Statistical Analysis on Manifolds and Its Applications to Video Analysis

Pavan Turaga, Ashok Veeraraghavan, Anuj Srivastava, Rama Chellappa

TR2010-022 June 2010

Abstract

The analysis and interpretation of video data is an important component of modern vision applications such as biometrics, surveillance, motionsynthesis and web-based user interfaces. A common requirement among these very different applications is the ability to learn statistical models of appearance and motion from a collection of videos, and then use them for recognizing actions or persons in a new video. These applications in video analysis require statistical inference methods to be devised on non-Euclidean spaces or more formally on manifolds. This chapter outlines a broad survey of applications in video analysis that involve manifolds. We develop the required mathematical tools needed to perform statistical inference on manifolds and show their effectiveness in real video-understanding applications.

Edited Book: Video Search and Mining

Statistical Analysis on Manifolds and its applications to Video Analysis

Pavan Turaga, Ashok Veeraraghavan, Anuj Srivastava, Rama Chellappa

Abstract The analysis and interpretation of video data is an important component of modern vision applications such as biometrics, surveillance, motion-synthesis and web-based user interfaces. A common requirement among these very different applications is the ability to learn statistical models of appearance and motion from a collection of videos, and then use them for recognizing actions or persons in a new video. These applications in video analysis require statistical inference methods to be devised on non-Euclidean spaces or more formally on manifolds. This chapter outlines a broad survey of applications in video analysis that involve manifolds. We develop the required mathematical tools needed to perform statistical inference on manifolds and show their effectiveness in real video-understanding applications.

1 Introduction

Applications in computer vision often involve the study of geometric scenes and their interplay with physical phenomena such as illumination and motion. When these scenes are imaged using cameras, the observed appearances obey certain mathematical constraints that are induced by the underlying physical constraints. Examples include the observation that images of a convex object under all possible illumination conditions lie on the so called 'illumination-cone' [17]. Images taken under a stereo-pair are constrained by the epipolar geometry of the cameras [22]. Similarly, the 3D pose of the human head is

Pavan Turaga and Rama Chellappa are with the Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 e-mail: {pturaga,rama}@umiacs.umd.edu. Ashok Veeraraghavan is with the Mitsubishi Electric Research Labs, Cambridge MA 02139 e-mail: {veerarag@merl.com}. Anuj Srivastava is with the Department of Statistics, Florida State University, Tallahasse FL e-mail: {anuj@stat.fsu.edu}. This work was partially supported by the Office of Naval Research under the Grant n00014-09-10664.

parameterized by three angles – hence, under constant illumination and expression, the observed face of a human under different viewing directions lies on a three-dimensional manifold. In a particular application, if the physical and mathematical constraints are well-understood, such as in epipolar geometry and illumination modeling, then one can design accurate modeling and inference techniques derived from this understanding.

In several applications of video analysis such as gait-based human ID, activity recognition, shape-based dynamics modeling, and video-based face recognition, some of the constraints that arise have a special form. These special constraints can often be expressed in the form of an equation with some smoothness criterion. Such constraints can be formally defined as manifolds. Before we give precise definitions of what is meant by a manifold, let us first consider some simple problems that illustrate why special attention is needed to study them. To enable this discussion, we shall for the time-being assume that a 'manifold' is defined as a set of points in $\mathbb{R}^n$ that satisfy an equation $f(x) = 0$ (with appropriate conditions on $f()$ that shall be spelt out in a later section). For example, the set of points that satisfy the equation $f(x) = x^T x - 1 = 0$ is the unit hyper-spherical manifold in $\mathbb{R}^n$.

Now one might ask, what is special about these constraints that require new mathematical tools from differential geometry and topology ? Can we not use the classical Euclidean methods and multi-variate statistics, with perhaps some loss of accuracy ? To answer these questions, we will consider a very simple engineering problem. Suppose, a highway construction engineer is laying out a road between two cities which are far apart. Given two cities on the earth, the engineer wants to know a) what is the length of the road required (so as to estimate the amount of building material that needs to be ordered), b) where should a rest-area that is mid-way between the two cities be placed.

Given two points x_1 and x_2 on the earth, he/she would like to compute the shortest distance between them. If the curvature of the earth were not taken into account, and all he/she knew was that the points are in $\mathbb{R}^3$, he/she might choose to use the standard Euclidean norm $\|x_1 - x_2\|$. Unfortunately, this would lead to the engineer underestimating the distance between the two cities. But equipped with the additional knowledge that the earth is well-approximated as a sphere in $\mathbb{R}^3$, we can interpret the Euclidean norm as the 'chordal-length' between these points. The knowledge of this geometry of the constraint set also shows that the Euclidean distance is not *intrinsic* i.e. if we sample points along the shortest straight line path, the samples do not lie on the sphere. This distance is thus meaningless for the engineer since it would require him to lay a tunnel underneath the surface instead of a road on the surface of the earth.

Similarly, given two cities/points that lie on the unit-hypersphere as before, we wish to compute a mean-point where a rest-area may be constructed. Once again, if we did not know the nature of the constraint set, we might use the arithmetic-mean as the mean-point. Now given the extra information that these points need to lie on a hypersphere, it is obvious that the arithmetic mean is not intrinsic either since it does not lie on the hypersphere. The arithmetic mean of these points would lie *under* the surface of the earth rendering it physically meaningless. A much more complicated situation arises when we want to place say 3 rest areas in the midst of 10 cities with some optimality criterion such as reducing the overall length of road to be laid. This requires solving an optimization problem with manifold-valued constraints.

Even though these are fairly simple applications, they illustrate the need to understand the underlying constraints to obtain geometrically meaningful distances and statistics. Naturally, in the presence of such constraints, classical Euclidean geometry fails to provide meaningful solutions. This motivates the need to study such non-Euclidean spaces via methods from differential geometry.

Organization: We begin the chapter by motivating the study of manifold analysis for video processing applications. We then provide an introduction to manifold theory and describe relevant manifolds and provide an introduction to differential geometry on these manifolds. In Section 4, we present methods to perform statistical inference on these manifolds. In Section 5, we present several applications of the presented theory to problems in video understanding.

2 Motivation for studying Manifolds in Video Analysis

Let us first consider some real applications in video understanding that require appreciating the geometry of some non-Euclidean manifolds. Once again, we shall for the time-being assume that a 'manifold' is defined as a set of points in $\mathbb{R}^n$ that satisfy an equation $f(x) = 0$ (with appropriate conditions on $f()$ that shall be spelt out in the next section).

The problem of video understanding can be studied from three widely differing perspectives a) The feature space, b) The model space and c) The transformation space. Even though the specific nature of these spaces can be quite different, a large class of these spaces can be described mathematically as manifolds. Traditionally, 'manifold-learning' methods have been at the forefront of these applications where an analytical characterization of these spaces cannot be found. In the past few years, computer vision researchers have made significant advancement in the analytical and geometric understanding of these varied spaces. This marks an important development in computer vision by moving away from data-driven approaches to geometry-driven approaches for characterizing videos. We provide specific examples of various analytical manifolds found in different applications of computer vision below.

1. **Feature Spaces:** Video understanding typically begins with the extraction of some specific features from the videos. Examples of these features include background subtracted images, shapes, intensity features, motion

vectors etc. These features extracted from the videos might satisfy certain geometric and photometric constraints. The feature space deals with understanding and characterizing the geometry of features that can be extracted from videos. The study of this space then enables appropriate modeling methodologies to be designed. Consider the example of the shape feature. Shapes in images are commonly described by a set of landmarks on the object being imaged. After appropriate translation, scale and rotation normalization it can be shown that shapes reside on a complex spherical manifold. Further, by factoring out all possible affine transformations, it can be shown that shapes reside on a Grassmann Manifold.

2. **Model Spaces:** After features are extracted from each frame of the video, the next step in video analysis, is to describe a sequence of such features using appropriate spatio-temporal models. One specific example of this is modeling the feature sequence as realizations of dynamical systems. Examples include dynamic textures, human joint angle trajectories and silhouette sequences. One popular dynamical model for such time-series data is the autoregressive and moving average (ARMA) model. The space spanned by the columns of the observability matrix of the ARMA model can be identified as a point on the Grassmann manifold. Time-varying and switching linear dynamical systems can then be interpreted as paths on the Grassmann manifold.
3. **Transformation Spaces:** Finally, the transformation space encompasses all possible manifestations of the same semantic activity. The study of this space is important to achieve invariance to factors such as view-changes and execution-rate changes. In this chapter we consider the specific instance of execution-rate variations in human activities, which is modeled as temporal warps of feature trajectories. The space of these warps is the space of positive and monotonically increasing functions mapping the unit-interval to the unit-interval. The derivatives of warping functions can be interpreted as probability density functions. The square-root form of pdf's can then be described as a sphere in the space of functions. Variability in sampling closed planar curves gives rise to variations in observed feature points on shapes. This variability can also be modeled as a sphere in the space of functions (also known as a Hilbert sphere).

As these examples illustrate, manifolds arise quite naturally in several vision-based applications.

2.1 Manifold theory in Vision

There has been an increasing awareness of the need to perform statistical inferences on non-Euclidean domains for a variety of reasons. There has been a significant amount of work in this area in several disciplines. Here, we will review some of these works. This treatment by no means should be considered exhaustive. The use of certain groups, e.g. Euclidean groups, have been fundamental in physics since Einstein and perhaps earlier. The Euclidean motion group plays a fundamental role in rigid body dynamics and uncertainties in modeling dynamic systems have been characterized using probability measures on this group. Another community that has combined the strengths of geometry and statistics is stochastic control [11, 10] where system variables and controls are constrained to be on certain non-Euclidean manifolds.

To the best of our knowledge, the first major effort in using geometry and statistics in pattern recognition was introduced by Ulf Grenander in the early 70s [19, 18]. Grenander created the field of pattern theory which had the following important components: (i) represent the systems of interest using algebraic structures that favor rule-based compositions, (ii) capture variability in these systems using probabilistic super-structures, and (iii) develop efficient algorithms for inferences using geometries of underlying spaces. Over the last three decades, this philosophy has been implemented in a number of contexts with explicit involvement of statistics on non-Euclidean manifolds. We list a few here: The work on analyzing anatomical variability using non-invasive imaging (such as MRI, PET, etc) involved probabilistic structures on high-dimensional deformation groups – this area has recently been labeled as *computational anatomy* [20, 31]. An algebraic pattern theoretic approach has not been exclusive to medical imaging only. It has also been used in addressing computer vision and image analysis problems. For example, in the problem of recognizing objects in images, the variability due to viewing angle of the camera is very important. [21] deals with the problem of estimating the pose as an element of $SO(3)$ and that of bounding the estimating error using statistical bounds. [40] studies the problem of using Markov Chain Monte Carlo methods for performing estimation on some matrix Lie groups e.g. $SO(n)$, and their quotient spaces, e.g. a Grassmann manifold, while [41] studies the problem of subspace tracking (in signal processing) as a problem of nonlinear filtering on a complex Grassmann manifold. While these papers involve statistical inferences on manifolds, there is a strong literature on more general optimization problems. For example, a major work in the area of optimization algorithms on Grassmann and Stiefel manifolds was presented by Edelman et al. [16, 1].

Another prominent area that employed statistical models and inferences on non-Euclidean manifolds is shape analysis. Starting with a trend-setting paper by Kendall [26, 28], there has been a remarkable literature on representing and analyzing shapes of objects, in images or otherwise, using a "landmark-based" approach. In terms of statistical analysis, this is perhaps the most mature area involving manifolds as domains [15, 36]. In more recent years, there has been an extension of Kendall's shape theory to infinite-dimensional representations of shapes of curves and surfaces [27, 39, 32].

The area of statistics and inference on manifolds has seen a large growth in recent years. Many of the ideas have been formally introduced and advanced through the efforts of many researchers. One of the landmark works in establishing mean estimation and central limit theorems for manifold-valued variables is Bhattacharya and Patrangenaru [5, 4]. Another important piece of work comes from Pennec [33] who has applied these notions for detection and classification of anatomical structures in medical images. Recent applications in computer vision have included study of Kendall's shape spaces for human gait analysis [48], and hilbert sphere modeling of time warp functions for human activities in [49]. Other applications include classification over Grassmann manifolds for shape and activity analysis [3, 45], and face recognition [30]. A recently developed formulation of using the covariance of features in image-patches has found several applications such as texture classification [46], and pedestrian detection [47]. Mean-shift clustering was extended to general Riemannian manifolds in [42].

3 Introduction to Manifolds

We shall first start with the topological definition of a manifold in terms of charts and atlases. Using them, we will show that $\mathbb{R}^n$ is indeed a differentiable manifold. Then, we state a theorem that defines sub-manifold of a manifold as a solution of an equation. This shall be specialized to the case of manifolds that are actually sub-manifolds of $\mathbb{R}^n$, arising as solutions of an equation in $\mathbb{R}^n$ with some conditions. Furthermore, we will establish the notions of tangent vectors and tangent spaces on non-Euclidean manifolds. This will then allow the use of classical statistical methods on the tangent planes via the exponential map and its inverse. We shall provide specific examples to illustrate these notions.

3.1 General Background from Differential Geometry

We start by considering the definition of a general differentiable manifold. The material provided here is brief and by no means comprehensive. We refer the interested readers to two excellent books [9][38] for a more detailed introduction to differential geometry and manifold analysis. A topological space M is called a **differentiable manifold** if, amongst other properties, it is *locally Euclidean*. This means that for each $p \in M$, 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; for any point $y \in U$, one can view the Euclidean coordinates $\phi(y) = (\phi_1(y), \phi_2(y), \ldots, \phi_n(y))$ as the coordinates of y. The dimension of the manifold M is n. This is a way of flattening the manifold locally. Using ϕ and ϕ^{-1}, one can move between the sets U and $\phi(U)$ and perform calculations in the more convenient Euclidean space. If there exists multiple such charts, then they are compatible, i.e. their compositions are smooth. We look at the some simple manifolds as examples.

Example 1. **($\mathbb{R}^n$ is a manifold)**

1. The Euclidean space $\mathbb{R}^n$ is an n-dimensional differentiable manifold which can be covered by the single chart $(\mathbb{R}^n, \phi)$, $\phi(x) = x$.
2. Any open subset of a differentiable manifold is itself a differentiable manifold. A well known example of this idea comes from linear algebra. Let $M(n)$ be the set of all $n \times n$ matrices; $M(n)$ can be identified with the set $\mathbb{R}^{n \times n}$ and is, therefore, a differentiable manifold. Define the subset $GL(n)$ as the set of non-singular matrices, i.e. $GL(n) = \{A \in M(n) | \det(A) \neq 0\}$, where $\det(\cdot)$ denotes the determinant of a matrix. Since $GL(n)$ is an open subset of $M(n)$, it is also a differentiable manifold.

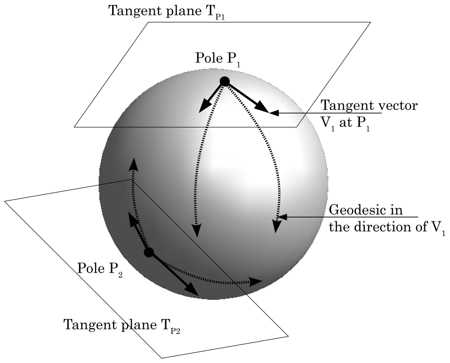


Fig. 1 Figure illustrating the notions of tangent spaces, tangent vectors, and geodesics

In order to perform differential calculus, i.e. to compute gradients, directional derivatives, critical points, etc., of functions on manifolds, one needs to understand the tangent structure of those manifolds. Although there are several ways to define tangent spaces, one intuitive approach is to consider differentiable curves on the manifold passing through the point of interest, and to study the velocity vectors of these curves at that point. To help visualize these ideas, we illustrate the notions of tangent planes, geodesics in figure 1. More formally, let M be an n-dimensional manifold and, for a point $p \in M$, consider a differentiable curve $\gamma : (-\epsilon, \epsilon) \to M$ such that $\gamma(0) = p$. The velocity $\dot{\gamma}(0)$ denotes the velocity of γ at p. This vector has the same dimension as the manifold M itself and is an example of a **tangent vector** to M at p. The set of all such tangent vectors is called the **tangent space** to M at p. Even though the manifold M maybe nonlinear, the tangent space $T_p(M)$ is always linear and one can impose probability models on it using more traditional approaches.

Example 2. 1. In case of the Euclidean space $\mathbb{R}^n$, the tangent space $T_p(\mathbb{R}^n) = \mathbb{R}^n$ for all $p \in \mathbb{R}^n$.

2. For $GL(n)$, the space of non-singular matrices and for an $A \in GL(n)$, let $\gamma(t)$ be a path in $GL(n)$ passing through $A \in GL(n)$ at $t = 0$. The velocity vector at p, $\dot{\gamma}(0)$, is an element of $M(n)$, the set of all $n \times n$ matrices.

Next we introduce the notion of a differential which is important in defining the submanifolds of interest to us. Several of the spaces we will study can be viewed as submanifolds of larger manifolds such as $\mathbb{R}^n$ and $GL(n)$. The differential of a smooth mapping $f : M \to N$ at $p \in M$, denoted by df_p, is a linear map $df_p : T_p(M) \to T_{f(p)}(N)$ specified as follows. Let $g : N \to \mathbb{R}$ be a smooth function. Then, for any $v \in T_p(M)$, define $(df_p(v))(g) = v(f \circ g)(p)$. A point $p \in M$ is said to be a **regular point** if df_p is onto, and its image under f is said to be a **regular value**.

Theorem 1. *Suppose M and N are manifolds of dimensions m and n respectively, and let $f : M \to N$ be a smooth map, with a regular value $y \in N$. Then $f^{-1}(y)$ is a submanifold of M of dimension $m - n$.*

This theorem states that the pullback sets of certain types of points under smooth maps have the submanifold structure. Important examples of such pullback sets include spheres in Euclidean spaces.

Example 3. 1. **Unit Sphere**: Using this theorem, let us check if $\mathbb{S}^n$ is indeed a submanifold of $\mathbb{R}^{n+1}$. Let $f : \mathbb{R}^{n+1} \to \mathbb{R}$ be a map given by $f(p) = \sum_{i=1}^{n+1} p_i^2$, where $p = (p_1, \ldots, p_{n+1})$. The differential of f is given $df_p(u) = 2\langle p, u\rangle$, which is clearly onto for all $p \in f^{-1}(1)$. Thus, 1 is a regular value of f and the set $f^{-1}(1)$ given by $\mathbb{S}^n$ is an n-dimensional submanifold of $\mathbb{R}^{n+1}$. Also, the tangent space $T_p(\mathbb{S}^n)$ is just the orthogonal complement of $p \in \mathbb{R}^{n+1}$.

2. **Orthogonal Matrices**: We now consider the set $O(n)$ of orthogonal matrices, which is a subset of the manifold $GL(n)$. We define $O(n)$ to be the set of all $n \times n$ invertible matrices O that satisfy $OO^T = I$. Define $S(n)$ to be the set of $n \times n$ symmetric matrices, and then define $f : GL(n) \to S(n)$ by $f(O) = OO^T$. It can easily be shown that I is a regular value of f and, hence, $f^{-1}(I) = O(n)$ is a submanifold of $GL(n)$. Note that $O(n)$ is not connected, but has two components: those orthogonal matrices with determinant $+1$, and those with determinant -1. The set of orthogonal matrices with determinant 1 is called the **special orthogonal group**, and denoted by $SO(n)$. The dimension of $O(n)$ can be determined by the above theorem; it is $n^2 - n(n+1)/2 = n(n-1)/2$. One can show that $T_O O(n) = \{OX | X \text{ is an } n \times n \text{ skew-symmetric matrix}\}$.

We now consider the task of measuring distances on a manifold. This is accomplished using a Riemannian metric defined as follows.

Definition 1. A **Riemannian metric** on a differentiable manifold M is a map $\langle \cdot, \cdot \rangle$ that smoothly associates to each point $p \in M$ a symmetric, bilinear, positive definite form on the tangent space $T_p(M)$.

A differentiable manifold with a Riemannian metric on it is called a **Riemannian manifold**.

Example 4. 1. $\mathbb{R}^n$ is a Riemannian manifold with the Riemannian metric $\langle v_1, v_2 \rangle = v_1^T v_2$, the standard Euclidean product.

2. We have earlier examined the manifold $O(n)$ and stated that its tangent space is: $T_O O(n) = \{OX : X \text{ is skew-symmetric}\}$. Define the inner product for any $Y, Z \in T_O O(n)$ by $\langle Y, Z \rangle = trace(YZ^T)$, where *trace* denotes the sum of diagonal elements. With this metric $O(n)$ becomes a Riemannian manifold.
3. Similarly, for the unit sphere $\mathbb{S}^n$ and a point $p \in \mathbb{S}^n$, the Euclidean inner product on the tangent vectors make $\mathbb{S}^n$ a Riemannian manifold. That is, for any $v_1, v_2 \in T_p(\mathbb{S}^n)$, we used the Riemannian metric $\langle v_1, v_2 \rangle = v_1^T v_2$.

Using the Riemannian structure, it becomes possible to define lengths of paths on a manifold. Let $\alpha : [0,1] \mapsto M$ be a parameterized path on a Riemannian manifold M that is differentiable everywhere on $[0,1]$. Then $\frac{d\alpha}{dt}$, the velocity vector at t, is an element of the tangent space $T_{\alpha(t)}(M)$, with length given by $\sqrt{\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \rangle}$. The length of the path α is then given by:

$$L[\alpha] = \int_0^1 \sqrt{\left(\left\langle \frac{d\alpha(t)}{dt}, \frac{d\alpha(t)}{dt} \right\rangle \right)} dt \ . \tag{1}$$

For any two points $p, q \in M$, one can define the distance between them as the infimum of the lengths of all smooth paths on M which start at p and end at q:

$$d(p,q) = \inf_{\{\alpha:[0,1]\mapsto M | \alpha(0)=p, \alpha(1)=q\}} L[\alpha] \ . \tag{2}$$

A path $\hat{\alpha}$ which achieves the above minimum, if it exists, is a **geodesic** between p and q on M.

Example 5. 1. Geodesics on a unit sphere $\mathbb{S}^n$ are great circles [9]. The distance minimizing geodesic between two points p and q is the shorter of the two arcs of a great circle joining them between them. As a parameterized curve, this geodesic is given by:

$$\alpha(t) = \frac{1}{\sin(\theta)} \left[\sin(\theta - t)p + \sin(t)q \right] \tag{3}$$

where $\theta = \cos^{-1}(\langle p, q \rangle)$.

2. To define geodesics on $SO(n)$, we introduce the notion of matrix exponential. For a matrix $A \in M(n)$, define its matrix exponential $\exp(A)$ by:

$$\exp(A) = I + \frac{A}{1!} + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots \tag{4}$$

Using the matrix exponential, one can define geodesics on $SO(n)$ (with respect to the Riemannian metric defined earlier) as follows: for any $O \in SO(n)$ and any skew-symmetric matrix X, $\alpha(t) \equiv O \exp(tX)$, is the unique geodesic in $SO(n)$ passing through O with velocity OX at $t = 0$ [9].

An important tool in studying statistics on a manifold is an exponential map. If M is a Riemannian manifold and $p \in M$, the **exponential map**

$\exp_p : T_p(M) \to M$, is defined by $\exp_p(v) = \alpha_v(1)$ where α_v is a constant speed geodesic whose velocity vector at p is v. For $\mathbb{R}^n$, under the Euclidean metric, since geodesics are given by straight lines, the exponential map is a simple addition: $\exp_p(v) = p + v$, for $p, v \in \mathbb{R}^n$. The exponential map on a sphere, $\exp : T_p(\mathbb{S}^n) \mapsto \mathbb{S}^n$, is given by $\exp_p(v) = \cos(\|v\|)p + \sin(\|v\|)\frac{v}{\|v\|}$. In case of $SO(n)$, the exponential is given by $\exp_O(X) = O\exp(X)$, where the exponential on the right side is defined in Eqn. 4. We illustrate the notions of the exponential map in figure 2.

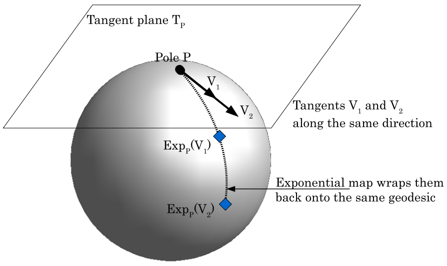

Tangent plane T$_P$
Pole P
V$_1$
V$_2$
Tangents V$_1$ and V$_2$ along the same direction
Exp$_P$(V$_1$)
Exponential map wraps them back onto the same geodesic
Exp$_P$(V$_2$)

Fig. 2 Figure illustrating the notion of exponential maps and inverse exponential maps.

3.2 Special Manifolds of Interest

We are interested in quotient spaces of the special orthogonal group $SO(n)$ studied earlier. We start by introducing the notion of a quotient space of a group. A group G is a set having an associative binary operation, denoted by $\cdot$, such that: (i) there is an identity element e in G, and (ii) for each element, there exists a unique inverse. Let H be a subgroup of G. For any element $g \in G$, define a left **coset** of H in G by $gH = \{g \cdot h \big| h \in H\}$. In general, the cosets are not subgroups and the only coset that is a subgroup of G is H itself (eH). For different elements g_1 and g_2, the cosets g_1H and g_2H will either be identical or disjoint. They will be identical when $g_2^{-1}g_1$ is an element of H; otherwise they will be disjoint. This is similar to an equivalence relation that partitions a set into disjoint equivalence classes. In fact, one can define an equivalence relation using membership of these cosets: we define $g_1 \sim g_2$ if $g_1 \in g_2H$, i.e. $g_1 = g_2h$ for some $h \in H$. In the notation of equivalence classes, we have $[g] = gH$. The quotient space $G/\sim$, also denoted by G/H to emphasize the role of H in defining $\sim$, is the set of all left cosets of H in G. The quotient space G/H is also called the space G *modulo* H, or the space that results when H *is removed from* G.

Now we consider three specific manifolds that are important in our analysis of features in videos.

1. **Stiefel Manifold**: Let the set of all $n \times d$ orthogonal matrices be $\mathcal{S}_{n,d}$,

$$\mathcal{S}_{n,d} = \{U \in \mathbb{R}^{n\times d} | U^T U = I_d\} \quad \subset GL(n,d). \tag{5}$$

$\mathcal{S}_{n,d}$ is called a **Stiefel** manifold, and each element of $\mathcal{S}_{n,d}$ provides an orthonormal basis for a d-dimensional subspace of $\mathbb{R}^n$. $\mathcal{S}_{n,d}$ can also be viewed as a quotient space of $SO(n)$ as follows. First, consider $SO(n-d)$ as a subgroup of $SO(n)$ using the embedding: $\phi_1 : SO(n-d) \mapsto SO(n)$, defined by

$$\phi_1(V) = \begin{bmatrix} I_d & 0 \\ 0 & V \end{bmatrix} \quad \in \quad SO(n)\ . \tag{6}$$

With this embedding, we can generate left cosets of $SO(n)$: for an $O \in SO(n)$, a coset is given by $O\phi_1(S(n-d))$. This defines an equivalence relation $\sim$ in $SO(n)$ according to: for $Q_1,\ Q_2 \in SO(n)$,

$$Q_1 \sim Q_2, \quad \text{if and only if} \quad Q_1 = Q_2\phi_1(V), \quad \text{for some} \quad V \in SO(n-d)\ .$$

In other words, $Q_1 \sim Q_2$ if and only if their first d columns are identical, irrespective of the remaining columns. Therefore, $\mathcal{S}_{n,d}$ can be viewed as the quotient space

$$\mathcal{S}_{n,d} = SO(n)/\sim \quad \text{or} \quad SO(n)/\phi_1(SO(n-d)) \quad \text{or simply} \quad SO(n)/SO(n-d)\ .$$

2. **Grassmann Manifold**: If one is interested only in the subspace spanned by the columns of U, and not in a particular basis, then the required space is reduced further. Let $SO(d) \times SO(n-d)$ be a subset of $SO(n)$ using the embedding $\phi_2 : (SO(d) \times SO(n-d)) \mapsto SO(n)$:

$$\phi_2(V_1, V_2) = \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} \quad \in \quad SO(n)\ , \quad V_1 \in SO(d), \quad V_2 \in SO(n-d). \tag{7}$$

As for $\mathcal{S}_{n,d}$, define an equivalence relation as a coset of $SO(n)$ generated by the subgroup $\phi_2(SO(d) \times SO(n-d))$ and let $\mathcal{G}_{n,d}$ be the quotient space $SO(n)/\phi_2(SO(d) \times SO(n-d))$, or simply $SO(n)/(SO(d) \times SO(n-d))$. We will use the square-brackets to denote elements of $\mathcal{G}_{n,d}$:

$$[U] = \{UO | U \in \mathcal{S}_{n,d},\ O \in SO(d)\}\ .$$

3. **Kendall's Shape Manifold**: Kendall [25] provided a mathematical theory for the description of landmark based shapes. Bookstein [8] and later Dryden and Mardia [14] have furthered the understanding of such landmark based shape descriptions. Kendall's representation of shape describes the shape configuration of n landmark points in an d-dimensional space as a $n \times d$ matrix containing the coordinates of the landmarks. Pre-shape is the geometric information that remains when location and scale effects are filtered out. Let the configuration of a set of n landmark points be given by a n-dimensional complex vector containing the positions of landmarks. Let us denote this configuration as X. The centered pre-shape is obtained by subtracting the mean from the configuration and then scaling to norm

one. The centered pre-shape is given by

$$Z_c = \frac{CX}{\parallel CX \parallel}, \quad where \quad C = I_n - \frac{1}{n} 1_n 1_n^T, \tag{8}$$

where I_n is a $n \times n$ identity matrix and 1_n is a n-dimensional vector of ones. The pre-shape vector that is extracted by the method described above lies on a spherical manifold. Let us denote this pre-shape space as $\mathcal{P}_{n,d}$. The shape space is now the quotient of the preshape space obtained by removing all rotations of the shape i.e. $\mathcal{P}_{n,d}/SO(d)$.

Example 6. **(Kendall Shape Metrics)** Several distance metrics have been defined in [14] to measure distances between shapes using the Kendall's shape representation. Here, we shall describe some of them. Consider two complex configurations X and Y with corresponding preshapes α and β. The full Procrustes distance between the configurations X and Y is defined as the Euclidean distance between the full Procrustes fit of α and β. Full Procrustes fit is chosen so as to minimize

$$d(Y, X) = \parallel \beta - \alpha s e^{j\theta} - (a + jb)1_n \parallel, \tag{9}$$

where s is a scale, θ is the rotation and $(a + jb)$ is the translation. The Full Procrustes distance is the minimum Full Procrustes fit i.e.,

$$d_{Full}(Y, X) = \inf_{s,\theta,a,b} d(Y, X). \tag{10}$$

We note that the preshapes are actually obtained after filtering out effects of translation and scale. Hence, the translation value that minimizes the full Procrustes fit is given by $(a + jb) = 0$, while the scale $s \ = 1$. The rotation angle θ that minimizes the Full Procrustes fit is given by $\theta \ = arg(|\alpha^* \beta|)$. The partial Procrustes distance between configurations X and Y is obtained by matching their respective preshapes α and β as closely as possible over rotations, but not scale. So,

$$d_{Partial}(X, Y) = \inf_{\mathit{\Gamma} \epsilon SO(d)} \parallel \beta - \alpha \mathit{\Gamma} \parallel . \tag{11}$$

It is interesting to note that the optimal rotation θ is the same whether we compute the full Procrustes distance or the partial Procrustes distance. The Procrustes distance $\rho(X, Y)$ is the closest great circle distance between α and β on the preshape sphere. The minimization is done over all rotations. Thus ρ is the smallest angle between complex vectors α and β over rotations of α and β. The three distance measures defined above are all trigonometrically related as

$$d_{Full}(X, Y) = \sin \rho(X, Y), \quad d_{Partial}(X, Y) = 2 \sin(\frac{\rho(X, Y)}{2}). \tag{12}$$

3.2.1 Tangent Structure of the Special Manifolds

If M/H is a quotient space of M under the action of a group $H \subset M$ (assuming H acts on M), then, for any point $p \in M$, a vector $v \in T_p(M)$ is also tangent to M/H as long as it is perpendicular to the tangent space $T_p(pH)$. Here, $T_p(pH)$ is considered as a subspace of $T_p(M)$. We will use this idea to find tangent spaces on $\mathcal{S}_{n,d}$, $\mathcal{G}_{n,d}$, and Kendall's shape space, using the tangent structure of $SO(n)$.

1. **Tangent Structure of $\mathcal{S}_{n,d}$**: Since $\mathcal{S}_{n,d} = SO(n)/\phi_1(SO(n-d))$, set $M = SO(n)$ and $H = \phi_1(SO(n-d))$, with ϕ_1 as defined in Eqn. 6. Let $J \in \mathbb{R}^{n \times d}$ be a tall-skinny matrix, made up of the first d columns of I_n; J acts as the "identity" element in $\mathcal{S}_{n,d}$. A vector in $T_{I_n}(SO(n))$, that is perpendicular to $T_{\phi_1(I_{n-d})}(I_n SO(n-d))$, when multiplied on right by J results in a tangent to $\mathcal{S}_{n,d}$ at J. This gives:
$$T_J(\mathcal{S}_{n,d}) = \{ \begin{bmatrix} C \\ -B^T \end{bmatrix} | C = -C^T, C \in \mathbb{R}^{d \times d}, B \in \mathbb{R}^{d \times (n-d)} \} \ . \tag{13}$$
For any other point $U \in \mathcal{S}_{n,d}$, let $Q \in SO(n)$ be a matrix that rotates the columns of U to align with the columns of J, i.e. let $U = Q^T J$. Note that the choice of Q is not unique. It follows that the tangent space at U is given by: $T_U(\mathcal{S}_{n,d}) = \{Q^T G | G \in T_J(\mathcal{S}_{n,d})\}$.
2. **Tangent Structure of $\mathcal{G}_{n,d}$**: In this case, set $M = SO(n)$ and $H = \phi_2(SO(d) \times SO(n-d))$, with ϕ_2 as given in Eqn. 7. Using the same argument made before, the vectors tangent to $SO(n)$ and perpendicular to the space $(T_{I_d}(SO(d)) \times T_{I_{n-d}}(SO(n-d)))$, will also be tangent to $\mathcal{G}_{n,d}$ after multiplication on right by J. Thus, the tangent space at $[J] \in \mathcal{G}_{n,d}$ is given by:
$$T_{[J]}(\mathcal{G}_{n,d}) = \{ \begin{bmatrix} 0 \\ -B^T \end{bmatrix} \ | \ B \in \mathbb{R}^{d \times (n-d)} \} \tag{14}$$
For any other point $[U] \in \mathcal{G}_{n,d}$, let $Q \in SO(n)$ be a matrix such that $U = Q^T J$. Then, the tangent space at $[U]$ is given by $T_U(\mathcal{G}_{n,d}) = \{Q^T G | G \in T_J(\mathcal{G}_{n,d})\}$.
3. **Tangent Structure of Kendall's Shape Space**: The pre-shape formed by n points lie on a $n-1$ dimensional complex hypersphere of unit radius. The Procrustes tangent coordinates of a preshape α are given by
$$v(\alpha, \mu) = \alpha \alpha^* \mu - \mu |\alpha^* \mu|^2. \tag{15}$$
where μ is the Procrustes mean shape of the data.

So far we have introduced several manifolds of interest – namely $\mathbb{S}^n$, $\mathcal{S}_{n,d}$ and $\mathcal{G}_{n,d}$ – and have defined their geometries, including their tangent spaces, Riemannian metrics, geodesics and exponential maps. Now we consider the task of studying statistics on these manifolds.

4 Statistical Inference on Manifolds

What are the challenges in performing a statistical analysis if the underlying state space is non-Euclidean? Take the case of the simplest statistic, the sample mean, for a sample set $(x_1, x_2, \ldots, x_n)$ on $\mathbb{R}^n$:

$$\bar{x}_k = \frac{1}{k} \sum_{i=1}^{k} x_i, \qquad x_i \in \mathbb{R}^n \ . \tag{16}$$

Since $\bar{x}_k$ is a widely used and studied statistic, one already knows the pros and cons of using $\bar{x}_k$ as an estimate of the population mean. For example, we know that $\bar{x}_k$ is an unbiased and efficient estimator, but it is susceptible to the outliers. Now what if the underlying space is not $\mathbb{R}^n$ but a non-Euclidean manifold instead? To answer this question we consider an n-dimensional Riemannian manifold M. Let $d(p,q)$ denote the length of the shortest geodesic between arbitrary points $p, q \in M$. To facilitate a general discussion, we will assume that there exists an embedding $\varepsilon : M \to V$ where V is an m-dimensional Hilbert space ($n \leq m$). We have chosen V to be a vector space so that we can perform a statistical analysis in V using standard techniques from multivariate calculus. The distance between any two elements $p, q \in M$ is the geodesic distance $d(p,q)$ when the geodesic is restricted to be in M and it is $\|\varepsilon(p) - \varepsilon(q)\|$, with the norm of V, when the geodesic is allowed to be in V. The latter distance, of course, depends on the choice of the embedding ε. We start the analysis by assuming that we are given a probability density function f on M. This function, by definition, satisfies the properties that $f : M \to \mathbb{R}_{\geq 0}$ and $\int_M f(p)dp = 1$, where dp denotes the reference measure on M with respect to which the density f is defined. We can extend f to the larger set V by simply setting:

$$\tilde{f}(x) = \begin{cases} f(p) & \text{if } \ x = \varepsilon(p), \ p \in M \\ 0 & \text{if } \ x \notin \varepsilon(M) \end{cases} . \tag{17}$$

Naturally, $\tilde{f}$ is a probability density function on V. There are two possibilities for computing statistics on M − intrinsic and extrinsic. We describe them next.

4.1 Intrinsic Statistics

The first question that we consider is: What is a suitable notion of mean on the Riemannian manifold M? A popular method for defining a mean on a manifold was proposed by Karcher [24] who used the centroid of a density as its mean.

Definition 2 (Karcher Mean [24]). The Karcher mean μ_{int} of a probability density function f on M is defined as local minimizer of the cost function: $\rho : M \to \mathbb{R}_{\geq 0}$, where

$$\rho(p) = \int_M d(p,q)^2 f(q)\ dq\ . \tag{18}$$

dq denotes the reference measure used in defining the probability density f on M. The value of function ρ at the Karcher mean is called the **Karcher variance**. How does the definition of Karcher mean adapt to the sample set, i.e. a finite set of points drawn from an underlying probability distribution? Let $q_1, q_2, \ldots, q_k$ be independent random samples from the density f. Then, the sample Karcher mean of these points is defined to be the local minimizer of the function:

$$\rho_k(p) = \frac{1}{k} \sum_{i=1}^{k} d(p, q_i)^2\ . \tag{19}$$

An iterative algorithm for computing the sample Karcher mean is as follows. Let μ_0 be an initial estimate of the Karcher mean. Set $j = 0$.

1. For each $i = 1, \ldots, k$, compute the tangent vector v_i such that the geodesic from μ_j, in the direction v_i, reaches q_i at time one, i.e. $\psi_1(\mu_j, v_i) = q_i$ or $v_i = \exp_{\mu_j}^{-1}(q_i)$.
2. Compute the average direction $\bar{v} = \frac{1}{k}\sum_{i=1}^{k} v_i$.
3. If $\|\bar{v}\|$ is small, then stop. Else, update μ_j in the update direction using

   $$\mu_{j+1} = \psi_\epsilon(\mu_j, \bar{v}),$$

   where $\epsilon > 0$ is small step size, typically 0.5. $\psi_t(p, v)$ denotes the geodesic path starting from p in the direction v parameterized by time t. In other words, $\mu_{j+1} = \exp_{\mu_j}(\epsilon\bar{v})$.
4. Set $j = j + 1$ and return to Step 1.

It can be shown that this algorithm converges to a local minimum of the cost function given in Eqn. 19 which by definition is μ_{int}. Depending upon the initial value μ_0 and the step size ϵ, it converges to the nearest local minimum.

We exploit the fact that the tangent spaces of M are vector spaces and can provide a domain for defining covariances. We can transfer the probability density f from M to a tangent space $T_p(M)$, using the inverse exponential map, and then use the standard definition of central moments in that vector space. For any point $p \in M$, let $p \to v \equiv \exp_\mu^{-1}(p)$ denote the inverse exponential map at μ from M to $T_\mu(M)$. The point μ maps to the origin $\mathbf{0} \in T_\mu(M)$ under this map. Now, we can define the Karcher covariance matrix as:

$$K_{int} = \int_{T_\mu(M)} vv^T f_v(v) dv, \quad v = \exp_\mu^{-1}(q)\ ,$$

where f_v is the induced probability density on the tangent space. For a finite sample set, the sample Karcher variance is given by

$$\hat{K}_{int} = \frac{1}{k-1} \sum_{i=1}^{k} v_i v_i^T, \quad \text{where} \quad v_i = \exp_\mu^{-1}(q_i) \ . \tag{20}$$

4.2 Extrinsic Statistics

The other possibility for performing statistics is to use the vector space structure of V to simplify calculations. In this case one transfers the probability measure to V, computes the pertinent statistical quantities in V and projects the final results back to M. Let $\Pi : V \to M$ be a projection map defined in such a way that

$$\Pi(v) = argmin_{p \in M} \|v - \varepsilon(p)\|^2 \ . \tag{21}$$

The existence and the uniqueness of Π, of course, depend on the nature of M, p and ε. Now, the extrinsic mean of a density f on M is defined as follows.

Definition 3 (Extrinsic Mean). The extrinsic mean of density f on M, specified with respect to an embedding ε of M in a larger vector space V, is given by

$$\mu_{ext} = \Pi(\nu),$$

where:

- Π is the projection defined in Eqn. 21,
- $\nu = \int_V v\tilde{f}(v)dv$ is the standard mean of $\tilde{f}$ in V, and
- $\tilde{f}$ is the unique extension of f from M to V (given by Eqn. 17).

Once the embedding ε has been chosen, and a mechanism for projection Π has been established, the rest of the process is quite straightforward. It requires computing the mean of $\tilde{f}$ in V and projecting it down to M. In case M is a Euclidean space, the projection is simply the identity operation and the extrinsic mean coincides with the classical mean. Additionally, in this case, if the Euclidean metric is chosen as the Riemannian metric, then the intrinsic mean also coincides with the classical mean.

What about the covariance analysis in an extrinsic framework? An extrinsic covariance can be defined similar to the extrinsic mean. Let $\pi : V \to T_\nu(M)$ be any linear map. Since it is a linear map, it can be written as a $n \times m$ matrix A so that $\pi(v) = Av$. Define the covariance

$$K_v = \int_V (v - \nu)(v - \nu)^t \tilde{f}(v)dv \ ,$$

in the vector space V and project it using:

$$K_{ext} = AK_vA^T \ . \tag{22}$$

The advantages and disadvantages of an extrinsic mean, with respect to the Karcher mean, are straightforward. The main advantage is its computational simplicity. Once an embedding ε is chosen, the rest of the analysis is quite standard and typically very fast. In contrast, computation of the Karcher mean requires repeated computations of the exponential and inverse exponential maps. The disadvantage is that the result $\Pi(\nu)$ depends on the choice of embedding ε which is quite arbitrary. Different embeddings will result in different solutions, and the projection Π itself may not be unique.

Example 7. **(Extrinsic Mean of Subspaces)** As discussed in section 3.2, the Grassmann manifold can be viewed as a quotient space of the set of full-rank $n \times d$ orthonormal matrices. We can also associate to each d-dimensional subspace an $n \times n$ idempotent projection matrix P of rank d (not to be confused with the projection operation Π), such that $P = YY^T$, where Y is a point on the $\mathcal{S}_{n,d}$ whose columns span the subspace. The space of $n \times n$ projectors of rank d, denoted by $\mathbb{P}_{n,d}$ can be embedded into the set of all $n \times n$ matrices – $\mathbb{R}^{n\times n}$ – which is a vector space. The projection Π from $\mathbb{R}^{n\times n}$ to $\mathbb{P}_{n,d}$ is given by

$$\Pi(M) = UU^T, \text{where } M = USV^T \text{ is the d-rank SVD of } M. \tag{23}$$

Using this embedding, we can define an extrinsic distance metric on the Grassmann manifold using the distance metric inherited from $\mathbb{R}^{n\times n}$.

$$d^2(P_1, P_2) = tr(P_1 - P_2)^T(P_1 - P_2) \tag{24}$$

Given a set of sample points on the Grassmann manifold represented uniquely by projectors $\{P_1, P_2, \ldots, P_N\}$, we can compute the extrinsic mean by first computing the mean of the P_i's and then projecting the solution to the manifold by means of equation (23). i.e.

$$\mu_{ext} = \Pi(P_{avg}), \text{ where } P_{avg} = \frac{1}{N}\sum_{i=1}^{N} P_i \tag{25}$$

4.3 Learning Distributions from Data

In addition to sample statistics such as the mean and covariance, it is possible to define parametric probability distribution functions on manifolds. The intrinsic distributions are defined on the manifolds of interest directly without embedding them into a vector space. Examples of such distributions include the Langevin distribution for spherical data. Another intrinsic way of defining probability distributions is to project parametric distributions onto the manifold of interest. In addition to intrinsic methods such as these, we can estimate extrinsic distributions as well.

Example 8. **(Intrinsic Density Estimation)** Suppose, we have n sample points, given by $q_1, q_2, ...q_n$ from a manifold $\mathcal{M}$. Then, we first compute their Karcher mean $\bar{q}$ as discussed before. The next step is to define and compute a sample covariance for the observed q_i's. The key idea here is to use the fact that the tangent space $T_{\bar{q}}(q)$ is a vector space. For a d-dimensional manifold, the tangent space at a point is also d dimensional. Using a finite-dimensional approximation, say $V \subset T_{\bar{q}}(q)$, we can use the classical multivariate calculus for this purpose. The resulting sample covariance matrix is given by:

$$\bar{\Sigma} = \frac{1}{n-1} \sum_{i=1}^{n} v_i v_i^T$$

where each v_i is a d-dimensional sample of the function $exp_{\bar{q}}^{-1} q_i$. Note that by definition, the mean of v_is should be zero. In cases where the number of samples n is smaller than d, one can apply an additional dimension-reduction tool to work on a smaller space. For instance, we can use the singular value decomposition (SVD) of the sample covariance matrix $\bar{\Sigma}$ and retain only the top m significant singular values and the corresponding singular vectors. In such cases, the covariance matrix is indirectly stored using $\lambda_1, \lambda_2, ...\lambda_m$ singular values and their corresponding singular vectors $u_1, u_2, ...u_m$.

The exponential map: $\exp_{\bar{q}} : T_{\bar{q}}(q) \rightarrow \mathcal{M}$ maps this covariance back to $\mathcal{M}$. Specifically, this approach is widely used to define wrapped-Gaussian densities on a given manifold. In general, one can define arbitrary pdfs on the tangent plane such as mixtures of Gaussians, Laplace etc and project it back to the manifold via the exponential map. This allows us to experiment with and choose an appropriate pdf that works well for a given problem domain.

Example 9. **(Extrinsic Densities using Kernels)** Here we discuss density estimation over the Grassmann manifold using extrinsic methods proposed by [12]. Given two orthonormal bases Y_1 and Y_2 we define the distance between the subspaces as the smallest squared Euclidean distance between their corresponding equivalence classes on the Stiefel manifold. Hence,

$$d^2([Y_1],[Y_2]) = \min_{R \in SO(d)} tr(Y_1 - Y_2 R)^T (Y_1 - Y_2 R) \tag{26}$$

This distance is called the Procrustes distance [12]. This minimization can be solved in closed form. It is possible to relax the constraint that $R \in SO(d)$ to $R \in GL(d)$. In this case, the minimum is attained at $R = A$ and the distance is given by $d^2(Y_1, Y_2) = tr(I_k - A^T A)$, where $A = Y_1^T Y_2$. We refer the reader to [12] for derivations and other cases. Using this interpretation, we can define extrinsic statistics on the Grassmann manifold. Here, we discuss a non-parametric method for estimation of pdfs. Given several samples from a pdf, represented by orthonormal basis $(Y_1, Y_2, \ldots, Y_n)$, the density can be estimated using extrinsic methods and the Procrustes metric [12] as

$$\hat{f}(Y;M) = \frac{1}{n}C(M)\sum_{i=1}^{n} K[M^{-1/2}(I_k - Y_i^T Y Y^T Y_i)M^{-1/2}] \quad (27)$$

where $K(T)$ is the kernel function, M is a $k \times k$ positive definite matrix which plays the role of the kernel width or a smoothing parameter. $C(M)$ is a normalizing factor chosen so that the estimated density integrates to unity.

5 Applications and Experiments

In this section, we present several examples where an understanding of the manifold that the data lies on can provide a principled means of solving the problem. The examples we discuss include 1. human gait analysis, 2. activity analysis via state-space modeling and 3. modeling execution-rate variations in human activities.

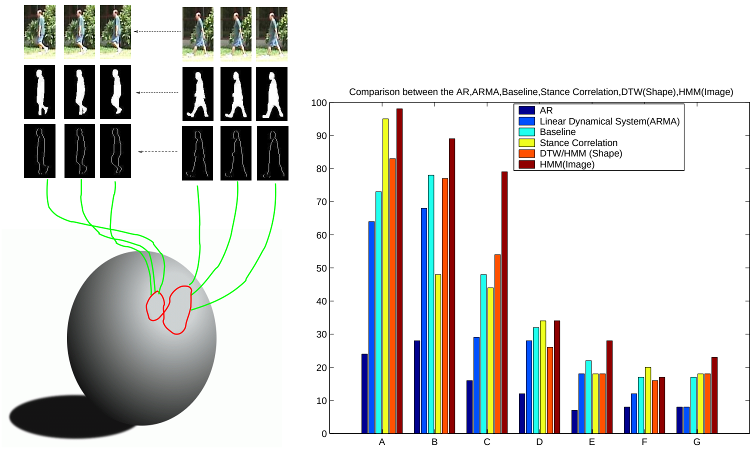
5.1 Feature Space Manifold: Kendall's Shape Sphere for Human Gait Analysis

Shape analysis plays a very important role in object recognition, matching and registration. There has been substantial work in shape representation and on defining a feature vector which captures the essential attributes of the shape. A description of shape must be invariant to translation, scale and rotation. The Kendall's shape space is a natural feature to use in such cases. Given a binary image consisting of the silhouette of a person, we extract the shape from this binary image. The procedure for obtaining shapes from the video sequence is graphically illustrated in Figure 3(a). Note that each frame of the video sequence maps to a point on the spherical shape manifold.

Consider a situation where there are two shape sequences and we wish to compare how similar these two shape sequences are. One may want to use non-parametric sequence matching such as Dynamic-Time warping or a parametric approach such as state-space modeling. In either case, we need to take into account the geometry of the shape-manifold for matching. Consider dynamic time warping, which has been successfully used by the speech recognition [34] community for performing non-linear time normalization. Pre-shape, as we have already discussed lies on a spherical manifold. In our experiments, we use the Procrustes shape distance described in section 3.2 during the DTW distance computations. For state-space modeling such as autoregressive (AR) or ARMA, we use the tangent structure of the manifold. We project a given sequence to the tangent plane constructed at the mean-point. The AR and ARMA model parameters are then estimated on the tangent-planes. The tangent structure for Kendall's shape manifold was discussed in 3.2.1. Once the model parameters are estimated, computing similarity between two sequences can be performed by computing the distance between the model parameters. We refer the reader to [48] for details of model fitting and computing similarity between the model-parameters. Next, we present some experiments that demonstrate the utility of these methods.

5.1.1 Gait Recognition Experiment on the USF Gait Database

The USF database [35] consists of 71 people in the Gallery. Various covariates such as camera position, shoe type, surface and time were varied in a controlled manner to design a set of challenge experiments[1][35]. On the USF database we conducted experiments on recognition performance using these methods- Stance Correlation, DTW on shape space, Stance based AR (a slight modification of the AR model [48]) and the ARMA model. Gait recognition experiments were designed for challenge experiments A-G. These experiments featured and tested the recognition performance against various covariates like the camera angle, shoe type, surface change etc. Refer to [35] for a detailed description of the various experiments and the covariates in these experiments. Figure 3(b) shows a comparison of the identification rate (rank 1) of the various shape and kinematics based algorithms. It is clearly seen that shape-based algorithms perform better than purely kinematics-based algorithms.



(a) (b)

Fig. 3 (a)Graphical illustration of the sequence of shapes obtained during a walking cycle, (b)Bar Diagram comparing the identification rate of various algorithms.

[1] Challenge Experiments:Probes A-G in increasing order of difficulty.

5.2 Model Space Manifold: Grassmann manifold for Human Activity Analysis

Modeling of human activities is an important problem in video-understanding. Applications of activity recognition include activity-based indexing, biometrics, motion synthesis, and anomaly detection. Human activity analysis typically proceeds in a hierarchical fashion. At lower-levels, some features pertaining to motion of the human are extracted from video sequences such as optical flow or background subtracted masks. Then, a model is imposed on the feature evolution such as Hidden Markov Models (HMMs) or Linear Dynamic Systems (LDS). Given training data, the goal is to estimate the model parameters. Here we study ARMA models and show that the study of these models can be formulated as a study of the geometry of the Grassmann manifold. A wide variety of time series data such as dynamic textures, human joint angle trajectories, shape sequences, video based face recognition etc are frequently modeled as ARMA models [37, 6, 48, 2]. The ARMA model equations are given by

$$f(t) = Cz(t) + w(t) \quad w(t) \sim N(0, R) \tag{28}$$

$$z(t+1) = Az(t) + v(t) \quad v(t) \sim N(0, Q) \tag{29}$$

where, z is the hidden state vector, A the transition matrix and C the measurement matrix. f represents the observed features while w and v are noise components modeled as normal with 0 mean and covariance R and Q respectively.

The model parameters (A, C) learned as above do not lie on a Euclidean space. The transition matrix A is constrained to be stable with eigenvalues inside the unit circle. The observation matrix C is constrained to be an orthonormal matrix. Now, starting from an initial condition $z(0)$, it can be shown that the *expected* observation sequence is given by

$$E \begin{bmatrix} f(0) \\ f(1) \\ f(2) \\ . \\ . \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ . \\ . \end{bmatrix} z(0) = O_\infty(M) z(0) \tag{30}$$

Thus, the expected observation sequence generated by a time-invariant model (A, C) lies in the column space of the extended *observability* matrix given by $O_\infty = [C^T, (CA)^T, (CA^2)^T, ...]^T$. However, motivated by the fact that human actions are of a finite-duration in time and not infinitely extending in time, we can simplify the study of the model by considering only an n-length expected observation sequence instead of the infinite sequence as above. Let the size of the temporal window be n. Thus, the n-length expected observation sequence generated by the model (A, C) lies in the column space of the *finite* observability matrix given by

$$O_n^T = \left[C^T, (CA)^T, (CA^2)^T, \dots (CA^{n-1})^T\right] \tag{31}$$

We can thus identify a dynamical model by a point on the Grassmann manifold, corresponding to the subspace spanned by the columns of the observability matrix. Since, the geometry of the Grassmann manifold is known we can use its geometry as discussed in sections 3.2 and 3.2.1 to define distances, exponential maps, and statistics (section 4) for video classification.

5.2.1 INRIA iXMAS Activity Recogntion Experiment

We performed a recognition experiment on the publicly available INRIA dataset [50]. The dataset consists of 10 actors performing 11 actions, each action executed 3 times at varying rates while freely changing orientation. We used the view-invariant representation and features as proposed in [50]. Specifically, we used the $16 \times 16 \times 16$ circular FFT features proposed by [50]. Each activity was modeled as a linear dynamical system. Testing was performed using a round-robin experiment where activity models were learnt using 9 actors and tested on 1 actor. In table 1, we show the recognition results obtained using four methods. The first column shows the results obtained using dimensionality reduction approaches of [50] on $16 \times 16 \times 16$ features. [50] reports recognition results using a variety of dimensionality reduction techniques (PCA, LDA, Mahalanobis) and here we choose the row-wise best performance from their experiments (denoted 'Best Dim. Red.') which were obtained using $64 \times 64 \times 64$ circular FFT features. The third column presents results using the method of using subspace angles based distance between dynamical models [13]. This is closely related to the geodesic on the Grassmann manifold for finite observability matrices. Column 4 shows the nearest-neighbor classifier performance using Procrustes metric on the Grassmann manifold ($16 \times 16 \times 16$ features). We see that the manifold Procrustes distance performs as well as subspace angles. But, statistical modeling of class conditional densities for each activity using parametric and non-parametric methods, leads to a significant improvement in recognition performance. In addition to activity analysis and ARMA modeling, we refer the reader to [45] for more example applications of statistical modeling on the Grassmann manifold in computer vision applications.

5.2.2 Activity based Summarization

The ARMA model described above in conjunction with statistical models on the Grassmann manifold can be used to summarize long videos. Towards this purpose, we describe long videos as outputs of time-varying ARMA models given by

Activity	**Dim. Red.** [50] 16^3 volume	**Best Dim. Red.** [50] 64^3 volume	**Subspace Angles** 16^3 volume	**Procrustes Metric** 16^3 volume	**Wrapped Normal** 16^3 volume	**Extrinsic Kernel** 16^3 volume
Check Watch	76.67	86.66	93.33	90	100	100
Cross Arms	100	100	100	96.67	96.67	100
Scratch Head	80	93.33	76.67	90	100	96.67
Sit Down	96.67	93.33	93.33	93.33	90	93.33
Get Up	93.33	93.33	86.67	80	96.67	96.67
Turn Around	96.67	96.67	100	100	96.67	100
Walk	100	100	100	100	100	100
Wave Hand	73.33	80	93.33	90	90	100
Punch	83.33	96.66	93.33	83.33	100	100
Kick	90	96.66	100	100	93.33	100
Pick Up	86.67	90	96.67	96.67	93.33	100
Average	88.78	93.33	93.93	92.72	96.06	98.78

Table 1 Comparison of view invariant recognition of activities in the INRIA dataset using a) Best DimRed [50] on $16 \times 16 \times 16$ features, b) Best Dim. Red. [50] on $64 \times 64 \times 64$ features, c) Nearest Neighbor using Subspace angles ($16 \times 16 \times 16$ features) d) Nearest Neighbor using Procrustes distance ($16 \times 16 \times 16$ features), e) Maximum likelihood using wrapped Gaussian($16 \times 16 \times 16$ features) f) Maximum likelihood using Parzen windows on the Grassmann manifold ($16 \times 16 \times 16$ features)

$$f(t) = C(t)z(t) + w(t) \quad w(t) \sim N(0, R(t)) \tag{32}$$

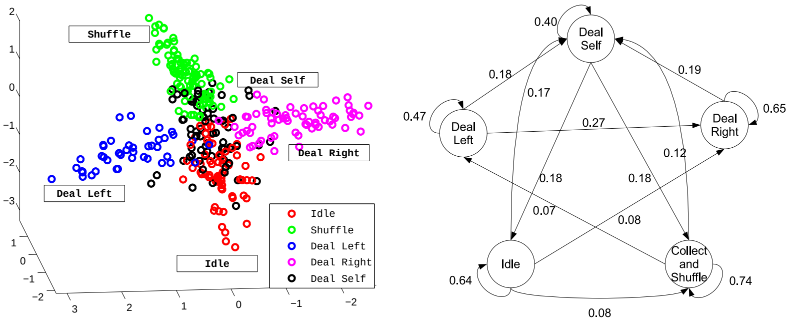
$$z(t+1) = A(t)z(t) + v(t) \quad v(t) \sim N(0, Q(t)) \tag{33}$$

Note that here the model parameters (A, C, Q, R) are allowed to vary with time. Further, we assume that the model parameters change slowly with time so that they can be approximated as locally constant. Thus, parameter estimation is done in short-temporal windows (say of length 20 frames). This gives rise to a sequence of model parameters $M_t = (A_t, C_t)$. Each element in the sequence can be considered to be a point on the Grassmann manifold arising due to the time-varying observability matrix.

$$O_n(M_t) = \left[C_t; C_tA_t; \ldots; C_tA_t^{n-1} \right] \tag{34}$$

Thus, the time-varying model can be viewed as a sequence of subspaces S_t, where each subspace is spanned by the columns of the observability matrix at the corresponding time instant. Thus, the sequence of subspaces can be seen as a trajectory on the Grassmann manifold. To compactly represent the subspace variations, we parametrize the trajectory using a switching model akin to the HMM on the Grassmann manifold. This representation can be used to provide a visual summarization of long videos [43]. The clusters of the HMM represent the distinct actions in the video e.g. spins, leaps, glides for the case of skating. The transition structure between the clusters represents how the overall activity in the video proceeds. In this experiment we show the results of summarizing a long video containing a complex activity – the game of Blackjack. For this, we used the dataset reported in [51]. A few sample frames from the dataset are shown in figure 4. The game of Blackjack consists of a few elements such as dealing cards, waiting for bids, shuffling the cards etc. We try to estimate a Grassmann switching model for the entire video of Blackjack. The Grassmann switching model would then represent a ‘summary’ of the game, where the clusters of the model represent various elements of the game and the switching structure represents how the game progresses. This video consists of about 1700 frames. We extracted the motion-histogram features as proposed in [51] for each frame of the video. The time-varying model parameters are estimated in sliding windows of size 10. The dimension of the state vector is chosen to be $d = 5$. To estimate the Grassmann switching model for the game of Blackjack, we manually set the number of clusters to 5. In figure 5(a), we show an embedding of the video obtained from the model parameters using Laplacian eigenmaps. Each point corresponds to a time-invariant model parameter (A, C) pair or equivalently a point on the Grassmann manifold. Each cluster was found to correspond to a distinct element of the game as shown. The switching structure between the clusters is encoded in the transition matrix and is shown in figure 5(b). Similar ARMA models were also used in [44] for summarizing a long skating video sequence.

Fig. 4 A few sample frames from the Blackjack dataset of [51].



(a) (b)

Fig. 5 (a) An embedding of the entire Blackjack video sequence. Figure best viewed in color. (b) Estimated structure of the game of Blackjack. (For the sake of clarity arcs with low weights have not been shown)

5.3 Transformation Space Manifold: Hilbert Sphere for modeling execution-rate variations in activities

In activity recognition, different instances of the same activity may consist of varying relative speeds at which the actions are executed, in addition to other intra- and inter- person variabilities. Most existing algorithms for activity recognition are not very robust to intra- and inter-personal changes of the same activity, and are sensitive to warping of the temporal axis due to variations in speed profile. Results on gait-based person identification shown in [7] indicate that it is very important to take into account the temporal variations in the person's gait. In [49], it was shown that accounting for execution rate enhances recognition performance for action recognition. Typical approaches for accounting for variations in execution rate are either directly based on the dynamic time warping (DTW) algorithm [34] or some variation of this algorithm [49].

For now, let us assume that for each frame of the video, an appropriate feature has been extracted and that the video data has now been converted into a feature sequence given by $f^1, f^2, ...$, for frames $1, 2, ...$ respectively. We will use $\mathcal{F}$ to denote the feature space associated with the chosen feature. Let γ be a diffeomorphism (A diffeomorphism is a smooth, invertible function with a smooth inverse) from $[0,1]$ to itself with $\gamma(0) = 0$ and $\gamma(1) = 1$. Also, let $\mathbf{\Gamma}$ be the set of all such functions. We will use elements of $\mathbf{\Gamma}$ to denote time warping functions. Our model for an activity consists of an average activity sequence given by $a : [0,1] \to \mathcal{F}$, a parameterized trajectory on the feature space. Any time-warped realization of this activity is then obtained using:

$$r(t) = a(\gamma(t)), \quad \gamma \in \mathbf{\Gamma} \ . \tag{35}$$

Equation (35) actually defines an action of $\mathbf{\Gamma}$ on $\mathcal{F}^{[0,1]}$, the space of all continuous activities. In our model, the variability associated with γ in each class will be modeled using a distribution P_γ on $\mathbf{\Gamma}$. For the convenience of analysis and computation, we prefer to work with $\psi = +\sqrt{\dot{\gamma}}$ instead of γ directly. There is a bijection between γ and ψ and the probability models on ψ directly relate to equivalent models on γ. Thus, we will introduce probability distributions P_ψ on the set of all ψs, for each activity class.

The parameters of this model are $a(t)$, the nominal activity trajectory, and P_ψ, the probability distribution on square-root representations of time warping functions. In general, the nominal activity trajectory $a(t)$ can also be chosen to be random. Here, we restrict our analysis to cases where the nominal activity trajectory $a(t)$ is deterministic but unknown. We will consider parametric forms of densities for P_ψ and reduce the problem of learning P_ψ to one of learning the parameters of the distribution P_ψ.

Let the space of all square-root density forms be given by

$$\mathbf{\Psi} = \{\psi : [0,1] \to \mathbb{R} | \psi \geq 0, \int_0^1 \psi^2(t) dt = 1\} \ . \tag{36}$$

This is the positive orthant of a unit hypersphere in the Hilbert space of all square-integrable functions on $[0,1]$. Let $T_\psi(\mathbf{\Psi})$ be the tangent space to $\mathbf{\Psi}$ at any given point ψ. Then, for any v_1 and v_2 in $T_\psi(\mathbf{\Psi})$, the Fisher-Rao metric is given by

$$\langle v_1, v_2 \rangle = \int_0^1 v_1(t)v_2(t)dt. \tag{37}$$

Since $\mathbf{\Psi}$ is a sphere, its geometry is well known and we can directly use known expressions for geodesics, exponential maps, and inverse exponential maps on $\mathbf{\Psi}$ as discussed in sections 3.2 and 3.2.1. Consequently, the algorithms for computing sample statistics, defining probability density functions, and generating inferences also become straightforward.

5.3.1 Common Activities Dataset

We used the UMD common activities dataset [49], a dataset of common activities to perform preliminary experiments to validate our model. The dataset consists of 10 activities and 10 different instances of each activity. We partition the dataset into 10 disjoint sets each containing 1 instance of every activity. In order to test the recognition performance for each set, we first learn the model parameters from the remaining nine sets and then perform recognition for the test sequences. We repeat the process for each of the 10 sets. Thus we ensure that there is no overlap between the training set and the test sequences. Figure 6 shows the 10 × 100 similarity matrix for using the function space algorithm with the uniform distribution on the space of temporal warps. Each column corresponds to a different test sequence while each row corresponds to a different activity. The strongly block diagonal nature of the similarity matrix indicates that the recognition algorithm performs well. In fact, on this database we obtained 100% recognition using both our algorithms.

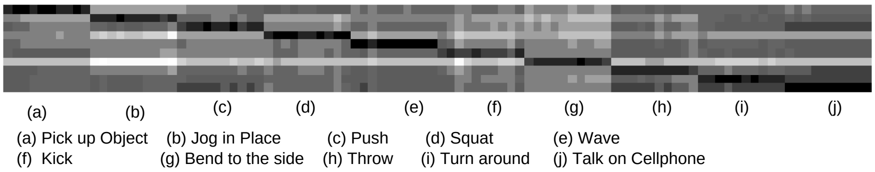


Fig. 6 10 X 100 Similarity matrix of 100 sequences and 10 different activities using the function space algorithm.

5.3.2 USF Gait Database

Since the model for learning the function space of time-warpings is not explicitly dependent on the choice of features, one could potentially use the same model to learn individual specific function spaces in order to perform activity-based person identification. The only difference would be that we would choose a feature that is person-specific (e.g., silhouette). The nominal activity trajectory would be individual specific in this case. Various external conditions (such as surface, shoe) induce systematic time-warping variations within the gait signatures of each individual. The function space of temporal warpings for each individual amounts to learning the class of person specific warping functions. By learning the function space of these variations we are able to account for the effects of such external conditions.

In order to compare the performance of our algorithm with the current state of the art algorithms, we also performed a gait-based person identification experiment on the publicly available USF gait database [35]. The USF database consists of 71 people in the Gallery. Various covariates like camera position, shoe type, surface and time were varied in a controlled manner to design a set of challenge experiments[35]. We performed a round-robin recognition experiment in which one of the challenge sets was used as test while the other seven were used as training examples. The process was repeated for each of the seven challenge sets on which results have been reported. Table 2 shows the identification rates of our algorithm with a uniform distribution on the space of warps (P_{Unif}), our algorithm with a wrapped Gaussian distribution on the tangent space of warps with shape as a feature and with binary image feature (P_{Gauss} and $P_{GaussIm}$). For comparison the table also shows the baseline algorithm [35], simple DTW on shape features [48] and the image-based HMM [23] algorithm on the USF dataset for the 7 probes A-G. Since most of these other algorithms could not account for the systematic variations in time-warping for each class the recognition experiment they performed was not round robin but rather used only one sample per class for learning. Therefore, to ensure a fair comparison, we also implemented a round-robin experiment using the linear warping (P_LW).

The average performance of our algorithms P_{Unif} and P_{Gauss} are better than all the other algorithms that use the same feature, (DTW/HMM (Shape)[48] and Linear warping P_{LW}) and is also better than the baseline[35] and HMM[23] algorithms that use the image as a feature. The improvement in performance while using binary image as a feature is shown in the last column ($P_{GaussIm}$). The experimental results presented here clearly show that using multiple training samples per class and learning the distribution of their time warps makes significant improvement to gait recognition results. While most algorithms based on learning from a single sample led to overfitting and therefore performed much better when the gallery was similar to the probe (Probe A-C), they also performed very poorly when the gallery

Table 2 Comparison of Identification rates on the USF dataset. Note that the experimental results reported in this table contain varying amounts of training data. While columns 2-6 (Baseline - pHMM) used only the gallery sequences for training, the results reported in columns 7-10 (P_{LW} - $P_{GaussIm}$) used all the probes except the test probe during training.

Pr- obe	Base- line	DTW Shape	HMM Shape	HMM Image	pHMM [29]	P_{LW}	P_{Unif}	P_{Gauss}	$P_{GaussIm}$
Avg.	42	42	41	50	65	51.5	59	59	64
A	79	81	80	96	85	68	70	78	82
B	66	74	72	86	89	51	68	68	78
C	56	52	56	74	72	51	81	82	76
D	29	29	22	32	57	53	40	50	48
E	24	20	20	28	66	46	64	51	54
F	30	19	20	17	46	50	37	42	56
G	10	19	19	21	41	42	53	40	55

and the probes were significantly different. But, since our algorithm has good generalization ability the performance of our algorithm did not suffer from overfitting and therefore did not drop as much when moving from probes A-C to Probes D-G.

6 Conclusions

In this chapter we provided a brief overview of the usefulness and effectiveness of statistical analysis on manifolds to specific applications in video analysis. Typical video analysis is usually composed of three stages of processing - feature extraction, building models and accounting for transformation invariance. We highlight three different applications of manifold analysis, one for each of the three stages in a typical video analysis framework. We describe Kendall shape manifold for shape feature representation. We show the applicability of the Grassmann manifold for understanding dynamical models. Finally, we show the space of time-warp transformations as a spherical manifold of functions. In all applications, we show experiments that illustrate the superior performance of algorithms that exploit the geometric properties of the underlying manifold.

References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton, NJ (2008)
2. Aggarwal, G., Roy-Chowdhury, A., Chellappa, R.: A system identification approach for video-based face recognition. International Conference on Pattern Recognition (2004)
3. Begelfor, E., Werman, M.: Affine invariance revisited. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2087–2094 (2006)
4. Bhattacharya, R., Patrangenaru, V.: Nonparametric estimation of location and dispersion on Riemannian manifolds. Journal for Statistical Planning and Inference **108**,

23–36 (2002)

5. Bhattacharya, R., Patrangenaru, V.: Large sample theory of intrinsic and extrinsic sample means on manifolds- I. The Annals of Statistics **31**(1), 1–29 (2003)
6. Bissacco, A., Chiuso, A., Ma, Y., Soatto, S.: Recognition of human gaits. In: IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 52–57 (2001)
7. Bobick, A., Tanawongsuwan: Performance analysis of time-distance gait parameters under different speeds. In: Audio- and Video-Based Biometric Person Authentication (AVBPA) (2003)
8. Bookstein, F.: Size and shape spaces for landmark data in two dimensions. Statistical Science **1**, 181–242 (1986)
9. Boothby, W.M.: An introduction to differentiable manifolds and Riemannian geometry. Academic Press Inc (1975)
10. Brockett, R.: Notes on Stochastic Processes on Manifolds. Systems and Control in the Twenty-First Century: Progress in Systems and Control, Volume 22. Birkhauser (1997)
11. Brockett, R.W.: System theory on group manifolds and coset spaces. SIAM Journal on Control **10**(2), 265–84 (1972)
12. Chikuse, Y.: Statistics on special manifolds, Lecture Notes in Statistics. Springer, New York. (2003)
13. Cock, K.D., Moor, B.D.: Subspace angles and distances between ARMA models. Proceedings of the Intl. Symposium of Mathematical Theory of Networks and Systems (MTNS) (2000)
14. Dryden, I., Mardia, K.: Statistical shape analysis. John Wiley and sons (1998)
15. Dryden, I.L., Mardia, K.V.: Statistical Shape Analysis. John Wiley & Son (1998)
16. Edelman, A., Arias, T., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM Journal of Matrix Analysis and Applications **20**(2), 303–353 (1998)
17. Georghiades, A.S., Kriegman, D.J., Belhumeur, P.N.: Illumination cones for recognition under variable lighting: Faces. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 52–59 (1998)
18. Grenander, U.: Probabilities on Algebraic Structures. Wiley (1963)
19. Grenander, U.: General Pattern Theory. Oxford University Press (1993)
20. Grenander, U., Miller, M.I.: Computational anatomy: An emerging discipline. Quarterly of Applied Mathematics **LVI**(4), 617–694 (1998)
21. Grenander, U., Miller, M.I., Srivastava, A.: Hilbert-Schmidt lower bounds for estimators on matrix Lie groups for ATR. IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(8), 790–802 (1998)
22. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, second edn. (2004)
23. Kale, A., Sundaresan, A., Rajagopalan, A., Cuntoor, N., Roy Cowdhury, A., Krueger, V., Chellappa, R.: Identification of humans using gait. IEEE Transactions on Image Processing **13**(9), 1163–1173 (2004)
24. Karcher, H.: Riemannian center of mass and mollifier smoothing. Communications on Pure and Applied Mathematics **30**, 509–541 (1977)
25. Kendall, D.: Shape manifolds, procrustean metrics and complex projective spaces. Bulletin of London Mathematical society **16**, 81–121 (1984)
26. Kendall, D.G.: Shape manifolds, procrustean metrics and complex projective spaces. Bulletin of London Mathematical Society **16**, 81–121 (1984)
27. Klassen, E., Srivastava, A., Mio, W., Joshi, S.: Analysis of planar shapes using geodesic paths on shape spaces. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(3), 372–383 (March, 2004)
28. Le, H.L., Kendall, D.G.: The Riemannian structure of Euclidean shape spaces: a novel environment for statistics. Annals of Statistics **21**(3), 1225–1271 (1993)
29. Liu, Z., Sarkar, S.: Improved gait recognition by gait dynamics normalization. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(6), 863–876 (2006)

30. Lui, Y.M., Beveridge, J.R.: Grassmann registration manifolds for face recognition. In: European Conference on Computer Vision, pp. 44–57. Marseille, France (2008)
31. Miller, M.I., Younes, L.: Group actions, homeomorphisms, and matching: A general framework. International Journal on Computer Vision **41**(1/2), 61–84 (2001)
32. Mio, W., Srivastava, A., Joshi, S.: On shape of plane elastic curves. International Journal on Computer Vision **73**(3), 307–324 (2007)
33. Pennec, X., Ayache, N.: Uniform distribution, distance and expectation problems for geometric features processing. Journal of Mathematical Imaging and Vision **9**(1), 49–67 (1998)
34. Rabiner, L., Juang, B.: Fundamentals of speech recognition. Prentice Hall (1993)
35. Sarkar, S., Phillips, P., Liu, Z., Vega, I., Grother, P., Bowyer, K.: The humanid gait challenge problem: data sets, performance, and analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 162–177 (2005)
36. Small, C.G.: The Statistical Theory of Shape. Springer (1996)
37. Soatto, S., Doretto, G., Wu, Y.N.: Dynamic textures. In: IEEE International Conference on Computer Vision, vol. 2, pp. 439–446 (2001)
38. Spivak, M.: A Comprehensive Introduction to Differential Geometry, Volume 1. Publish or Perish, Inc (1970)
39. Srivastava, A., Joshi, S., Mio, W., Liu, X.: Statistical shape analysis: Clustering, learning and testing. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(4), 590–602 (2005)
40. Srivastava, A., Klassen, E.: Monte Carlo extrinsic estimators for manifold-valued parameters. IEEE Trans. on Signal Processing **50**(2), 299–308 (2001)
41. Srivastava, A., Klassen, E.: Bayesian, geometric subspace tracking. Journal for Advances in Applied Probability **36**(1), 43–56 (2004)
42. Subbarao, R., Meer, P.: Nonlinear mean shift over riemannian manifolds. International Journal on Computer Vision **84**(1), 1–20 (2009)
43. Turaga, P., Chellappa, R.: Locally time-invariant models of human activities using trajectories on the grassmannian. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2435–2441 (2009)
44. Turaga, P., Veeraraghavan, A., Chellappa, R.: Unsupervised view and rate invariant clustering of video sequences. Computer Vision and Image Understanding **113**(3), 353–371 (2009)
45. Turaga, P.K., Veeraraghavan, A., Chellappa, R.: Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision. In: IEEE International Conference on Computer Vision and Pattern Recognition (2008)
46. Tuzel, O., Porikli, F., Meer, P.: Region covariance: A fast descriptor for detection and classification. In: European Conference on Computer Vision, pp. 589–600. Graz, Austria (2006)
47. Tuzel, O., Porikli, F., Meer, P.: Pedestrian detection via classification on Riemannian manifolds. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(10), 1713–1727 (2008)
48. Veeraraghavan, A., Roy-Chowdhury, A., Chellappa, R.: Matching shape sequences in video with an application to human movement analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(12), 1896–1909 (2005)
49. Veeraraghavan, A., Srivastava, A., Roy Chowdhury, A.K., Chellappa, R.: Rate-invariant recognition of humans and their activities. IEEE Transactions on Image Processing **18**(6), 1326–1339 (2009)
50. Weinland, D., Ronfard, R., Boyer, E.: Free viewpoint action recognition using motion history volumes. Computer Vision and Image Understanding **104**(2), 249–257 (2006)
51. Zhong, H., Shi, J., Visontai, M.: Detecting unusual activity in video. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 819–826 (2004)