Automatic motion learning in the presence of anomalies using coefficient feature space representation of trajectories

KHALID Shehzad[1]   NAFTEL Andrew[2]

Abstract   Techniques for understanding video object motion activity are becoming increasingly important with the widespread adoption of CCTV surveillance systems. Motion trajectories provide rich spatiotemporal information about an object's activity. This paper presents a novel technique for clustering of object trajectory-based video motion clips using basis function approximations. Motion cues can be extracted using a tracking algorithm on video streams from video cameras. In the proposed system, trajectories are treated as time series and modelled using orthogonal basis function representation. Various function approximations have been compared including least squares polynomial, Chebyshev polynomials, Piecewise Aggregate Approximation, Discrete Fourier Transform (DFT) and Modified DFT (DFT-MOD). A novel framework,namely Iterative Hierarchical Semi-Agglomerative ClusTering using Learning Vector Quantization (Iterative HSACT-LVQ), is proposed for learning of patterns in the presence of significant number of anomalies in training data. In this context, anomalies are defined as atypical behavior patterns that are not represented by sufficient samples in training data and are infrequently occurring or unusual. The proposed algorithm does not require any prior knowledge about the number of patterns hidden in unclassified dataset. Experiments using complex real life trajectory datasets demonstrate the superiority of our proposed Iterative HSACT-LVQ based motion learning technique compared with other recent approaches.

Key words   Object trajectory, dimensionality reduction, trajectory clustering, event mining, anomaly detection.

1   Introduction

Techniques for understanding video object motion activity are becoming increasingly important with the widespread adoption of CCTV surveillance systems. Content-based visual data management techniques are now urgently required for tasks such as motion data search and retrieval, discovery and grouping of similar motion patterns, detection of anomalous behaviour, motion understanding and prediction. These techniques are essential for the development of next generation 'actionable intelligence' surveillance systems.

Much of the earlier research focus in motion analysis has been on high-level object trajectory representation schemes that are able to produce compressed forms of motion data [1][2][3][4][5][6][7][8][9]. The literature on trajectory-based motion understanding and pattern discovery is less mature but advances using Learning Vector Quantization (LVQ)[10], Self-Organising Maps (SOMs) [11][12], Hidden Markov Models (HMMs) [13][14], and fuzzy neural networks [15] have all been reported. Most of these techniques attempt to learn high-level motion behaviour patterns from sample trajectories using discrete point sequences as input to a machine learning algorithm. For realistic motion sequences, convergence of these techniques is slow and the learning phase is usually carried out offline due to the high dimensionality of the input data space.

Related work within the data mining community on approximation schemes for indexing time series data is highly relevant to the parameterisation of object trajectories. Trajectories are defined as set of points representing the ordered observation of location of moving object taken at different points in time. Trajectories can therefore be represented as time series data which makes the indexing techniques for time series applicable to motion data. However, computer vision researchers have been slow to realize the potential of this work. For example, Discrete Fourier Transforms (DFT) [16], Discrete Wavelet Transforms (DWT) [17], Adaptive Piecewise Constant Approximations (APCA) [18], and Chebyshev polynomials [19] have been used to conduct similarity search in time series data.

Previous work on learning of motion patterns mainly took a supervised learning approach where complete information about normal motion patterns is available. Supervised approaches to motion learning are less useful in video surveillance applications since labelled training data in not usually available, or is impractical to obtain. Techniques are required that learn motion activity patterns in an unsupervised manner. Training data will often contain anomalies or outliers, that is unusual or infrequently occuring motion behaviour patterns. The learning algorithm must adapt to the presence of anomalies and also must be robust in the presence of noise and occlusion.

In this paper, we apply time series modeling of spatiotemporal data to the problem of unsupervised learning of motion activity patterns using object trajectories. Trajectories are modelled using function approximation techniques and motion patterns are learnt in the coefficient subspace. A novel Iterative HSACT-LVQ algorithm has been proposed for learning of motion patterns whilst filtering anomalous samples from training data.

The remainder of the paper is organized as follow. We review some relevant background material in section 2. In section 3 we present some function approximation approaches to trajectory representation. In section 4, an iterative learning algorithm has been proposed to learn patterns in the presence of anomalies. Experiments have been performed to show the effectiveness of proposed system for trajectory-based learning of motion patterns in the presence of anomalous motion samples. These experiments are reported in section 5. The paper concludes with a discussion and proposals for further work.

2   Background and related work

Motion trajectory descriptors are known to be useful candidates for video indexing and retrieval schemes. Previous work has sought to represent moving object trajectories through piecewise linear or quadratic interpolation functions [4][20], motion histograms [1] or discretised direction-based schemes [5][21][9]. Spatiotemporal representations using piecewise-defined polynomials were proposed by Hsu [6], although consistency in applying a trajectory-splitting scheme across query and searched trajectories can be prob-

---

1.Department of Computer Science and Engineering, Bahria University, Islamabad, 44000, Pakistan Email: shehzad_khalid@hotmail.com   2.Department of Computer Science, University of Manchester, Manchester, M601QD, UK

lematic. Affine and more general spatiotemporally invariant schemes for trajectory retrieval have also been presented [2][3][7]. The importance of selecting the most appropriate trajectory model and similarity search metric has received relatively scant attention [8]. In addition to polynomial models, a wide variety of basis functions have been used to approximate object trajectories [22][19][17].

It is surprising to find that many of these candidate time series indexing schemes have not yet been applied to the problem of motion data mining and trajectory clustering. Recent work has either used probabilistic models such as HMMs [23] or discrete point-based trajectory flow vectors (PBF) [11][15][10] as a means of learning patterns of motion activity. An agglomerative clustering algorithm based on the Longest Common Subsequence (LCSS) approach is proposed in [24][25] for grouping similar motion trajectories. The problem with PBF vector-encoded trajectory representation is the heavy computational burden making prospects for online learning of motion patterns remote.

Learning of patterns from trajectory data to extract high level information has gained interest quite recently. Earlier work rely upon labelled training data for model training [14][26][27]. There exists some work on learning from unclassified training data such as [28][24][29][10][12][30][31][32][25]. A number of eigenspace clustering techniques have been proposed recently [39][45]. However, these approaches normally require known number of clusters. Given an unclassified dataset, the number of motion classes are normally unknown. Some approaches, based on spectral clustering, attempts to approximate the number of clusters [38][37]. Affinity propagation-based approaches have also been proposed recently [42]. Affinity Propagation (AP) operates by exchanging messages between training data points until a high quality set of clusters (represented by identified exemplars) emerges. However, AP requires the specification of two important parameters: preference parameter and the damping factor. It is very hard to know the value of these parameters that will yield optimal clustering results. The solution to this problem is provided by Wang et al. [43]. They proposed an adaptive affinity propagation method for clustering to automatically select the preference parameter to identify the correct number of clusters and finding the optimal clustering solution. However, these approaches can not cater for the presence of anomalies in training data although it is very difficult to be sure of clean training data when the trajectory dataset in unlabelled.

The contribution of this paper is to present a novel mechanism for learning of motion patterns whilst filtering anomalous samples from training data. The proposed technique does not require any prior knowledge about the number and type of patterns hidden in datasets. Trajectory clustering is carried out in coefficient feature space that results in efficient discovery of patterns of similar object motion behaviour.

3 Feature space representation of trajectory

Without loss of generality, we consider the projection of a moving object O in the (x, y) image plane. O registers its location (x_i, y_i) in (x, y, t) space at each instant of time $t = t_i$. The object trajectory $T(O)$ is defined by the point sequence

$$T(O) = (x_1, y_1, t_1), (x_2, y_2, t_2), ..., (x_n, y_n, t_n) \quad (1)$$

where n is the sequence length. In tracking applications, observations are recorded at regular time intervals and hence we assume $t_i = i$ where i is the frame index. In the proposed work, the lower dimensional representation of trajectories is generated by splitting the trajectories into 1D time series in x and y space, represented as $X = x_i, Y = y_i$, $i = 1, ..., n$. We consider five alternative techniques to achieve dimensionality reduction of trajectories including Least Squares (LS) [33], Chebyshev (CS)[19], Piecewise Aggregate Approximation (PAA) [34], Discrete Fourier Transform (DFT) [33] and Modified DFT (DTF-MOD).

DFT-MOD is an extension of DFT. DFT-MOD is generated by augmenting the DFT coefficients-based feature vector with some extra information regarding the length and starting location of the trajectory. This important information is not modelled correctly by DFT as it selects only top few DFT coefficients which simply models the mean and trend of motion in the trajectory. DFT assumes signal to be periodic which causes poor approximation at the borders of a time series. All these factors may contribute to the fall-off in retrieval accuracies, using simple DFT based dimensionality reduction, where starting point and duration of motion are important features for distinguishing different trajectories. The derivation of DFT-MOD based feature space representation of trajectories, using DFT coefficients, is specified as follows:

The n-point DFT of $\{x_i\}$, defined as a sequence $\{X_f\}$ of n complex numbers ($f = 0, ..., n - 1$), is given by eq. (2). A similar expression can be defined for $\{y_i\}$ as given in eq. (3).

$$X_f = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} x_i exp(-j2\pi fi/n) \quad f = 0, 1, ..., n-1 \quad (2)$$

$$Y_f = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} y_i exp(-j2\pi fi/n) \quad f = 0, 1, ..., n-1 \quad (3)$$

where j is the imaginary unit $j = \sqrt{-1}$, and X_f, Y_f are complex numbers with the exception of X_0, Y_0 which are real. Typically, the DFT sequence is truncated after m terms, $f = 0, ..., m - 1$. More formally, let a_i and $\widehat{a_i}$ be the real and imaginary part of X and b_i and $\widehat{b_i}$ be the real and imaginary part of Y. Trajectories can be represented in the coefficient feature space by a $2(2m - 1)$ dimensional vector of DFT coefficients $\mathbf{F}_{DFT}$, where

$$\mathbf{F}_{DFT} = [a_0, a_1, \widehat{a_1}, ..., a_{m-1}, \widehat{a_{m-1}}, b_0, b_1, \widehat{b_1}, ..., b_{m-1}, \widehat{b_{m-1}}] \quad (4)$$

The DFT-MOD based feature space representation of trajectory is then represented as:

$$\mathbf{F}_{DFT-MOD} = [\omega n, \omega x_0, \omega y_0, \mathbf{F}_{DFT}] \quad (5)$$

where (x_0, y_0) is the starting location approximated by taking the mean of first 10 points of the trajectory, n is the length of trajectory and ω is the scaling factor. When padding the starting point and length information to the DFT coefficients, it is important that these information be scaled down so that they do not dominate the trend information captured by the DFT coefficients. In all of our experiments, we used default scaling factor of $\omega = 0.5$. Trajectories can now be represented in the coefficient feature space by a $2(2m - 1) + 3$ dimensional vector of DFT-MOD coefficients $\mathbf{F}_{DFT-MOD}$.

Comparative evaluation of different trajectory representation schemes is presented in section 5.

4 Learning of Motion Trajectories in Presence of Anomalies

Given the difficulties in motion learning from corrupted datasets, we now present a novel algorithm for learning patterns in the presence of anomalies in training data. The proposed clustering mechanism is a cooperative learning algorithm that combines Learning Vector Quantization (LVQ) with Hierarchical Semi-Agglomerative ClusTering (HSACT). Bearing in mind the nature of motion trajectory data and the drawbacks associated with its high dimensionality, learning of motion patterns is done in coefficient feature space representation of trajectory. The proposed unsupervised learning algorithm does not require any prior information about the number of patterns present in the unclassified dataset. No labelled training data is required for learning the distribution of each motion pattern in the trajectory data. In addition to this, the proposed algorithm does not require a huge set of training data. It can efficiently learn patterns with small training set with at least some samples from each unknown pattern.

4.1 Framework for Learning Motion Patterns

This section describes the framework to robustly discover motion patterns from corrupted and unclassified trajectory datasets. The proposed framework comprises of two distinct components namely LVQ component and HSACT component. The architecture chosen for the LVQ consists of a single layer of input neurons connected directly to a single 1-dimensional layer of output neurons. LVQ component is responsible for extracting fine groupings in trajectory dataset. HSACT component works on the fine clusters, extracted by LVQ component, to generate coarse clusters and, in the process, discovers the actual number of groupings in the trajectory dataset. In a pure agglomerative clustering mechanism [24][25], the initial number of clusters is equivalent to the number of instances in the data set. Each input data is assigned to a separate cluster making each cluster a singleton grouping. This approach is not scalable to large datasets. Therefore, in the proposed framework for motion learning, the training data is sub-sampled by identifying mutually disjunctive sub-clusters using the LVQ component. The initial number of clusters in the HSACT component is then equivalent to the number of sub-clusters generated by the LVQ component which is far less than the size of the training data. This results in making the "semi" agglomerative mechanism scalable to large datasets.

The proposed mechanism for learning of motion patterns from corrupted training dataset is an iterative approach and in each iteration, some anomalous trajectories are filtered from the training data. This results in reducing the adverse affects, caused by the presence of anomalies in training data, on learning of normal motion patterns. The process continues till no trajectory is identified and filtered as anomalous. Fig. 1 depicts the iterative nature of the proposed learning mechanism to cater for the presence of anomalies in training data. The curve at the top of Fig. 1, represents the distribution of training samples in a given state-space. High peaks in distribution curve represent training samples belonging to normal motion pattern where as low peaks represent samples which are not common and are anomalous.

4.2 Learning Algorithm

The learning algorithm, for unsupervised learning of motion patterns from corrupted training data, comprises of the following steps:

1. Initialize the LVQ network with greater number of output neurons than the number of clusters we want to identify in the motion trajectory data DB using:

$$\#_{output} = \begin{cases} \xi & if\,\xi < 100 \\ 100 & otherwise \end{cases} \quad (6)$$

where

$$\xi = size(DB)/4 \quad (7)$$

2. Estimate a single multivariate Gaussian (PDF) using the training data DB as:

$$PDF = \frac{1}{\sqrt{2\pi\Sigma}} exp\left[-\frac{(X-\mu)^2}{2\Sigma}\right] \quad (8)$$

where $X \in DB$, μ is the mean and Σ is the covariance estimate associated to DB. Generate $\#_{output}$ samples from the PDF $N(\mu, \Sigma)$ and use them to initialize the weight vectors associated with each of the output neurons.

3. Determine the winning output node k (indexed by c) such that the Euclidean distance between the current input vector B and the weight vector W_k is minimum:

$$c = arg\,min\,_k\|F - W_k(t)\| \qquad \forall k \quad (9)$$

4. Train LVQ network by adjusting the weight vector of winning output node. The weight vector of winning node c is updated using

$$W_c(t+1) = W_c(t) + \alpha(t)(F - W_c(t)) \quad (10)$$

where $\alpha(t)$ is the learning rate of LVQ and t is the training cycle index.

5. Decrease the learning rate $\alpha(t)$ exponentially over time using:

$$\alpha(t) = 1 - e^{t-t_{max}} \quad (11)$$

where t_{max} is the maximum number of training iterations.

6. Repeat steps 3-5 for all the training iterations.

7. Ignore output neurons with no training data associated to them.

8. Calculate Cluster Validity Index (CVI) to check the quality of current state of cluster. To ignore the effect of anomalies, CVI is calculated only for those clusters with significant memberships. Criteria for clusters to be included in calculation of CVI is specified as:

$$\mathbf{\Gamma}_{valid} = \{\Gamma_i \in \mathbf{\Gamma} \,|\, |\Gamma_i| \geq \kappa\} \quad \forall i \quad (12)$$

where $\mathbf{\Gamma}$ is the set of all clusters, $|\Gamma_i|$ is the number of training samples associated to cluster Γ_i and κ is the threshold constant. For the current number of valid clusters, the mathematical expression of CVI is given as:

$$CVI(k) = \left(\frac{1}{k} \times \frac{E_1}{E_k} \times D_k\right)^2 \quad (13)$$

$$E_k = \sum_{j=1}^{k} \sum_{X \in \Gamma_j \,\wedge\, \Gamma_j \in \mathbf{\Gamma}_{valid}} \|X - W_j\| \quad (14)$$

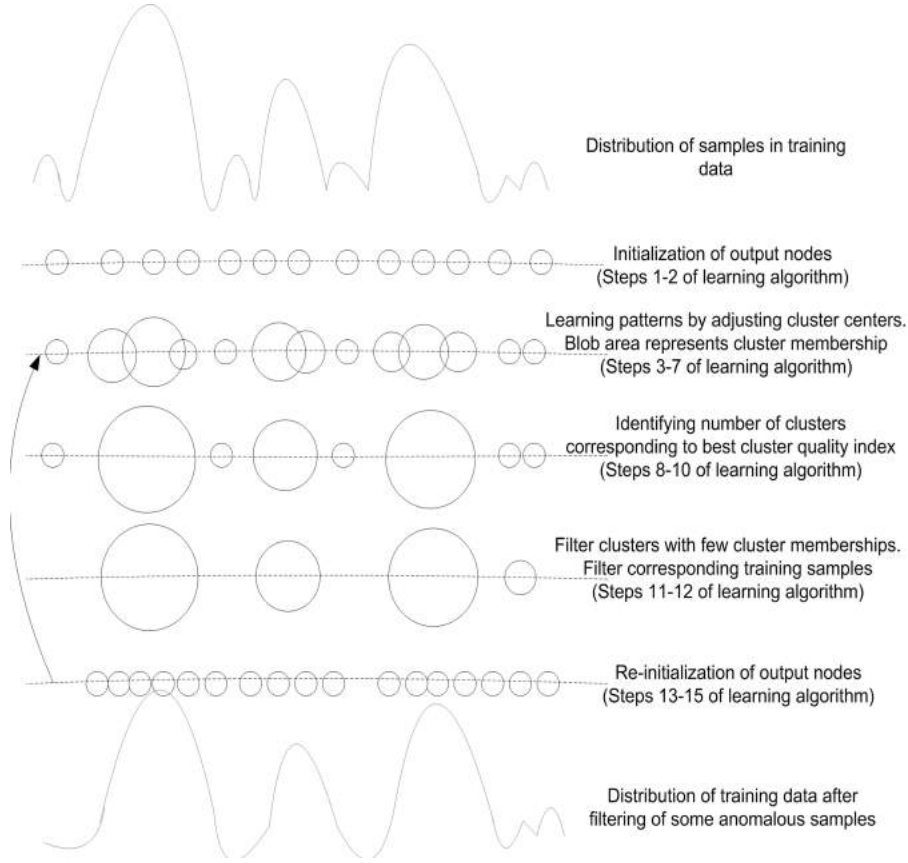$$D_k = max_{i,j=1}^{k}\|W_i - W_j\| \quad (15)$$

Figure 1: Depiction of single iteration in an iterative algorithm for learning patterns in presence of anomalies

where k represents number of clusters, X represents a sample training data associated to valid clusters and W_j represents weight vector associated to cluster Γ_j. In eq. (13), the factor $\frac{1}{k}$ will decrease CVI index as k is increased. On the other hand, $\frac{E_1}{E_k}$ increases CVI index as E_1 is a constant and E_k decreases with increase in k. The third factor D_k will increase with the value of k. These three factors tend to balance each other nicely. Values of k, resulting in higher values of CVI index, indicate better clustering.

9. Identify the closest pair of cluster (i, j) (indexed by (a, b)) given by the condition

$$(a, b) = arg\, min_{(i,j)} dist(i, j) \quad \forall i, j \wedge i \neq j \quad (16)$$

where

$$dist(i, j) = [(W_i - W_j)^T (W_i - W_j)]^{\frac{1}{2}} \quad (17)$$

After finding the most similar pair of clusters, the two clusters are merged into one using

$$W_{ab} = \frac{mW_a + nW_b}{m + n} \quad (18)$$

where m, n are the number of sample trajectories mapped to clusters a and b respectively.

10. Iterate through steps 8-9 till the number of clusters get equivalent to 1. Identify the number of clusters corresponding to highest CVI value.

11. Validate the stability of clustering process (convergence). This is done by identifying and filtering the clusters with fewer cluster memberships as:

$$\mathbf{\Gamma}_{anomal} = \{\Gamma_i \in \mathbf{\Gamma} ||\Gamma_i| < \kappa\} \quad \forall i \quad (19)$$

If no cluster has been identified as anomalous, the resulting clusters are considered to be stable without having any negative effect caused by the presence of anomalies in training data. On the other hand, if some clusters have been identified as anomalous, re-initialize the LVQ network as specified in eq. (6).

12. Identify and filter the anomalous trajectories present in the training data DB using:

$$DB_{filtered} = \{X \in DB | X \in \Gamma_i \wedge \Gamma_i \in (\mathbf{\Gamma} - \mathbf{\Gamma}_{anomal})\} \quad \forall i \quad (20)$$

13. Approximate Gaussian PDF of weight vectors, associated to valid clusters Γ_{valid} , as:

$$PDF_{valid} = \frac{1}{\sqrt{2\pi \Sigma_{valid}}} exp\left[-\frac{(W - \mu_{valid})^2}{2\Sigma_{valid}}\right] \quad (21)$$

where W is the weight vector, μ_{valid} is the mean and Σ_{valid} is the covariance estimate associated to the set of weight vectors representation of valid clusters.

14. Re-initialize the weight vectors associated to output neurons. Let $\#_{output}$ is the number of output neurons with which the network is initialized and $\#_{valid}$ is

the number of normal patterns identified in the previous learning iteration. The new network is initialized by using $\#_{valid}$ weight vectors identified in the previous learning iteration along with $(\#_{output} - \#_{valid})$ weight vectors obtained randomly from the PDF $N(\mu_{valid}, \Sigma_{valid})$ as approximated using eq. (21).

15. Go to step 3 for learning the patterns, using the training data $DB_{filtered}$, with anomalous trajectories filtered from it.

5   Experimental Results

This section analyzes the performance of proposed Iterative HSACT-LVQ based algorithm for unsupervised learning of motion patterns in the presence of anomalies in training data.

5.1   Experimental Datasets

Experiments are conducted on five different synthetic and real life motion trajectory datasets. These include CAV-CORR, CAV-FRNT, LAB, SIM_3 and SIM_5 datasets. The characteristics of these datasets are summarized in Table 1.

5.2   Experiment 1: Performance evaluation of dimensionality reduction techniques

The performance of five different trajectory representation schemes has been compared. The purpose of the experiment is to investigate the robustness of various dimensionality reduction mechanisms to the real life problem of noise and occlusion in motion data. Experiments have been performed using CAV-FRNT dataset.

The dataset provides ground truth (i.e. manually labelled) trajectory data and so we try to simulate the effects of noise and occlusion. The dataset is corrupted with noise by adding gaussian noise, at levels corresponding to the standard deviation σ, to every point in the original trajectory dataset. We set $\sigma = \{1.0, 2.0, 3.0\}$ to simulate different noise levels. Feature space vectors of trajectories in S and S_C are generated separately using LS, CS, DFT, DFT-MOD and PAA. Each corrupted trajectory in S_C is then selected as an example query Q_C and we search for its closest match Q in the original dataset S. This is defined by $argmin_{Q \in S} ED(Q_C, Q)$. A set of rankings $\forall Q_C \in S_C$ is produced. In the absence of noise, the closest match to Q_C should be its corresponding uncorrupted version in S which produces a rank value of unity. For ease of comparison we record the proportion of times (as a percentage) the query trajectory is ranked correctly as unity when taken over all S_C. This is repeated for different number of coefficients in LS, CS, PAA, DFT and DFT-MOD and for various values of σ. The results are summarised in Fig. 2.

For small amounts of noise ($\sigma < 2.0$), the choice of dimensionality reduction (DR) scheme or number of coefficients does not appear to be too critical, except that there is a fall-off in the performance of LS with increasing feature vector size. For higher noise levels (i.e. $\sigma >= 2.0$), DFT-MOD and DFT perform better than the other representation schemes, followed by PAA. CS does not give good retrieval accuracies and its performance degrades with increasing noise levels as compared to DFT-MOD and DFT. This may be explained by the fact that Euclidean distance defined over Fourier coefficients is more noise resistant in the frequency domain. DFT-MOD further caters for the disadvantage of DFT that it assumes signal to be periodic which cause poor approximation at the boundaries of x and y time series. Appending the scaled information about the

starting point and length further strengthens the feature space representation of trajectories using the DFT coefficients.

The retrieval experiments are now repeated but this time the trajectory dataset is corrupted with occlusion. The occlusion is simulated by replacing the occluded subsequence with the predicted location of object during occlusion, based on the motion trend of the object before the occlusion. Let $(\Delta x, \Delta y)$ be the difference between two points immediately before the occluded sub-sequence. Assume (x_p, y_p) is the first point in the occluding sequence and k is the subsequence length. Then

$$(\hat{x_i}, \hat{y_i}) = (x_{i-1} + \Delta x, \ y_{i-1} + \Delta y) , \quad i = p, ..., p + k \quad (22)$$

where

$$\Delta x = x_{p-1} - x_{p-2} \quad (23)$$

and

$$\Delta y = y_{p-1} - y_{p-2} \quad (24)$$

The occluded subsequence is then replaced with the estimated vector in eq. (22). Average retrieval accuracies, obtained using different types of representation schemes with varying numbers of function coefficients and different levels of occlusion are shown in Fig. 3. The retrieval accuracies were averaged over 10 random subsequence removals to reduce bias due to choice of starting position of occluding subsequence. The results from Fig. 3 show that CS and DFT-MOD gives good retrieval accuracies. PAA, in contrast to its performance in the presence of noise, does not give good retrieval accuracies in the presence of occlusion. The results demonstrated in this section have shown that DFT-MOD is robust to different levels of noise and occlusion in the trajectory data. DFT-MOD has, therefore, been selected as dimensionality reduction mechanism to have a feature space representation of trajectory data.

5.3   Experiment 2: Learning Motion Patterns in Presence of Anomalies

The purpose of this experiment is to demonstrate the effectiveness of Iterative HSACT-LVQ algorithm for learning of patterns while catering for the presence of anomalies in unclassified training data. The experiment has been conducted on real life CAV-CORR and LAB motion datasets, as shown in Fig. 4, that contain anomalous trajectories within the datasets themselves. CAV-CORR dataset is an unclassified dataset that contains 126 trajectories with no information about the number of patterns and the class labels associated to individual trajectories. CAV-CORR dataset is also suffering from the problem of perspective effects due to the presence of depth in CAV-CORR scene. To avoid this problem, we transform the observation obtained from the camera into the form which is closer to the actual occurrence. We use ground plane homography to map image coordinates to ground plane for CAV-CORR dataset. The calibration is available from [35]. The homography matrix is estimated using the method outlined in [36]. On the other hand, LAB dataset is a classified dataset generated in the laboratory controlled environment for testing purposes. There are 152 trajectories in the LAB dataset of which 140 trajectories are normal and 12 trajectories are abnormal. The normal trajectories are classified into 4 classes. The number of trajectories for class 0, class 1, class 2 and class 3 are 28, 28, 26 and 58 respectively.

Trajectories are modelled using DFT-MOD based coefficient feature vectors. Training samples from the dataset are passed through HSACT-LVQ network one by one and

Dataset	Description	# of trajectories	Extraction method	Labelled (Y/N)
SIM_3/ SIM_5	Simulated datasets comprising of two dimensional coordinates generated from Gaussian distributions to form 3 or 5 clusters.	arbitrary	Simulation.	Y
CAV-FRNT	A manually annotated video sequences of moving people from corridor view in a shopping centre. Object tracking coordinates are generated using interactive program and stored in XML files.	126	Parsing XML files containing motion coordinates.	N
CAV-CORR	A manually annotated video sequences of moving people from corridor view in a shopping centre. Object tracking coordinates are generated using interactive program and stored in XML files.	126	Parsing XML files containing motion coordinates.	N
LAB	Realistic dataset generated in the laboratory controlled environment for testing purposes. Trajectories can be categorized into 4 classes.	152	Tracking moving object and storing motion coordinates.	Y

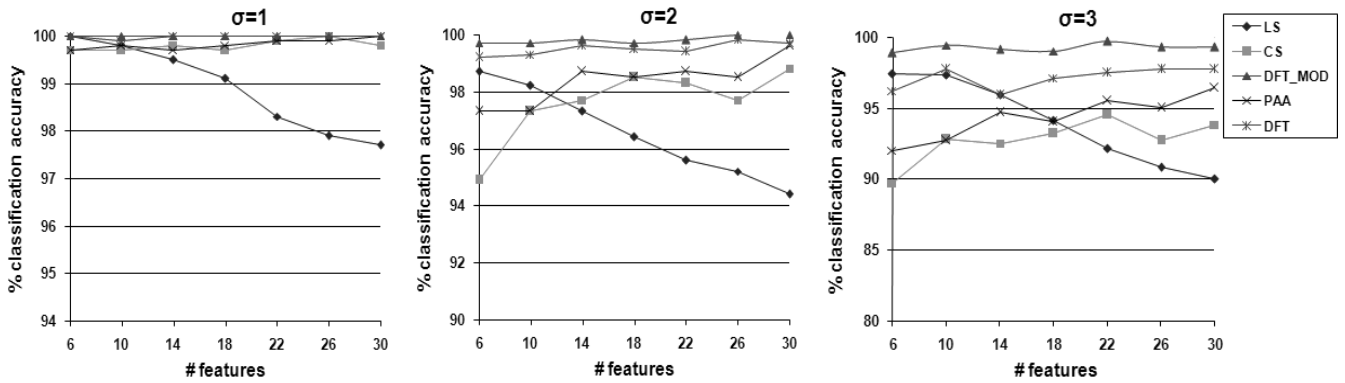Table 1: Overview of datasets used for experimental evaluation

Figure 2: Effect of different levels of noise on trajectory retrieval accuracy using CAV-FRNT dataset
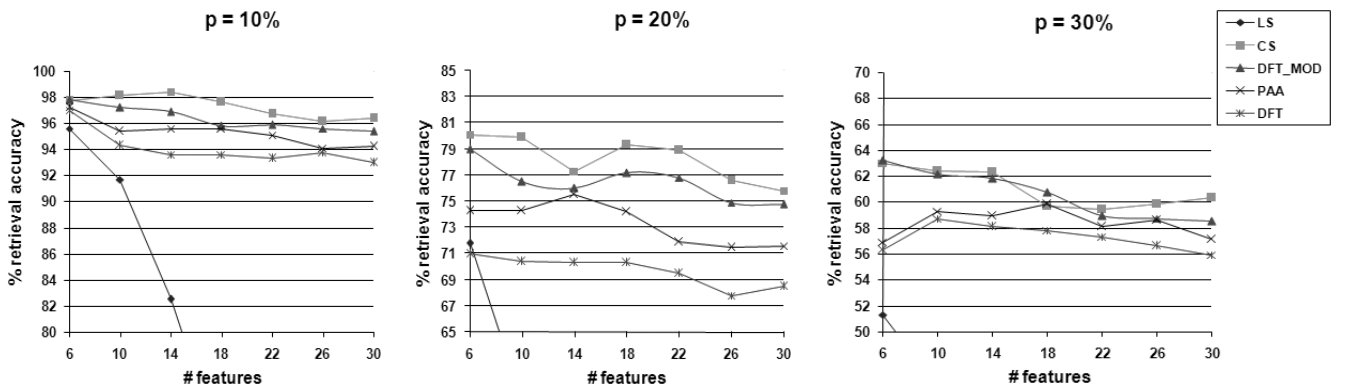

Figure 3: Effect of different lengths of occluding sub-sequences on trajectory retrieval accuracy using CAV-FRNT dataset

the network is trained for $t_{max} = 10000$ number of iterations. The clusters with fewer cluster memberships are identified as anomalous and are filtered. To ignore the effect of anomalous trajectories on the quality of clustering, CVI is calculated only for normal clusters. We assume

$\kappa = 0.05 \times |DB|$ in eq. (19) where $|DB|$ is the total number of samples in training data. If some clusters are identified as anomalous, training samples associated to such clusters are removed from training dataset. The learning process is repeated again, but now using the filtered training data.

This process continues till clustering process gets stable and no cluster is identified as anomalous.
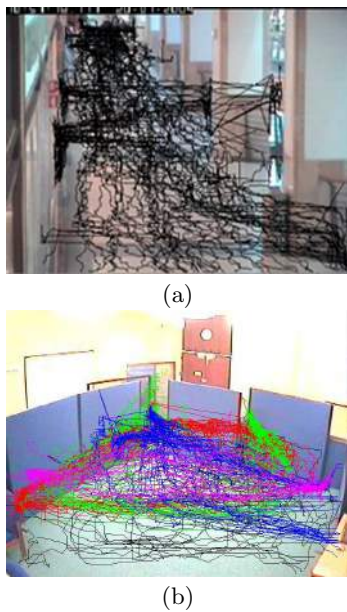

(a)

(b)

Figure 4: Background scene with overlaid trajectories from (a) CAV-CORR dataset (b) LAB dataset

The clustering results obtained, by applying the Iterative HSACT-LVQ methodology on CAV-CORR dataset, are shown in Fig. 5. The red trajectory in each class represents the trajectory that is closest to the class mean and the green points represent the starting point of each trajectory. As CAV-CORR dataset is an unclassified dataset with no ground truth, we can simply evaluate the quality of clustering obtained using Iterative HSACT-LVQ algorithm by having a visual inspection. Qualitatively, similar motion trajectory patterns appear to have been grouped together quite successfully. Trajectories that are distant from all the identified normal patterns and are filtered out as anomalous during learning process are shown in Fig. 6. 10 trajectories are filtered as anomalous and each anomalous trajectory is represented by separate colour. It is clear from Fig. 6 that anomalous trajectories are dissimilar from normal motion patterns as shown in Fig. 5. The experiment is also repeated on LAB dataset and the clustering results obtained are shown in Fig. 7. The green points represent the starting point of each trajectory. Trajectories, identified as anomalous using Iterative HSACT-LVQ methodology, are shown in Fig. 8. LAB dataset is a classified dataset and matching of clustering results with the ground truth showed that all the trajectories are clustered correctly giving 100% accuracy. All the 12 abnormal trajectories, as specified in the ground truth, are correctly identified and filtered as anomalous. Clustering results on CAV-CORR and LAB dataset provide support to the claim that Iterative HSACT-LVQ is an effective approach for learning patterns in complex real life motion datasets that are not free of anomalies.

5.4 Comparison of Iterative HSACT-LVQ with Competitive Techniques

The purpose of this experiment is to compare the performance of proposed Iteratvie HSACT-LVQ algorithm with the adaptation of spectral clustering [37][26][38] and adaptive affinity propagation (AAP) [43]. The matlab code for imple-
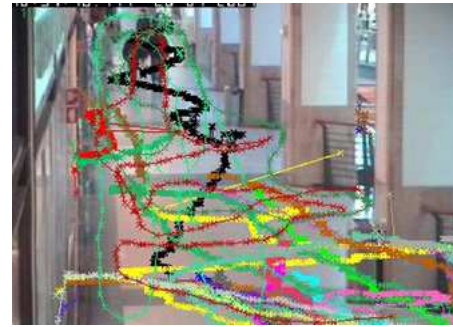

Figure 6: Trajectories, identified as anomalous, during learning of patterns by applying Iterative HSACT-LVQ algorithm on CAV-CORR dataset
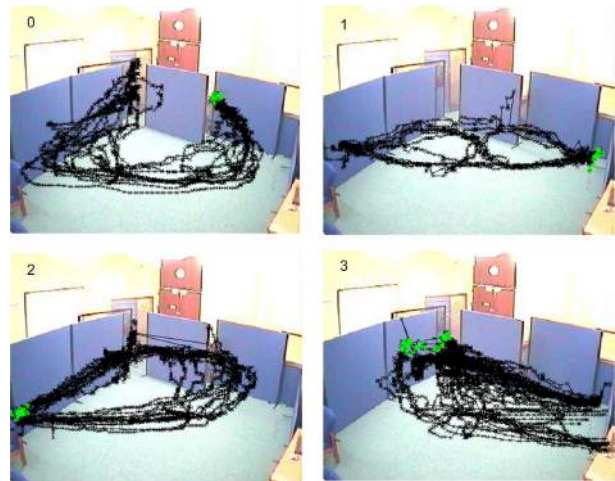

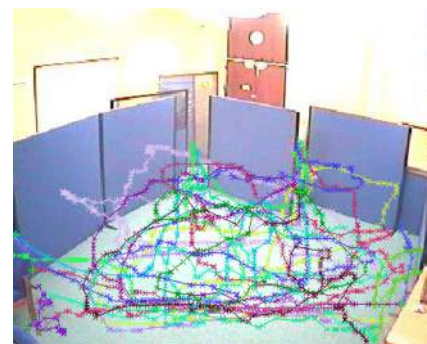Figure 7: Classification of test trajectories from LAB dataset

Figure 8: Trajectories, identified as anomalous, during learning of patterns by applying Iterative HSACT-LVQ algorithm on LAB dataset
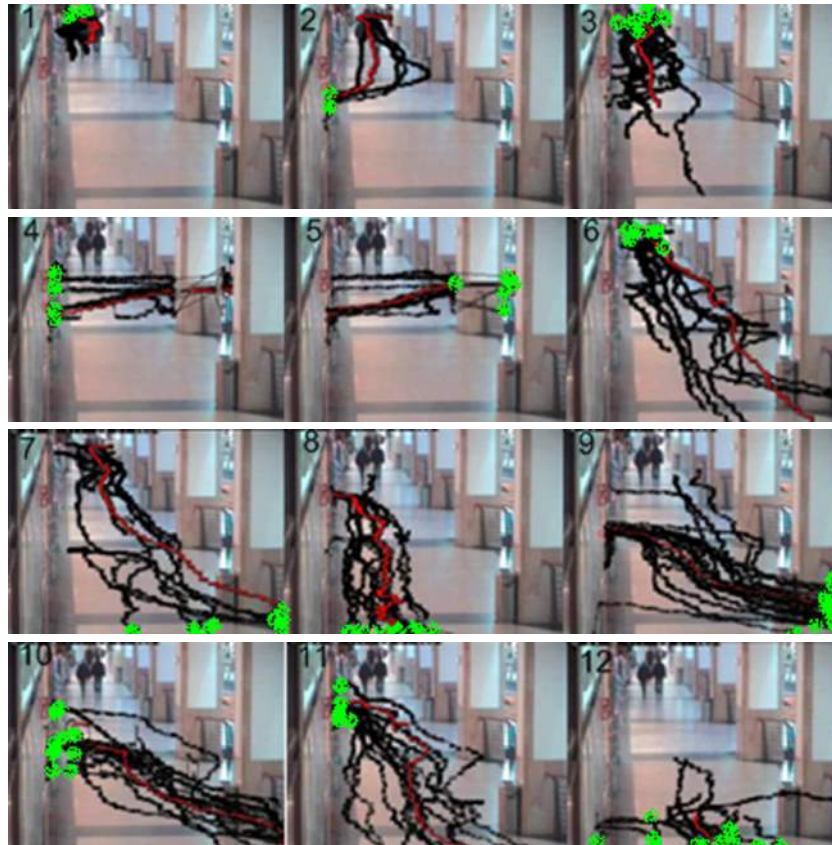
Figure 5: Motion trajectory clustering of CAV-CORR dataset after applying homography

mentation of adaptive affinity propagation is obtained from [44]. Comparative evaluation is provided in terms of detection of correct number of clusters, quality of clustering, robustness to the presence of different number of anomalous samples in training data and response time.

The experiment has been conducted on simulated SIM_3 and SIM_5 datasets. Learning of patterns is conducted separately using Iterative HSACT-LVQ, AAP and spectral clustering, and two cluster validity indices are employed to evaluate the quality of clustering results. These include $Dunn$'s index [40] and Calinski-Harabasz (CH) index [41].

The experiment is performed on clean (without anomalies) as well as corrupted (with anomalies) training data to investigate the effect of presence of anomalies on the performance of learning algorithms. The number of clusters identified by different clustering algorithms along with the quality of clustering using clean training data, as indicated by three different cluster validity indices, are presented in Table 2. Similar results using corrupted training data are presented in Table 3. Higher values for $Dunn$ and CH index indicate better clustering.

As evidenced from Table 2 and 3, Iterative HSACT-LVQ algorithm performs consistently better than AAP and spectral clustering for all the datasets in the presence of clean and corrupted training data. Comparison of results from Table 2 and 3 show good consistency in the performance of Iterative HSACT-LVQ in the presence of clean and corrupted training data. However, the performance of AAP and spectral clustering degrades significantly in the presence of anomalies in training data. Moreover, the number
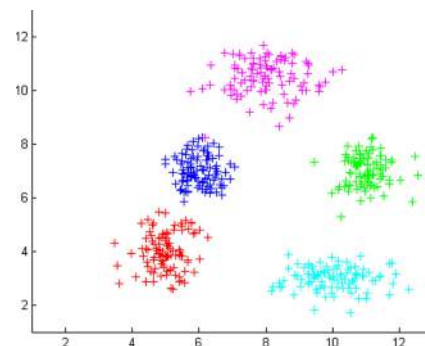

Figure 9: SIM_5 dataset sampled from five Gaussians

of clusters identified using our proposed approach is consistent with the number of groupings hidden in the dataset. On the other hand, AAP and spectral clustering is not able to identify the correct number of clusters. This is verified by matching the identified number of clusters with the actual number of groupings hidden in classified SIM_3 and SIM_5 datasets.

Effectiveness of Iterative HSACT-LVQ algorithm, as compared to competitive clustering algorithms, is now demonstrated graphically for SIM_5 dataset. The training data from SIM_5 dataset is shown in Fig. 9. Gaussian parameters used to generate each of the clusters in Fig. 9 are presented in Table 4. The training data is obtained by generating 70 samples from each of the Gaussian distribu-
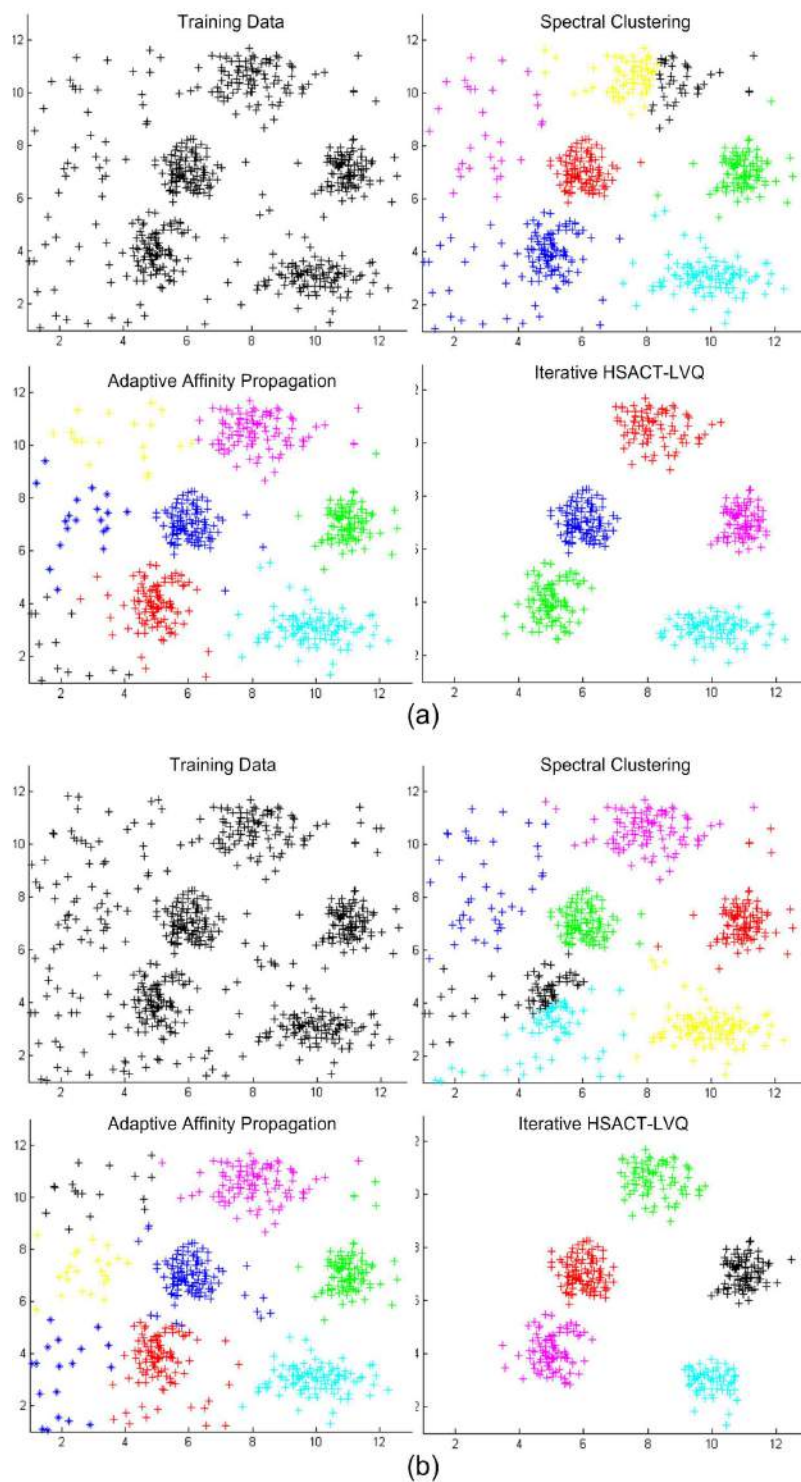
Figure 10: Learning of patterns from SIM$_5$ dataset using Iterative HSACT-LVQ, AAP and spectral clustering in the presence of different number of anomalies (#*anomalies*) in training data (a) #*anomalies* = 60 (b) #*anomalies* = 90

Datasets		HSACT-LVQ			Spectral			AAP		
	Ground Truth	# of clusters	CH	Dunn	# of clusters	CH	Dunn	# of clusters	CH	Dunn
SIM$_3$	3	3	1667.3	3.17	2	1173.7	2.98	2	1278.1	3.19
SIM$_5$	5	5	1891.3	2.78	4/5	1463.4	1.85	4	1254.6	2.74

Table 2: Comparison of Iterative HSACT-LVQ, AAP and spectral clustering based on the number and quality of clusters using clean training data from simulated datasets

Datasets		HSACT-LVQ			Spectral			AAP		
	Ground Truth	# of clusters	CH	Dunn	# of clusters	CH	Dunn	# of clusters	CH	Dunn
SIM$_3$	3	3	1516.7	2.91	6/7	823.9	1.16	3	1092.9	2.11
SIM$_5$	5	5	1974.5	2.97	7	1105.6	1.27	8	1165.7	1.89

Table 3: Comparison of Iterative HSACT-LVQ, AAP and spectral clustering based on the number and quality of clusters using corrupted training data (with anomalies) from simulated datasets

tion. The anomalies are induced in SIM$_5$ dataset by generating different number of data points from a uniform distribution such that $(x, y) \in (U(1, 12), U(1, 12))$. The data points lying within 3 standard deviations of normal clusters are then removed from the set of anomalous data points.

Cluster Colour	Mean	Covariance
Blue	(6,7)	$\begin{pmatrix} 0.2 & 0 \\ 0 & 0.3 \end{pmatrix}$
Green	(11,7)	$\begin{pmatrix} 0.2 & 0 \\ 0 & 0.3 \end{pmatrix}$
Red	(5,5)	$\begin{pmatrix} 0.3 & 0 \\ 0 & 0.4 \end{pmatrix}$
Magenta	(8,10)	$\begin{pmatrix} 0.7 & 0 \\ 0 & 0.4 \end{pmatrix}$
Cyan	(10,3)	$\begin{pmatrix} 0.6 & 0 \\ 0 & 0.2 \end{pmatrix}$

Table 4: Gaussian parameters used to generate 5 clusters

Results of learning patterns using Iterative HSACT-LVQ, AAP and spectral clsutering is demonstrated graphically in Fig. 10. Fig. 10(a) presents learning result for SIM$_5$ dataset in the presence of 60 anomalies in training data and Fig. 10(b) presents learning results in the presence of 120 anomalies. Data points belonging to same class are represented with similar colour and marker to ease the visualization of learned clusters. Comparing clustering results with the ground truth for SIM$_5$ dataset shows that Iterative HSACT-LVQ identifies the right number of clusters even in the presence of significant number of anomalies. The overall distribution of normal clusters learned using proposed algorithm, remains unaffected by the presence of anomalies in training data. On the other hand, quality of AAP and spectral clustering is significantly affected by the presence of anomalies. The proposed technique is also robust to the presence of higher number of anomalous samples in training data.

Comparison of different clustering algorithms is now provided by investigating the scalability of these algorithms to the number of training data samples and the number of options from which to identify the correct number of clusters

present in the dataset. We have implemented these algorithms using MATLAB 7 and running times are noted on an Intel Pentium IV 1.73 GHz machine with 504 MB of RAM. Experiment has been conducted on SIM$_5$ dataset and the response time of clustering algorithms, for different number of candidate clusters, are presented in Table 5. It is evident from the results in Table 5 that our proposed approach is scalable to number of options from which to select the right number of patterns hidden in the dataset. This is one of the important advantages of incorporating HSACT component with LVQ based learning that the proposed clustering algorithm takes same amount of time for any number of cluster options. On the other hand, spectral clustering exhibits that as the number of cluster options increases, there is an increase in time complexity. Adaptive affinity propagation exhibits consistent time complexity for different number of cluster options but its response time is significantly higher as compared to Iterative HSACT-LVQ. Comparison of different clustering algorithm based on the response time for different number of samples in training dataset is presented in Table 6. It is evident from the results in Table 6 that our proposed approach is scalable to the increasing number of samples in training dataset. On the other hand, spectral clustering exhibits that as the number of training samples increases, there is an increase in time complexity.

6 Discussion and conclusions

In this paper, we have presented a detailed discussion on unsupervised learning of patterns with anomalies present in the training data. A novel Iterative HSACT-LVQ algorithm has been proposed for learning of motion patterns whilst filtering anomalous samples from training data. The proposed methodology automatically determines the number of hidden patterns by evaluating the quality of clustering with different number of patterns and selecting the one which gives best result. Trajectory clustering is carried out in coefficient feature space. We have compared the performance of different dimensionality reduction techniques. DFT-MOD has been selected as dimensionality reduction mechanism as it gives overall best results in real time situation where motion data is susceptible to high level of noise and occlusion. Mapping trajectories from point sequence vectors to DFT-MOD coefficient feature space improves learning efficiencies.

# of cluster options	Response Time of Iterative HSACT-LVQ (sec.)	Response Time of Spectral (sec.)	Response Time of Adpative Affinity Propagation (sec.)
2	11.74	6.91	273.45
4	11.62	11.93	273.45
6	11.61	15.70	275.12
8	11.32	21.31	275.50
10	11.31	28.85	271.37
12	11.46	36.51	273.72
14	11.10	45.98	274.27
16	11.57	52.43	269.97
18	11.27	59.64	274.91
20	11.67	74.36	275.15

Table 5: Comparison of clustering algorithms based on the response time for different number of cluster options

# of of Training Samples	Response Time of Iterative HSACT-LVQ (sec.)	Response Time of Spectral (sec.)	Response Time of AAP (sec.)
300	9.54	63.91	117.81
500	9.61	130.37	469.92
900	9.71	217.23	1706.12

Table 6: Comparison of different clustering algorithm based on the response time for different number of samples in training dataset.

Experimental results are presented to show that Iterative HSACT-LVQ based learning mechanism gives better clustering results than competitive techniques such as spectral clustering. Learning of patterns, using proposed approach, is unaffected by anomalies in training data until the presence of very high number of anomalies results in development of groupings within anomalous data, thus resulting in identification of ghost clusters. The approach is also scalable to number of options from which to select the right number of patterns hidden in the dataset. Our techniques have been validated using variety of synthetic and real-life datasets.

By definition, trajectories are low level motion features as they simply represent the set of positions through which the object has moved. Trajectory-based motion understanding is therefore too limited for representing more complex behaviours to describe the overall event occurring in the scene. It is a requirement for a practical system to identify and model complex events and multiple objects interactions such as people exchanging objects in the scene, group of people fighting, a person leaving suspicious object in the scene. In further work, we intend to focus on identifying linkages between different motion patterns and try to learn the behaviour presented by different motion trajectories as a group.

References

1 Z. Aghbari, K. Kaneko, A. Makinouchi, Content-trajectory approach for searching video databases, IEEE Transanction on Multimedia, December 2003, **5**(4): 516-531.

2 F.I. Bashir, A.A. Khokhar, D. Schonfield, Segmented Trajectory based Indexing and Retrieval of Video Data, IEEE International Conference on Image Processing, Barcelona, Spain, September 2003, **3**: II-623-6.

3 F.I. Bashir, A.A. Khokhar, D.Schonfield, A hybrid system for affine-invariant trajectory retrieval, Proc. MIR'04, 2004, 235-242.

4 S.F. Chang, W. Chen, J.M. Horace, H. Sundaram, D. Zhong, A Fully Automated Content based Video Search Engine Supporting Spatiotemporal Queries, IEEE Transactions on Circuits and System for Video Technology, September 1998, **8**(5): 602-615.

5 S. Dagtas, W. Ali-Khatib, A. Ghafor, R.L. Kashyap, Models for motion-based video indexing and retrieval, IEEE Transactions on Image Processing, 2000, **9**(1): 88-101.

6 C.T. Hsu, S.J. Teng, Motion trajectory based video indexing and retrieval, IEEE International Conference on Image Processing, 2002, **1**: 605-608.

7 Y. Jin, F. Mokhtarian, Efficient video retrieval by motion trajectory, Proceedings of British Machine Vision Conference, Kingston, September 2004, 667-676.

8 S. Khalid, A. Naftel, Evaluation of matching metrics for trajectory based indexing and retrieval of video clips, Proceedings of IEEE WACV, Colorado, USA, January 2005, 242-249.

9 C. Shim, J. Chang, Trajectory based video retrieval for multimedia information systems, Proceedings of ADVIS, 2004, 372-382.

10 N. Johnson, D. Hogg, Learning the distribution of object trajectories for event recognition, Proceedings of British Conference on Machine Vision, 1995, 582-592.

11 W. Hu, X. Xiao, D. Xie, T. Tan, S. Maybank, Traffic accident prediction using 3-D model based vehicle tracking, IEEE Transactions on Vehicular Tech, May 2004, **53**(3): 677-694.

12 J. Owens, A. Hunter, Application of the Self-Organising Map for Trajectory Classification, Proceedings of Third IEEE International Workshop on Visual Surveillance, Dublin, Ireland, July 2000, pp. 77.

13 F.I. Bashir, A.A. Khokhar, D.Schonfeld, View-invariant motion trajectory based activity classification and recognition, ACM Multimedia Systems, special issue on Machine Learning Approaches to Multimedia Information retrieval, 2006, 45-54.

14 F.I. Bashir, A.A. Khokhar, D.Schonfeld, HMM based Motion Recognition System using Segmented PCA, IEEE International Conference on Image Processing, Genova, Italy, Sept. 11-14, 2005, 1288-1291.

15 W. Hu, D. Xie, T. Tan, S. Maybank, Learning activity patterns using fuzzy self-organizing neural networks, IEEE Transactions on Systems, Man & Cybernetic, June 2004, **34**(3): 1618-1626.

16 C. Faloutsos, M. Ranganathan, Y. Manolopoulos, Fast Sub-sequence Matching in Time-Series Databases, Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data, 1994, 419-429.

17 K. Chan, A. Fu, Efficient time series matching by wavelets, Proc. of International Conference on Data Engineering, Sydney, March 1999, 126-133.

18 E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrota, Locally adaptive dimensionality reduction for indexing large time series databases, Proc. ACM SIGMOD Conference, 2001, 151-162.

19 Y. Cai, R. Ng, Indexing Spatio-Temporal Trajectories with Chebyshev Polynomials, ACM SIGMOD/PODS Conference, France, June 13-18, 2004, 599-610.

20 S. Jeanin, A. Divakaran, MPEG-7 visual motion descriptors, IEEE Trans. Circuits Syst. Video Technology, June 2001, **11**(6): 720-724.

21 C. Shim, J. Chang, Content based retrieval using trajectories of moving objects in video databases, Proceedings of IEEE 7th International Conference on Database Systems for Advanced Applications, 2001, 169-170.

22 R. Agarwal, C. Faloutsos, A. Swami, Efficient Similarity Search in Sequence Databases, 4th International Conference of Foundations of Data Organization and Algorithms, Evanston, Illinois, USA, October 1993, pp. 69-84.

23 J. Alon, S. Sclaroff, G. Kollios, V. Pavlovic, Discovering clusters in motion time-series data, Proc. IEEE CVPR, June 2003, **1**: 375-381.

24 D. Buzan, S. Sclaroff, G. Kollios, Extraction and Clustering of Motion Trajectories in Video, International Conference on Pattern Recognition, Cambridge, UK, 2004, 521-524.

25 M. Vlachos, G. Kollios, D. Gunopulos, Discovering Similar Multidimensional Trajectories, Proceedings of the International Conference on Data Engineering, San Jose, CA, 2002, 673-684.

26 F.I. Bashir, A.A. Khokhar, D.Schonfeld, Object Trajectory-Based Activity Classification and Recognition Using Hidden Markov Models, IEEE Transactions on Image Processing, 2007, **16**(7): 1912-1919.

27 D. Zhang, Gatica-Perez, S. Bengio, I. McCowan, Semi-supervised adapted hmms for unusual event detection, Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2005, 611-618.

28 C. Abraham, P-A. Cornillon, E. Matzner-Lober, N. Molinari, Unsupervised curve clustering using b-splines, Scandinavian Journal of Statistics,September 2003, **30**(3): 581-595.

29 W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, S. Maybank, A system for learning statistical motion patterns, IEEE Transactions on Pattern Analysis and Machine Learning, September 2006, **28**(9): 1450-1464.

30 J. Owens, A. Hunter, Novelty Detection in Video Surveillance Using Hierarchical Neural Networks, Proceedings of ICANN, Madrid, Spain, August 28-30, 2002, 1249-1254.

31 C. Stauffer, E. Grimson, Learning Patterns of Activity Using Real-Time Tracking, IEEE Transactions on Pattern Recognition and Machine Intelligence, 2000, **22**(8): 747-757.

32 N. Sumpter, A.J. Bulpitt, Learning spatio-temporal patterns for predicting object behaviour, Image and Vision Computing, 2000, **18**: 697-704.

33 W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in C: The Art of Scientific Computing, 2nd edition, Cambridge University Press, Cambridge, England, 1992.

34 E.J. Keogh, M.J. Pazzani, A simple dimensionality reduction technique for fast similarity search in large time series databases, Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2000, 122-133.

35 CAVIAR test sequences [Online]. Available: http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/

36 A. Zisserman, R. Harley, Multiple View Geometry in Computer Vision, Cambridge University Press, 2003.

37 F. Porikli, T. Haga, Event Detection by Eigenvector Decomposition using Object and Frame Features, International Conference on Computer Vision and Pattern Recognition, 2004.

38 F.I. Bashir, A.A. Khokhar, D.Schonfeld, Real-time motion trajectory based indexing and retrieval of video sequences, IEEE Transactions on Multimedia, January 2007, **9**(1): 58-65.

39 D.L. Davie, D.W. Bouldin, A Cluster Separation Index, IEEE Trans. Pattern Analysis and Machine Intelligence, 1979, **1**: 224-227.

40 J.C. Dunn, A Fuzzy Relative of ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, Journal of Cybernetics, 1973, **3**: 32-57.

41 R.B. Calinski, J. Harabasz, A Dendrite Method for Cluster Analysis, Comm. in Statistics, 1974, **3**: 1-27.

42 B.J. Frey, D. Dueck, Clustering by Passing Messages between Data Points, Science, 2007, **315**: 972-976.

43 K. Wang, J. Zhang, D. Li, X. Zhang, T. Guo, Adaptive Affinity Propagation Clustering, Acta Automatica Sinica, 2007, **33**(12): 1242-1246.

44 http://www.mathworks.com/matlabcentral/fileexchange/loadAuthor.do?objectType=authorobjectId=1095267

45 Y.N. Andrew , I.J. Micahel, Y. Weiss, On spectral clustering: Analysis and an algorithm, In Advances in Neural Information and Processing Systems, 2001, **14**.

Shehzad Khalid   graduated from Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Pakistan, in 2000. He received the M. Sc. degree from National Unviersity of Science and Technology, Pakistan in 2003 and the Ph.D. degree from the University of Manchester, U.K., in 2009. He is currently an Assistant professor at the Bahria University of Management and Computer Sciences, Pakistan. His research interests includes: dimensionality reduction, indexing and retrieval, profiling and classification, trajectory-based motion learning profiling and classification, computer vision, machine learning.

Andrew Naftel   received the B.Sc. degree in mathematics and the Ph.D. degree in photogrammetric engineering from the University of Bradford, U.K., in 1985 and 1989, respectively. During 1990 to 1999, he was a Lecturer, and then a Senior Lecturer, in the Department of Mathematics and Statistics, University of Central Lancashire, Preston, U.K. In 2000, he joined the Department of Computing, University of Bradford. Between 2002-08, he was a Lecturer in Computer Science at the University of Manchester. His research interests include computer vision, image understanding and visual information retrieval. Dr Naftel currently works as a guidance and control engineer in the aerospace industry