Object Trajectory-Based Activity Classification and Recognition Using Hidden Markov Models

Faisal I. Bashir, *Member, IEEE*, Ashfaq A. Khokhar, *Senior Member, IEEE*, and Dan Schonfeld, *Senior Member, IEEE*

***Abstract*—Motion trajectories provide rich spatiotemporal information about an object's activity. This paper presents novel classification algorithms for recognizing object activity using object motion trajectory. In the proposed classification system, trajectories are segmented at points of change in curvature, and the subtrajectories are represented by their principal component analysis (PCA) coefficients. We first present a framework to robustly estimate the multivariate probability density function based on PCA coefficients of the subtrajectories using Gaussian mixture models (GMMs). We show that GMM-based modeling alone cannot capture the temporal relations and ordering between underlying entities. To address this issue, we use hidden Markov models (HMMs) with a data-driven design in terms of number of states and topology (e.g., left-right versus ergodic). Experiments using a database of over 5700 complex trajectories (obtained from UCI-KDD data archives and Columbia University Multimedia Group) subdivided into 85 different classes demonstrate the superiority of our proposed HMM-based scheme using PCA coefficients of subtrajectories in comparison with other techniques in the literature.**

***Index Terms*—Activity recognition, Gaussian mixture models (GMMs), hidden Markov models (HMMs), trajectory modeling.**

I. Introduction

OBJECT motion-based analysis and recognition has gained significant interest in scientific circles lately. This trend can be attributed mainly to unprecedented advances in hardware and software technologies that allow spatiotemporal data of objects to be easily derived from video and nonvideo sources. Also, novel applications employing analysis of motion trajectory are emerging due to enhanced interest in homeland security as well as due to prevalence of multimedia gadgets in commercial and scientific endeavors. Examples of the motion trajectory include tracking results from video trackers, sign language data measurements gathered from wired glove interfaces fitted with sensors, Global Positioning System (GPS) coordinates of satellite phones, etc. An important application area in this domain is automatic video surveillance which is used, for example, in real-time observation of people and vehicles, in a busy environment, leading to a description of actions and mutual interactions. The research challenge here is to quickly learn the permitted activities and set an alarm at any illegal or abnormal activity being performed. We emphasize that in light of psychological studies reported in the literature [23], it is clear that object motion plays a key role in the domain of activity analysis in general and in video surveillance, in particular [37].

An object trajectory is typically modeled as a sequence of consecutive locations of the object on a coordinate system resulting in a vector in 2-D or 3-D Euclidean space. In different trajectory-based applications, there are two major cornerstones for successful system development: a compact and robust representation of the trajectories to capture the spatiotemporal movement patterns; and a semantically meaningful high-level description of the activities, actions and events based on this trajectory data. This paper is focused on both of these issues and introduces a novel method employing Gaussian mixture models (GMM)-based representations and hidden Markov model (HMM)-based classifiers for motion trajectory representation and analysis. We segment the object trajectories into perceptually similar atomic units that we shall refer to as subtrajectories. The representation of these subtrajectories in the principal component analysis (PCA) subspace is then used to learn the statistical models for each class of object motion. Finally, we represent trajectories as temporal sequences of subtrajectories analogous to the characterization of words as a sequence of phonemes. We subsequently propose the representation of motion trajectories using ergodic HMMs. Experiments on two large datasets of trajectories are reported.

The remaining sections of this paper are organized as follows. Section II surveys related work on trajectory representation and activity analysis. Section III briefly describes our trajectory segmentation and PCA-based representation. Section IV presents the model-based representation of object trajectories for complex action recognition using both GMMs and HMMs. In order to analyze the quality of the proposed classifier, we also discuss the Kullback–Leibler divergence between GMMs and HMMs. Section V provides a comparison of the model-based trajectory representation methods described above with other methods reported in the literature in connection with face recognition. Finally, in Section VI, we present a brief summary and conclusion with an outline of future research in this area.

II. Related Work

This section provides a survey of the related work from recent literature in the areas of trajectory representation, statistical modeling and applications of trajectory-based representation and learning. Object motion is an important feature for the representation and discrimination of an object and its activities from others in video applications. Earlier approaches in motion-based methods focused on object tracking from raw and compressed domain videos [14], [38], [39]. Indexing and

Manuscript received April 6, 2005; revised November 19, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Manuel Samuelides.

F. I. Bashir was with the University of Illinois at Chicago, Chicago, IL 60607 USA. He is now with the Retica Systems, Inc., Waltham, MA 01801 USA (e-mail: fbashir@gmail.com.)

A. A. Khokhar and D. Schonfeld are with the Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: ashfaq@uic.edu; dans@uic.edu).

Digital Object Identifier 10.1109/TIP.2007.898960

searching based on object motion as the dominant cue has attracted a lot of research activity in the past few years [13]. Our previous work on trajectory indexing and retrieval [4] segments the trajectories based on dominant sign changes in curvature data. We represent the subtrajectories using PCA coefficients. We have addressed the view-invariant representation of trajectories for scenarios where similar trajectories are captured from different view points [1]. View-invariant representation has also been addressed in [35] for modeling and recognizing actions performed by individuals in video sequences. Their approach, though compact in representation, cannot be used for partial trajectory processing or generic trajectory representation.

Recently, semantics-based processing of trajectory data to extract high-level information has gained significant interest [7]. Yacoob *et al.* [48] have presented a framework for modeling and recognition of human motions based on principal components. Each activity is represented by eight motion parameters recovered from five body parts of the human walking scenario. In [36], a semantic event detection technique for snooker videos is presented. Trajectory of the while ball is generated using a color-based particle filter. The evolution of the white ball position is modeled using a discrete HMM. In [21], the issue of recognizing a set of plays from American football videos is considered. Using a set of classes each representing a particular game plan and computation of perceptual features from trajectories, the propagation of uncertainty paradigm is implemented using automatically generated Bayesian network. The problem with above approaches is that they are highly domain-dependant, with domain knowledge and sensor dependence on video data being intimately woven into the systems. A sensor-independent approach towards modeling activity performed by a group of objects (persons, cars, etc.) is presented in [44]. Objects in scene are taken as points and they consider the "shape" formed by a configuration of point objects at a given time instant. This "shape" is tracked over time, normal shape is learned and abnormality is detected as perturbation in this shape. Although robust for multiagent abnormal activity detection, this approach can not be applied for single object trajectories. De la Torre *et al.* [15] use PCA and HMM for lip-tracking and eye-tracking. Martin *et al.* [27] model the trajectories for gesture recognition using multidimensional histogram of gestures. In their approach, no segmentation to obtain subtrajectories is performed; only the recent history is taken into account. Starner and Pentland [41] address the issue of American sign language recognition from video sequences. An eight-element feature vector is obtained consisting of each hand's x and y positions, angle of axis of least inertia, and eccentricity of bounding ellipse. Bettinger *et al.* [5] address the problems of learning a person's facial behaviors from video sequences and synthesizing sequences demonstrating the same behavior using HMMs. It is important to point out that the notion of trajectory, the process of segmentation and representation used in [5] are entirely different than the method presented in this paper. Another approach, which has originally been proposed in the context of face recognition by Moghaddam *et al.* [28] can be directly modified for trajectory processing as outlined in Section V. We have reported some preliminary findings towards trajectory-based activity recognition using GMMs [2], HMMs [3], and neural networks [33]. In the forthcoming sections, we present our model-based recognition system that uses GMMs and HMMs for trajectory modeling based on optimal representation provided by PCA. While GMMs and HMMs have been used as a tool in recognition tasks such as speech recognition, face recognition, etc., their use in motion trajectory representation and classification through PCA of subtrajectories presented in this paper is novel.

III. PCA-Based Subtrajectory Representation

A trajectory in our work is a 2-D N-tuple corresponding to the x and y-axes projections of the object's centroid location at each instant of time, $\{(X_k, Y_k), k = 1, \ldots, N\}$. We classify the trajectories into separate classes. The word "class" refers to a type of activity (represented by its full trajectory) for which we have a sufficient number of samples to train the system. The activity classification scheme presented here relies on our robust trajectory segmentation and PCA-based subtrajectory representation [4]. Our trajectory segmentation scheme looks for sharp changes in object's velocity and acceleration through 1st and 2nd order derivatives. Consequently, trajectories are segmented at points of maximum change in curvature of the trajectory. We represent the subtrajectories using PCA because of its optimal energy compaction properties resulting from custom bases derived from the data [24]. The x and y data of each subtrajectory are concatenated into a single vector and all the vectors of subtrajectories from all the classes are stacked to form one data matrix. The principal components of this data matrix are then computed using eigenspace decomposition of the estimated covariance matrix [24]. More details can be found in [4]. The set of PCA coefficients of all subtrajectories for each class are subsequently used to train a stochastic model for each class as explained in the next section.

IV. Model-Based Recognition of Object Trajectories

Model-based recognition and classification has been extensively used in applications such as action recognition [7], sports video analysis [47], speech/speaker recognition [12], etc. In this section, we present the GMM-based modeling, wherein the multimodal probability density function (PDF) corresponding to PCA coefficients of subtrajectories in each class is represented using Gaussian mixtures. The successful static PDF estimation using GMMs is further extended to robustly model temporal variations using continuous-density HMMs.

A. Gaussian Mixture Models

Given the PCA-based representation of subtrajectories for each class, we wish to model the underlying class probability distribution from the training set data. The training set is made as diverse as possible so the recognition system learns all possible data variations. This diversity in training set results in the underlying PDF to be increasingly complex. Hence, the statistical properties of PDF of the class become increasingly nontrivial to model. In the training phase, we estimate the parameters of Gaussian mixtures using the expectation-maximization (EM) algorithm. Once the training phase has been completed, new trajectories are categorized as one of the learned classes of object motion based on the maximum likelihood (ML) principle.

For the parameter estimation problem, we first form the set of training set PCA feature vectors of subtrajectories for an individual class. Note that, since the PCA feature vectors for each subtrajectory are M-dimensional, all of the individual multivariate Gaussian distributions will be M-dimensional. Let the set of PCA coefficients of subtrajectories for the c^{th} class be denoted by Y_c. The class PDF $p(Y_c|\Theta)$ can be modeled to an arbitrary accuracy using a mixture of Gaussians

$$P(Y_c|\Theta_c) = \sum_{i=1}^{N_c} c_i \mathbb{N}(Y_c; \mu_i, \Sigma_i) \tag{1}$$

where $\mathbb{N}(Y_c; \mu_i, \Sigma_i)$ is the M-dimensional Gaussian density with mean vector μ_i and covariance matrix Σ_i, and c_i are the mixing parameters of the Gaussian components, satisfying $\sum_{i=1}^{N_c} c_i = 1$. The mixture is completely specified by the parameter $\Theta_c = \{c_i, \mu_i, \Sigma_i\}_{i=1}^{N_c}$. Since the parameter estimation phase is identical for each class and the training is performed on the disjoint dataset of these classes, we drop the class indexing subscript from our notation for brevity. Now, given a training set of subtrajectories for a particular class $\{y^t\}_{t=1}^{N_T}$, represented by their M-dimensional PCA coefficients, the mixture parameters can be estimated using the ML principal; i.e.,

$$\overset{*}{\Theta} = \arg\max \left[\prod_{t=1}^{N_T} P\left(y^t \middle| \Theta\right) \right]. \tag{2}$$

This estimation problem can be solved using the EM algorithm [16]. A major problem in GMM-based modeling is the reliable estimation of the number of modes to be used. We automatically estimate the number of modes from training set data using a string of pruning, merging and mode-splitting processes. We initialize the number of modes as twice the maximum number of subtrajectories in all of the trajectories for the class. The mixing weight of a mode c_i multiplied by the number of input data samples N determines how many input data samples are effectively used to estimate the mode parameters. This is the simple measure of "value" of each mode. As long as this product is sufficiently high, the mode is estimated accurately. If c_i is too low, the mode is eliminated or merged with another. The weighted skew (third-order moment) and kurtosis (fourth-order moment) for each mode are also monitored. If the sum of these values exceeds a threshold for any mode, that mode is split in two. Finally, we keep track of the distances between all the modes in order to keep the modes far apart from each other. If two modes are found to be too close to each other, they are merged. Merging involves forming a weighted sum of the two modes (weighted by c_1 and c_2).

Once the GMMs for all classes have been trained, the classification of new trajectories can be performed by computing the likelihood for each GMM. For this purpose, the PCA coefficient vectors of the input trajectory after segmentation are posed as an observation sequence to each GMM. During this computation, the likelihood is computed for each individual mode, and the corresponding weights are applied to generate the likelihood of the Gaussian mixture. The trajectory is declared to belong to the class represented by the GMM with the highest likelihood.

B. Hidden Markov Models

The Gaussian mixture-based modeling, as outlined in the previous subsection, represents a robust way of estimating the PDF for each motion pattern class. This method can be helpful in modeling classes where contents are time-invariant and do not have a strong dependence on temporal ordering. Our subtrajectory-based representation approach models the trajectories as a sequence of subtrajectories. This approach requires a scheme that takes the temporal dependence among subtrajectories into account. We propose to adopt the use of HMMs for trajectory classification and recognition applications. HMMs allow the system to stay in the same state or to transit to the next state at any given time according to state transition probabilities learned from training data. This allows modeling of temporal variations, where the duration of the state is a variable. In this context, we are interested in modeling a class of object motions based on the temporal ordering of subtrajectories. Since subtrajectories represent segments of atomic motions between points of change in motion pattern, the resulting process can be modeled as first-order Markov chain. We also observe that mixture of Gaussians is a robust method of estimating the PDF in the absence of temporal variations. We, therefore, propose to use continuous density HMMs, where each state of the HMM is modeled by a mixture of Gaussians.

The first parameter specified for an HMM is the number of states. For each class, represented by a separate HMM, we set the number of states equal to the maximum number of subtrajectories in all the training set trajectories for that class. Once the number of states is fixed, the complete set of model parameters describing the HMM are given by the triplet

$$\lambda = \{\pi_j, a_{ij}, b_j\} \tag{3}$$

where π_j is the probability of the j^{th} subtrajectory being the first subtrajectory among all the trajectories, a_{ij} denotes the probability of the j^{th} subtrajectory occurring immediately after the i^{th} subtrajectory, and b_j denotes the PDF of j^{th} state. We use a Gaussian mixture-based representation for the state PDF.

Once the set of training trajectories for a class are segmented and the number of states decided, the HMM's parameter triplet in (3) can be estimated. For a given trajectory, let there be T subtrajectories. Then, the state variable q_t which corresponds to the t^{th} subtrajectory, takes one of N values $q_t \in \{S_1, \ldots, S_N\}$. Since we assume a Markovian process, the probability distribution of q_{t+1} depends only on q_t. This is described by the state transition probability matrix A whose elements a_{ij} represent the probability that q_{t+1} corresponds to state S_j given that q_t corresponds to S_i. The initial state probabilities are denoted by π_j, the probability that q_1 equals S_i. The observational data O_t from each state of the HMM is generated according to a PDF dependent on the state at the instant of t^{th} subtrajectory, denoted by $b_j(O_t)$. This state-conditional observation PDF is modeled as a Gaussian mixture given by (4), shown at the bottom of the next page, where c_{jm}, μ_{jm} and Σ_{jm} denote the scalar mixing parameter, P-dimensional mean vector and $P \times P$ covariance matrix of the m^{th} Gaussian component in the j^{th} state. Here, each Gaussian component is

a multivariate normal distribution of the same dimensionality as the PCA coefficients representing the subtrajectories. The parameters of the HMM are initialized to random values and the Baum–Welch algorithm is used for estimation using the forward-backward procedure [34].

Once the HMMs for all classes have been trained, the classification of new trajectories can be performed by computing the likelihood that HMM i best describes the test trajectory. For this purpose, the PCA coefficient vectors of input trajectories after segmentation are posed as an observation sequence to each HMM. Given HMMs for the L classes, $\lambda_1, \lambda_2, \ldots, \lambda_L$, and the set of PCA coefficient vectors of input subtrajectories (i.e., the observation sequence) $O_1, O_2, \ldots, O_m$, we assign class label k as the HMM that maximizes the likelihood given subtrajectories [29], [34]

$$k = \arg \max_{i \in [1,\ldots,L]} \sum_j P\left(O_{t+1:m} \left| q_t^i = j, O_{1:t}\right.\right) P\left(q_t^i = j, O_{1:m}\right). \tag{5}$$

This computation is efficiently performed using the forward recursion procedure in the Baum–Welch algorithm [34]. Observe that the ML estimate is equivalent to the maximum *a posteriori* (MAP) estimate for the same model parameters when the prior distribution of the HMMs is uniform.

C. Analysis of GMM and HMM Classifiers

Recently, investigators have chosen to compare between classifiers by measuring the average distance between the classes represented by the individual classifiers [32], [43]. Some of the most common measures used in classification analysis include likelihood-based measures [43], Kullback–Leibler distance (KLD) measure [26], etc. The KLD, or relative entropy, is defined as the average discrimination information per observation between PDFs f_1 and f_2 [26]

$$D_{\mathrm{KLD}}\left(f_1 \| f_2\right) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)} dx. \tag{6}$$

Closed-form expressions of KLD exist for a small family of distributions such as Gaussian and generalized Gaussian density functions. However, exact closed-form expressions for GMMs and HMMs are still not known. Vasconcelos [43] has addressed the problem of finding an approximate closed-form expression for the KLD between GMMs called *asymptotic likelihood approximation* (ALA). Conditions under which it converges to KLD are also given in [43]. Do [17] derives an upper-bound on the Kullback–Leibler distance rate (KLDR) between HMMs and computes the asymptotic approximation based on their parametric representation. Silva *et al.* [40] propose the *average divergence distance* (ADD) based on a representation of the divergence calculated at the observation distribution level of the models.

In the following, we first give the ALA-based approximation of KLD for GMMs and HMMs. Let us consider a GMM-based representation of the PDF of a class, given by

$$P_i(x) = \sum_{k=1}^{M_i} c_{ik}.\mathbb{N}(x, \mu_{i,k}, \Sigma_{i,k}). \tag{7}$$

We use the ALA-based approximation of KLD between GMMs [43] given by (8), shown at the bottom of the page.

Next, we formulate the KLD between continuous density HMMs where each state is modeled using mixture of Gaussians. From [40] we know that the KLD between continuous density HMMs is given by

$$\mathrm{KLD}_H\left(\lambda_i \| \lambda_j\right) = \sum_{l=1}^{m} E_{s_l^i\, s_l^j}\left[\mathrm{KLD}_G\left(P_i^{s_l^i} \middle\| P_j^{s_l^j}\right)\right] \tag{9}$$

where λ_i and λ_j denote the two HMMs; $s_l^i \in \langle V_l^i, \ldots, V_N^i \rangle$ is the l^{th} state in the state sequence of λ_i generated according to the model's initial and transition probabilities; $P_i^{sl^i}$ represents the Gaussian mixture for state s_l^i of λ_i; and $E_{s_l^i s_l^j}(\cdot)$ denotes the expectation operator used to compute the expected value under all possible state mappings. Here, a key assumption is that the two HMMs have the same number of states, m which in our case is the number of subtrajectories. We retain this assumption to simplify the ensuing analysis. We can now extend the ALA-based approximation of the KLD between continuous density

$$b_j(O_t) = \sum_{m=1}^{M} c_{jm} \frac{1}{(2\pi)^{P/2} \left|\Sigma_{jm}\right|^{1/2}} \exp\left\{-\frac{1}{2}(O - \mu_{jm})^T \Sigma_{jm}^{-1}(O - \mu_{jm})\right\} \tag{4}$$

$$\mathrm{KLD}_G\left(P_i \| P_j\right) = -\sum_{k=1}^{M_i} c_{ik}\left[\log c_{j,\beta(k)} + \log \mathbb{N}\left(\mu_{i,k}, \mu_{j,\beta(k)}, \Sigma_{j,\beta(k)}\right) - \frac{1}{2}\mathrm{trace}\left\{\Sigma_{j,\beta(k)}^{-1}\Sigma_{i,k}\right\}\right]$$

where

$$\beta(k) = r \Leftrightarrow \|\mu_{i,k} - \mu_{j,r}\|_{\Sigma_{j,r}}^2 - \log c_{jr} < \|\mu_{i,k} - \mu_{j,s}\|_{\Sigma_{j,s}}^2 - \log c_{js}, \forall s \neq r \tag{8}$$

HMMs where each state is modeled using mixture of Gaussians [see (8) and (9)] to obtain (10), shown at the bottom of the page.

Here, M_i^l refers to the number of Gaussian components (modes) in the GMM of the i^{th} HMM's l^{th} state; $c_{ik}^{s_l^i}$, $\mu_{i,k}^{s_l^i}$ and $\Sigma_{i,k}^{s_l^i}$ represent the weight, mean and variance of the k^{th} Gaussian component in the l^{th} state of the i^{th} HMM; $c_{j\beta(k)}^{s_l^j}$, $\mu_{j,\beta(k)}^{s_l^j}$, and $\Sigma_{j,\beta(k)}^{s_l^j}$ represent the weight, mean and variance of the Gaussian component in the l^{th} state of the j^{th} HMM, which is closest to the k^{th} component in the i^{th} HMM, as defined by the mapping $\beta(k)$ given in (17). We note here that our approximation of the KLD depends on the transition probabilities through expectation of all possible states. We shall now present sufficient conditions under which the interclass distances generated by HMMs are greater than those using GMMs. Subsequently, we validate this claim under the operating conditions in our problem domain and present numerical experiments confirming this result.

It can be easily shown that the ALA-based approximation of KLD for HMMs where each state is modeled using mixture of Gaussians is greater than that based on GMMs; i.e.,

$$\text{KLD}_H\left(\lambda_i \| \lambda_j\right) > \text{KLD}_G\left(P_i \| P_j\right) \tag{11}$$

under certain conditions. In our PCA-based representation, where the covariance matrices are set to be diagonal, these conditions simplify to

$$\begin{aligned}
&\sum_{l=1}^{m} E\left[\sum_{k=1}^{M_i^l} c_{ik}^{s_l^i} \cdot \log c_{j\beta(k)}^{s_l^j}\right] < \sum_{k=1}^{M_i} c_{ik} \cdot \log\, c_{j\beta(k)} \\
&\quad \sum_{l=1}^{m} E\left[\sum_{k=1}^{M_i^l} c_{ik}^{s_l^i} \cdot \sum_{l=1}^{D} \frac{\left(\mu_{i,k,l}^{s_l^i} - \mu_{j,\beta(k),l}^{s_l^j}\right)^2}{\sigma_{j,\beta(k),l}^{s_l^j}}\right] \\
&> \sum_{k=1}^{M_i} c_{ik} \cdot \sum_{l=1}^{D} \frac{\left(\mu_{i,k,l} - \mu_{j,\beta(k),l}\right)^2}{\sigma_{j,\beta(k),l}} \\
&\quad \sum_{l=1}^{m} E\left[\sum_{k=1}^{M_i^l} c_{ik}^{s_l^i} \cdot \sum_{l=1}^{D} \log \sigma_{j,\beta(k),l}^{s_l^j}\right] > \sum_{k=1}^{M_i} c_{ik} \cdot \sum_{l=1}^{D} \log \sigma_{j,\beta(k),l} \\
&\quad \sum_{l=1}^{m} E\left[\sum_{k=1}^{M_i^l} c_{ik}^{s_l^i} \cdot \sum_{l=1}^{D} \frac{\sigma_{i,k,l}^{s_l^i}}{\sigma_{j,\beta(k),l}^{s_l^j}}\right] > \sum_{k=1}^{M_i} c_{ik} \cdot \sum_{l=1}^{D} \frac{\sigma_{i,k,l}}{\sigma_{j,\beta(k),l}}.
\end{aligned} \tag{12}$$

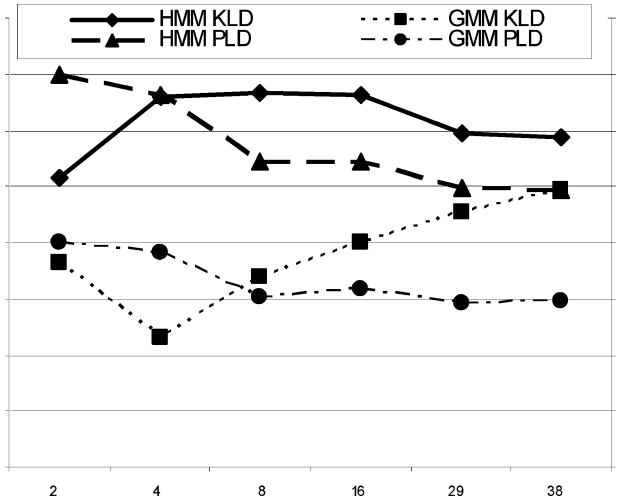


Fig. 1. Interclass distances for HMM- and GMM- based representations using KLD and PLD.

The second distance measure we compute is the *posterior likelihood-based distance* (PLD) between classes for each input trajectory from the test set given the model. Given two trajectories X_i and X_j and their corresponding models P_i and P_j we compute two terms: self-fitness and cross-fitness. We then sum up the individual contributions by all trajectories for a distance measure between the two models as proposed in [32]

$$\begin{aligned}
&D_{\text{PLD}}\left(X_i, X_j\right) \\
&\quad = \left|L\left(X_i \,|P_i\right) + L\left(X_j \,|P_j\right) - L\left(X_i \,|P_j\right) - L\left(X_j \,|P_i\right)\right|.
\end{aligned} \tag{13}$$

Whereas the KLD-based distance computation, as defined in (8) and (10), takes into account the model parameters after training, the PLD measure is computed based on the performance of classification system on a test corpus. We have computed both the KL and posterior likelihood-based distances for GMM and HMM-based representation. This experiment is performed on the ASL dataset with several different class sizes and the results are displayed in Fig. 1. As shown in the figure, the HMM-based representation results in wider interclass distances as compared to its GMM counterpart.

V. Comparison and Analysis

This section compares the performance of our GMM and HMM-based trajectory modeling approaches proposed in the

$$\text{KLD}_H = -\sum_{l=1}^{m} E\left[\sum_{k=1}^{M_i^l} c_{ik}^{s_l^i} \left\{\log c_{j\beta(k)}^{s_l^j} + \log \mathbb{N}\left(\mu_{i,k}^{s_l^i}, \mu_{j,\beta(k)}^{s_l^j}, \Sigma_{j,\beta(k)}^{s_l^j}\right) - \frac{1}{2}\text{trace}\left(\Sigma_{j,\beta(k)}^{s_l^{j-1}} \Sigma_{i,k}^{s_l^i}\right)\right\}\right]$$

where

$$\beta(k) = r \Leftrightarrow \left\|\mu_{i,k}^{s_l^i} - \mu_{j,r}^{s_l^j}\right\|_{\Sigma_{j,r}^{s_l^j}}^2 - \log\, c_{jr}^{s_l^j} < \left\|\mu_{i,k}^{s_l^i} - \mu_{j,t}^{s_l^j}\right\|_{\Sigma_{j,t}^{s_l^j}}^2 - \log\, c_{jt}^{s_l^j} \tag{10}$$

previous section. We also compare our results with an adaptation of the PCA-based density estimation approach outlined in [28]. The approach in [28] is derived for parametric density estimation in high-dimensional spaces using eigenspace decomposition applied to face recognition. The training phase comprises estimation of the mean $\overline{X}$ and covariance Σ of the distribution from the given training set $\{X^t\}$. The sufficient statistic for characterizing the likelihood based on Gaussian density assumption uses the *Mahalanobis* distance which is estimated from the M principal components. Based on this formulation, the likelihood estimate can be written as

$$\hat{P}(X|\Omega) = \left[\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{M/2}\prod_{i=1}^{M}\lambda_i^{1/2}}\right]\left[\frac{\exp\left(-\frac{\varepsilon^2(x)}{2\rho}\right)}{(2\pi\rho)^{(N-M)/2}}\right] \quad (14)$$

where the first term is the true marginal density in the principal feature space and the second term is the estimated marginal density in the orthogonal complement space [28]. For the classification of a new trajectory X, (14) is computed for all classes Ω. The trajectory is declared to belong to the class generating the ML.

A. Datasets

We have used two datasets in our simulations: The Australian Sign Language (ASL) dataset is obtained from UCI's KDD archives [20]. These trajectories are obtained by registring hand coordinates at each successive instant of time by using a Power Glove interfaced to the system. We extract the x and y locations of hand at each sampling instant as five professional signers sign around 95 words in multiple sessions. For each word, there are 69 recordings of signing activity across all signers. We use trajectories from around 83 classes (5727 trajectories) for training and recognition. The second dataset in our experiments has been provided to us by Columbia University's Digital Video and Multimedia Group (DVMM) [13] and contains object trajectories tracked from video clips of sports activities, like high jump, slalom skiing, etc. This dataset, HJSL (108), contains around 40 trajectories of high jump, and 68 trajectories of slalom skiing objects.

B. Simulation Results

We train the system using 50% of the trajectory samples from all classes under consideration, while testing is performed on all the trajectories of each class. We have already compared the GMM and HMM-based models in terms of their interclass separations using KLD and PLD in Section IV-C. Here, we first report the performance of the two systems in terms of average ROC curves and three related scalar metrics of performance for a dataset of 1104 trajectories with 16 classes. The *area under curve* is a convenient way of comparing classifiers, and varies from 0.5 (random classifier) to 1.0 (ideal classifier) [18]. The other metric is the *optimal operating point* based on *equal error rate* criterion. This metric yields an optimal trade-off between false positives and true positives under the equal cost assumption between false acceptance and correct acceptance. The last

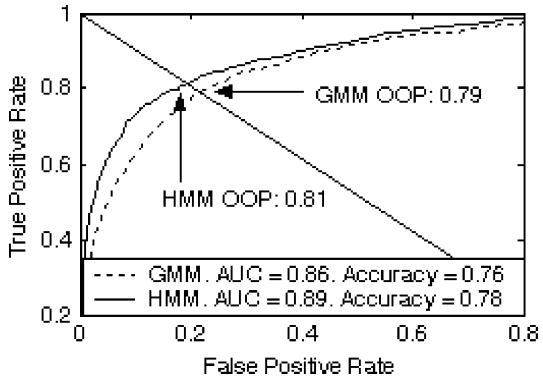


Fig. 2. ROC curves showing the performance of our GMM- and HMM-based classifiers. Area under curve, optimal operating point, and accuracy are shown for the two classifiers.

TABLE I
PROBABILITY OF ACCURACY VALUES FOR VARIOUS CLASS SIZES FROM THE ASL AND THE HJSL DATASETS

Datasets	ASL						HJSL
	#Classes : #Trajectories						
	2: 138	4: 276	8: 552	16: 1104	29: 2001	38: 2622	
HMM	0.96	0.92	0.86	0.78	0.69	0.66	0.91
GMM	0.98	0.89	0.85	0.74	0.67	0.64	0.90
PCA-DE [28]	0.94	0.93	0.83	0.73	0.56	0.62	0.45

scalar metric used to compare between classifiers is the classification accuracy across all the classes defined as

$$P_{\text{Accuracy}} = 1 - \frac{|F|}{|S|} \quad (15)$$

where $|F|$ represents the cardinality of the false positives set, while $|S|$ represents the cardinality of the total dataset. It should be noted here that the ROC curve in Fig. 2 represents an average of 16 individual curves for a total of 1104 queries for classification. The same comment applies to the probability of accuracy value as well.

Finally, we compare our proposed GMM and HMM-based classification systems with the PCA-based density estimation (PCA-DE) approach in [28]. We report the results on a wide range of dataset sizes from the ASL dataset, as well as the HJSL dataset. The results are reported in terms of the average probability of accuracy of the classifiers in Table I.

Based on these results, we see that the PCA-based approaches yield a superior representation for trajectory modeling. From Table I, we also note that the relative accuracy of the HMM-based classifier compared with GMM and Moghaddam *et al.* [28] increases with an increase in the number of classes, thus making it more scalable for large number of classes. Moreover, we note that much higher probability of accuracy values would have been attained provided the size of the training datasets would have been increased proportionally. Furthermore, the HMM-based trajectory modeling system proposed in this paper where individual states are modeled by mixtures of Gaussians has been shown to

perform consistently better than the other trajectory modeling techniques used in all of our experiments.

VI. Summary and Conclusion

In this paper, we have presented a novel framework for motion trajectory-based statistical modeling and classification of data captured from any form of object tracking. We first outline a PDF representation approach using GMMs for motion trajectory identification. The strength of this technique is robust time-invariant modeling of the PDF, but its drawback is the lack of temporal modeling in the formulation. We solve this problem by employing Gaussian mixture-based HMMs for trajectory modeling. While GMMs and HMMs have been used in various recognition tasks, their use in motion trajectory representation and classification in this paper presents a novel new approach to trajectory analysis. We have based our experiments on various measures of performance including Kullback–Leibler divergence for HMMs. The classification systems were tested on two standard datasets in different application domains with 2000 + trajectories. Future research must focus on motion trajectory-based modeling and classification of video sequences that are robust to camera orientation and movement. Generalization of the proposed PCA representation to nonlinear transformations (e.g., Kernel PCA or Kernel discriminant analysis), are needed to deal with nonlinear classification. We plan to explore the estimation of the number of modes for GMMs using the method presented by Gassiat [19] as an alternative to the pruning, merging and mode-splitting process. On theoretical analysis part, the HMM-based formulation proposed in this presentation can be proved to be a particular case of the so-called triplet Markov chain [30]. An important extension of our approach would be required to perform multiple motion trajectory-based classification for "semantic" retrieval from video sequences.

References

[1] F. Bashir, A. Khokhar, and D. Schonfeld, "A hybrid system for affine-invariant trajectory retrieval," presented at the ACM SIGMM Multimedia Information Retrieval Workshop, New York, 2004.

[2] F. Bashir, A. Khokhar, and D. Schonfeld, "Automatic object trajectory-based motion recognition using Gaussian mixture models," presented at the IEEE Int. Conf. Multimedia Expo, Amsterdam, The Netherlands, Jul. 6–8, 2005.

[3] F. Bashir, W. Qu, A. Khokhar, and D. Schonfeld, "HMM-based motion recognition system using segmented PCA," presented at the IEEE Int. Conf. Image Processing, Genoa, Italy, Sep. 11–14, 2005.

[4] F. Bashir, A. Khokhar, and D. Schonfeld, "Real-time motion trajectory-based indexing and retrieval of video sequences," *IEEE Trans. Multimedia*, vol. 9, no. 1, pp. 58–65, Jan. 2007.

[5] F. Bettinger, J. Cootes, and C. J. Taylor, "Modelling facial behaviours," presented at the Brit. Machine Vision Conf., 2002.

[6] A. Braffort and R. Gherbi, "Video tracking and recognition of pointing gestures using hidden Markov models," presented at the IEEE Int. Conf. Intelligent Engineering Systems, 1998.

[7] M. Brand, N. Oliver, and A. Pentland, "Coupled hidden Markov models for complex action recognition," in *Proc. Conf. Computer Vision and Pattern Recognition*, 1997, pp. 994–994.

[8] K. P. Burnham and D. R. Anderson, *Model Selection and Multi-Model Inference – A Practical Information Theoretic Approach*. New York: Springer, 2002.

[9] T. Caelli, A. McCabe, and G. Briscoe, "Shape tracking and production using hidden Markov models," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 15, no. 1, pp. 197–221.

[10] G. Celeux, D. Chauveau, and J. Diebolt, "Some stochastic versions of the EM algorithm," in *J. Statist. Comput. Simul.*, 2002, vol. 55, pp. 287–314.

[11] S. F. Chang, W. Chen, H. J. Meng, H. Sundaram, and D. Zhong, "A fully automated content-based video search engine supporting spatiotemporal queries," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 602–615, Sep. 1998.

[12] T. Chen, C. Huang, C. Chang, and J. Wang, "On the use of Gaussian mixture model for speaker variability analysis," presented at the Int. Conf. SLP, Denver, CO, 2002.

[13] W. Chen and S. F. Chang, "Motion trajectory matching of video objects," presented at the SPIE Conf., San Jose, CA, Jan. 2000.

[14] S. Cheung and A. Zakhor, "Fast similarity search on video sequences," presented at the IEEE Int. Conf. Image Processing, 2003.

[15] F. De La Torre, Y. Yacoob, and L. Davis, "A probabilistic framework for rigid and non-rigid appearance based tracking and recognition," in *Proc. 4th IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2000, pp. 491–498.

[16] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B*, vol. 39, no. 1, pp. 1–38, 1977.

[17] M. N. Do, "Fast approximation of Kullback–Leibler distance for dependence trees and hidden Markov models," *IEEE Signal Process. Lett.*, vol. 10, no. 4, pp. 115–118, Apr. 2003.

[18] T. Fawcett, ROC Graphs: Notes and practical considerations for researchers, " Tech. Rep. HPL-2003–4, HP Labs, 2004.

[19] E. Gassiat, "Likelihood ratio inequalities with applications to various mixtures," *Ann. Inst. Henri Poincare*, vol. 38, pp. 897–906, 2002.

[20] 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

[21] S. S. Intille and A. F. Bobick, "Recognizing planned, multiperson action," *Comput. Vis. Image Understand.*, vol. 81, pp. 414–445, 2001.

[22] M. Isard and A. Blake, "A mixed-state CONDENSATION tracker with automatic model-switching," in *Proc. Int. Conf. Computer Vision*, 1998, pp. 107–112.

[23] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Percept. Psychophys.*, vol. 14, no. 2, pp. 201–211, 1973.

[24] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.

[25] R. Kass and A. Raferty, "Bayes factors," *J. Amer. Statist. Assoc.*, vol. 90, pp. 773–795, 1995.

[26] S. Kullback, *Information Theory and Statistics*. New York: Wiley, 1958.

[27] J. Martin, D. Hall, and J. Crowley, "Statistical gesture recognition through modeling of parameter trajectories," in *Proc. Int. Gesture Workshop on Gesture-based Communication in Human-Computer Interaction*, 1999, pp. 129–140.

[28] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 696–710, Jul. 1997.

[29] K. P. Murphy, "Learning Markov processes," in *The Encyclopedia of Cognitive Science*, L. Nadal, Ed. *et al.* New York: Nature Macmillan, 2002.

[30] W. Pieczynski and F. Desbouvries, "On triplet Markov chains," presented at the Int. Symp. Applied Stochastic Models and Data Analysis, Brest, France, May 2005.

[31] F. Porikli and T. Haga, "Event detection by eigenvector decomposition using object and frame features," presented at the Int. Conf. Computer Vision and Pattern Recognition, 2004.

[32] F. M. Porikli, "Trajectory distance metric using hidden Markov model based representation," presented at the Eur. Conf. Computer Vision, May 2004.

[33] W. Qu, F. Bashir, A. Khokhar, and D. Schonfeld, "A motion trajectory based video retrieval system using parallel adaptive self organizing maps," presented at the Int. Joint Conf. Neural Networks, Montréal, QC, Canada, Aug. 4, 2005.

[34] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[35] C. Rao, A. Yilmaz, and M. Shah, "View-invariant representation and recognition of actions," *Int. J. Comput. Vis.*, vol. 50, no. 2, pp. 203–226, 2002.

[36] N. Rea, R. Dahyot, and A. Kokaram, "Semantic event detection in sports through motion understanding," presented at the Proc. Conf. Image and Video Retrieval, Dublin, Ireland, Jul. 21–23, 2004.

[37] J. M. Rubin and W. A. Richards, "Boundaries of visual motion," Tech. Rep. AIM-835 Artif. Intell. Lab., Mass. Inst. Technol., Cambridge, 1985.

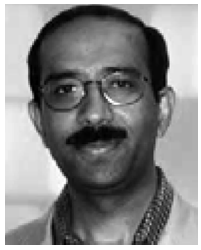
[38] E. Sahouria and A. Zakhor, "A trajectory based video indexing system for street surveillance," presented at the IEEE Int. Conf. Image Processing, 1999.

[39] D. Schonfeld and D. Lelescu, "VORTEX: Video retrieval and tracking from compressed multimedia databases—multiple object tracking from MPEG-2 bitstream," *J. Vis. Commun. Image Represent.*, vol. 11, pp. 154–182, 2000, Invited Paper.
[40] J. Silva and S. Narayanan, "A statistical discrimination measure for hidden Markov models based on divergence," presented at the Int. Conf. Spoken Language Processing, Oct. 4–8, 2004.
[41] T. Starner and A. Pentland, "Visual recognition of American sign language using hidden Markov models," presented at the Int. Workshop Automatic Face and Gesture Recognition, Zurich, Switzerland, 1995.
[42] C. M. Taskiran, C. A. Bouman, and E. J. Delp, "Discovering video structure using the pseudo-semantic trace," in *Proc. SPIE Conf.*, San Jose, CA, 2001, vol. 4315, pp. 571–578.
[43] N. Vasconcelos, "On the efficient evaluation of probabilistic similarity functions for image retrieval," *IEEE Trans. Inf. Theory*, vol. 50, no. 7, pp. 1482–1496, Jul. 2004.
[44] N. Vaswani, A. R. Chowdhury, and R. Chellappa, "Shape activity: A continuous state HMM for moving/deforming shapes with application to abnormal activity detection," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1603–1616, Oct. 2005.
[45] A. Vinciarelli and S. Bengio, "Offline cursive word recognition using continuous density hidden Markov models trained with PCA or ICA features," in *Proc. 6th Int. Conf. Pattern Recognition*, vol. 3, pp. 81–84.
[46] A. A. Wilson and A. F. Bobick, "Hidden Markov models for modelling and recognizing gesture under variation," *Hidden Markov Models: Appl. Comput. Vis.*, pp. 123–160, 2001.
[47] L. Xie, S. F. Chang, A. Divakaran, and H. Sun, "Structure analysis of soccer video with hidden Markov models," presented at the IEEE Inter. Conf. Acoustic, Speech, Signal Processing, Orlando, FL, May 2002.
[48] Y. Yacoob and M. J. Black, "Parameterized modeling and recognition of activities," *Comput. Vis. Image Understand.*, vol. 73, no. 2, pp. 232–247, Feb. 1999.

Faisal I. Bashir (S'05–M'06) received the B.S. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Chicago (UIC) in 2000 and 2006, respectively.

From 2001 to 2005, he was Research Assistant in the Multimedia Systems Lab, UIC. He was a Computer Vision Consultant at Mitsubishi Electric Research Labs (MERL), Cambridge, MA, during the first half of 2006. He is currently with Retica Systems, Inc., Waltham, MA, as Senior Scientist. His research interests include content-based multimedia indexing and retrieval, computer vision, machine learning, and biometric identification.

Dr. Bashir was a recipient of the National Talent Scholarship from 1995 to 2000 and the Provost Award for Graduate Research in Spring 2005.



Ashfaq A. Khokhar (SM'99) received the M.S. degree in computer engineering from Syracuse University, Syracuse, NY, in 1989, and the Ph.D. degree in computer engineering from the University of Southern California, Los Angeles, in 1993.

After receiving the Ph.D. degree, he spent two years as a Visiting Assistant Professor in the Department of Computer Sciences and the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN. In 1995, he joined the Department of Electrical and Computer Engineering, University of Delaware, Newark, where he first served as an Assistant Professor and then as an Associate Professor. In Fall 2000, he joined the Department of Computer Science and the Department of Electrical and Computer Engineering, University of Illinois at Chicago, where he currently serves as a Full Professor. He has published over 100 technical papers in refereed conferences and journals in the area of parallel computing, image processing, computer vision, and multimedia systems. His research interests include wireless and sensor networks, multimedia systems, datamining, and high-performance computing.

Dr. Khokhar was a recipient of the NSF CAREER award in 1998. His paper entitled "Scalable S-to-P Broadcasting in Message Passing MPPs" won the Outstanding Paper award at the International Conference on Parallel Processing in 1996. He served as the Program Chair of the 17th Parallel and Distributed Computing Conference (PDCS), 2004, and Vice Program Chair for the 33rd International Conference on Parallel Processing (ICPP), 2004. He was nominated for IEEE fellow in 2006.

Dan Schonfeld (M'90–SM'05) was born in Westchester, PA, in 1964. He received the B.S. degree in electrical engineering and computer science from the University of California, Berkeley, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Johns Hopkins University, Baltimore, MD, in 1986, 1988, and 1990, respectively.

In 1990, he joined the University of Illinois at Chicago, where he is currently an Associate Professor in the Department of Electrical and Computer Engineering. He has authored over 100 technical papers in various journals and conferences. His current research interests are in signal, image, and video processing; video communications; video retrieval; video networks; image analysis and computer vision; pattern recognition; and genomic signal processing.

Dr. Schonfeld was coauthor of a paper that won the Best Student Paper Award in Visual Communication and Image Processing 2006. He was also coauthor of a paper that was a finalist in the Best Student Paper Award in Image and Video Communication and Processing 2005. He has served as an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (Nonlinear Filtering) as well as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (Multidimensional Signal Processing and Multimedia Signal Processing).