Low complexity differential geometric computations with applications to activity analysis

by

Rushil Anirudh

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved June 2012 by the
Graduate Supervisory Committee:

Pavan Turaga, Chair
Andreas Spanias
Baoxin Li

ARIZONA STATE UNIVERSITY

August 2012

ABSTRACT

In this thesis, we consider the problem of fast and efficient indexing techniques for time sequences which evolve on manifold-valued spaces. Using manifolds is a convenient way to work with complex features that often do not live in Euclidean spaces. However, computing standard notions of geodesic distance, mean etc. can get very involved due to the underlying non-linearity associated with the space. As a result a complex task such as manifold sequence matching would require very large number of computations making it hard to use in practice. We believe that one can device smart approximation algorithms for several classes of such problems which take into account the geometry of the manifold and maintain the favorable properties of the exact approach. This problem has several applications in areas of human activity discovery and recognition, where several features and representations are naturally studied in a non-Euclidean setting. We propose a novel solution to the problem of indexing manifold-valued sequences by proposing an intrinsic approach to map sequences to a symbolic representation. This is shown to enable the deployment of fast and accurate algorithms for activity recognition, motif discovery, and anomaly detection. Toward this end, we present generalizations of key concepts of piece-wise aggregation and symbolic approximation for the case of non-Euclidean manifolds. Experiments show that one can replace expensive geodesic computations with much faster symbolic computations with little loss of accuracy in activity recognition and discovery applications. The proposed methods are ideally suited for real-time systems and resource constrained scenarios.

DEDICATION

To Papa & Mummy,

ACKNOWLEDGEMENTS

This thesis would have been far from complete if not for the continuous help and patience of my advisor Dr. Pavan Turaga. I would like to thank him for all the support and guidance in helping me complete this work. I would also like to thank the Department of Arts, Media & Engineering for providing me with all the facilities and a conducive research environment. Last but not the least, this would not have been possible with out the support of family and friends, who have always been there. Thanks guys!

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 MOTIVATION

Manifold theory has gained popularity in vision over the past few years because it provides us with tools to work in feature spaces that are not necessarily Euclidean. Many features used in vision can naturally be studied as objects belonging to these manifolds. While studying shapes, for example, manifolds allow us to perform shape algebra and hence enabling physical interpretations to notions such as "average shapes" etc. Extending well defined concepts and techniques in Euclidean spaces to manifolds allows us to easily interpret and work with complex data. However this is not trivial because we do not fully understand the semantics of non Euclidean spaces yet. For manifolds, standard notions of distance, statistics, quantization etc. need modification to account for the non-linearity of the underlying space. As a result, basic computations such as geodesic distance, mean computation etc are highly involved in terms of computational complexity, and often result in long iterative procedures further increasing the computational load. Not surprisingly, many standard approaches for sequence modeling and indexing which are designed for vector-spaces need significant generalization to enable application to these non-Euclidean spaces. Many use extrinsic methods as an alternative to avoid these problems, but these methods are approximate and work under the assumption that the data points lie close to each other on the manifold. The advantage of using extrinsic methods, which operate with the data embedded into Euclidean space and project the end result onto the manifold, is that they are quick and easy to compute as compared to intrinsic methods, which operate on a tangent space associated with the manifold, and are usually iterative.

In this thesis, we consider the problem of manifold sequence matching which is the generalization of time series matching to manifold spaces, applications of which include activity recognition, motif discovery in videos and anomaly detection. Given the aforementioned limitations, such a complex problem would require several geodesic computations, rendering it impractical to work with. However, we believe that one can device smart approximation algorithms for several classes of such problems which take into account the geometry of the manifold as well as maintain the favorable properties of the exact approach - if there exists one. In light of this, this thesis explores low complexity approximation algorithms for the task of sequence matching on manifolds. We demonstrate its utility on the problem of activity recognition and discovery. We show that using these low complexity approximations, it is possible to retain most of the recognition accuracy while enjoying sufficiently low computations. We test the robustness of the technique on three of the most commonly used features for activity analysis - Histograms of oriented optical flow (HOOF), region covariances and shape silhouettes. These features are extracted every frame and can be studied as points lying on well defined Riemannian manifolds. Therefore the problem of approximating these activities becomes one of discretizing a manifold valued time series. To this end, we propose the generalization of a popular technique - Symbolic Approximation [19]. Applying this to human activity has significant potential in resource constrained environments like robotic platforms, where one needs algorithms with reduced computational and communication requirements.

1.2 RELATED WORK

Activity analysis is one of the biggest challenges of computer vision with important applications such as surveillance[26] in public places such as airpots and

subway stations, Robotics [2] where systems allow robots to perform actions based on what they see and more recently in human-computer interactions (HCIs) [27]. Methods today do not allow us to operate on a low complexity basis because they involve performing recognition on a set of rich features which invariably have very expensive computations. Sensor based recognition systems can be made to work in such a way, but they perform well for the specific applications that they were designed for and cannot be easily generalized. Examples of manifold valued sequences arise quite frequently in the field of human activity analysis, where many features such as shape contours, stick figures, histograms of oriented optical flow etc. have been studied in a non-Euclidean setting. As an example, shape spaces have long been considered Riemannian manifolds - Kendall's shape space is a complex spherical manifold [17], and affine shape spaces are Grassmann manifolds [1]. Further examples of such manifold representations abound in computer vision literature. The space of $d \times d$ covariance matrices or tensors [29] which appear both in medical imaging [29] as well as texture analysis [37] is a Riemannian manifold. The space of linear subspaces also called the Grassmann manifold, occurs in image set-modeling [14], video modeling by linear dynamic systems [36], and tensor decomposition [22]. Long-term complex activities are often modeled as time-varying linear dynamical systems [35], which can be interpreted as a sequence of points on a Grassmann manifold, providing another motivating application for the problem of indexing of manifold sequences. Fig 1.1 shows some of these features and their manifold representations. Here C denotes the set of square-root velocity functions (SRVFs) representing planar closed curves and S_1 denotes the shape space of such curves obtained after removing the reparameterization and rotation groups. Landmarks can be obtained in $\mathbb{R}^3$, S_2 is the shape space obtained after removing translation, scale and rotation for

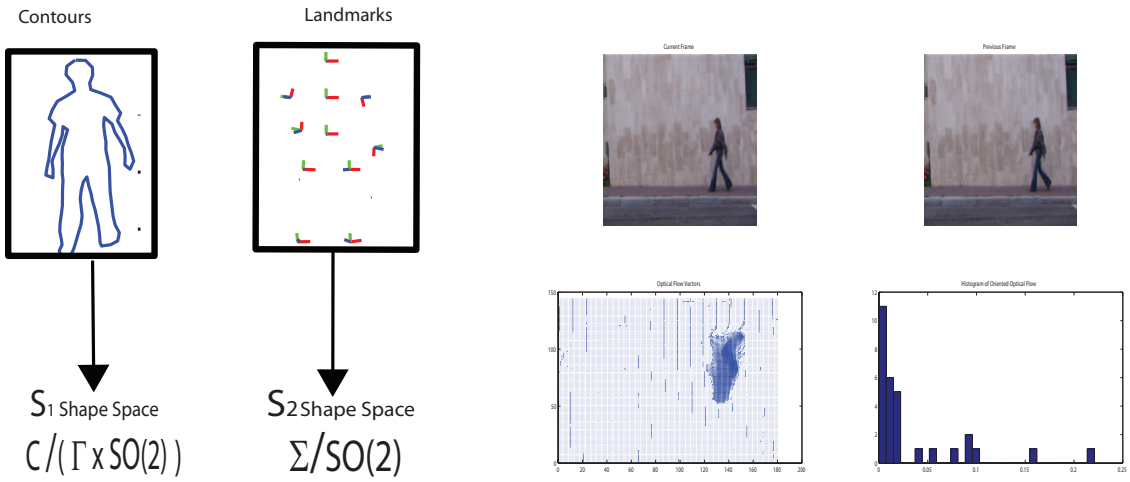


Figure 1.1: Non Euclidean representations of features commonly used in human activity analysis.

'k' landmarks (where $\Sigma = \{y \in \mathbb{R}^{3xk} \mid\mid y \mid\mid = 1, \sum_{j=1}^{k} y_{:,j} = 0\}$. Also shown is the Histogram of Oriented Optical Flow. The histograms are points on a B-dimensional Hypersphere or $\mathbb{S}^{B-1}$.

The problem of indexing of static data on manifolds is related, and recent hashing based approaches have been proposed [5, 34]. This problem is quite distinct from indexing of sequences on manifolds, where we are not interested in indexing individual points, but sequences of points. Another related line of work in recent years has been advances in Riemannian metrics for sequences on manifolds [32]. These approaches consider a sequence as an equivalent vector-field on the manifold. A distance function is imposed on such vector-fields in a square-root elastic framework. This is applied to the special case of curves in $2D$-Euclidean space [32] and also nD Euclidean space [15]. While such a distance function could be utilized for the purposes of indexing and approximation of sequences, it is offset by the computational load required in computing the distance function for long sequences. Further, the generalization of these metrics to sequences on manifolds is non-trivial and

not fully explored. We propose a novel approximation method for manifold sequences, which is consistent with the underlying geometry of the manifold, which also lends itself to fast algorithms for sequence indexing, motif discovery etc. among other applications.

In the case of scalar time-series, the pattern discovery problem is defined in [9] as "Given a real-valued time series T, an integer k, to find the k-most frequent patterns occurring in T". A recent approach to tackle this problem has been to discretize the sequence in a way such that the representative symbolic form contains most of the information as the original sequence, but enabling much faster computations in the symbolic space. This class of approaches are broadly termed as **S**ymbolic **A**ggregate Appro**x**imation (SAX) [19]. Several problems of indexing and motif discovery from time series have been addressed using this framework [19, 24, 7]. Multidimensional extensions to these algorithms have also been proposed such as [33, 23, 38] We argue that this class of approaches has several appealing characteristics for manifold-valued time-series, as they enable us to replace highly non-linear distance function computations with much faster and simpler symbolic distance computations. however, to enable this generalization one needs to extend several key concepts to the manifold setting.

1.3 CONTRIBUTIONS AND ORGANIZATION

The main contributions of this thesis are the following:

1. We present the first formalization of indexing time series evolving on non-Euclidean spaces.

2. We propose an intrinsic generalization of symbolic aggregate approximation to the case of manifolds.

3. We show that the recognition and discovery of human activity patterns can be achieved very efficiently without significant loss of accuracy using the proposed techniques, while avoiding expensive geodesic distance computations during run-time.

1.4 OVERVIEW

This thesis is organized into five Chapters. In chapter 2 we briefly describe the required notations and concepts to enable formalizing the sequence indexing problem on manifolds. In Chapter 3 we describe the proposed representation for manifold sequences which allows efficient algorithms to be deployed for a variety of tasks such as motif discovery, low-complexity activity recognition, and anomaly detection. Toward this end, we focus on the piecewise aggregate and Symbolic approximation (PAA, SAX) [3, 19] formulation, and present an intrinsic method to extend it to manifolds. Chapter 4 provides a discussion on the expeiments and results using the proposed technique on popular datasets. The final chapter concludes the work and discusses possible directions for future work.

Chapter 2

INTRODUCTION TO MANIFOLDS

2.1 TOPOLOGY AND MANIFOLDS

A topological space is a set $\mathcal{M}$, with a specified class of subsets or neighborhoods ϕ such that 1) ϕ & $\mathcal{M}$ are open, 2) The intersection of any two open sets is open and 3) The union of any number of open sets is open. A topological space is called *Hausdorff* if any two points of $\mathcal{M}$ possess non-intersecting neighborhoods. A continuous function $f : \mathcal{M} \to$ S is one that maps open sets on $\mathcal{M}$ to open sets on to the space S, that may or may not be the same as $\mathcal{M}$. If the function f has an inverse that is also continuous then $\mathcal{M}$ & S are said to be homeomorphic.

Finally, a manifold $\mathcal{M}$ of dimension N, is a topological - Hausdorff space that is locally homeomorphic to $\mathbb{R}^N$ i.e. for each $p \in \mathcal{M}$, there exists an open neighborhood U of p and a mapping $\phi : U \to \mathbb{R}^n$ such that $\phi(U)$ is open in $\mathbb{R}^n$ and $\phi : U \to \phi(U)$ is a diffeomorphism. The pair (U, ϕ) is called a *coordinate chart* for the points that fall in U.

The Euclidean space $\mathbb{R}^d$ is studied as a manifold using the identity chart. The complex coordinate space $\mathbb{C}^n$ becomes a $2n$-dimensional manifold via the chart $\mathbb{C}^n \to \mathbb{R}^{2n}$ replacing every complex coordinate z_j by a pair of real coordinates Re z_j, Im z_j. The sphere $S^n = \{x \in \mathbb{R}^{n+1} : \sum_{i=0}^{n} x_i^2 = 1\}$ is made into a smooth manifold of dimension n, by means of the two stereographic projections onto $\mathbb{R}^n \cong \{x \in \mathbb{R}^{n+1} : x_0 = 0\}$, from the North and South poles $(\pm 1, 0, \ldots, 0)$. The corresponding change of coordinates is given by $(x_1, \ldots, x_n) \to (x_1/|x|^2, \ldots, x_n/|x|^2)$.

In computer vision, the Grassmann and the Stiefel manifolds are used in several applications as described earlier. The Grassman manifold is the

space of d-dimensional subspaces in $\mathbb{R}^n$ and the Stiefel manifold is the space of d orthonormal vectors in $\mathbb{R}^n$.

2.1.1 MANIFOLD SEQUENCES

As shown in fig 2.1, like in Euclidean space, a sequence of points that evolve over time on the manifold can be studied as a time series. To analyze sequences or curves on manifolds, one needs to take recourse to understanding tangent-space and exponential mappings. A tangent-space at a point of a manifold $\mathcal{M}$ is obtained by considering the velocities of differentiable curves passing through the given point. i.e. for a point $p \in \mathcal{M}$, a differentiable curve passing through it is represented as $\beta : (-\delta, \delta) \to \mathcal{M}$ such that $\beta(0) = p$. The velocity $\dot{\beta}(0)$ refers to the velocity of the curve at p. This vector has the same dimension as the manifold and is a tangent vector to $\mathcal{M}$ at p. The set of all such tangent vectors is called the tangent space to $\mathcal{M}$ at p. The tangent space $T_p(\mathcal{M})$ is always a vector-space.

2.1.2 RIEMANNIAN METRIC

The distance between two points on a manifold is measured by means of the 'length' of the shortest curve connecting the points. The notion of length is formalized by defining a **Riemannian metric**, which is a map $\langle \cdot, \cdot \rangle$ that associates to each point $p \in \mathcal{M}$ a symmetric, bilinear, positive definite form on the tangent space $T_p(\mathcal{M})$. The Riemannian metric allows one to compute the infinitesimal length of tangent-vectors along a curve. The length of the entire curve is then obtained by integrating the infinitesimal lengths of tangents along the curve. i.e. given $p, q \in \mathcal{M}$, the distance between them is the infimum of the lengths of all smooth paths on $\mathcal{M}$ which start at p and end at q:

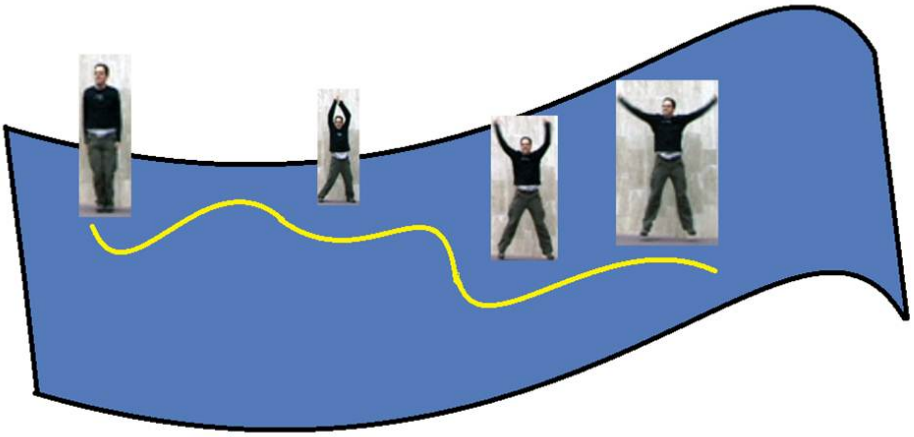


Figure 2.1: Features are extracted on each image/frame of a video depicting human activity resulting in a sequence of features evolving over time, or a manifold valued time series. The idea is shown here using sample data from the Wiezmann Data set for human action [13]

$$d(p,q) = \inf_{\{\beta:[0,1]\mapsto\mathcal{M}|\beta(0)=p,\beta(1)=q\}} L[\beta], \text{where,} \tag{2.1}$$

$$L[\beta] = \int_0^1 \sqrt{\left(\left\langle \dot{\beta}(t), \dot{\beta}(t) \right\rangle\right)} dt \tag{2.2}$$

If $\mathcal{M}$ is a Riemannian manifold and $p \in \mathcal{M}$, the **exponential map** $\exp_p : T_p(\mathcal{M}) \to \mathcal{M}$, is defined by $\exp_p(v) = \beta_v(1)$ where β_v is a specific geodesic in the direction of the tangent-vector v. The inverse mapping $\exp_p^{-1} : \mathcal{M} \to T_p$ called the inverse exponential map at a 'pole', takes a point on the manifold and returns a point on the tangent space of the pole.

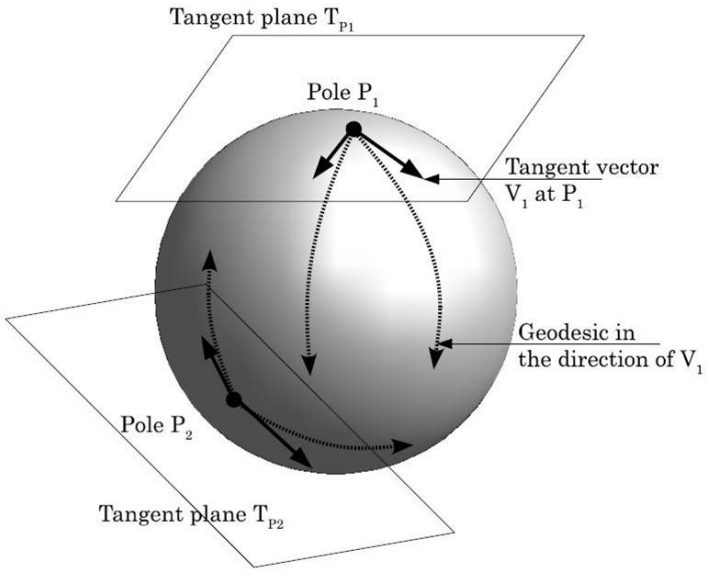


Figure 2.2: Exponential, Inverse exponential maps and the Tangent Space.

Chapter 3

SYMBOLIC AGGREGATE APPROXIMATION

Briefly, the PAA and SAX formulation consist of the following principal ideas. A given 1D scalar time-series is first divided into windows and the sequence in each window is represented by its mean value. This process is referred to as **piece-wise aggregation** and is illustrated in fig 3.1. Then, a set of 'break-points' is chosen which correspond to dividing the range of the time-series into equi-probable bins. These break-points comprise the symbols into which the given time-series will be translated into. For each window, the mean value is then replaced by the closest symbol. This step is referred to as **symbolic approximation**, as shown in fig 3.2. This representation has been shown to enable efficient solutions to scalar time-series indexing, retrieval, and analysis problems [19].

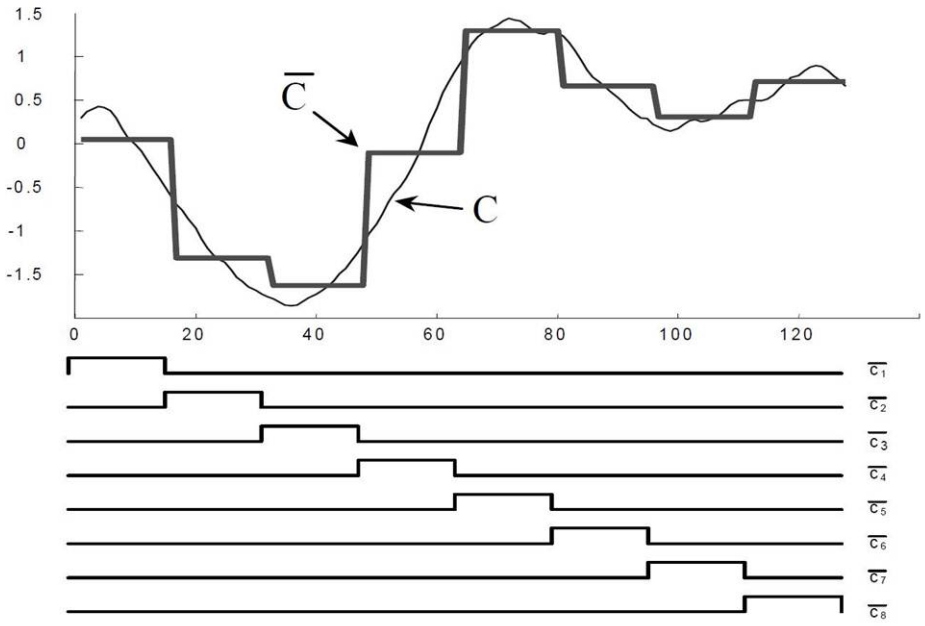


Figure 3.1: Piecewise Aggregation: C is the original time series and $\bar{C}$ represents the sequence after replacing the sequence in each window by its mean.

In the manifold setting, to enable us to exploit the advantages offered

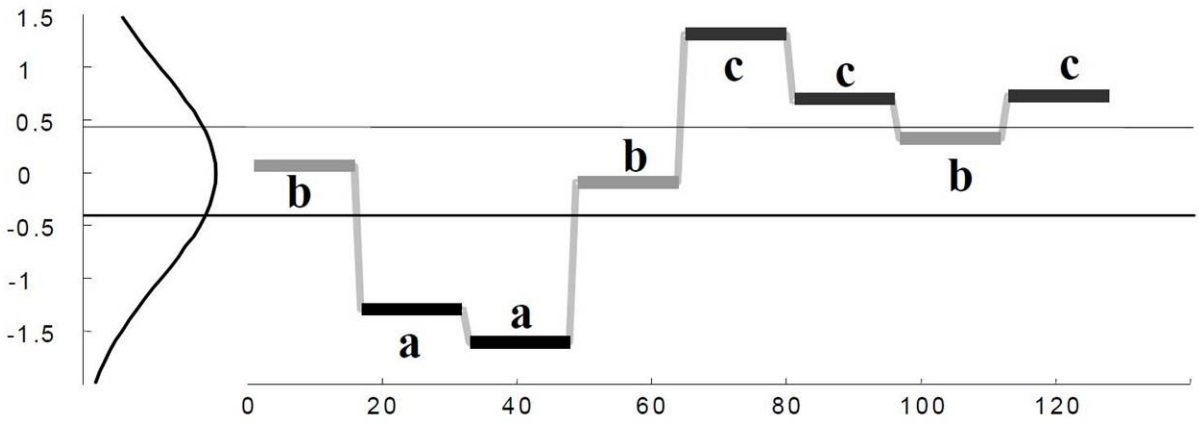


Figure 3.2: Symbolic Approximation: after choosing suitable break-points, each level is assigned a label based on its proximity to break-points. For example, the sequence shown here will then be approximated to "b a a b c c b c ".

by the symbolic representation of sequences, we need solutions to the following main problems: a) piece-wise aggregation: which can be achieved by appropriate definitions of the mean of a windowed sequence on a manifold, and b) symbolic approximation: which requires choosing a set of points that represent equi-probable regions on the manifold to serve as the 'symbols'. Here, we discuss how to generalize these concepts for the case of manifolds.

3.1 PIECE-WISE AGGREGATION OF MANIFOLD SEQUENCES

Given a sequence $\beta(t) \in \mathcal{M}$, we define its piece-wise approximation in terms of local-averages in small time-windows. To do this, we first need a notion of a mean of points on a manifold. Given a set of points on a manifold, a commonly used definition of their mean is the **Karcher mean** [16], which is defined as the point μ that minimizes the sum of squared-distance to all other points: $\mu = \arg\min_{x \in \mathcal{M}} \sum_{i=1}^{N} d(x, x_i)^2$, where d is the geodesic distance on the manifold.

Computing the mean is not usually possible in a closed form, and is unique only for points that are close together [16]. An iterative procedure

is popularly used in estimation of means of points on manifolds [28]. Since in local time windows, points are not very far away from each other, the algorithm always converges. Thus, given a manifold-valued time series $\beta(t)$, and a window of length w, we compute the mean of the points in the window and this gives rise to the piece-wise aggregate approximation for manifold sequences.

3.2 SYMBOLIC APPROXIMATION OF MANIFOLD SEQUENCES

As discussed above, one of the key-steps in performing symbolic approximation for manifold-valued time-series is to obtain a set of discrete symbols which represent equi-probable regions on the manifold. One approach would be to try clustering approaches and use the cluster centers as symbols. However, it is well-known that standard clustering approaches do not necessarily result in equi-probable distributions over symbols [40, 18, 30]. The same holds true for manifold-valued data and we shall later illustrate this with comparative examples. When symbols are not equiprobable, there is a possibility of inducing a probabilistic bias in the process [20]. Further, data structures such as suffix trees that are useful for anomaly detection applications produce optimal results when symbols are equi-probable [19, 8].

Generation of equi-probable symbols is a hard problem in general. It has been observed that a 'conscience' based competitive learning approach does result in symbols that are much more equi-probable [10] than those obtained from clustering approaches. The algorithm of [10] is devised only for vector-spaces. In this section, we present a generalization of this approach to account for non-Euclidean geometries. The tools that we build upon include computation of geodesic distances, exponential maps and inverse-exponential maps. These are known for many standard manifolds commonly occurring in

computer vision applications.

The conscience mechanism starts with a set of initial symbols/exemplars. When an input data-point is presented, a competition is held to determine the symbol closest in distance to the input point. Here, we use the geodesic distance on the manifold for this task. Let us denote the current set of K symbols as $\{S_1, S_2, \ldots, S_K\}$, where each $S_i \in \mathcal{M}$. Let the input data point be denoted as $X \in \mathcal{M}$. The output y_i associated with the i^{th} symbol is described as

$$y_i = 1, \text{ if } d^2(S_i, X) \leq d^2(S_j, X), \forall j \neq i \tag{3.1}$$

$$y_i = 0, \text{otherwise} \tag{3.2}$$

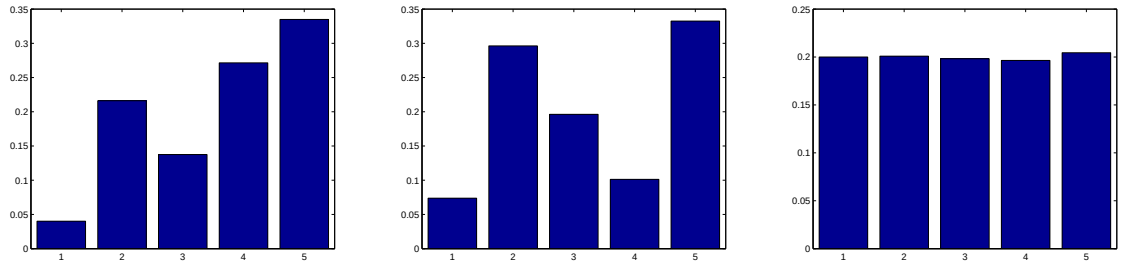
where, $d()$ is the geodesic distance on the manifold. Since this version of competition does not keep track of the fraction of times each symbols wins, it is modified by means of a bias term to promote more equitable wins among the symbols. A bias b_i is introduced for each symbol based on the number of times it has won in the past. Let p_i denote the fraction of times symbol i wins the competition. This is updated after each competition as

$$p_i^{new} = p_i^{old} + B(y_i - p_i^{old}) \tag{3.3}$$

where $0 < B << 1$. The bias b_i for each symbol is computed as $b_i = C(\frac{1}{K} - p_i)$, where C is a scaling factor chosen to make the bias update significant enough to change the competition (see below). The modified competition is given by

$$z_i = 1, \text{ if } d^2(S_i, X) - b_i \leq d^2(S_j, X) - b_j, \forall j \neq i \tag{3.4}$$

$$z_i = 0, \text{otherwise.} \tag{3.5}$$



(a) K-Means (b) Affinity Propagation (c) Equiprobable Clustering

Figure 3.3: Probability Density Functions of the labels generated using (a) K-Means clustering, (b) Affinity Propagation and (c) Equi-Probable Clustering are shown. As seen above, equiprobable clustering assigns all clusters with almost equal probability.

Finally, the winning symbol is adjusted by moving it partially towards the input data point. The partial movement of a symbol towards a data-point can be achieved by means of the exponential and inverse-exponential map as

$$S_i^{new} = \exp_{S_i^{old}}[\alpha \exp^{-1}_{S_i^{old}}(X) z_i]. \tag{3.6}$$

This competition based approach is a data-driven iterative algorithm and thus obtains symbols that are equi-probable for the given dataset. In practice, this means that one needs to provide a small training-set to the system a priori before the system can be deployed. This is in contrast to the 1D time-series symbolic approximation where one assumes that the values of the time-series can be approximated as a Gaussian distribution[19], which allows one to obtain equiprobable symbols in closed-form without the need for any training. However, this approach does not generalize to manifolds and thus the need for a training phase. The proposed algorithm for conscience based equiprobable symbol learning is summarized in algorithm 1.

Here, we illustrate the strength of this approach in obtaining equiprobable symbols on manifolds. For this experiment we chose the UMD human

Algorithm 1 Equi-probable symbol generation on manifolds.

Input: Dataset $\{X_1, \ldots, X_n\} \in \mathcal{M}$. Initial set of symbols $\{S_1, \ldots, S_k\}$.
Parameters: Biases $b_i = 0$, learning rate α, win update factor B, conscience factor C.
while $iter \leq maxiter$ **do**
  for $j = 1 \rightarrow n$ **do**
    $\tilde{i} \leftarrow \min_i d^2(X_j, S_i) - b_i$
    $z_{\tilde{i}} = 1,\ z_i = 0, i \neq \tilde{i}$
    $S_i \leftarrow \exp_{S_i}[\alpha \exp_{S_i}^{-1}(X_j) z_i]$
    $p_i \leftarrow p_i + B(z_i - p_i)$
    $b_i \leftarrow C(1/k - p_i)$
  end for
end while

activity dataset [39] and preprocessed it such that we obtain the outer contour of the human. A detailed discussion of the dataset, processing, choice of shape metrics etc. appears in the experiments section. Here, we performed clustering of the shapes into 5 clusters and used the centroids as symbols. We show the histograms of the symbols as obtained in fig 3.3. We also used the conscience based approach to obtain our symbols and the histograms of the symbols are shown as well. As can be seen, both k-means and affinity propagation result in symbols that are far from equi-probable. The proposed approach results in symbols which are much closer to a uniform distribution.

3.3 ACTIVITY RECOGNITION AND DISCOVERY

The applications considered in this paper are recognition and discovery of human activities. For recognition, a very commonly used approach involves storing labeled sequences for each activity, and performing recognition using a distance-based classifier, a nearest-neighbor classifier being the simplest one. When activity sequences involve manifold-valued time-series, distance computations are quite intensive depending on the choice of shape metric. We explore here the utility of the symbolic approximation as an alternative way

for approximate yet fast recognition of activities that can replace the expensive geodesic distance computations during testing. This is especially applicable in real-time deployments and in cases where recognition occurs remotely and there is a need to reduce the communication requirements between the sensor and the analysis engine.

For each N-length activity, we extract a symbolic representation in windows of size w (where typically $w << N$). To replace geodesic distance computations for recognition, we will consider subsequences in their symbolic representations to calculate the distance between activities. Let $\mathrm{p_{sub}}$ (eg: 'b c c d e a') and $\mathrm{q_{sub}}$ (eg: 'a f f f e c') be two such subsequences of length l, then the distance metric $\mathrm{d_{symbol}}$, defined on symbols, is:

$$d_{symbol}(p_{sub}, q_{sub}) = \sum_{i=1}^{l} d\Big(\mathbf{S}\big(p_{sub}(i)\big),\ \mathbf{S}\big(q_{sub}(i)\big)\Big) \tag{3.7}$$

where $\mathbf{S}$ is the set of symbols and $\mathbf{S}(a)$ is the manifold point corresponding to the symbol a etc. Since the symbols are known a priori, the distance between them can be pre-computed and stored as a look-up table of pairwise geodesic distances between symbols. This enables us to compute distances between activities in near constant time.

For activity discovery, we consider the problem as one of mining for motifs in time-series. In finding motifs, it is important to consider only **non-trivial matches**, because if we have a *match* at the i^{th} location there is a good chance that we will have one at the $(i \pm 1)^{th}$ locations as well. Sequences p and q are said to be a non-trivial match when the following conditions are met :

1. $d_{symbol}(p_{sub}, q_{sub}) \leq R$, where R is a threshold.

2. p_{sub} and q_{sub} are separated by at least **n** symbols, where **n** is the symbol length of each activity.

3. There is no *better* match for p_{sub} in the neighborhood of q_{sub} and vice versa. We defined a neighborhood as a window of length n centered around the current location.

This idea is shown in fig 3.4. Next, for every non-trivial match we store its location and find the top k motifs. For each of the k motifs, we define a *center* for the motif as the sequence which is at minimum distance to all the sequences similar to it. These centers are the k most recurring patterns in the manifold valued time series.

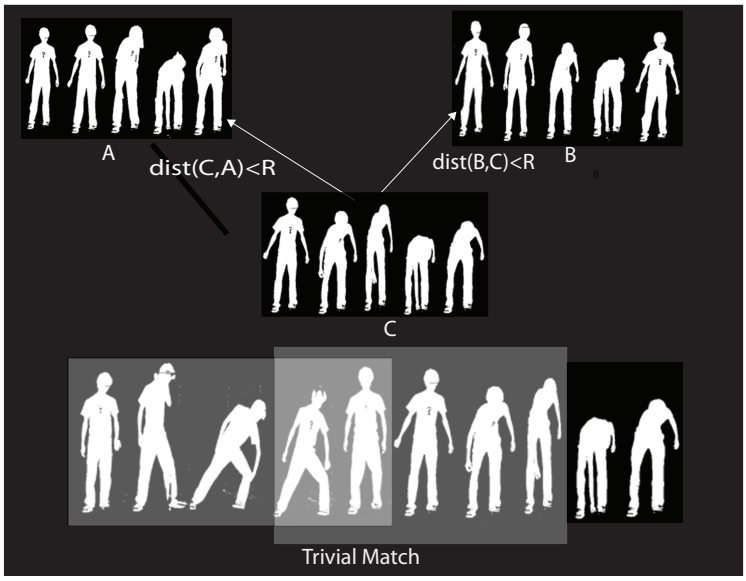


Figure 3.4: An illustration of trivial and non-trivial matches is shown. Here, C is a motif and subsequences A ,B represent valid matches to C. R is the threshold for a match. An intermediate sub-sequence is a trivial match and is rejected by the threshold.

3.4 SPECIAL CASES AND LIMITATIONS

In this section, we discuss the limitations and some special cases of the proposed formulation. The overall approach assumes that a training set can be

easily obtained from which we can extract the symbols for sequence approximation. In the 1D scalar case, this is not an issue, and one assumes that data distribution is a Gaussian, thus the choice of symbols can be obtained in closed-form without any training. If data is not Gaussian, a simple transformation/normalization of the data can be easily performed. In the manifold case, there is no simple generalization of this idea, and we are left with the option of finding symbols that are tuned for the given dataset.

For the special case of $\mathcal{M} = \mathbb{R}^n$, the approach boils down to familar notions of piece-wise aggregation and symbolic approximation. For the case of manifolds implicitly specified using samples, we suggest the following approach. One can obtain an embedding of the data into a Euclidean space and apply the special case of the algorithm for $\mathcal{M} = \mathbb{R}^n$. The requirement for the embedding here is to preserve geodesic distances between local pairs of points, since we are only interested in ensuring that data in small windows of time are mapped to points that are close together. Any standard dimensionality reduction approach can be used for this task. However, recent advances have resulted in algorithms for estimating exponential and inverse exponential maps numerically from sampled data points [21]. This would make the proposed approach directly applicable for such cases, without significant modifications. Thus the proposed formalism is applicable to manifolds with known geometries as well as to those whose geometry needs to be estimated from data.

Chapter 4

EXPERIMENTS AND RESULTS

In this chapter, we demonstrate the utility of the proposed algorithms for symbolic approximation and its application to activity recognition and discovery. We first describe the datasets, and choice of features.

4.1 DATASETS

We performed experiments on three different datasets namely: the UMD Human Activity Dataset [39], the Weizmann Dataset for Human Actions [13] and the UCSD Traffic Database [4].

The UMD database consists of 10 different activities like bend, jog, push, squat etc., each activity was repeated 10 times, so there were a total of 100 sequences in the dataset.

The Weizmann Dataset consists of 93 videos of 10 different actions each performed by 9 different persons. The classes of actions include running, jumping, walking, side walking etc.

The traffic video database consists of 254 video sequences of daytime highway traffic in Seattle in three patterns i.e. heavy, medium and light traffic. It was collected from a single stationary traffic camera over two days.

4.2 CHOICE OF FEATURE REPRESENTATION AND METRICS

To show the robustness of our technique, we used a different feature on each data set such as landmarks on the silhouette of the subject for the UMD Dataset, Histogram of Oriented Optical Flow for the Weizmann Dataset and Region Covariance on the Traffic Dataset achieving near state of the art recognition performance. All these features lie on well defined Manifolds. We will

describe their Manifold representations along with the choice of metrics used here.

4.2.1 LANDMARKS ON THE SILHOUETTE

The background within the UMD Dataset is relatively static which allows us to perform background subtraction. From the extracted foreground, we perform morphological operations and extract the outer contour of the human. We sampled a fixed number of points on the outer contour of the silhouette to yield landmarks. The specific choice of shape space, and associated metrics are described next.To represent the points sampled on the outer contour of the human, we use an affine invariant representation of landmarks as follows. The set of landmark points is given by a $m \times 2$ matrix $L = [(x_1, y_1); (x_2, y_2); \ldots; (x_m, y_m)]$, of the set of m landmarks of the centered shape. The *affine shape space* [12] is useful to remove the effects of small variations in camera location or small changes in the pose of the subject. Affine transforms of the base shape L_{base} can be expressed as $L_{affine}(A) = L_{base} * A^T$, and this multiplication by a full-rank matrix on the right preserves the column-space of the matrix L_{base}. Thus, the 2D subspace of $\mathbb{R}^m$ spanned by the columns of the matrix L_{base} is an *affine-invariant* representation of the shape. i.e. $span(L_{base})$ is invariant to affine transforms of the shape. Subspaces such as these can be identified as points on a Grassmann manifold [36].

We used extrinsic approaches for Grassmann manifold computations which are conceptually simpler and implemented more easily. A given d-dimensional subspace of $\mathbb{R}^m$, $\mathcal{Y}$ can be associated with a idempotent rank-d projection matrix $P = YY^T$, where Y is a $m \times d$ orthonormal matrix such as $span(Y) = \mathcal{Y}$. The space of $m \times m$ projectors of rank d, denoted by $\mathbb{P}_{m,d}$ can be embedded into the set of all $m \times m$ matrices - $\mathbb{R}^{m \times m}$- which is a vector space.

Using the embedding $\Pi : \mathbb{R}^{m\times m} \to \mathbb{P}_{m,d}$ we can define a distance function on the manifold using the metric inherited from $\mathbb{R}^{m\times m}$.

$$d^2(P_1, P_2) = tr(P_1 - P_2)^T(P_1 - P_2) \tag{4.1}$$

The projection $\Pi : \mathbb{R}^{m\times m} \to \mathbb{P}_{m,d}$ is given by:

$$\Pi(M) = UU^T \tag{4.2}$$

where $M = USV^T$ is the d-rank SVD of M.

Given a set of sample points on the Grassmann manifold represented uniquely by projectors $\{P_1, P_2, ...P_N\}$, we can compute the extrinsic mean by first computing the mean of the P_i's and then projecting the solution to the manifold as follows:

$$\mu_{ext} = \Pi(P_{avg}), \text{where } P_{avg} = \frac{1}{N}\sum_{i=1}^{N} P_i \tag{4.3}$$

4.2.2 HISTOGRAMS OF ORIENTED OPTICAL FLOW (HOOF)

As described in [6],optical flow is a natural feature for motion sequences. Directions of optical flow are computed for every frame, then binned according to their primary angle with the horizontal axis and weighted according to their magnitudes. Using magnitudes alone is susceptible to noise and can be very sensitive to scale. Thus all optical flow vectors, $v = [x, y]^T$ with direction $\theta = tan^{-1}(\frac{y}{x})$ in the range

$$-\frac{\pi}{2} + \pi\frac{b-1}{B} \leq \theta < -\frac{\pi}{2} + \pi\frac{b}{B} \tag{4.4}$$

will contribute by $\sqrt{x^2 + y^2}$ to the sum i bin b, $1 \leq b \leq B$, out of a total of B bins. Finally, the histogram is normalized to sum up to 1. Each frame is represented by one histogram and hence a sequence of histograms

are used to describe an activity. The histograms $h_t = [\ h_{t;1},\ \ldots\ ,\ h_{t;B}]$ can be re-parameterized to the *square root representation* for histograms, $\sqrt{h_t} = [\sqrt{h_{t;1}},\ldots,\sqrt{h_{t;B}}]$ such that $\sum_{i=1}^{B}(\sqrt{h_{t;i}})^2$=1. The Riemannian metric between two points R_1 and R_2 on the hypersphere is $d(\mathrm{R}_1, \mathrm{R}_2) = cos^{-1}(\mathrm{R}_1^{\mathrm{T}}\mathrm{R}_2)$. This projects every histogram onto the unit B-dimensional hypersphere or $\mathbb{S}^{B-1}$.

The geodesic distance between any two points ψ_i, ψ_j on a unit sphere is simply the angle between them, i.e.

$$\mathrm{dist}(\psi_i, \psi_j) = \cos^{-1}\langle\psi_i, \psi_j\rangle = \cos^{-1}\left(\int_0^T \psi_i(s)\psi_j(s)ds\right) \tag{4.5}$$

where $\langle . \rangle$ is the normal dot product between points in the sphere under the $\mathbb{L}^2$ metric. From the differential geometry of the sphere, the exponential map is defined as

$$\exp_{\psi_i}(\upsilon) = \cos(||\upsilon||_{\psi_i})\psi_i + \sin(||\upsilon||_{\psi_i})\frac{\upsilon}{||\upsilon||_{\psi_i}} \tag{4.6}$$

Where $\upsilon \in T_{\psi_i}(\Psi)$ is a tangent vector at ψ_i and $||\upsilon||_{\psi_i} = \sqrt{\langle\upsilon, \upsilon\rangle_{\psi_i}} = (\int_0^T \upsilon(s)\upsilon(s)ds)^{\frac{1}{2}}$. In order to ensure that the exponential map is a bijective function, we restrict $||\upsilon||_{\psi_i} \in [0, \pi]$. The logarithmic map from ψ_i to ψ_j is then given by

$$\overrightarrow{\psi_i\psi_j} = \log_{\psi_i}(\psi_j) = \frac{\mathbf{u}}{(\int_0^T\ \mathbf{u}(s)\ \mathbf{u}(s)ds)^{\frac{1}{2}}}\cos^{-1}\langle\psi_i, \psi_j\rangle\,, \tag{4.7}$$

with $\mathbf{u} = \psi_i - \langle\psi_i, \psi_j\rangle\,\psi_j$.

4.2.3 REGION COVARIANCE

Tuzel et al [37] describe a region descriptor and apply it to the problem of texture classification. They use the *covariance* of *d*- features like the three-dimensional color vector, the norm of the first and second derivatives of intensity with respect to x and y etc. Since moving traffic can be studied as dynamic textures, we use the region covariance as the feature for the traffic database. We used norm of first and second derivatives of intensities with

Activity Type	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5
Jogging	**7**	0	0	0	0
Squatting	0	**7**	0	0	0
Bending Knees	0	0	**8**	0	0
Waving	0	0	0	**9**	0
Throwing	0	0	0	0	**8**

Table 4.1: Confusion Matrix for the Discovered Motifs

respect to x and y to calculate a covariance matrix. Therefore each pixel in the image is converted to a 4 dimensional feature vector:

$$F(x,y) = \left[\frac{\partial I(x,y)}{\partial x}, \frac{\partial I(x,y)}{\partial y}, \frac{\partial^2 I(x,y)}{\partial x^2}, \frac{\partial^2 I(x,y)}{\partial y^2}\right] \tag{4.8}$$

We also use the distance measure proposed in [11] to measure the dissimilarity of two covariance matrices:

$$\rho(\ \mathbf{C1,\ C2}) = \sqrt{\sum_{i=1}^{n} \ln^2 \lambda_i(\ C_1, C_2)} \tag{4.9}$$

where $\lambda_i(\ \mathbf{C1,\ C2})_{i=1\ldots n}$ are the generalized eigenvalues of C_1 and C_2 computed from

$$\lambda_i C_1 x_i - C_2 x_i = 0\ ,\ i = 1 \ldots d \tag{4.10}$$

and $x_i \neq 0$ are the generalized eigenvectors.

Pennec et al [29] describe the exponential and logarithmic maps for tensors, or positive semi-definite matrices as shown in equations (4.11) and (4.12) respectively.

$$\exp_\Sigma(W) = \Sigma^{\frac{1}{2}} \exp(\Sigma^{-\frac{1}{2}} W \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}} \tag{4.11}$$

$$\log_\Sigma(\Lambda) = \Sigma^{\frac{1}{2}} \log(\Sigma^{-\frac{1}{2}} \Lambda \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}} \tag{4.12}$$

Where Σ is the pole, Λ - a point in tensor space, and W is the tangent vector at the pole (i.e., symmetric matrices but not necessarily definite or positive.)

4.3 ACTIVITY DISCOVERY EXPERIMENT

For this experiment, we randomly concatenated 10 repetitions of 5 different activities of the UMD dataset to create a sequence that was 50 activities long. Each activity consists of 80 frames which were sampled by a sliding window of size 20 frames with step size of 10 frames. After symbolic approximation, this resulted in 6 symbols per activity, chosen from an alphabet of 25 symbols, as shown in fig 4.2. The *motifs* or repeating patterns, in five activities - *Jogging, Squatting, Bending Knees, Waving and Throwing* were discovered automatically using the proposed method, the confusion matrix is shown in table 4.1 and some sample motifs are shown in fig 4.1. Each of the discovered motifs was validated manually to obtain the confusion matrix. As can be seen, it shows a strong diagonal structure, which indicates that the algorithm works fairly well. Even though all executions of the same activity are not found, we do not find any false matches either.

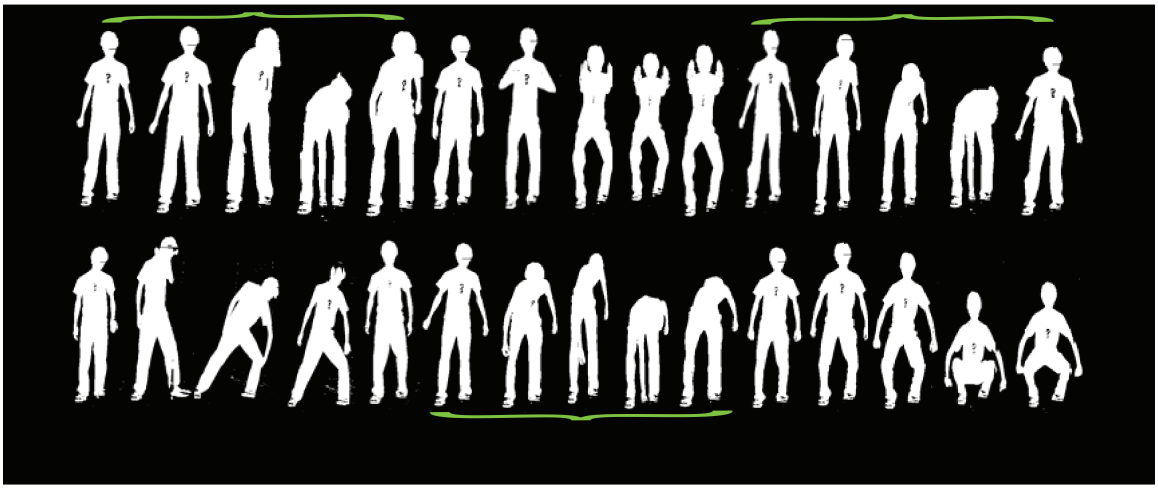


Figure 4.1: Sample images from the dataset used for the activity discovery experiment. A few recurring action patterns are highlighted.

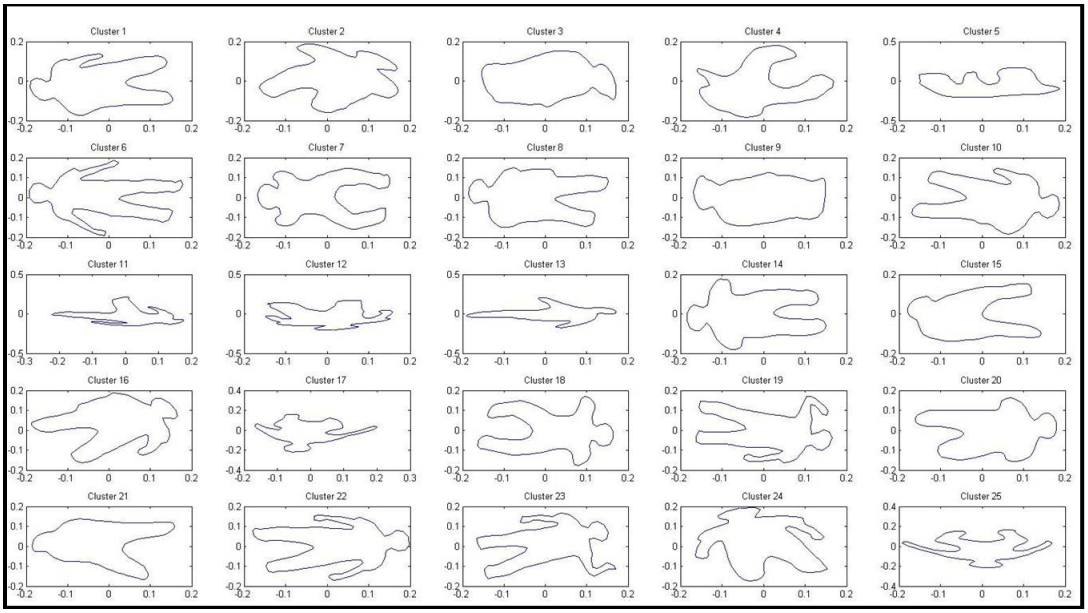


Figure 4.2: Symbol alphabet for the UMD activity data set. The cluster centers obtained are shown here where each center represents one member of the class of symbols.

4.4 ACTIVITY RECOGNITION USING MANIFOLD SYMBOLIC APPROXIMATION:

Symbolic approximation plays a significant role in reducing computational complexity since it allows us to work with symbols instead of working with high dimensional feature sets. In this experiment, we test the utility of the proposed symbolic approximation method for fast and approximate recognition of activities over three datasets. For the UMD dataset, we performed the experiment on all the activities using a leave one-execution-out test in which we trained on 9 executions and tested on the remaining execution the results are shown in Table 4.2. For the Weizmann Dataset we performed the experiment on all the 9 subjects performing 10 activities each with a total of 90 activities. We used a sliding window of length 3 frames and 50 clusters to form the alphabet of symbols. The results for the leave-one-execution-out recognition test for different degrees of approximation and the confusion matrix are shown in Table 4.3 & Table 4.3, as it can be seen, the performance does not decrease

by much for approximated sequences. For the Traffic Database, we performed the symbolic approximation by sampling the original video database such that each video was represented by $13 - 14$ frames. This combined with using 50 clusters to represent the symbols resulted in each video being approximated to 13 symbols. We performed the recognition experiment on 4 different test sets which contained 25% of the total videos. The results are shown in Table 4.5 & 4.6.

4.5 NON-EQUIPROBABLE SYMBOLS

While the proposed formalism supports arbitrary symbols to be used, equi-probable symbols are a desirable feature in several applications [19]. However, at the expense of such inefficiencies, one can often acquire higher accuracy by using symbols that are not equi-probable. Here, we test the increase in recognition accuracy when we use an intrinsic manifold version of k-means clustering to generate our symbols. As seen in table 4.2, it is possible to achieve close to optimal recognition accuracies with a non-optimal choice of symbols. However, for the Traffic Dataset it can be seen in Table 4.5 that the equiprobable clustering performs better. This is mostly because nearly 50% of the videos in the dataset belong to the class of Light Traffic, which will result in obtaining symbols that favor the class. A detailed analysis of these considerations needs to be discussed in the context of a particular application. Instead, here we simply demonstrate the flexibility and generality of the proposed formalism.

Activity	Pick Object	Jog	Push	Squat	Wave	Kick	Bend	Throw	Turn	Talk	Avg.
Veeraraghavan et al[39]	100	100	100	100	100	100	100	100	100	100	**100%**
Equi-Probable	90	70	70	100	100	90	80	100	50	100	**85%**
Non-Equi Probable	90	100	100	90	100	100	100	90	90	100	**96%**

Table 4.2: Accuracies for the recognition experiment using symbolic approximation compared to an oracle geodesic distance based nearest neighbor classifier. In general, recognition with symbols performs as well as full-fledged geodesic distance computations. Further, for this dataset non equi-probable symbols obtained by manifold k-means clustering results in negligible loss of accuracy.

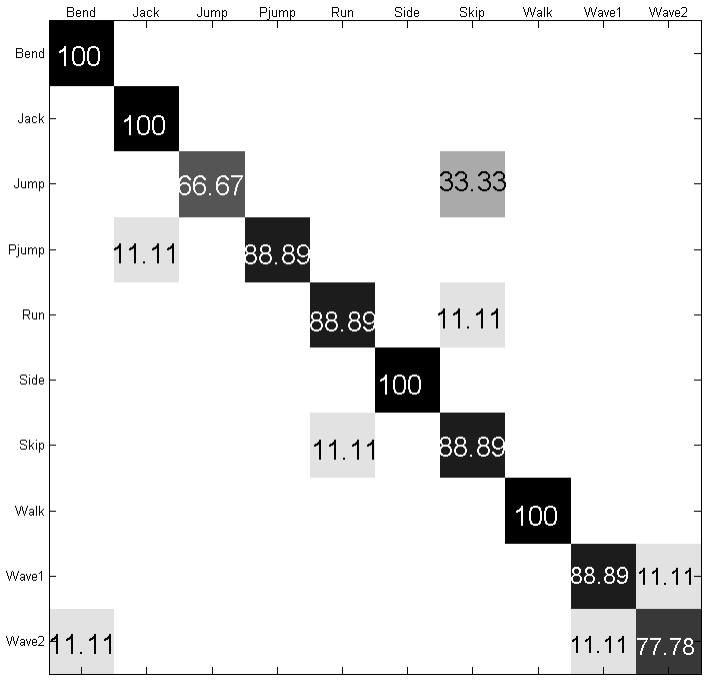


Table 4.4: Confusion Matrix for the Weizmann Dataset.

Proposed Method - 1 Symbol/Frame	90%
Proposed Method - 0.5 Symbols/Frame	86.67%
Proposed Method - 0.33 Symbols/Frame	81.11%
Chaudry et al [6]	95.66%
Gorelick et al [13]	97.83%
Niebles et al [25]	90.00%

Table 4.3: Recognition Performance for the Weizmann Dataset.

	Expt 1	Expt 2	Expt 3	Expt 4
SAX	84%	87.5%	87.5 %	78.5%
SAX non-equiprobable	85.71%	84.38%	87.5%	74.6%
Oracle LDS [31]	84.12%	85.93%	87.5 %	92.06

Table 4.5: Comparison in recognition performance on 4 different test sets.

	Heavy	Medium	Light
Heavy	37 (37)	5(7)	2(0)
Medium	13 (4)	24 (39)	8 (2)
Light	3(0)	11(1)	151 (164)

Table 4.6: Confusion Matrix for the Traffic Videos Dataset. Results achieved by Chan et al [4] are shown in brackets.

Chapter 5

CONCLUSIONS AND FUTURE WORK

In this thesis we presented a formalization of manifold-valued time-series approximation for efficient and low-complexity activity discovery, activity recognition, and anomaly detection applications. We presented novel algorithms to extend symbolic aggregate approximation techniques to non-Euclidean manifolds. The results show that it is possible to significantly reduce Riemannian computations during run-time by an intrinsic indexing and approximation algorithm - allowing us to work with symbols instead of complex features. This is the first formalization of its kind for manifold valued sequences which opens up several avenues for future work.

A natural extension to this work would be in picking better symbolic representatives for each class, i.e. perform clustering that fits the data better. Understanding how features for activities are distributed on manifolds will enable us to do this. A related problem is that of understanding what these symbols physically mean, which ultimately asks the question - could we break down actions into a set of simpler motion primitives? The effect of equiprobable symbols is another aspect of the problem that could be studied in more detail. We have demonstrated the flexibility in using this technique for both kinds of symbols. As it can be seen with the traffic dataset, in situations where the data is heavily biased towards a particular class, equiprobable symbols will perform better. Finally, a theoretical and empirical analysis of the advantages of the proposed formalism on resource-constrained systems such as robotic platforms would be another avenue of research.

BIBLIOGRAPHY

[1] E. Begelfor and M. Werman. Affine invariance revisited. *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[2] M.A.-A. Bhuiyan, M.E. Islam, N. Begum, M. Hasanuzzaman, Chang Hong Liu, and H. Ueno. Vision based gesture recognition for human-robot symbiosis. In *Computer and information technology, 2007. iccit 2007. 10th international conference on*, pages 1 –6, dec. 2007.

[3] Kaushik Chakrabarti, Eamonn J. Keogh, Sharad Mehrotra, and Michael J. Pazzani. Locally adaptive dimensionality reduction for indexing large time series databases. *ACM Trans. Database Syst.*, 27(2):188–228, 2002.

[4] A.B. Chan and N. Vasconcelos. Classification and retrieval of traffic video using auto-regressive stochastic processes. In *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, pages 771–776, June 2005.

[5] R. Chaudhry and Y. Ivanov. Fast approximate nearest neighbor methods for non-euclidean manifolds with applications to human activity analysis in videos. In *European Conference on Computer Vision*, Crete, Greece, September 2010.

[6] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1932 –1939, June 2009.

[7] Bill Yuan chi Chiu, Eamonn J. Keogh, and Stefano Lonardi. Probabilistic discovery of time series motifs. In *KDD*, pages 493–498, 2003.

[8] Maxime Crochemore, Thierry Lecroq, Artur Czumaj, Leszek Gasieniec, Stefan Jarominek, Wojciech Plandowski, and Wojciech Rytter. Speeding up two string-matching algorithms. In *STACS*, pages 589–600, 1992.

[9] Josenildo Costa da Silva and Matthias Klusch. Privacy-preserving discovery of frequent patterns in time series. In *Industrial Conference on Data Mining*, pages 318–328, 2007.

[10] D. Desieno. Adding a conscience to competitive learning. *IEEE International Conference on Neural Networks*, 1:117–124, 1988.

[11] Moonen B. Forstner, W. A metric for covariance matrices. Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart University, 1999.

[12] C. R. Goodall and K. V. Mardia. Projective shape analysis. *Journal of Computational and Graphical Statistics*, 8(2), 1999.

[13] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri. Actions as space-time shapes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(12):2247 –2253, dec. 2007.

[14] J. Ham and D. D. Lee. Extended grassmann kernels for subspace-based learning. *Neural Information Processing Systems*, pages 601–608, 2008.

[15] Shantanu H. Joshi, Eric Klassen, Anuj Srivastava, and Ian Jermyn. A novel representation for riemannian analysis of elastic curves in r^n. In *CVPR*, 2007.

[16] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977.

[17] D.G. Kendall. Shape manifolds, Procrustean metrics and complex projective spaces. *Bulletin of London Mathematical society*, 16:81–121, 1984.

[18] T. Kohonen. *Self-Organizing Maps*. Berlin: Springer - Verlag., 1995.

[19] Jessica Lin, Eamonn J. Keogh, Stefano Lonardi, and Bill Yuan chi Chiu. A symbolic representation of time series, with implications for streaming algorithms. In *DMKD*, pages 2–11, 2003.

[20] Jessica Lin and Yuan Li. Finding approximate frequent patterns in streaming medical data. In *CBMS*, pages 13–18, 2010.

[21] T. Lin and H. Zha. Riemannian manifold learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:796–809, 2008.

[22] Yui Man Lui, J. Ross Beveridge, and Michael Kirby. Action classification on product manifolds. In *CVPR*, pages 833–839, 2010.

[23] David Minnen, Charles L. Isbell, Irfan A. Essa, and Thad Starner. Detecting subdimensional motifs: An efficient algorithm for generalized multivariate pattern discovery. In *ICDM*, pages 601–606, 2007.

[24] Abdullah Mueen, Eamonn J. Keogh, Qiang Zhu, Sydney Cash, and M. Brandon Westover. Exact discovery of time series motifs. In *SDM*, pages 473–484, 2009.

[25] Juan Carlos Niebles, Hongcheng Wang, and Li Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. *International Journal of Computer Vision*, 79(3):299–318, 2008.

[26] Wei Niu, Jiao Long, Dan Han, and Yuan-Fang Wang. Human activity detection and recognition for video surveillance. In *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, volume 1, pages 719 –722 Vol.1, june 2004.

[27] Branislav Kisacanin; Vladimir Pavlovic and Thomas S. Huang;. *Real-Time Vision for Human-Computer Interaction*. Springer Science+Business Media, Inc, 2005.

[28] X. Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.

[29] X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

[30] B. D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press., 1996.

[31] Aswin C. Sankaranarayanan, Pavan K. Turaga, Richard G. Baraniuk, and Rama Chellappa. Compressive acquisition of dynamic scenes, 2010.

[32] Anuj Srivastava, Eric Klassen, Shantanu H. Joshi, and Ian H. Jermyn. Shape analysis of elastic curves in euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:1415–1428, 2011.

[33] Yoshiki Tanaka, Kazuhisa Iwamoto, and Kuniaki Uehara. Discovery of time-series motif from multi-dimensional data based on mdl principle. *Machine Learning*, 58(2-3):269–300, 2005.

[34] P. Turaga and R. Chellappa. Nearest-neighbor algorithms on non-euclidean manifolds for computer vision applications. In *Indian Conference on Computer Vision, Graphics, and Image Processing (ICVGIP)*, December 2010.

[35] Pavan K. Turaga and Rama Chellappa. Locally time-invariant models of human activities using trajectories on the grassmannian. In *CVPR*, pages 2435–2441, 2009.

[36] Pavan K. Turaga, Ashok Veeraraghavan, Anuj Srivastava, and Rama Chellappa. Statistical computations on grassmann and stiefel manifolds for image and video-based recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33(11):2273–2286, 2011.

[37] O. Tuzel, F. M. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. *European Conference on Computer Vision*, pages II: 589–600, 2006.

[38] Alireza Vahdatpour, Navid Amini, and Majid Sarrafzadeh. Toward unsupervised activity discovery using multi-dimensional motif detection in time series. In *IJCAI*, pages 1261–1266, 2009.

[39] A. Veeraraghavan, A. Roy-Chowdhury, and R. Chellappa. Matching shape sequences in video with an application to human movement analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(12):1896–1909, 2005.

[40] P. Zador. Asymptotic quantization error of continuous signals and the quantization dimension. *Information Theory, IEEE Transactions on*, 28(2):139 – 149, March 1982.