AUTOMATIC OBJECT TRAJECTORY-BASED MOTION RECOGNITION USING GAUSSIAN MIXTURE MODELS

Faisal Bashir[1], Ashfaq Khokhar[2], Dan Schonfeld[3]

University of Illinois at Chicago,
851 S. Morgan St., Chicago, IL, 60607.
{[1]fbashir, [2]ashfaq,[3]ds}@ece.uic.edu

ABSTRACT

In this paper, we propose a novel technique for model-based recognition of complex object motion trajectories using *Gaussian Mixture Models* (GMM). We build our models on *Principal Component Analysis* (PCA)-based representation of trajectories after segmenting them into small units of perceptually similar pieces of motions. These subtrajectories are then fitted with automatically-learnt mixture of Gaussians to estimate the underlying class probability distribution. Experiments are performed on two data sets; the ASL data set (from UCI's KDD archives) consists of 207 trajectories depicting signs for three words, from Australian Sign Language (ASL); the HJSL data set contains 108 trajectories from sports videos. Our experiments yield an accuracy of 85+% performing much better than existing approaches.

1. INTRODUCTION AND RELATED WORK

Object trajectory-based analysis and recognition has gained significant interest in the scientific circles lately. This is primarily due to unprecedented advances in hardware and software technologies that allow spatio-temporal data of objects to be easily derived from video sequences and other motion sensor devices. Examples of the object trajectory include tracking results from video trackers, sign language data measurements gathered from wired glove interfaces fitted with sensors, GPS coordinates from animal mobility experiments, etc. Even though there has been a lot of research effort recently towards generation of this trajectory data through video tracking, representation and analysis of this spatio-temporal data for modeling and recognition is still in its initial stages. In our previous works [1][2], we demonstrated the effectiveness of PCA-based representation for object motion-based indexing and retrieval tasks. Yacoob et al [7] 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. The high-dimensional trajectory using all the eight parameters of object motion is reduced using PCA. In [6], a semantic event detection technique for snooker videos is presented. Trajectory is generated by tracking the white ball using a color-based particle filter.

Gaussian mixtures have been used in speech modeling for speaker identification, accent classification and much more. In [8] the issue of speech and cross-talk detection in multi-channel audio is addressed. GMM-based classifier is used to classify the speech in different combinations of local speech and cross-talk for multi-speakers multi-microphones setting. Chen et al [3] propose an enhancement of speech/speaker recognition by modeling the speaker variability. They use PCA and independent component analysis (ICA) to extract the sources of dominant speaker variability. This variability is then modeled by GMM which results in superior performance as compared to those systems that don't take this variability in account. In [4], they address another dimension of the same problem, namely the effect of accent on speech/speaker identification. They use GMMs for Mandarin accent identification. They argue that GMM method can avoid building the models for phoneme or phoneme class, which is not economic for many applications. This paper presents a novel approach for model-based recognition of complex object trajectories using GMMs.

The rest of the paper is organized as follows: section 2 briefly describes our PCA-based trajectory representation scheme; GMM-based modeling of object motions is described in section 3; experimental setup along with results is presented in section 4, and section 5 rounds up with conclusions.

2. PCA-BASED SEGMENTED TRAJECTORY REPRESENTATION

This section provides a very brief overview of our trajectory representation scheme based on trajectory segmentation and PCA. We recognize that more often

0-7803-9332-5/05/$20.00 ©2005 IEEE

than not, full trajectory information is unavailable in video tracking applications due to occlusion, etc. This calls for the trajectory representation methods that can perform even in the case of partial trajectory information. We address this problem by segmenting the trajectories at the points of perceptual discontinuities. The other concern in trajectory modeling is its compact representation for efficient distance computation. For this purpose, we use the PCA-based representation of subtrajectories.

2.1. Trajectory Segmentation and Normalization

Our subtrajectory-based representation is motivated by motion perception in humans which is highly piece-wise based on atomic units of actions. The other substantial advantage of this trajectory segmentation approach is that it facilitates the modeling and recognition of trajectories which only have the partial trajectory information available. The discontinuities in the trajectory are detected with the help of velocity (1st derivative) and acceleration (2nd derivative). From the x- and y- projections of trajectory data, we compute the curvature which measures the sharpness of a bend in a 2-D curve and captures derivatives up to 2nd order. It is given by:

$$\kappa[k] = \frac{x'[k]y''[k] - y'[k]x''[k]}{[x'[k]^2 + y'[k]^2]^{3/2}} \quad (1)$$

Here $x'[k]$ refers to first derivative of x- projection of trajectory, and similar notation holds for other variables in equation. We perform a hypothesis testing-based process to locate the points of maximum change on curvature data. These inflection points are detected with the help of a likelihood ratio test-based approach. More details can be found in [1].

2.2. PCA-based Representation

We represent the subtrajectories using PCA because of its optimal energy compaction properties resulting from custom bases derived from the data. We concatenate the x- and y- data of each subtrajectory into one x-y vector for combined representation. All these similar vectors of trajectories from all the classes are then stacked to form one data matrix. The principal components of this data matrix are then estimated using Eigenspace decomposition of the estimated covariance matrix. To achieve dimensionality reduction, only the first M PCs are retained to form the transformation matrix Φ_M. The pool of subtrajectories is finally represented by their PCA coefficients using the transformation:

$$B = \Phi_M^T [A - Avg] \quad (2)$$

where A is the data matrix of subtrajectories, Avg is the vector containing mean of the data set and B is the matrix containing PCA coefficients of all the subtrajectories. The set of PCA coefficients of all the subtrajectories for each class are then used to train one GMM per class as explained in the next section.

3. GAUSSIAN MIXTURE MODELING

Given the PCA-based representation of subtrajectories for each class, we wish to model the underlying class probability distribution function (PDF) from training set data. The training set is made as diverse as possible so the recognition system learns all the possible variations in data. This diversity in training set causes the underlying PDF to be more complex. Hence as we make the recognition system more robust and tolerant to noise and variations, the statistical properties of class PDF's become more and more non-trivial to model. The class PDF $P(y)$ can be modeled to an arbitrary accuracy using mixture of Gaussians:

$$P(y|\Theta) = \sum_{i=1}^{N_c} \pi_i \mathbb{N}(y; \mu_i, \Sigma_i) \quad (3)$$

where $\mathbb{N}(y; \mu, \Sigma)$ is the M-dimensional Gaussian density with mean vector μ and covariance matrix Σ, and π_i are the mixing parameters of the Gaussian components, satisfying $\sum \pi_i = 1$. The mixture is completely specified by the parameter $\Theta = \{\pi_i, \mu_i, \Sigma_i\}_{i=1}^{N_c}$. Given a training set of subtrajectories $\{y_t\}_{t=1}^{N_T}$, represented by their M-dimensional PCA coefficients, the mixture parameters can be estimated using the ML principal:

$$\overset{*}{\Theta} = \text{argmax}\left[\prod_{t=1}^{N_T} P(y^t|\Theta)\right] \quad (4)$$

This estimation problem is best solved using the Expectation-Maximization algorithm which consists of the following two-step iterative process:

- E-Step:

$$h_i^k(t) = \frac{\pi_i^k \mathbb{N}(y^t; \mu_i^k, \Sigma_i^k)}{\sum_{j=1}^{N_c} \pi_j^k \mathbb{N}(y^t; \mu_j^k, \Sigma_i^k)} \quad (5)$$

- M-Step:

$$\pi_i^{k+1} = \frac{\sum_{t=1}^{N_T} h_i^k(t)}{\sum_{i=1}^{N_c}\sum_{t=1}^{N_T} h_i^k(t)} \quad (6)$$

$$\mu_i^{k+1} = \frac{\sum_{t=1}^{N_T} h_i^k(t) y^t}{\sum_{t=1}^{N_T} h_i^k(t)} \quad (7)$$

$$\Sigma_i^{k+1} = \frac{\sum_{t=1}^{N_T} h_i^k(t)(y^t - \mu_i^{k+1})(y^t - \mu_i^{k+1})^T}{\sum_{t=1}^{N_T} h_i^k(t)} \quad (8)$$

The EM algorithm is monotonically convergent in likelihood and is thus guaranteed to find a local maximum in the total likelihood of training data.

A major problem in the GMM-based modeling is the reliable estimation of 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 the trajectories for the class. The mixing weight of a mode π_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 high enough, the mode is estimated accurately. If π_i is too low, the mode is eliminated or merged with another. The weighted *skew* (3rd-order moment) and *kurtosis* (4th-order moment) for each mode are also monitored. If the sum of these values goes above a threshold for any mode, that mode is split in two. Figure 1 shows the 1-sigma contour of GMM's for all three classes in first two PCs.

Once the GMM's for all the classes have been trained, the classification of new trajectories can be performed by computing the log-likelihoods. For this purpose, the PCA coefficients vectors of input trajectory after segmentation are posed as observation sequence to each GMM. The trajectory is declared to belong to the class represented by GMM with the highest log-likelihood.

4. COMPARISON AND RESULTS

The experiments are carried out on two data sets. The first one is Australian Sign Language (ASL) data set obtained from UCI's KDD archive[1]. We use the x- and y-trajectories of signer's hands as they sign three different words. This data set has 207 total trajectories from three words signed by five professional signers. The hand locations are captured by the Power Glove sensor worn by the signers. The other data set (HJSL), donated by Columbia University's multimedia group, contains trajectories of athletes performing high jump and slalom skiing. This data set has 108 trajectories with 40 high jumps and 68 skiing trajectories.

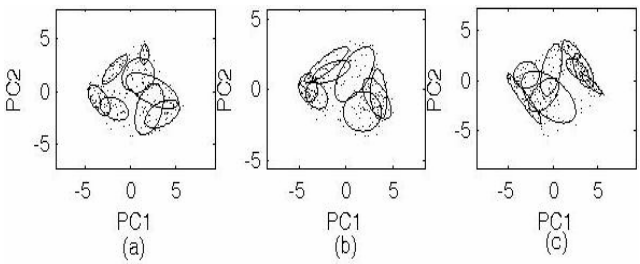
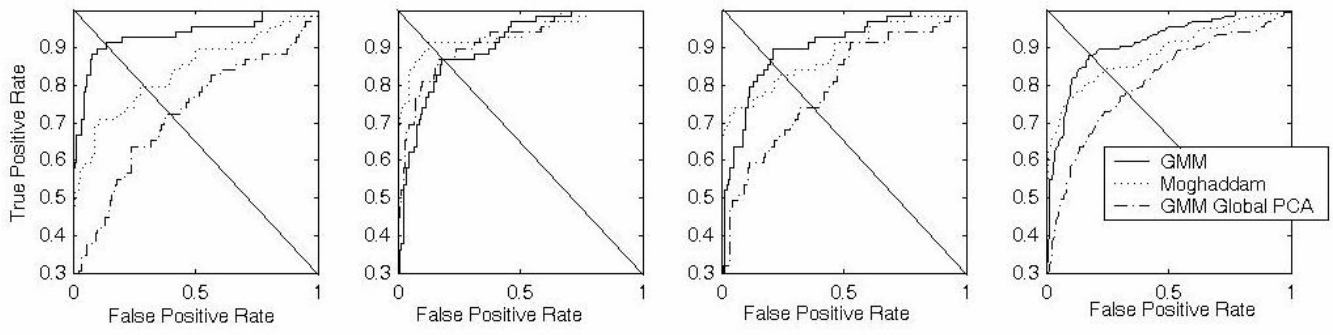


Figure 1: 1-Sigma contours of GMM's learnt from three classes. (a) 'Norway'. (b) 'Alive'. (c) 'Crazy'.

To establish a base case, we have implemented two different systems for comparison. One is the PCA-based Gaussian PDF estimation approach by Moghaddam et al [5] developed for face recognition. We improvise on their technique for trajectory recognition purposes. In this approach, the PCA is performed on the set of full trajectories, in a global sense, without segmentation. The resulting PCA coefficients are used to estimate the underlying PDF's for each class of trajectories. The other approach that we report for comparison is based on the same GMM based implementation, but it uses full trajectories instead of segmented subtrajectories. This experiment helps us test the validity of our trajectory segmentation based approach. For experiments, we divide the two data sets in terms of training- and test- sets in two configurations; I) both training- and test- sets have half the trajectories from each class; II) training set having half the trajectories and test set containing all the trajectories from each class. This results in four scenarios for two data sets labeled ASL I,II and HJSL I,II in table 1. We perform the ROC analysis on the three systems to measure their stability across the classes at varying thresholds of decision. This experiment is carried out in ASL II setting. The resulting ROC curves are shown in figure 2 which depicts the performance on all three classes as well as the 'average' curve depicting the average overall behavior of the classifier. Diagonal lines are superimposed on the curves to indicate the Equal Error Rate (EER) criterion. For the purpose of comparison in terms of ROC analysis, a classifier *A* is declared superior to another classifier *B*, if *A*'s ROC curve is above that of *B*'s, i.e. towards the upper left hand corner. An inspection of the ROC curves makes it clear that the GMM based system's performance is better than other classifiers considered. Only in class 2 (trajectory for word 'alive') does our GMM-based classifier show a dip in performance. But in other two classes and in the average case, the GMM classifier has better performance. Finally, we compare all the three classification systems in terms of accuracy. In this context, all the test set trajectories are posed to the classifiers at once and resulting labels retrieved. These labels are then matched against the ground truth and total 'false alarms' are counted in each case. The accuracy is computed as:

[1] http://kdd.ics.uci.edu/databases/auslan/auslan.html



(a) (b) (c) (d)

Figure2: ROC curves using ASL II settings for: (a) Class 1 'Norway'. (b) Class 2 'Alive'. (c) Class 3 'Crazy'. (d) Average.

$$\text{accuracy} = 1 - \frac{|\text{false alarms}|}{|\text{test set}|} \quad (9)$$

The results for these trajectory recognition experiments in terms of accuracy are reported in table 1.

These results show the superiority of our segmented PCA-based approach using GMM over the global PCA-based approach using Gaussian PDF estimation and a global PCA-based GMM method. This can be attributed to several factors. We segment the trajectories at perceptually significant points of change in curvature, and represent the resulting subtrajectories using the optimal representation of PCA. The set of subtrajectories are then represented by a mixture of Gaussians whose parameters including the number of modes are automatically learnt from the training set data. This highly reliable method of class PDF estimation results in a highly accurate motion recognition system. We are also experimenting with a Hidden Markov Model (HMM) based system for trajectory classification and we have obtained encouraging preliminary results. Due to space limitations those results will be reported elsewhere.

Method	ASL I	ASL II	HJSL I	HJSL II
GMM	85.29	92.75	79.63	89.81
Moghaddam [5]	86.27	93.24	38.88	45.37
GMM Global	69.61	73.91	62.96	63.89

Table 1: Object motion-based trajectory recognition results

5. CONCLUSIONS

In this paper, we have presented a GMM based novel trajectory modeling approach for object motion recognition. The trajectories are segmented using a hypothesis testing approach based on curvature. The resulting subtrajectories are used for eigenspace decomposition and represented by their resulting PCA coefficients. The training set trajectories for each class in terms of their PCA coefficients are then modeled by GMM's. The parameters of these GMM's including the number of modes are automatically learnt from training set data. The models are tested on two data sets; the non-visual Australian Sign Language data set and the video tracking-based sports video data set. Comparisons are reported with a face recognition-based approach in the literature and a global PCA-based GMM implementation. Recognition results for our system show a marked improvement in recognition, yielding accuracy rates of around 85+%.

6. REFERENCES

[1] Bashir F., Schonfeld D., Khokhar A., "Segmented trajectory based indexing and retrieval of video data", IEEE International Conference on Image Processing, ICIP 2003, Barcelona, Spain.

[2] Bashir F., Khokhar A., Schonfeld D., "A Hybrid System for Affine-Invariant Trajectory Retrieval", 6th ACM SIGMM Multimedia Information Retrieval workshop, MIR 2004.

[3] Chen T., Huang C., Chang C., Wang J., "On the Use of Gaussian Mixture Model for Speaker Variability Analysis", ICSLP'2002, Denver, USA. 2002.

[4] Chen T., Huang C., Chang C., Wang J., "Automatic Accent Identification using Gaussian Mixture Model," IEEE workshop on ASRU'2001, Italy, 2001.

[5] Moghaddam B., Pentland A., "Probabilistic Visual Learning for Object Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-19(7):696-710, July 1997.

[6] Rea N., Dahyot R., Kokaram A., "Semantic Event Detection in Sports through motion understanding", Proceedings of Conference on Image and Video Retrieval (CIVR) 2004, Dublin, Ireland, July 21-23, 2004.

[7] Yacoob Y., Black M. J., "Parameterized Modelling and Recognition of Activities", Computer Vision and Image Understanding, Vol. 73 (2), Feb. 1999. pp. 232 – 247.

[8] Wrigley S.N., Brown G.J., Wan V., Renals S., "Speech and Crosstalk Detection in Multichannel Audio", IEE Transactions on speech and audio processing, Vol. 13 (1), Jan. 2005.