^{ij}, \quad L_{jk} > L_{ijk}^{jk}, \quad L_{ik} > L_{ijk}^{ik}. \quad (11)$$

Therefore, a sufficient condition for (10) to be satisfied is that

$$-\log L_{ij} - \log L_{jk} - \log L_{ik} + 2\log L_{jk} + 2\log L_{il} - 2\log L_l > 0. \quad (12)$$

By simple substitution, this can be rewritten as

$$d(i,j) - d(j,k) + d(i,k) - d(i,l) > \frac{1}{2}K_0 \log N. \quad (13)$$

Therefore, if (13) is satisfied, merging of i, j, k as one group can be rejected according to (6). Similar exclusion conditions have been derived for categories (a)–(m) and are shown in Table I. 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.

As these exclusion conditions are just based on pairwise dissimilarities, $d(i,j), d(j,k), d(i,k)$, etc., they can be checked quickly before performing HMM training for trajectory triplets. If the hierarchical clustering is regarded as a tree searching process, these exclusion conditions enable an early pruning of branches. Our 2-depth search (with pruning) algorithm is summarized in Fig. 3. Note that the 2-depth search is still a greedy approach. The reason that we do not consider a D-depth ($D > 2$) search is that the computational load becomes very large as D increases.

III. NORMAL CLUSTER IDENTIFICATION AND ABNORMALITY DETECTION

Suppose that all trajectories in the training dataset are clustered into C groups with C corresponding HMMs. Out of these groups, normal (dominant) trajectory groups need to be identified and their corresponding HMMs are kept as normal models, which is used to detect unusual trajectories from new video. As described in Section I, some researchers determine the dominant groups as groups with large number of samples or groups with high intercluster similarities. However, we believe that dominant groups can be better determined in a probabilistic framework as explained next.

Normally, each trajectory is associated with one of the C HMMs. However, in fact at the reclassfication step in our clustering algorithm (step 3) in Fig. 1 and step 4) in Fig. 3), the likelihood of each trajectory being generated from every HMM, i.e., $L_c^i = P(i \mid \theta_c)$, is calculated, where $i = 1, 2, \ldots, N$ denotes any trajectory in the dataset and θ_c ($c = 1, 2, \ldots, C$) the trained HMMs. In other words, each trajectory i has probability L_c^i of being generated by model θ_c. Hence, we may consider each trajectory as being generated by a mixture model, with each component being one of the C HMMs, i.e.,

$$P(i) = \sum_{c=1}^{C} \left[\pi(c) \cdot L_c^i\right] \quad (14)$$

where $\pi(c)$ is the prior probability of the HMM component c, which can be estimated by the EM iteration. Initially, we assume equal prior

0) *Initialization:* each trajectory in the dataset forms a group and is fitted with an HMM. There are N groups with N HMMs;
1) *Pairwise Group Measurement:* calculate $d(i,j)$ of any two groups i and j in the dataset using (5);
2) *Triple Group Measurement:* calculate only $d(i,j,k)$ for those three groups i,j,k that do not satisfy any of the conditions in Table I;
3) *Merging:* merge two group pairs or three groups together, based on which results in the greatest decrease of BIC. The clustering process terminates if no merge can decrease BIC;
4) *Reclassification:* new HMMs are trained for the merged groups; then based on the $(N-2)$ HMMs, all trajectories are reclassified into $(N-2)$ groups by the maximum likelihood (ML) criterion;
5) *Retraining:* $(N-2)$ HMMs are retrained based on the updated $(N-2)$ data groups;
6) *Update:* $N = N-2$; go back to step 1).

Fig. 3. 2-depth greedy search algorithm.

probabilities, i.e., $\pi(c) = 1/C$. In the E-step, the posterior probability of component c given trajectory i is estimated by Bayes' rule

$$P(c \mid i) = \frac{\pi(c) \cdot L_c^i}{\sum_{r=1}^{C} [\pi(r) \cdot L_r^i]}. \tag{15}$$

In the M-step, the prior probability of each component can be computed by averaging $P(c \mid i)$ for all trajectories, i.e.,

$$\pi(c) = \frac{1}{N} \sum_{i=1}^{N} P(c \mid i). \tag{16}$$

These updated $\pi(c)$ are substituted into (15) for another round of iterations. The iteration continues until $\pi(c)$ converges.

Based on this mixture model calculated above, HMMs of normal events can be determined as those with high $\pi(c)$ (e.g., above average). We denote the normal models by

$$\mathcal{N}_c = \left\{ c \,\middle|\, \pi(c) > \frac{1}{C} \right\}. \tag{17}$$

In addition, all the normal trajectories, i.e., the training trajectories associated with normal models are denoted by

$$\mathcal{N}_i = \Big\{ i \mid \exists\, c \in \mathcal{N}_c, \text{s.t.}, L_c^i = \max_r L_r^i, \quad r = 1,2,\ldots,C \Big\}. \tag{18}$$

Given any query trajectory q, its corresponding probability defined by (14) can be divided into two parts, the probability of q being generated by normal models $P(q \mid \mathcal{N}_c)$ and the probability of q being generated by other models $P(q \mid \overline{\mathcal{N}_c})$, that is

$$\begin{aligned} P(q) &= \sum_{c \in \mathcal{N}_c} [\pi(c) \cdot L_c^q] + \sum_{c \in \overline{\mathcal{N}_c}} [\pi(c) \cdot L_c^q] \\ &= P(q \mid \mathcal{N}_c) + P(q \mid \overline{\mathcal{N}_c}). \end{aligned} \tag{19}$$

Trajectory q is classified as unusual if $P(q \mid \mathcal{N}_c)$ is less than a threshold A. We define the threshold as the minimum probability of a normal trajectory (from $\mathcal{N}_i$) being generated by normal models $(\mathcal{N}_c)$, i.e.,

$$A = \min_i P(i \mid \mathcal{N}_c), \quad i \in \mathcal{N}_i \tag{20}$$

where

$$P(i \mid \mathcal{N}_c) = \sum_{c \in \mathcal{N}_c} \Big[\pi(c) \cdot L_c^i \Big]. \tag{21}$$

IV. Experimental Results

A. Experiments With Traffic Surveillance Trajectories

The proposed method was first tested with a real traffic surveillance scene. This video scene showed a road crossing on a school campus, where many pedestrians were moving along different paths in different directions. Most people were walking along the road (i.e., normal events), while some people did not follow the normal paths (i.e., unusual events). The motion trajectories were extracted from video by background subtraction and object tracking. We simply use 2-D coordinates of the object center at every frame as trajectory features. With a video frame rate of 15 fps, the temporal trajectory length varies from 80 to 200. Our database includes 1000 trajectories in total, with 898 normal ones and 102 unusual ones. Experiments were performed on a leave-one-out basis, i.e., each time 900 trajectories are randomly chosen from the database for training, and the other 100 were used for testing. The average performance of several training and testing cycles was recorded.

Since an HMM is used to characterize trajectories, the parameters of the HMM need to be determined properly. According to the BIC-based dissimilarity measure (5), an HMM with the same number of parameters is used for each trajectory group. In our experiments, we used an HMM with a single Gaussian emission probability and a constant number of states. A mixture of Gaussians was not used, considering that most trajectories in our database had simple shape and not much variation locally. The number of states is determined using a data-driving method similar to the one in [8]. We randomly selected some trajectories from the database, and segmented each one of them into sub-trajectories by searching for local maxima of the spatial curvature of the 2-D trajectory. The number of states was then determined as the typical number of sub-trajectories for these trajectories, because a state can be viewed as a basic pattern of the trajectory. In this experiment, we used five states as determined by the data.

In order to evaluate the proposed dynamic hierarchical clustering (DHC) method in detail, we performed unusual event detection tests at different levels of clustering. To be specific, based on different number of clusters in the hierarchical process $(C = 900, 800, \ldots, 100, 22)$, normal cluster identification and unusual event detection was performed as described in Section III. Abnormal detection results for all these levels were collected, with the false alarm rate (FAR) and false rejection rate (FRR) shown in Fig. 4(a)–(b), respectively. The lines denoted by "DHC1" and "DHC2" correspond to the dynamic hierarchical clustering algorithm using 1-depth and 2-depth search strategies, respectively. The clustering was terminated when the number of clusters reached 22 since no additional clustering could reduce the BIC value.

In addition, we implemented two baseline methods for comparison: the classical hierarchical clustering algorithm (1-depth search, with no dynamics) and the spectral clustering method using CLR-based dissimilarity measure [2], [3] (described in Section I). Results of these two baseline methods are also shown in Fig. 4, denoted by "HC1" and "SC", respectively. Spectral clustering (SC) is not a hierarchical process, thus it only has one result, shown as horizontal lines in the two graphs.

There is a clear improvement in the performance of the proposed DHC algorithms. At the termination point, the 1-depth search strategy achieves an average FAR of 18% and an average FRR of 12%, and the 2-depth search strategy achieves an average FAR of 10% and an average FRR of 7%. Out of the final 22 trajectory clusters, ten clusters

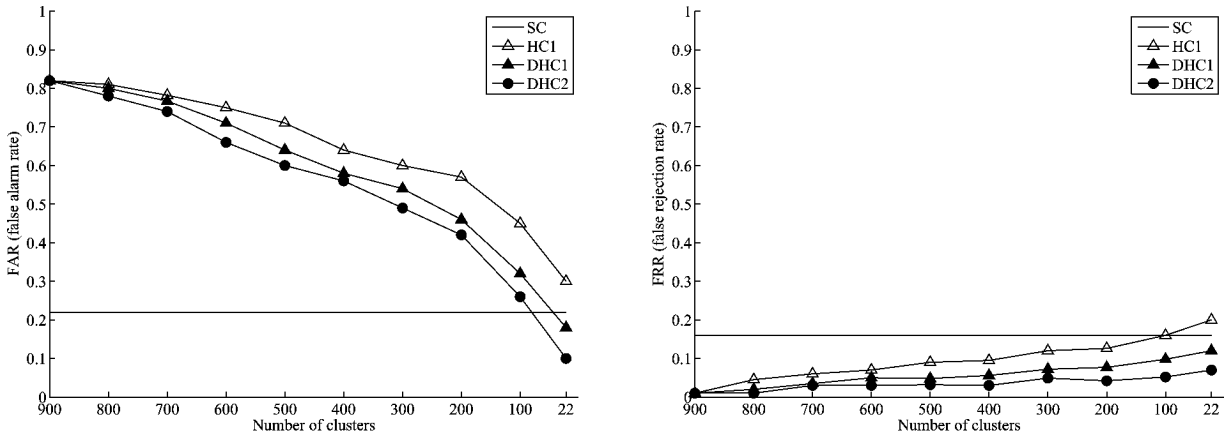

Fig. 4. Comparison results at different level of clustering.

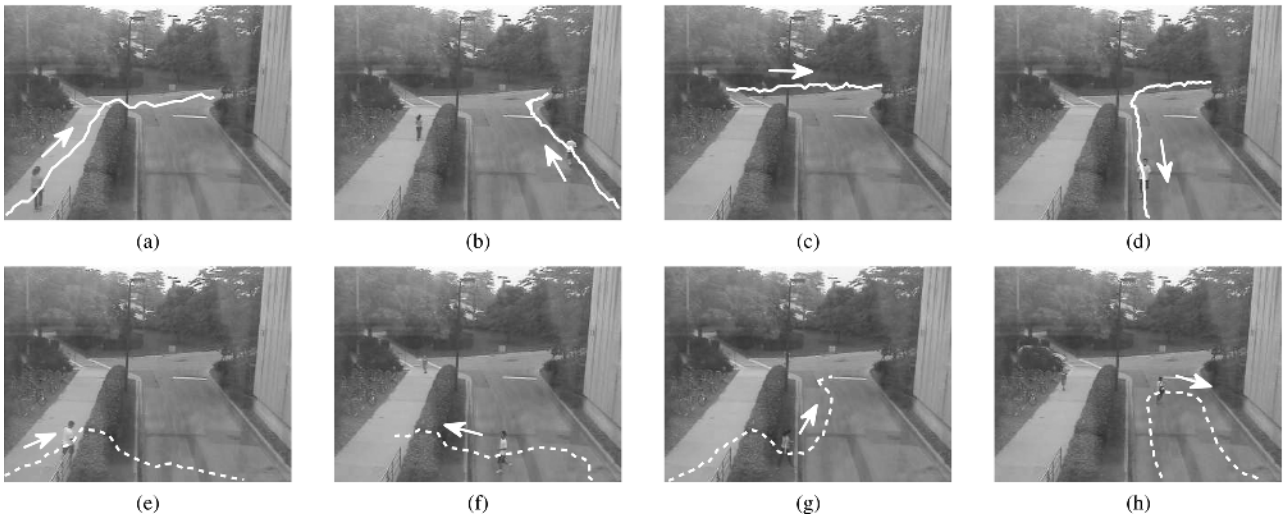

Fig. 5. Examples of normal (a)–(d) and unusual (e)–(h) trajectories.

were identified as normal ones, with 4 examples shown in Fig. 5(a)–(d) (solid lines). The other 12 clusters correspond to various unusual trajectories, with examples shown in Fig. 5(e)–(h) (dashed lines). This is because we add procedures to address the overfitting problem in the clustering process, while classical hierarchical clustering (HC1) does not. Spectral clustering (SC) is a very good clustering algorithm; however, if used with a model-based dissimilarity measure, its nonhierarchical nature does not help in dealing with overfitted models.

B. Experiments With Other Time Series

Objects trajectories extracted from surveillance video represent a time series of features. Therefore, the proposed unusual event detection method should also be applicable on other kind of multivariate time series. We have performed experiments with the Australian Sign Language (ASL) dataset, which is obtained from UCI's KDD archive [19]. Time series in this dataset are obtained by collecting the positions of the language signer's hand at each sampling instant when five professional signers sign around 95 words in multiple sessions. There are no unusual signs collected. However, our definition of unusualness is based on its distinctness. If some specific words are signed for many times, while some other words are signed for only a few times, we can regard the latter case as unusual. Therefore, to construct our database, we select from the original dataset 1072 samples of 15 sign words (72 samples for each word on average) as normalities, and 106 samples of another ten words (11 samples for each word on average) as unusual (see Table II). Each sample is a time series of x and y hand locations, and its time length varies from 60 to 150.

TABLE II
DATABASE OF SIGN LANGUAGE TRAJECTORIES USED

# of normal samples	# of unusual samples
1072 (from 15 sign words)	106 (from 10 sign words)

TABLE III
COMPARISON RESULTS FOR EXPERIMENTS ON OTHER TIME SERIES

	FAR	FRR
Classic hierarchical clustering (HC1)	0.33	0.21
CRL-based spectral clustering (SC)	0.27	0.17
DHC by 1-depth search (DHC1)	0.22	0.15
DHC by 2-depth search (DHC2)	0.20	0.13

Experiments of unusual trajectory detection were performed with the dynamic hierarchical clustering (1-depth/2-depth search) and the two baseline algorithms, similarly to the ones in Section IV-A. We only show the final error rates (FAR and FRR) at the clustering termination point in Table III. All four clustering algorithms use a model-based dissimilarity measure which suffers from the overfitting problem. The proposed dynamic hierarchical clustering (DHC) method that addresses

this problem has been shown to perform better. In addition, the 2-depth search strategy has achieved better results than the 1-depth search, because it decreases BIC more efficiently.

V. Conclusion

We aim to solve the following video mining problem: given some video event data (i.e., object trajectories) without labels, including mostly normal events and a few outliers (unusual events), we need to identify the unusual events from other events and detect more unusual events in other unseen video. Clustering-based approaches try to identify unusualness of events by clustering all events into groups. Models of normal events can be trained based on the dominant clusters. These normal models are used for unusual event detection.

In this correspondence, we have proposed a dynamic hierarchical clustering (DHC) method to address three problems with the current clustering-based approaches. First, the model overfitting problem is suppressed by incorporating into the standard agglomerative hierarchical clustering an iterative cycle of trajectory reclassification and model retraining. Second, a 2-depth search strategy for clustering has been designed that works better than the typical 1-depth greedy search. Its computational complexity is properly handled by early pruning of certain search paths. Finally, the normal and unusual cluster identification is based on a probabilistic framework instead of intuition. These advantages have also been demonstrated experimentally over two baseline methods.

References

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

[2] F. Porikli and T. Haga, "Event detection by eigenvector decomposition using object and frame features," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops*, Jun. 2004, vol. 7, p. 114.

[3] T. Xiang and S. Gong, "Video behaviour profiling and abnormality detection without manual labelling," in *Proc. IEEE Int. Conf. Computer Vision*, Oct. 2005, vol. 2, pp. 1238–1245.

[4] Z. Fu, W. Hu, and T. Tan, "Similarity based vehicle trajectory clustering and anomaly detection," in *Proc. IEEE Int. Conf. Image Processing*, Sep. 2005, vol. 2, pp. 602–605.

[5] H. Zhou and D. Kimber, "Unusual event detection via multi-camera video mining," in *Proc. IEEE Int. Conf. Pattern Recognit.*, Aug. 2006, vol. 3, pp. 1161–1166.

[6] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan, "Semi-supervised adapted hmms for unusual event detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Jun. 2005, vol. 1, pp. 611–618.

[7] Y. A. Ivanov and A. F. Bobick, "Recognition of visual activities and interactions by stochastic parsing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 852–872, Aug. 2000.

[8] F. I. Bashir, A. A. Khokhar, and D. Schonfeld, "Object trajectory-based activity classification and recognition using hidden Markov models," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1912–1919, Jul. 2007.

[9] T. Duong, H. Bui, D. Phung, and S. Venkatesh, "Activity recognition and abnormality detection with the switching hidden semi-Markov model," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Jun. 2005, vol. 1, pp. 838–845.

[10] N. Cuntoor and R. Chellappa, "Epitomic representation of human activities," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Jun. 2007, pp. 1–8.

[11] B. Juang and L. Rabiner, "A probabilistic distance measure for hidden Markov modeling," *AT&T Tech. J.*, vol. 64, no. 2, pp. 391–408, 1985.

[12] S. S. Chen and P. S. Gopalakrishnan, "Clustering via the Bayesian information criterion with applications in speech recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 1998, vol. 2, pp. 645–648.

[13] L. Zelnik-Manor and M. Irani, "Event-based analysis of video," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Dec. 2001, vol. 2, pp. 123–130.

[14] I. Junejo, O. Javed, and M. Shah, "Multi feature path modeling for video surveillance," in *Proc. IEEE Int. Conf. Pattern Recognition*, Aug. 2004, vol. 2, pp. 716–719.

[15] W. Grimson, C. Stauffer, R. Romano, and L. Lee, "Using adaptive tracking to classify and monitor activities in a site," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Jun. 1998, pp. 22–29.

[16] T. Syeda-Mahmood and F. Wang, "Unsupervised clustering using multi-resolution perceptual grouping," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Jun. 2007, pp. 1–8.

[17] D. Makris and T. Ellis, "Learning semantic scene models from observing activity in visual surveillance," *IEEE Trans. Syst., Man, Cybern. B*, vol. 35, no. 3, pp. 397–408, Jun. 2005.

[18] F. Jiang, Y. Wu, and A. K. Katsaggelos, "Abnormal event detection from surveillance video by dynamic hierarchical clustering," in *Proc. IEEE Int. Conf. Image Processing*, Sep. 2007, vol. 5, pp. 145–148.

[19] S. Hettich and S. D. Bay, The UCI KDD Archive, Univ. California, Dept. Inf. Comput. Sci., Irvine, CA, 1999 [Online]. Available: http://kdd.ics.uci.edu