

HMM-BASED MOTION RECOGNITION SYSTEM USING SEGMENTED PCA

Faisal Bashir¹, Wei Qu², Ashfaq Khokhar³, Dan Schonfeld⁴

University of Illinois at Chicago,
851 S. Morgan St., Chicago, IL, 60607.
{¹fbashir,²wqu,³ashfaq,⁴ds}@ece.uic.edu

ABSTRACT

In this paper, we propose a novel technique for model-based recognition of complex object motion trajectories using *Hidden Markov Models* (HMM). 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 grouped using *spectral clustering* to decide on the number of states for each HMM representing a class of object motion. The hidden states of the HMMs are represented by *Gaussian Mixtures* (GM's). This way the HMM topology as well as the parameters are automatically derived from the training data in a fully unsupervised sense. 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 90+% 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. Vlachos et al [6] use longest common subsequence (LCSS) to measure similarity between two trajectories. They note the drawbacks of Euclidean

distance which is sensitive to small variations in time axis, and dynamic time warping (DTW) which pairs all elements of the sequences. 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 [5], a semantic event detection technique for snooker videos is presented. Trajectory is generated by tracking the white ball using a color-based particle filter. This paper presents a novel approach for model-based recognition of complex object trajectories using HMMs.

The rest of the paper is organized as follows: section 2 briefly describes our PCA-based trajectory representation scheme; HMM-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. For additional details, we refer to [1],[2]. This segmented trajectory-based approach allows us to represent and recognize partial trajectories that may result due to occlusion or noisy information. We segment the object trajectories into atomic units of actions separated by perceptual discontinuities in trajectory data. 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. Before computing the curvature and detecting segmentation points, the input trajectory is low-pass filtered as well as normalized for shift invariance.

For representation of subtrajectories, we concatenate the x- and y- data of each subtrajectory into one x-y vector. 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 for compact representation, and first m PCs are used for transformation as:

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

where Φ_m is the m -dimensional PCA transformation matrix, 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.

3. HMM-BASED MODELING

HMMs are finite state stochastic machines that allow dynamic time warping for modeling time series data that satisfies the Markovian property. Simply put, the 1st order Markovian property assumes the independence of current state with all past ones given the immediately previous state. The object trajectory data is a stochastic process with temporal continuity, just like speech which has been successfully modeled using HMMs for decades. But trajectory data has one inherent problem that could fail HMM-based modeling, namely it does not necessarily satisfy the 1st order Markovian property. When an object starts making a particular move, it goes along with the flow for a while before changing direction, etc. So if individual successive trajectory points are used to model the states, the resulting Markov chain is bound to be non 1st order. This is where our segmented trajectory-based approach fits things nicely together. Our segmented trajectory-based representation scheme allows the trajectory data to be modeled exactly on the same lines as speech is modeled. In this context, we are interested in modeling a class of object motions (words) based on loose temporal ordering of subtrajectories (phonemes). Based on this modeling, one subtrajectory can be used to model the state of the HMM. Since the subtrajectories represent the segments of atomic motions between points of change in motion pattern, the resulting Markov chain is 1st order.

The first parameter of specification for an HMM is the number of states. We propose a spectral clustering-based algorithm that learns the number of states automatically from the given training data. Once the number of states is fixed, the complete set of model parameters describing HMM are given by the triplet:

$$\lambda = \{\pi_j, a_{ij}, b_j\} \quad (3)$$

where π_j is the probability of j^{th} subtrajectory being the first subtrajectory among all the trajectories, a_{ij} denotes the probability of j^{th} subtrajectory occurring right after the i^{th} subtrajectory, and b_j denotes the PDF of state j . We use Gaussian Mixture-based representation for state PDF. Once the separate HMMs are trained for all the classes, classification of new trajectories can be performed based on likelihoods compute for such trajectories when they're posed to the individual HMM's. The following subsections describe this process in detail.

3.1. Spectral Clustering using K-Means

The segmentation process results in several similar subtrajectories within one class of trajectories. It will not be wise to use all these subtrajectories as distinct states. So we cluster the subtrajectories represented by their PCA coefficients using spectral clustering. These states are then modeled using Gaussian Mixtures (GM) as explained in next subsection. The spectral clustering process starts by computing the *affinity matrix*, a symmetric matrix that contains the distance of each subtrajectory (based on its PCA coefficients) with every other subtrajectory. The eigenspace decomposition of this symmetric matrix is then used to estimate the optimal number of clusters and to group the individual subtrajectories into clusters. Given the set of n subtrajectories in R^p represented in a matrix form S , we look for the number of clusters k and put the n subtrajectories into k clusters. Our decision for the number of clusters k computes the cluster validity score proposed in [4]:

$$\alpha_k = \sum_{c=1}^k \frac{1}{N_c} \sum_{i,j \in Z_c} Y_{ij} \quad (4)$$

where Z_c denotes the cluster c , N_c the number of items in cluster c , and Y_{ij} the normalized eigenvector matrix. Once the eigenspace decomposition of affinity matrix has been performed, we form the eigenvector matrix X by stacking k largest PCs $X = [x_1, x_2, \dots, x_k] \in R^{n \times k}$. From this, the normalized eigenvector matrix Y is formed by renormalizing each of X 's rows to have unit length:

$$Y_{ij} = X_{ij} / (\sum_j X_{ij}^2)^{1/2} \quad (5)$$

Then we perform k-means clustering on the subtrajectory matrix S to form k clusters and compute α_k for this value of k . This process is repeated for $k \in \{1, \dots, K\}$ and finally the value of k that corresponds to maxima in α_k is chosen as the optimal number of states.

3.2. HMM Parameter Estimation

Once the set of training trajectories for a class are segmented and number of states are decided, the HMM's parameter triplet in equation (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 $q_t \in \{S_1, \dots, S_N\}$. Since the Markovian assumption is valid, 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_i , 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 Gaussian mixtures:

$$b_j(O_t) = \sum_{m=1}^M c_{jm} N(\mu_{jm}, \Sigma_{jm}),$$

$$= \sum_{m=1}^M c_{jm} \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(O - \mu_{jm})^T \Sigma_{jm}^{-1} (O - \mu_{jm})\right\} \quad (6)$$

Here each Gaussian component is a multivariate normal distribution of same dimensionality as the PCA coefficients representing the subtrajectories. The parameters of the HMM are initialized to random values and then Baum-Welch algorithm is used to re-estimate the parameters using forward-backward procedure. The above discussion relates to training sequence of subtrajectories resulting from one trajectory. Given a set of trajectories corresponding to each class, we extend the training to multiple training set trajectories. At each iteration of Baum-Welch re-estimation, the contribution from all the individual training set trajectories are summed up in the forward-backward estimation parameters. If the change in parameter values stays less than a prefixed threshold for 10 successive iterations, the algorithm is declared to converge. After one set of iterations for parameter estimation, we perform the annealing by expanding the covariance matrices of PDF estimates and by pushing the state transition matrix and prior state probabilities closer to 'uniform'. This has the effect of nudging the solution away from bad starting point in search of a global maximum.

Once the HMM'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 HMM. The trajectory is declared to belong to the class represented by HMM with the greatest 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 archive¹. 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.

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 [3] 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 LAMSTAR neural network. Here, the vector quantized PCA coefficients for individual subtrajectories are concatenated to form one training vector per training set trajectory. The weights of the neural network are then learnt based on these vectors. Due to space limitations, we skip the details of this neural network-based system. 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 first 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>

Method	ASL I	ASL II	HJSL I	HJSL II
HMM	91.18	95.65	81.48	90.74
Moghaddam [3]	86.27	93.24	38.88	45.37
Neural Network	74.86	84.57	71.33	81.2

Table 1: Object motion-based trajectory recognition results

$$\text{accuracy} = 1 - \frac{|\text{false alarms}|}{|\text{test set}|} \quad (7)$$

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 HMM over the global PCA-based approach using Gaussian PDF estimation and neural network-based approaches. 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 grouped together using spectral clustering to decide the number of states for HMM. Finally, the dynamic time warping properties of HMM's combined with Gaussian mixture-based state PDF representation result in a highly accurate motion recognition system.

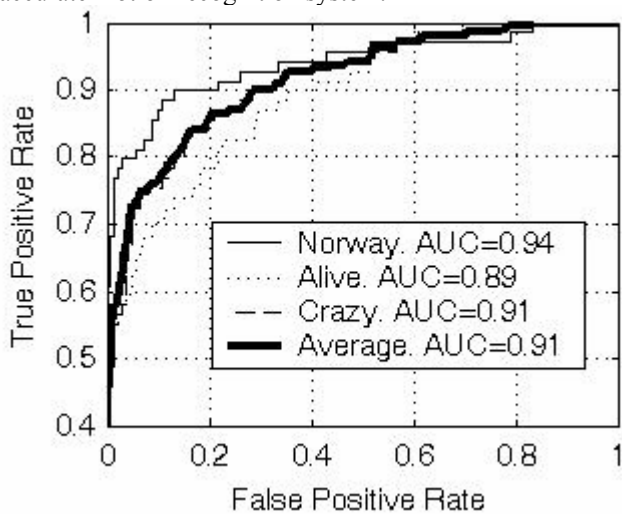


Figure 1: ROC curves for HMM based trajectory recognition on three classes as well as the average ROC curve.

After establishing the superior performance of our HMM based approach, we perform the ROC analysis on our system and report its stability across different classes at varying thresholds of decisions. This experiment is carried out in ASL II setting. The resulting ROC curves are shown in figure 1 which depicts the performance on all three classes as well as the ‘average’ curve depicting the average overall behavior of the classifier. A commonly

used metric for ROC analysis, namely the area under ROC curve is also computed and displayed for the three ROC curves as well as the average curve. The AUC results closely agree with our accuracy measures in table 1. These results establish the stability of our system for several classes at different operating characteristics.

5. CONCLUSIONS

In this paper, we have presented an HMM 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 grouped together using spectral clustering to automatically decide on the number of states. The training set trajectories for each class are then modeled by HMM's where the state conditional probability distributions are represented by Gaussian mixtures. 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 neural network-based implementation. Recognition results for our system show a marked improvement in recognition, yielding accuracy rates of around 90+%.

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] 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.
- [4] Porikli F., Haga T., “Event Detection by Eigenvector Decomposition using Object and Frame Features”, International Conference on Computer Vision and Pattern Recognition, CVPR 2004.
- [5] 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.
- [6] Vlachos M., Kollios G., Gunopulos D., “Discovering Similar Multidimensional Trajectories”, Proceedings of the 18th International Conference on Data Engineering (ICDE), 2002. pp. 673.
- [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.