

Statistical analysis on Stiefel and Grassmann Manifolds with applications in Computer Vision

Pavan Turaga, Ashok Veeraraghavan and Rama Chellappa
Center for Automation Research
University of Maryland, College Park
{pturaga, vashok, rama}@umiacs.umd.edu

Abstract

Many applications in computer vision and pattern recognition involve drawing inferences on certain manifold-valued parameters. In order to develop accurate inference algorithms on these manifolds we need to a) understand the geometric structure of these manifolds b) derive appropriate distance measures and c) develop probability distribution functions (pdf) and estimation techniques that are consistent with the geometric structure of these manifolds. In this paper, we consider two related manifolds - the Stiefel manifold and the Grassmann manifold, which arise naturally in several vision applications such as spatio-temporal modeling, affine invariant shape analysis, image matching and learning theory. We show how accurate statistical characterization that reflects the geometry of these manifolds allows us to design efficient algorithms that compare favorably to the state of the art in these very different applications. In particular, we describe appropriate distance measures and parametric and non-parametric density estimators on these manifolds. These methods are then used to learn class conditional densities for applications such as activity recognition, video based face recognition and shape classification.

1. Introduction

Many applications in computer vision such as dynamic textures [24, 9], human activity modeling and recognition [7, 30], video based face recognition [2], shape analysis [14, 20] involve learning and recognition of patterns from exemplars which lie on certain manifolds. Given a database of examples and a query, the following two questions are usually addressed – a) what is the ‘closest’ example to the query in the database ? b) what is the ‘most probable’ class to which the query belongs ? A systematic solution to these problems involves a study of the manifold on which the data lies. The answer to the first question involves study of the geometric properties of the manifold, which then leads to appropriate definitions of distance metrics on the man-

ifold (geodesics etc). The answer to the second question involves statistical modeling of inter- and intra-class variations on the manifold. The solution to the second problem goes far beyond simply defining distance metrics. Given several samples per class, one can derive efficient probabilistic models on the manifold by exploiting the class ensemble populations that are also consistent with the geometric structure of these manifolds. This offers significant improvements in performance over, say, a distance-based nearest-neighbor classifier. In addition to this, statistical models also provide us with generative capabilities via appropriate sampling strategies.

In this paper, we concern ourselves with two related manifolds that often appear in several applications in computer vision – the *Stiefel* Manifold and the *Grassmann* Manifold. The Stiefel manifold is the space of k orthonormal vectors in R^m , represented by an $m \times k$ matrix Y , such that $Y^T Y = I_k$. The Grassmann manifold is the space of k dimensional *subspaces* in R^m . While a point on the Grassmann manifold represents a subspace, a point on the Stiefel manifold also specifies exactly what frame (choice of basis vectors) is used in order to specify the subspace. The study of these manifolds has important consequences for applications such as dynamic textures [24, 9], human activity modeling and recognition [7, 30], video based face recognition [2] and shape analysis [14, 20] where data naturally lies either on the Stiefel or the Grassmann manifold.

The geometric properties of the Stiefel and Grassmann manifolds are well understood and we refer the interested reader to [13, 1] for specific details. The scope of this paper is not restricted to the differential geometry of these manifolds. Instead, we are interested in statistical modeling and inference tools on these manifolds and their applicability in vision applications. While a meaningful distance metric is an important requirement, statistical methods (such as learning probability density functions) provide far richer tools for pattern classification problems. In this context, we describe the Procrustes distance measures on the Stiefel and the Grassmann manifold [10]. Further, we describe parametric and non-parametric kernel based density functions

on these manifolds and describe learning algorithms for estimating the distribution from data.

Prior Work: Statistical methods on manifolds have been studied for several years in the statistics community [5, 21, 22]. A compilation of research results on statistical analysis on the Stiefel and Grassmann manifolds can be found in [10]. Their utility in practical applications has not yet been fully explored. Theoretical foundations for manifolds based shape analysis were described in [14, 20]. The Grassmann manifold structure of the affine shape space is exploited in [4] to perform affine invariant clustering of shapes. [26] exploited the geometry of the Grassmann manifold for subspace tracking in array signal processing applications. Statistical learning of shape classes using non-linear shape manifolds was presented in [27] where statistics are learnt on the manifold’s tangent space. Classification on Riemannian manifolds have also been explored in the vision community such as in [28, 29].

Organization of the paper: In section 2, we present motivating examples where Stiefel and Grassmann manifolds arise naturally in vision applications. In section 3, a brief review of statistical modeling theory on the Stiefel and Grassmann manifolds is presented. In section 4, we demonstrate the strength of the framework on several example applications including view invariant activity recognition, video based face recognition, shape matching and retrieval. Finally, section 5 presents some concluding remarks.

2. Motivating Examples

1. **Spatio-temporal dynamical models:** A wide variety of spatio-temporal data in computer vision are modeled as realizations of dynamical models. Examples include Dynamic textures [24], human joint angle trajectories [7] and silhouette sequences [30]. One popular dynamical model for such time-series data is the autoregressive and moving average (ARMA) model. For the ARMA model closed form solutions for learning the model parameters have been proposed in [19, 24] and are widely used. The parameters of the model are known to lie on the Stiefel manifold as noted in [23]. Given several instances, current approaches involve computing the distance between them using well-known distance measures [11] followed by nearest neighbor classification. Instead, given several instances of each class we can learn compact class conditional probability density functions over the parameter space – the Stiefel manifold in this case. Maximum likelihood and maximum a posteriori estimation can then be performed in order to solve problems such as video classification, clustering and retrieval.

2. **Shape Analysis:** Representations and recognition

of shapes is a well understood field [15, 12]. The shape observed in an image is a perspective projection of the original shape. In order to account for this, shape theory studies the equivalent class of all configurations that can be obtained by a specific class of transformation (e.g. linear, affine, projective) on a single basis shape. It can be shown that affine and linear shape spaces for specific configurations can be identified by points on the Grassmann manifold [20]. Given several exemplar shapes belonging to a few known classes, we are interested in estimating a probability distribution over the shape space for each of the classes. These can then be used for problems such as retrieval, classification or even to learn a generative model for shapes. The theory developed here may also impact several related problems such as video stabilization and image registration which amounts to finding the best warping based on mean affine shapes of a few landmark points.

3. **On-line Visual Learning via Subspace Tracking:** Applications involving dynamic environments and autonomous agents such as a mobile robot navigating through an unknown space cannot be represented by static models. In such applications it is important to adapt models, that have been learnt offline, according to new observations in an online fashion. One approach is to perform incremental PCA to dynamically learn a better representational model as the appearance of the target dynamically changes as in [3]. Incremental PCA has also been used to recognize abnormalities in the visual field of a robot as in [18]. In an unrelated domain, the theory of subspace tracking on the Grassmann manifold [26] has been developed for array signal processing applications. Since PCA basis vectors represent a subspace which is identified by a point on the Grassmann manifold, subspace tracking lends itself readily to statistical analysis for online visual learning applications.

3. Statistics on Manifolds

In this section, we discuss the problem of pattern recognition on Stiefel and Grassmann manifolds. Specifically, parametric and non-parametric distributions and parameter estimation are reviewed. Procrustes analysis and corresponding distance metrics on the manifolds is also presented. The discussion will focus mainly on the Stiefel manifold. Similar results extend to the Grassmann manifold [10].

3.1. Definitions

The Stiefel Manifold $V_{k,m}$ [10]: The Stiefel manifold $V_{k,m}$ is the space whose points are k -frames in R^m , where a set of k orthonormal vectors in R^m is called a k -frame in

$R^m (k \leq m)$. Each point on the Stiefel manifold $V_{k,m}$ can be represented as a $m \times k$ matrices X such that $X^T X = I_k$, where I_k is the $k \times k$ identity matrix.

The Grassmann Manifold $G_{k,m-k}$ [10]: The Grassmann manifold $G_{k,m-k}$ is the space whose points are k -planes or k -dimensional hyperplanes (containing the origin) in R^m . An equivalent definition of the Grassmann manifold is as follows. To each k -plane ν in $G_{k,m-k}$ corresponds a unique $m \times m$ orthogonal projection matrix P idempotent of rank k onto ν . If the columns of an $m \times k$ matrix Y spans ν , then, $Y Y^T = P$.

For the case of $k = 1$, the Stiefel manifold reduces to the unit hypersphere in m -dimensions. Each point on the manifold represents a vector of unit length. Similarly, for $k = 1$ the Grassmann manifold reduces to the real projective space which consists of all lines through the origin.

3.2. Statistical Modeling

Two of the most studied functional probability density forms are the Matrix Langevin and Matrix Bingham distributions.

The Matrix Langevin Distribution [10]: The matrix Langevin Distribution on $V_{k,m}$, denoted by $L(m, k; F)$ is given by

$$\frac{1}{{}_0H_1(\frac{1}{2}m; \frac{1}{4}F^T F)} \exp(\text{tr}(F^T X)) \quad (1)$$

where $\text{tr}()$ is the matrix trace. F is an $m \times k$ matrix which represents the parameters of the distribution. The normalizing factor ${}_0H_1()$ is a hyper-geometric function. The matrix Langevin distribution is of the exponential form. It reduces to the uniform distribution on $V_{k,m}$ for $F = 0$.

The Matrix Bingham Distribution [10]: The matrix Bingham Distribution on $V_{k,m}$, denoted by $B(m, k; A)$ is given by

$$\frac{1}{{}_1H_1(\frac{1}{2}k; \frac{1}{2}m; A)} \exp(\text{tr}(X^T A X)) \quad (2)$$

where $\text{tr}()$ is the matrix trace. A is an $m \times m$ symmetric matrix. The Bingham distribution is closely related to the matrix normal distribution [10] and reduces to the uniform distribution on $V_{k,m}$ for $A = 0$.

Parameter Estimation In pattern classification problems, one is interested in estimating parameters of class conditional distributions from sample populations for each class. Let the singular value decomposition of the Langevin distribution parameter F of known rank $p \leq k$ be given by $F = \Gamma \Lambda \Theta^T$. Let (X_1, X_2, \dots, X_n) be n samples from the matrix Langevin distribution. Let $\bar{X} = \frac{1}{n} \sum X_i$ be the arithmetic mean of the samples. Let $\bar{X} = U_X S_X V_X^T$ be the

reduced rank (rank = p) singular value decomposition of \bar{X} . The maximum likelihood estimators for $\hat{\Gamma}$, $\hat{\Theta}$ are given by $\hat{\Gamma} = U_X$, $\hat{\Theta} = V_X$. Solving for the parameter $\hat{\Lambda}$ is non-trivial. For the large sample asymptotic case, i.e. when n is large and $m \gg k$, $\hat{\Lambda} = m S_X$. We refer the reader to [10] for detailed derivations and proofs for the asymptotic and non-asymptotic cases. Corresponding results for the Bingham distribution can also be found in [10].

3.3. Distance Metrics on $V_{k,m}$

The Stiefel and Grassmann manifolds are endowed with a Riemannian structure that lends itself to computation of distances between points on the manifold via geodesics [13, 1]. Instead of geodesic computations, we adopt the Procrustes distance metric proposed in [10] which is defined in the ambient euclidean space. As will be discussed shortly, this choice results in efficient computation of the distance metrics and the class conditional probability density estimators on the manifolds.

Procrustes representations and corresponding distance metrics are defined to be invariant to specific classes of transformations depending on the problem domain. Examples include Procrustes representations for landmark points which are invariant to uniform scaling and rotation of the shape [15]. We seek similar representations on the Stiefel manifold. Two representations of points on the Stiefel manifold can be defined [10].

- Representation V_a : A point X on $V_{k,m}$ is an $m \times k$ matrix such that $X^T X = I_k$.
- Representation V_b : A point X on $V_{k,m}$ is identified with an equivalence class of $m \times k$ matrices $X R$ in $R_{m,k}$, for $R > 0$. This is also called the Procrustes representation of the Stiefel manifold.

Euclidean distance: The squared Euclidean distance for two given matrices X_1 and X_2 on $V_{k,m}$ based on representation V_a is given by

$$d_{V_a}^2(X_1, X_2) = \text{tr}(X_1 - X_2)^T (X_1 - X_2) \quad (3)$$

$$= 2\text{tr}[I_k - \frac{1}{2}(X_1^T X_2 + X_2^T X_1)] \quad (4)$$

Procrustes Distance: The squared Procrustes distance for two given matrices X_1 and X_2 on $V_{k,m}$, based on representation V_b , is the smallest squared Euclidean distance between any pair of matrices in the corresponding equivalence classes. Hence

$$d_{V_b}^2(X_1, X_2) = \min_{R>0} \text{tr}(X_1 - X_2 R)^T (X_1 - X_2 R) \quad (5)$$

$$= \min_{R>0} \text{tr}(R^T R - 2X_1^T X_2 R + I_k) \quad (6)$$

Lemma 3.3.1: Let A be a $k \times k$ constant matrix. Consider the minimization of the quadratic function $g(R) = \text{tr}(R^T R - 2A^T R)$ of a matrix argument R .

1. If R varies over the space $R_{k,k}$ of all $k \times k$ matrices, the minimum is attained at $R = A$.
2. If R varies over the space of all $k \times k$ positive semi-definite matrices, the minimum is attained at $R = B^+$, where B^+ is the positive semi-definite part of $B = \frac{1}{2}(A + A^T)$.
3. If R varies over the orthogonal group $O(k)$, the minimum is attained at $R = H_1 H_2^T = A(A^T A)^{-1/2}$, where $A = H_1 D H_2^T$ is the singular value decomposition of A .

The proof of this follows easily using the method of Lagrange multipliers. We refer the reader to [10] for alternate proofs. Thus, for the case of the first constraint, where R varies over the space $R_{k,k}$ of all $k \times k$ matrices, the distance is given by $d_{V_b}^2(X_1, X_2) = \text{tr}(I_k - A^T A)$, where $A = X_1^T X_2$. We have used this metric in all our experiments. A closer inspection reveals that these distance measures are not symmetric in their arguments, hence are not true distance metrics. This can be trivially solved by defining a new distance metric as the average of the distance between the 2 points taken in both forward and backward directions.

Note that the Procrustes representation defines an equivalence class of points on the Stiefel manifold which are related by a *right* transformation. This directly relates to the interpretation of the Grassmann manifold as the orbit-space of the Stiefel manifold. All points on the Stiefel manifold related by a right transformation map to a single point on the Grassmann manifold. Thus, for comparing two subspaces represented by two orthonormal matrices, say Y_1 and Y_2 , we compute their Procrustes distance on the Stiefel manifold. We do not explicitly use the representation of points on the Grassmann manifold as $m \times m$ idempotent projection matrices (section 3.1). Instead, the corresponding Procrustes representation on the Stiefel manifold is an equivalent one. This representational choice also leads to methods that are more computationally efficient as opposed to working with large $m \times m$ matrices.

3.4. Kernel density functions

Kernel methods for estimating probability densities have proved extremely popular in several pattern recognition problems in recent years [6] driven by improvements in computational power. Kernel methods provide a better fit to the available data than simpler parametric forms.

Given several examples from a class (X_1, X_2, \dots, X_n) on the manifold $V_{k,m}$, the class conditional density can be

estimated using an appropriate kernel function. We first assume that an appropriate choice of a divergence or distance measure on the manifold has been made (section 3.3). For the Procrustes distance metric $d_{V_b}^2$ the density estimate is given by [10] as

$$\hat{f}(X; M) = \frac{1}{n} C(M) \sum_{i=1}^n K[M^{-1/2}(I_k - X_i^T X X^T X_i)M^{-1/2}] \quad (7)$$

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. The matrix valued kernel function $K(T)$ can be chosen in several ways. We have used $K(T) = \exp(-\text{tr}(T))$ in all the experiments reported in this paper.

4. Applications and Experiments

In this section we present a few application areas and experiments that demonstrate the usefulness of statistical analysis on the manifolds.

4.1. Dynamical Models

Algorithmic Details: Linear dynamical systems represent a class of parametric models for time-series. For high-dimensional time-series data (dynamic textures etc), the most common approach is to first learn a lower-dimensional embedding of the observations via PCA, and learn temporal dynamics in the lower-dimensional space. The PCA basis vectors form the model parameters for the corresponding time-series. Thus, the estimated model parameters lie on the Stiefel Manifold. For comparison of two models we use the Procrustes distance on the Stiefel manifold. Moreover, we can also use kernel density methods to learn class conditional distributions for the model parameters.

4.1.1 ARMA model

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 autoregressive and moving average (ARMA) models [24, 7, 30, 2]. The ARMA model equations are given by

$$f(t) = Cz(t) + w(t) \quad w(t) \sim N(0, R) \quad (8)$$

$$z(t+1) = Az(t) + v(t) \quad v(t) \sim N(0, Q) \quad (9)$$

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 respec-

tively. Closed form solutions for learning the model parameters (A, C) from the feature sequence $(f_{1:T})$ are widely used in the computer vision community [24]. Let observations $f(1), f(2), \dots, f(\tau)$, represent the features for the time indices $1, 2, \dots, \tau$. Let $[f(1), f(2), \dots, f(\tau)] = U\Sigma V^T$ be the singular value decomposition of the data. Then $\hat{C} = U, \hat{A} = \Sigma V^T D_1 V (V^T D_2 V)^{-1} \Sigma^{-1}$, where $D_1 = [0 \ 0; I_{\tau-1} \ 0]$ and $D_2 = [I_{\tau-1} \ 0; 0 \ 0]$.

The model parameters (A, C) learned as above do not lie on a linear space. The transition matrix A is only constrained to be stable with eigenvalues inside the unit circle. The observation matrix C is constrained to be an orthonormal matrix. Thus, the C matrix lies on the Stiefel manifold. For comparison of models, the most commonly used distance metric is based on subspace angles between columns spaces of the observability matrices [11] (denoted as Subspace Angles). The extended observability matrix for a model (A, C) is given by

$$O_\infty^T = [C^T, (CA)^T, (CA^2)^T, \dots, (CA^n)^T \dots] \quad (10)$$

Thus, a linear dynamical system can be alternately identified as a point on the Grassmann manifold corresponding to the column space of the observability matrix. In experimental implementations, we approximate the extended observability matrix by the finite observability matrix as is commonly done [23].

$$O_n^T = [C^T, (CA)^T, (CA^2)^T, \dots, (CA^n)^T] \quad (11)$$

As already discussed in section 3.3, comparison of two points on the Grassmann manifold can be performed by using the Procrustes distance metric on the Stiefel manifold (denoted as NN-Pro-Stiefel) without explicitly using the projection matrix representation of points on the Grassmann manifold. Moreover, if several observation sequences are available for each class, then one can learn the class conditional distributions on the Stiefel manifold using kernel density methods. Maximum likelihood classification can be performed for each test instance using these class conditional distributions (denoted as Kernel-Stiefel).

4.1.2 Activity Recognition

We performed a recognition experiment on the publicly available INRIA dataset [31]. 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 [31]. Specifically, we used the $16 \times 16 \times 16$ circular FFT features proposed by [31]. Each activity was modeled as a linear dynamical system. Testing was performed using a round-robin experiment where activity models were learnt

Activity	Dim. Red. [31] 16^3 volume	Best Dim. Red. [31] 64^3 volume	Subspace Angles 16^3 volume	NN-Pro-Stiefel 16^3 volume	Kernel-Stiefel 16^3 volume
Check Watch	76.67	86.66	93.33	90	100
Cross Arms	100	100	100	96.67	100
Scratch Head	80	93.33	76.67	90	96.67
Sit Down	96.67	93.33	93.33	93.33	93.33
Get Up	93.33	93.33	86.67	80	96.67
Turn Around	96.67	96.67	100	100	100
Walk	100	100	100	100	100
Wave Hand	73.33	80	93.33	90	100
Punch	83.33	96.66	93.33	83.33	100
Kick	90	96.66	100	100	100
Pick Up	86.67	90	96.67	96.67	100
Average	88.78	93.33	93.93	92.72	98.78

Table 1. Comparison of view invariant recognition of activities in the INRIA dataset using a) Best DimRed [31] on $16 \times 16 \times 16$ features, b) Best Dim. Red. [31] on $64 \times 64 \times 64$ features c) Nearest Neighbor using Procrustes distance on the Stiefel manifold ($16 \times 16 \times 16$ features), d) Maximum likelihood using kernel density methods on the Stiefel manifold ($16 \times 16 \times 16$ features)

using 9 actors and tested on 1 actor. For the kernel method, all available training instances per class were used to learn a class-conditional kernel density as described in section 3.4. In table 1, we show the recognition results obtained using four methods. The first column shows the results obtained using dimensionality reduction approaches of [31] on $16 \times 16 \times 16$ features. [31] 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 corresponds to the method of using subspace angles based distance between dynamical models [11]. Column 4 shows the nearest-neighbor classifier performance using Procrustes distance measure on the Stiefel manifold ($16 \times 16 \times 16$ features). We see that the manifold Procrustes distance performs as well as system distance. But, statistical modeling of class conditional densities for each activity using kernel density methods on the Stiefel manifold, leads to a significant improvement in recognition performance. Note that even though the manifold approaches presented here use only $16 \times 16 \times 16$ features they outperform other approaches that use higher resolution ($64 \times 64 \times 64$ features) as shown in table 1. Moreover, computational complexity of the manifold Procrustes distance is extremely low since it involves computing a small $k \times k$ matrix R , whereas subspace angles [11] requires solving a high-dimensional discrete time Lyapunov equation.

4.1.3 Video-Based Face Recognition

Video-based face recognition (FR) by modeling the ‘cropped video’ either as dynamical models ([2]) or as a collection of PCA subspaces [16] have recently gained popularity because of their ability to recognize faces from low

	Test condition	System Distance	Procrustes	Kernel density
1	Gallery1,Probe2	81.25	93.75	93.75
2	Gallery2,Probe1	68.75	81.25	93.75
3	Average	75%	87.5%	93.75%

Table 2. Comparison of video based face recognition approaches a) ARMA system distance, b) Stiefel Procrustes distance, c) Manifold kernel density

resolution videos. However, in this case, we focus only on the C matrix of the ARMA model or PCA subspace as the distinguishing model parameter. This is because the C matrix encodes the appearance of the face, whereas the A matrix encodes the dynamic information. For video-based FR, only the facial appearance is important and not the facial dynamics. The C matrices are orthonormal, hence points on the Stiefel manifold. But, for recognition applications, the important information is encoded in the subspace spanned by the C matrix. Hence, we identify the model parameters (C 's) as points on the Grassmann Manifold. Therefore, both Procrustes distance and Kernel density methods are directly applicable. We tested our method on the dataset used in [2]. The dataset consists of face videos for 16 subjects with 2 sequences per subject. Subjects arbitrarily change head orientation and expressions. The illumination conditions differed widely for the 2 sequences of each subject. For each subject, one sequence was used as the gallery while the other formed the probe. The experiment was repeated by swapping the gallery and the probe data. The recognition results are reported in table 2. For kernel density estimation, the available gallery sequence for each actor was split into three distinct sequences. As seen in the last column, the kernel-based method outperforms the other approaches.

4.2. Affine Shape Analysis

Algorithmic Details: The representation and analysis of shapes has important applications in object recognition, gait recognition and image registration. Landmark based shape analysis is one of the most widely used approaches for representing shapes. A shape is represented by a set of landmark points on its contour. A shape is represented by a $m \times 2$ matrix $S = [(x_1, y_1); (x_2, y_2); \dots; (x_m, y_m)]$, of the set of m landmarks of the centered scaled shape. The *shape space* of a base shape is the set of equivalent configurations that are obtained by transforming the base shape by an appropriate spatial transformation. For example, the set of all affine transformations of a base shape forms the *affine shape space* of that base shape. More rigorously, let $\chi = (x_1, x_2, \dots, x_m)$ be a configuration of m points where each $x_i \in R^2$. Let γ be a transformation on R^2 . For example, γ could belong to the affine group, linear group, projec-

Algorithm	Rank 1	Rank 2	Rank 3	Rank 4
SC [17]	20/40	10/40	11/40	5/40
IDSC [17]	40/40	34/40	35/40	27/40
Hashing [8]	40/40	38/40	33/40	20/40
Grassmann Procrustes	38/40	30/40	23/40	17/40

Table 3. Retrieval experiment on articulation dataset. Last row is the results obtained using Grassmann manifold Procrustes representation. No articulation invariant descriptors were used.

tive group etc. Let

$$A(\gamma, (x_1, \dots, x_m)) = (\gamma(x_1), \dots, \gamma(x_m)) \quad (12)$$

be the *action* of γ on the point configuration.

In particular, the *affine shape space* [14] [25] is very important because the effect of the camera location and orientation can be approximated as affine transformations on the original base shape. The affine transforms of the shape can be derived from the base shape simply by multiplying the shape matrix S by a 2×2 full rank matrix on the right (translations are removed by centering). Multiplication by a full-rank matrix on the right preserves the column-space of the matrix S . Thus, all affine deformations of the same base shape, map to the same point on the Grassmann manifold [25]. Therefore, a systematic study of affine shape space essentially boils down to a study of the points on the Grassmann manifold. We can use both Procrustes distance and kernel density methods described earlier for several applications of affine invariant shape analysis such as shape retrieval and recognition.

4.2.1 Articulation Database

We conducted a retrieval experiment on the articulated shape database from [17]. We use the same test scheme proposed in [17]. The database consists of 8 object classes with 5 examples for each class. For each shape, 4 top matches are selected and the number of correct hits for ranks 1, 2, 3, 4 are reported. Table 3 summarizes the results obtained on this dataset. The proposed approach compares well with other approaches. It should be noted however, that this is not a fair comparison, as we do not use any articulation-invariant descriptors such as the ones used in [17] and [8]. In spite of this, manifold-based distance metrics perform very well.

4.2.2 Affine MPEG-7 Database

Since the strength of the approach lies in affine invariant representation of shapes, we conducted a synthetic experiment using the MPEG-7 database. We took one base shape from each of the 70 object classes and created 10 random affine warps of the shapes with varying levels of additive noise. This new set of shapes formed the gallery for the experiment. Sample shapes that were generated are shown

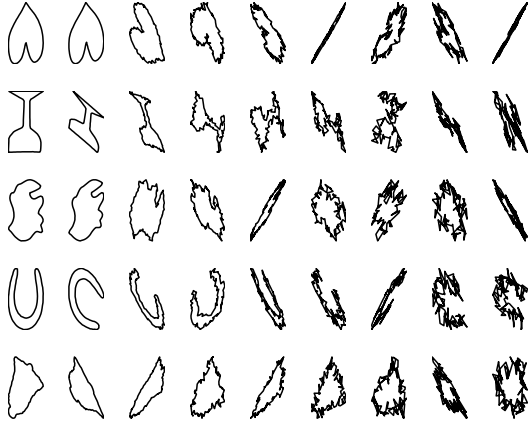


Figure 1. Synthetic data generated from the MPEG database. The first column shows base-shapes from the original MPEG dataset for 5 objects. The remaining columns show random affine warps for the base shapes with increasing levels of additive noise.

in figure 1. The test set was created by randomly picking a gallery shape and affine warping it with additive noise. The recognition experiment was performed using the Procrustes distance and the kernel density methods. For comparison, we used the popular shape Procrustes distance [15] as a baseline measure. We also used the ‘arc-length’ distance metric used in [4]. The arc-length distance metric is the Frobenius norm of the angle between two subspaces. In all cases, the experiments were repeated with 100 Monte-Carlo trials for each noise level in order to robustly evaluate the performance. The performance of the methods is compared in Figure 2 as a function of noise to signal ratio. It can be seen that manifold-based methods perform significantly better than straightforward shape Procrustes measures. Among the manifold methods, the kernel density method outperforms both the Procrustes and the arc-length distance measures. Since the Grassmann manifold based methods accurately account for the affine variations found in the shape, they outperform simple methods that do not account for affine invariance. Moreover, since the kernel methods learn a probability density function for the shapes on the Grassmann manifold, it outperforms distance based nearest neighbor classifiers using Grassmann arc-length and Grassmann Procrustes.

4.2.3 Sampling from Distributions

Generative capabilities of parametric probability densities can be exploited via appropriate sampling strategies. Once the distribution is learnt, one can synthesize samples from the distribution in a two step process. We first generate a sample from a proposal distribution (we used a matrix-variate normal centered around the class mean), then we use an accept-reject strategy to generate the final shape [10].

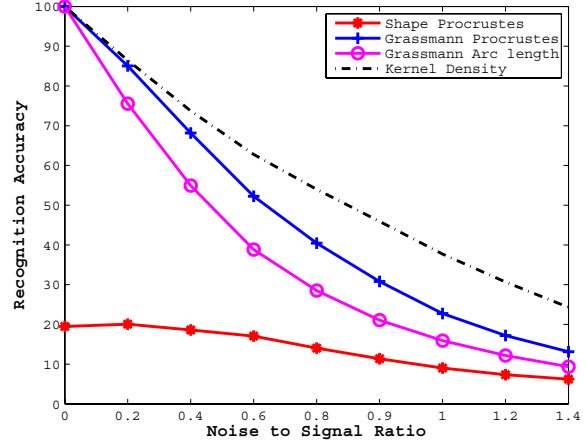


Figure 2. Comparison of recognition performance on MPEG-7 database. For comparison we used the shape Procrustes measure [15] and the Grassmann arc-length distance [4]. Manifold based methods perform significantly better than direct application of shape Procrustes measure. Among the manifold methods, statistical modeling via kernel methods outperforms the others.

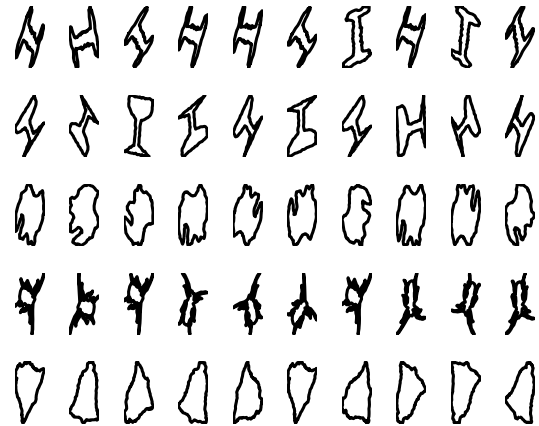


Figure 3. Samples generated from estimated class conditional densities for a few classes of the MPEG dataset

We show a sampling experiment using this technique. For this experiment, we took one shape from each of the object classes in the MPEG-7 database and corrupted it with additive noise to generate several noisy samples for each class. We used the Grassmann representation of points as idempotent projection matrices. Then, we learnt a parametric Langevin distribution on the Grassmann manifold for each class. Note that the distribution is learnt on the Grassmann manifold, hence, a sample from the distribution represents a subspace in the form of a projection matrix. To generate an actual shape we need to first choose a $2 - frame$ for the generated subspace which can be done via SVD of the projection matrix. Once the $2 - frame$ is chosen, actual shapes can be generated by choosing random coordinates in the $2 - frame$. We show sampling results in Figure 3.

5. Conclusion

In this paper we have studied statistical analysis on two specific manifolds – Stiefel and Grassmann. Matrix operators that lie on these manifolds arise naturally in several vision applications. Multi-class and multi-exemplar recognition and classification tasks require efficient statistical models to be estimated from data. We presented approaches from statistical literature which provide mathematically elegant and computationally efficient methods to perform statistical analysis. We demonstrated their utility in practical vision applications such as activity classification, video based face recognition and shape recognition, and showed how the same basic tools are applicable to a wide range of problem domains.

Acknowledgments: This research was funded (in part) by the US government VACE program.

References

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematicae*, 80(2):199–220, 2004.
- [2] G. Aggarwal, A. Roy-Chowdhury, and R. Chellappa. A system identification approach for video-based face recognition. *International Conference on Pattern Recognition*, 2004.
- [3] M. Artac, M. Jogan, and A. Leonardis. Incremental PCA for on-line visual learning and recognition. *International Conference on Pattern Recognition*, 2002.
- [4] E. Begelfor and M. Werman. Affine invariance revisited. *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.
- [5] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds-I. *Annals of Statistics*, 2003.
- [6] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, NJ, USA, 2006.
- [7] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto. Recognition of human gaits. *IEEE Conference on Computer Vision and Pattern Recognition*, 2:52–57, 2001.
- [8] S. Biswas, G. Aggarwal, and R. Chellappa. Efficient indexing for articulation invariant shape matching and retrieval. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [9] A. B. Chan and N. Vasconcelos. Modeling, clustering, and segmenting video with mixtures of dynamic textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence (Accepted for future publication)*, 2007.
- [10] Y. Chikuse. *Statistics on special manifolds, Lecture Notes in Statistics*. Springer, New York., 2003.
- [11] K. D. Cock and B. D. Moor. Subspace angles and distances between ARMA models. *Proc. of the Intl. Symp. of Math. Theory of networks and systems*, 2000.
- [12] I. Dryden and K. Mardia. *Statistical Shape Analysis*. Oxford University Press, 1998.
- [13] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal Matrix Analysis and Application*, 20(2):303–353, 1999.
- [14] C. R. Goodall and K. V. Mardia. Projective shape analysis. *Journal of Computational and Graphical Statistics*, 8(2), 1999.
- [15] D. Kendall. Shape manifolds, procrustean metrics and complex projective spaces. *Bulletin of London Mathematical Society*, 16:81–121, 1984.
- [16] K. C. Lee, J. Ho, M. H. Yang, and D. Kriegman. Video-based face recognition using probabilistic appearance manifolds. *IEEE Conference on Computer Vision and Pattern Recognition*, 2003.
- [17] H. Ling and D. W. Jacobs. Shape classification using the inner-distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(2), 2007.
- [18] H. V. Neto and N. Ulrich. Incremental PCA: An alternative approach for novelty detection. *Towards Autonomous Robotic Systems*, 2005.
- [19] P. V. Overschee and B. D. Moor. Subspace algorithms for the stochastic identification problem. *Automatica*, 29:649–660, 1993.
- [20] V. Patrangenaru and K. V. Mardia. Affine shape analysis and image analysis. *22nd Leeds Annual Statistics Research Workshop*, 2003.
- [21] B. Pelletier. Kernel density estimation on riemannian manifolds. *Statistics & Probability Letters*, 73(3):297–304, 2005.
- [22] X. Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.
- [23] P. Saisan, G. Doretto, Y. Wu, and S. Soatto. Dynamic texture recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.
- [24] S. Soatto, G. Doretto, and Y. N. Wu. Dynamic textures. *ICCV*, 2:439–446, 2001.
- [25] G. Sparr. Depth computations from polyhedral images. *European Conference on Computer Vision*, 1992.
- [26] A. Srivasatava and E. Klassen. Bayesian geometric subspace tracking. *Advances in Applied Probability*, 36(1):43–56, March 2004.
- [27] A. Srivastava, S. H. Joshi, W. Mio, and X. Liu. Statistical shape analysis: Clustering, learning, and testing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(4), 2005.
- [28] R. Subbarao and P. Meer. Nonlinear mean shift for clustering over analytic manifolds. *IEEE Conference on Computer Vision and Pattern Recognition*, 1:1168–1175, 2006.
- [29] O. Tuzel, F. Porikli, and P. Meer. Human detection via classification on riemannian manifolds. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [30] 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.
- [31] D. Weinland, R. Ronfard, and E. Boyer. Free viewpoint action recognition using motion history volumes. *Computer Vision and Image Understanding*, 104(2):249–257, 2006.