Image and Vision Computing 30 (2012) 380–388

Contents lists available at SciVerse ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Advances in matrix manifolds for computer vision ☆

Yui Man Lui *

Department of Computer Science, Colorado State University, Fort Collins, CO 80523, USA

ARTICLE INFO

Article history:
Received 28 February 2011
Received in revised form 24 June 2011
Accepted 4 August 2011

Keywords:
Lie groups
Stiefel manifolds
Grassmann manifolds
Riemannian manifolds

ABSTRACT

The attention paid to matrix manifolds has grown considerably in the computer vision community in recent years. There are a wide range of important applications including face recognition, action recognition, clustering, visual tracking, and motion grouping and segmentation. The increased popularity of matrix manifolds is due partly to the need to characterize image features in non-Euclidean spaces. Matrix manifolds provide rigorous formulations allowing patterns to be naturally expressed and classified in a particular parameter space. This paper gives an overview of common matrix manifolds employed in computer vision and presents a summary of related applications. Researchers in computer vision should find this survey beneficial due to the overview of matrix manifolds, the discussion as well as the collective references.

1. Introduction

The topology of a given space characterizes the proximity between data and plays a vital role in pattern recognition. Pattern analysis takes place in the context of data lying in some inherent geometrical structure. Simply ignoring the geometrical aspect, or naively treating the space as Euclidean, may cause undesired effects. Fortunately, there have been increasing efforts applied to pattern analysis on matrix manifolds. The aim of this article is to review some common matrix manifolds and summarize their applications as they relate to computer vision.

Despite the import of the geometry to the data, traditional approaches often quantify data in a vector space. This assumption may not always be valid for images [1]. To account for the geometry of the images, manifold learning techniques like ISOmetric Mapping (ISOMAP) [2] and Local Linear Embedding (LLE) [3] were introduced. These methods learn a mapping from the ambient space to the intrinsic space so that nearby points remain near each other after a projection. Although manifold learning can be effective in modeling the intrinsic structure of a manifold, they require a large amount of densely sampled training data; such rich training data may not be available for some real-world applications.

Another school of thought is to represent images in an underlying parametrized space. This gives rise to the representation of matrix manifolds. While ISOMAP and LLE model a manifold through training data, the use of matrix manifolds derives from the properties of differential geometry. In particular, data may be viewed as elements in some parameter space in which the idiosyncratic aspects of the geometry of the data can be characterized using algebraic operations. In fact, image data are often seen as the orbit of elements under the action of matrix manifolds, e.g. rotation group. Matrix manifolds may be the natural representation for some computer vision applications. In the present paper, the advances in this manifold representation arereviewed; the matrix manifolds of interest are Lie groups, Stiefel, Grassmann, and Riemannian manifolds.

The concept of matrix manifolds dates back to the 19th century [4], slowly gaining attention in the mathematics, physics, and other scientific communities. As we proceed in the computer age, more and more abstract concepts are found newly valuable in various applications, focusing particularly on matrix manifolds in this paper. Because matrix manifolds have proven to be instrumental in computer vision, we will summarize their applications to face recognition, action classification, clustering, visual tracking, and motion grouping and segmentation in Section 3. Before we survey these applications, a brief overview of matrix manifolds is given in Section 2. Finally, discussion and summary are provided in Section 4.

2. Matrix manifolds

The geometry of non-Euclidean spaces gives rise to the notion of manifolds. This section provides a brief summary of the matrix manifolds discussed in this paper. Details and rigorous treatments on these subjects can be found in [5–10].

2.1. Lie groups

A Lie group is both a group and a smooth manifold. A group G is defined as a nonempty set together with a binary operation $\circ$ satisfying the following axioms:

- Closure: If $g_i, g_j \in G$, then, $g_i \circ g_j \in G$.
- Associativity: If $g_i, g_j, g_k \in G$, then, $(g_i \circ g_j) \circ g_k = g_i \circ (g_j \circ g_k)$.

☆ This paper has been recommended for acceptance by special issue Guest Editor Rama Chellappa.

* Tel.: +1 970 4924181.
E-mail address: lui@cs.colostate.edu.

0262-8856/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2011.08.002

- Identity: If $g_i, e \in G$, then $g_i\, e = e\, g_i = g_i$.
- Inverse: If $g_i, e \in G$, then $g_i \circ g_i^{-1} = g_i^{-1} \circ g_i = e$.

The $(\mathbb{Z}, +)$ is an example of a group under a group action, multiplication.

A smooth manifold is a pair $(\mathcal{M}, \mathcal{F})$ where $\mathcal{M}$ is a set and $\mathcal{F}$ is a differentiable structure of $\mathcal{M}$. The differentiable structure provides a mechanism that $\mathcal{M}$ is entirely covered by a collection of charts $\{(\mathcal{U}, \phi)\}$ called atlas where ϕ is a bijective map of $\mathcal{U} \in \mathcal{M}$ such that $\phi: \mathcal{U} \rightarrow \mathcal{V}$. Here, $\mathcal{U}$ is an open subset in $\mathcal{M}$ and $\mathcal{V}$ is an open subset in $\mathbb{R}^n$. Formally, the smooth atlas satisfies

- $\cup_\alpha \mathcal{U}_\alpha = \mathcal{M}$,
- $\phi_\alpha : \mathcal{U}_\alpha \rightarrow \mathcal{V}_\alpha \in \mathbb{R}^n$ (homeomorphic),
- For any intersecting pair of open sets, there exists $\Phi_{\beta\alpha} := \phi_\beta \circ \phi_\alpha^{-1} \in C^\infty$.

The notion of smooth atlas is further depicted in Fig. 1. The maximal atlas of a set $\mathcal{M}$ is the differentiable (smooth) structure on $\mathcal{M}$. As such, the topology induced by $\mathcal{F}$ is Hausdorff, second countable, and locally Euclidean. A Lie group is defined as the connected elements of a continuous group with an analytic group action (differentiable mapping). A matrix Lie group G is the group of *nonsingular* matrices with a smooth manifold structure so that every element in the group is closed under a group action ($(X, Y) \mapsto XY$: $G \times G \mapsto G$) and has an inverse ($X \mapsto X^{-1}$: $G \mapsto G$) where the group action is a smooth map.

A set of orthogonal matrices denoted as $\mathcal{O}(n)$ is an example of a matrix Lie group. It is known that orthogonal matrices could have a determinant of either $+1$ or -1, however they are not connected on a manifold. Thus, a *special* subgroup of orthogonal matrices whose determinant is equal to $+1$ is usually specified. Formally, let $\mathcal{SO}(n)$ be a set of $n \times n$ orthogonal matrices defined as:

$$\mathcal{SO}(n) = \left\{ Y \in \mathbb{R}^{n \times n} : Y^T Y = I,\ det(Y) = 1 \right\} \tag{1}$$

The additional determinant constraint ($\det = 1$) ensures that all matrices in this group are rotation matrices. This set of matrices is called a special orthogonal group which is also an example of Lie groups.

2.1.1. Lie algebra

Geometrically, Lie groups can be locally considered equivalent to a vector space in $\mathbb{R}^n$. Any element in a Lie group can construct a tangent space. In particular, the tangent space at the identity forms a Lie algebra. Formally, a Lie algebra is a vector space together with a map such that elements in the Lie algebra are closed under the Lie bracket ($[A, B] = AB - BA$) and satisfy the Jacobi identity ($[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0$).

Taking a special orthogonal group as an illustration, the Lie algebra of $\mathcal{SO}(n)$ is denoted as $\mathfrak{so}(n)$. Let Y be an $n \times n$ matrix parametrized by a curve $f(t)$ in $\mathcal{SO}(n)$ such that $f(0) = I$, and $f'(0) = \Delta$. When we differentiate the expression, $f(t)^T f(t) = I$, with respect to t using the product rule, we have

$$f(t)^T f'(t) + f'(t)^T f(t) = 0 \tag{2}$$

At $t = 0$, we have $\Delta + \Delta^T = 0$. We can see that Δ is skew symmetric ($\Delta^T = -\Delta$) and it is closed in the Lie algebra under the Lie bracket. Thus, the space of a Lie algebra of $\mathcal{SO}(n)$ is a set of skew symmetric matrices expressed as:

$$\mathfrak{so}(n) = \left\{ \Delta \in \mathbb{R}^{n \times n} : \Delta + \Delta^T = 0 \right\}. \tag{3}$$

A known example of a Lie group and a Lie algebra is the 2D affine group where $\begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathcal{SO}(3)$ and $\begin{bmatrix} 0 & -\theta & 0 \\ \theta & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathfrak{so}(3)$. Hence, $\mathcal{SO}(3)$ is the natural state space for the rigid motion in $\mathbb{R}^n$.

2.1.2. Exponential mapping

While a Lie group and a Lie algebra have different geometric structures, there exists a map connecting these two spaces called the exponential map, $\exp: \mathfrak{g} \rightarrow G$ where G is a Lie group and $\mathfrak{g}$ is its Lie algebra. For $v \in \mathfrak{g}$, we have $\exp(v) = \gamma(1)$ where $\gamma : \mathbb{R} \rightarrow G$ is the one-parameter subgroup of G. The exponential map also establishes diffeomorphism of an open neighborhood V of 0 in $\mathfrak{g}$ onto an open neighborhood U of e (the identity element) in G.

Examples of exponential maps include the Euclidean space; that is $\exp(v) = v$ where G is $\mathbb{R}^n$. In the matrix Lie group, we can define the matrix exponential for any square matrix as the following power series.

$$\exp(\Delta) = \sum_{k=0}^{\infty} \frac{\Delta^k}{k!} = 1 + \Delta + \frac{\Delta^2}{2!} + \cdots \tag{4}$$

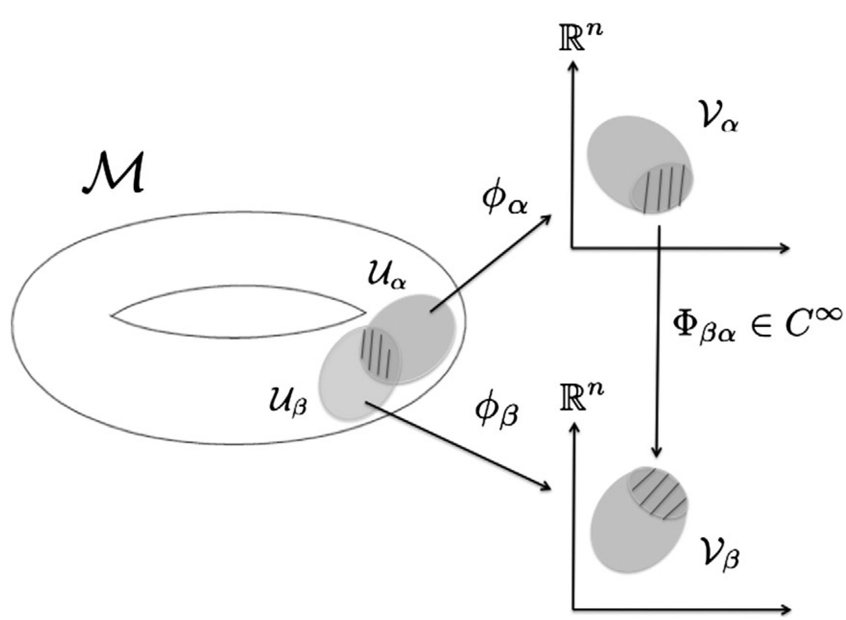

Fig. 1. The notion of smooth atlas: 1) The manifold $\mathcal{M}$ can be covered by a collection of charts $\{(\mathcal{U}, \phi)\}$; 2) there exists a bijective map ϕ from an open set of $\mathcal{U}$ to an open set of $\mathcal{V}$ in $\mathbb{R}^n$; 3) there exists a composite map between different open subsets in $\mathbb{R}^n$ when there is an intersection between open subsets on $\mathcal{M}$.

Eq. (4) reveals that $exp(0)$ is equal to the identity matrix I, and the exponential map yields absolute convergence for each matrix. Additionally, if $f(t)=exp(t\Delta)$, we can see that $f'(0)=\Delta$ and $f(0)=I$. From Eqs. (1,3,4), we can see that $exp(\Delta)$ is orthogonal if and only if Δ is skew-symmetric.

2.1.3. Logarithmic mapping

The inverse mapping from G to g is called the logarithmic map. In particular for the orthogonal group, the logarithmic map, $\log: \mathcal{SO}(n) \mapsto \mathfrak{so}(n)$, is defined as the power series.

$$\log(Y) = \sum_{k-1}^{\infty} \frac{(-1)^{k-1}(Y-I)^k}{k} = (Y-I) - \frac{(Y-I)^2}{2} + \frac{(Y-I)^3}{3} - \cdots \quad (5)$$

The series described in Eq. (5) would only converge when $\|Y-I\|<1$; thus, the logarithmic map is defined around the neighborhood of the identity matrix I. The exponential and logarithmic mappings provide a diffeomorphism (a bijective smooth function with a smooth inverse function) between an open neighborhood of 0 in the Lie algebra and the neighborhood near identity in the Lie group. Fig. 2 gives a pictorial summary of a Lie group, a Lie algebra, exponential and logarithmic mappings.

2.1.4. The Baker–Campbell–Hausdorff formula

The exponential function maps an element from the Lie algebra to its corresponding group element. Matrices from the Lie algebra may serve as exponential coordinates, i.e., $Y=exp(A)$. Since matrix multiplication is the group action in the matrix group, we need to consider the multiplication in exponential coordinates. For commutative groups, we have $AB=BA$ implying that the exponential function satisfies $exp(A)exp(B)=exp(A+B)$. However, this identity does not hold for non-commutative groups.

For non-commutative groups, it is possible to formulate $exp(A)exp(B)=exp(C)$, then the matrix C can be rewritten as $log(exp(A)exp(B))$. From Eqs. (4) and (5), we have

$$C = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \left\{ \left(\sum_{i=0}^{\infty} \frac{A^i}{i!} \right) \left(\sum_{j=0}^{\infty} \frac{B^j}{j!} \right) - I \right\}^k \quad (6)$$

$$= A + B + \frac{1}{2}[A, B] + O\left(|A, B|^3\right)$$

where $[A, B]$ is the Lie bracket which serves as a commutator operation; this series is called the Baker–Campbell–Hausdorff (BCH) formula [7]. Consequently, one may express $exp(A)exp(B)$ as $exp(\mathrm{BCH}(A, B))$; in practice, it is common to use the first order term from the BCH formula from Eq. (6).

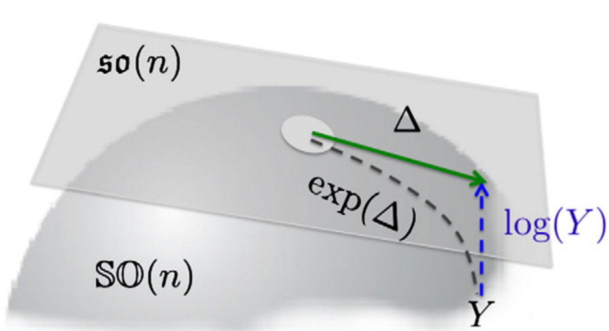

Fig. 2. An example of a special orthogonal group: mappings between a Lie group and a Lie algebra using the exponential and logarithmic maps.

2.2. Stiefel manifolds

A Stiefel manifold $\mathcal{V}_{n,p}$ is a set of of $n\times p$ orthonormal matrices endowed with a Riemannian structure. $\mathcal{V}_{n,p}$ can be considered a quotient space of $\mathcal{O}(n)$ so we can identify an isotropy subgroup H of $\mathcal{O}(n)$ expressed as $\left\{ \begin{bmatrix} I_p & 0 \\ 0 & Q_{n-p} \end{bmatrix} : Q_{n-p} \in \mathcal{O}(n-p) \right\}$ where the isotropy subgroup keeps the element unchanged.

To represent $\mathcal{V}_{n,p}$ as a quotient group, let W be an element of $\mathcal{O}(n)$ such that W can be factorized into $[Y|Y_\perp]$ where $Y \in \mathbb{R}^{n\times p}$ and $Y_\perp \in \mathbb{R}^{n\times(n-p)}$ is the orthogonal complement of Y. Then, each point on a Stiefel manifold corresponds to an equivalence class $\lfloor Y \rfloor$ with respect to the equivalence relation $\left\{ [Y|Y_\perp] \begin{bmatrix} I_p & 0 \\ 0 & Q_{n-p} \end{bmatrix} : Q_{n-p} \in \mathcal{O}(n-p) \right\}$. As such, $W_1 \sim W_2 \Leftrightarrow W_1 = W_2 h$ where W_1 and $W_2 \in \mathcal{O}(n)$, and $h \in H$. Because the equivalence class only concerns the first p columns of a matrix, an element on a Stiefel manifold is defined as:

$$\mathcal{V}_{n,p} = \{Y \in \mathbb{R}^{n\times p} : Y^T Y = I_p\}. \quad (7)$$

Hence, $\mathcal{V}_{n,\,p}$ is viewed as $\mathcal{O}(n)/H = \mathcal{O}(n)/\mathcal{O}(n-p)$. From a group theory point of view, $\mathcal{O}(n)$ is a Lie group and $\mathcal{O}(n-p)$ is its subgroup so that $\mathcal{O}(n)/\mathcal{O}(n-p)$ represents the orbit space. In other words, $\mathcal{V}_{n,\,p}$ is the quotient group of $\mathcal{O}(n)$ by $\mathcal{O}(n-p)$.

2.3. Grassmann manifolds

A Grassmann manifold $\mathcal{G}_{n,p}$ is a set of p-dimensional linear subspaces of $\mathbb{R}^n$. Similar to Stiefel manifolds, a Grassmann manifold can also be identified as a quotient space of $\mathcal{O}(n)$ and its isotropy subgroup composes all elements of $\left\{ \begin{bmatrix} Q_p & 0 \\ 0 & Q_{n-p} \end{bmatrix} : Q_p \in \mathcal{O}(p), Q_{n-p} \in \mathcal{O}(n-p) \right\}$ Thus, we have $[Y|Y_\perp] \begin{bmatrix} Q_p & 0 \\ 0 & Q_{n-p} \end{bmatrix} = [Y\,Q_p | Y_\perp Q_{n-p}]$ and the equivalence class consists of a $p\times p$ orthogonal matrix Q_p mapping one point onto the other. Considering the first p columns, the entire equivalence class can be represented as the subspace spanned by the columns of a given matrix Y defined as:

$$\lfloor Y \rfloor = \left\{ YQ_p : Q_p \in \mathcal{O}(p) \right\}. \quad (8)$$

The quotient representation of Grassmann manifolds is expressed as $\mathcal{G}_{n,p} = \mathcal{O}(n)/(\mathcal{O}(p)\times\mathcal{O}(n-p))$. Using the quotient group of Stiefel manifolds, we can state Grassmann manifolds more concisely as $\mathcal{G}_{n,p} = \mathcal{V}_{n,p}/\mathcal{O}(p)$. Putting it differently, the set of all orbits of $\mathcal{V}_{n,p}$ under the group action $\mathcal{O}(p)$ is a point on $\mathcal{G}_{n,p}$. This quotient representation of Grassmann manifolds establishes two matrices as being from the same equivalence class if their columns span the same p dimensional subspace, $W_1 \sim W_2 \Leftrightarrow \mathrm{span}(W_1) = \mathrm{span}(W_2)$.

The key distinctions between Stiefel and Grassmann manifolds are the order and the choice of basis. Because of the equivalence relation, there is no unique order or basis of a matrix representing an element from a Grassmann manifold. On the other hand, the order of the basis is important for the elements on Stiefel manifolds.

2.4. Riemannian manifolds

In the Riemannian framework, the tangent space $T_x\mathcal{M}$ at each point x of a manifold $\mathcal{M}$ is endowed with a smooth inner product $<\cdot,\cdot>_x$. In addition, a local coordinate $x=(x^1, x^2, \ldots, x^n)$ on an open set U of $\mathcal{M}$ induces a basis $\frac{\partial}{\partial x} = \left(\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \ldots, \frac{\partial}{\partial x^n} \right)$ of the tangent space. The Riemannian metric can then be expressed as $g_{ij}|_x = <\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}>$ from the tangent vectors. The component $g_{ij}|_x$ forms a matrix G_x so that

$g_{ij}|_x = \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \rangle = \left(\frac{\partial}{\partial x^i}\right)^T G_x \left(\frac{\partial}{\partial x^j}\right)$. The matrix G_x is the local representation of the Riemannian metric in the chart x. As such, G_x is viewed as a point in a vector bundle of $U \times \mathbb{R}^{\frac{n(n+1)}{2}}$ locally because G_x only has $\frac{n(n+1)}{2}$ unique entries; it is symmetric positive definite at every point on the manifold denoted as Sym_n^+. Formally, we have

$$Sym_n^+ = \{Y \in \mathbb{R}^{n \times n} : Y^T = Y, Y > 0\} \tag{9}$$

where $Y>0$ indicates $x^T Y x > 0$ for any $x \in \mathbb{R}^n \setminus \{0\}$. Furthermore, the space of Sym_n^+ is an open convex cone so that $X + tY \in Sym_n^+$ for $t>0$ given X, $Y \in Sym_n^+$. The simplest example of a Riemannian manifold is the Euclidean space where the inner product is $\mathbb{R}^n \times \mathbb{R}^n \simeq T\mathbb{R}^n$.

Once the Riemannian metric is specified, the Riemannian distance on a connected Riemannian manifold can be expressed in a closed form. In particular, one may induce an affine-invariant metric so that points in Sym_n^+ are invariant under the action of the isotropy subgroup and the quotient representation is expressed as $Sym_n^+ = GL_n^+/O(n)$. Also, the exponential and logarithmic mappings on Sym_n^+ can be efficiently computed using spectral decomposition. For details on these computations see [11].

Riemannian manifolds are not limited to Sym_n^+. The generalization of a Riemannian manifold is a semi-Riemannian manifold (pseudo-Riemannian manifold) where the Riemannian metric G_x does not need to be positive definite. Instead, G_x is a smoothly varying symmetric bilinear form on the tangent space and non-degenerate, i.e., $g_{ij}|_x \neq 0$. In addition, the tangent space $T_x\mathcal{M}$ can be factorized to $P \oplus N$ such that g is positive definite on P and negative definite on N.

Finally, the Riemannian metric can be inherited naturally from the embedded space. The Stiefel manifold and the Grassmann manifold are considered the Riemannian submanifold and the Riemannian quotient manifold, respectively.

3. Computer vision on matrix manifolds

Matrix manifolds have been exploited in many computer vision applications. The following subsections summarize these applications including face recognition, action recognition, clustering, visual tracking, and motion grouping and segmentation.

3.1. Face recognition

While Riemannian manifolds characterize a smooth inner product as a point in a vector bundle, Grassmann and Stiefel manifolds provide a geometric structure for subspaces. These manifolds have been utilized as a basis for face recognition in the context of still imagery, videos (a stream of face imagery), image-sets (a set of images collected from different times), and 3D imagery. These efforts are summarized in Table 1.

Optimization methods are often employed in conjunction with projection pursuit for machine learning and pattern recognition applications. Liu et al. [12] exploited discriminant analysis for face recognition using a learned projection to mapimages to a lower dimensional space where the projection was viewed as an element on a Grassmann manifold. The gradient flow on the Grassmann manifold was approximated and the locally optimal projection was found using gradient directed search. Lin et al. [14] proposed a Maximum Effective Information (MEI) criterion maximizing mutual information for face recognition. Because the MEI imposed the homogeneity condition, it gave rise to rotation invariance. The projection matrix was then optimized on a Grassmann manifold using the conjugate gradient method.

Fisher's Linear Discriminant Analysis (LDA) may be the most widely used framework for pattern recognition. Hamm and Lee [17,18] extended the set-based LDA to use kernel functions in Grassmannian. The projection metric and the Binet–Cauchy metric induced the Grassmann kernels. The LDA with Grassmann kernels was then introduced and applied to face recognition, outperforming some earlier techniques for image-set matching like MSM and DCCA [35,36]. Wang and Shi [29] attempted to kernelize geodesic distances from Grassmannian, i.e. arc-length and chordal distance. Nevertheless, experiments concluded that kernel PCA followed by the Grassmann discriminant analysis [17] has performed better than the kernelized geodesic distances in face recognition. Recently, Park and Savvides [32] have extended the use of Grassmann kernels to multilinear tensor representation. Harandi et al. [33] exploited discriminant analysis with graph embedding on Grassmann manifolds for face recognition from videos. The proposed Grassmannian kernel further improved the performance in image-set matching.

Mathematically, LDA can be formulated as a trace quotient problem with the solution obtained from a generalized eigenvalue decomposition. Yan and Tang [13] showed that the optimal solution to the trace quotient problem was constrained to lie on a Grassmann manifold. Specifically,

Table 1
The use of matrix manifolds for face recognition.

Authors & reference	Year	Data source	Manifold	Dataset
Liu et al. [12]	2004	Still imagery	Grassmann	CMU-PIE, ORL
Yan and Tang [13]	2006	Still imagery	Grassmann	CMU-PIE, XM2VTS, AR
Lin et al. [14]	2006	Still imagery	Grassmann	FERET, XM2VTS, PURDUE
Cheng et al. [15,16]	2006	Image set	Grassmann	CMU-PIE
Hamm and Lee [17,18]	2008	Image set, video	Grassmann	YaleB, Extended YaleB
Pham and Venkatesh [19]	2008	Still imagery	Stiefel	CMU-PIE, YaleB, FERET
Wang et al. [20]	2008	Video	Grassmann	Honda/UCSD, CMU-MoBo
Turaga et al. [21]	2008	Video	Grassmann	Honda/UCSD
Tron and Vidal [22]	2008	Still imagery	$SE(3)$	Weizmann
Pang et al. [23]	2008	Still imagery	Sym_n^+	AR, FERET
Zhao et al. [24]	2008	Still imagery	Riemannian	FRGC Exp 4
Lui et al. [25]	2008	Image set	Grassmann	FRGC Exp 4
Lui and Beveridge [26]	2008	Still imagery	Grassmann	FERET
Lui et al. [27]	2009	Still imagery	Stiefel	CMU-PIE, YaleB, Extended YaleB
Beveridge et al. [28]	2009	Image set	Grassmann	CMU-PIE, YaleB
Wang and Shi [29]	2009	Image set	Grassmann	CMU-PIE
Yu et al. [30]	2009	3D range imagery	Riemannian	UND Biometrics Database
Sirvalingam et al. [31]	2010	Still imagery	Sym_n^+	FERET
Park and Savvides [32]	2011	Still imagery	Grassmann	CMU-PIE, Extended YaleB
Harandi et al. [33]	2011	Image set, video	Grassmann	CMU-PIE, BANCA, CMU-MoBo
Turaga et al. [34]	Appear	Image set, video	Grassmann	CMU-PIE, MBGC

the derivative of the trace quotient problem was equivalent to the projection on the horizontal space on the Grassmann manifold where the projection vectors were the eigenvectors of a weighted difference matrix.

Whereas discriminant analysis learns a projection that maximizes the trace quotient, multivariate regression learns a projection by fitting the class labels in a regression space. Pham and Venkatesh [19] applied a lasso regression for face recognition. The loss function was formulated as a quadratic function with an L_1 norm regularization which was used to avoid over-fitting with small training size. Dual projections were employed by the regression process. While one was used for dimension reduction, the other was utilized to fit the projected data to the labels. Since the projection matrix was constrained to be orthogonal, it was found via the steepest descent method on a Stiefel manifold.

It is known that a set of convex objects under a fixed pose with a Lambertian reflectance surface forms a convex subspace called the illumination cone. Since illumination cones reside in a vector space, they can be approximated as points on Grassmann manifolds. Cheng et al. [15] studied face recognition using image-sets which were collected from various fixed pose illumination variants. The image-set classification was performed using the truncated [15] and smallest [28] canonical angles. This study was further extended to low resolution imagery in illumination spaces [16] and unconstrained image-sets [25].

Video provides a convenient way to collect image-sets. Turaga et al. [21] performed video-based face recognition where the appearance was obtained from the parameter of the AutoRegressive and Moving Average (ARMA) model and represented as an element in Grassmannian. Procrustes and kernel distances were employed for classification. This work was extended to the tangent space where data were mapped to the tangent space centered at the Karcher mean and the Gaussian fit was applied to the projected data for face recognition [34]. Tron and Vidal [22] performed face recognition using a sensor network where the state of each camera is a pose residing on a special Euclidean group $\mathcal{SE}(3)$ ($\mathcal{SO}(3) \times \mathbb{R}^3$). Neighboring cameras were used in joint estimation for face poses; as such, the global Karcher mean for the face pose was estimated.

Image-sets or videos can be naturally represented on special manifolds; however, they may not be available in some applications. To embed a single image on special manifolds, Lui et al. [27] employed a statistical illumination model to relight a single image to a set of fixed pose illumination variants. The relighted illumination variants were then projected on the tangent space on a Stiefel manifold. Because the order of the data plays an important role on Stiefel manifolds, the relighted illumination variants were more discriminative on Stiefel manifolds than Grassmann manifolds. To further relax the fixed pose constraint, Lui and Beveridge [26] employed affine transformations to sample the image manifold and selected a local neighborhood to form a tangent space. Due to a vector space structure in tangent spaces, they were embedded on a Grassmann manifold for face recognition.

While the geodesic distance can be computed based on the intrinsic geometry of a Grassmann manifold, heuristics have been applied to modify the use of canonical angles. Wang et al. [20] proposed the use of the weighted average between the exemplar distance and the variation distance as a manifold-to-manifold distance measure. The exemplar distance was defined as the correlation between the orthogonal exemplar samples and the variation distance was defined as using canonical angles from a subset of images based on local linearity.

Riemannian manifolds have also been studied for face recognition. The applications span both 2D and 3D imagery. Pang et al. [23] represented a face as a set of gabor features whose covariance matrix was expressed on Sym_n^+. Zhao et al. [24] performed unsupervised manifold learning on semi-Riemannian manifolds. The tensor metric was determined by discretized Laplacian smoothing and nullity of the semi-Riemannian space. Sirvalingam et al. [31] applied sparse decomposition to reformulate a symmetric positive definite matrix as a linear combination of a dictionary atoms. As a consequence, this led to a convex determinant maximization problem solved by an interior point method. On the other hand, Yu et al. [30] considered 3D range imagery and represented a three-dimensional vector on a unit sphere of S^2 which is a Riemannian space. The logarithmic map was then applied to transform points to a tangent plane T_pS^2 for discriminant analysis.

Age estimation and expression recognition from facial images are emerging areas in computer vision. Turaga et al. [37] showed that the geodesic velocity from an average face to the estimated face can be used for age estimation. Specifically, the space of landmarks was interpreted as a Grassmann manifold and all points on the manifold were projected onto the tangent space at the intrinsic mean. As such, the velocity of geodesic flows was used in regression for age estimation. Taheri et al. [38] performed facial expression analysis on the Grassmann manifold. To cope with facial deformation, the velocity vectors from different facial action units were transformed to the neutral face by applying parallel transport. Each action unit template was then modeled as a Gaussian distribution for expression analysis.

3.2. Action recognition

The characterization of human activities and actions using matrix manifolds has been a recent focus in the computer vision community as summarized in Table 2. Human actions can be represented as a sequence of silhouettes. Veeraraghavan et al. [39] modeled human shapes on a shape manifold where sequences were first normalized using Dynamic Time Warping (DTW). A sequence of shape changes was then extracted from the tangent space and an ARMA model was exploited to learn the dynamics of the human movement. The space spanned by the parameters of the linear dynamic system was identified as an element on a Grassmann manifold. This ARMA representation of human activity was further extended to clustering [40]. The cascade of ARMA models was viewed as a regular expression grammar, and grammatical inference was applied to action recognition.

Turaga et al. [21,34] investigated statistical modeling and Procrustes representation on special manifolds for human activity recognition. In

Table 2
The use of matrix manifolds for action recognition.

Authors & Reference	Year	Model	Manifold	Dataset
Veeraraghavan et al. [39]	2005	ARMA + DTW	Shape + Grassmann	CMU Activity, MOCAP
Turaga et al. [40]	2007	Cascade ARMA	Grassmann	UMD Common Activity
Turaga et al. [21,34]	2008	ARMA + Procrustes	Grassmann + Stiefel	IXMAS
Turaga and Chellappa [41]	2009	TV-LDS + DTW	Grassmann	UMD Common Activity
Veeraraghavan et al. [42]	2009	DTW	Shape	UMD Common Activity, IXMAS, USF Gait
Li and Chellappa [43]	2010	SIS + Alignment	Stiefel	USF Gait, KTH
Chaudhry and Ivanova [44]	2010	Spectral Hashing	Riemannian	KTH
Guo et al. [45,46]	2010	Covariance descriptor + Sparsity	Sym_n^+	Weizmann Action, KTH
Lui et al. [47]	2010	Product Manifold	Grassmann	Cambridge-Gesture, KTH
Lui and Beveridge [48]	2011	Tangent Bundle	Grassmann	Cambridge-Gesture, KTH, UCF Sport
Abdelkader et al. [49]	2011	Markov Chain	Sym_n^+	UMD Common Activity, UMD Body Gesture

this work, the Procrustes distance characterized by the kernel density function was used to compare two subspaces on a Stiefel manifold. The trajectories on a Grassmann manifold were also exploited using human activities modeled by Time-Varying Linear Dynamic Systems (TV-LDS). DTW was utilized to normalize two sequences of actions. Recently, Abdelkader et al. [49] modeled the contour shape as a point in a shape space of closed curves associated with Riemannian geometry. The trajectory on the shape space was characterized by a Markovian graphical model and used for action classification.

While ARMA models extract shapes and dynamics from the observability matrix, appearance, horizontal motion, and vertical motion can be directly obtained via a modified High Order Singular Value Decomposition (HOSVD). Lui et al. [47] modeled the appearance, horizontal motion, and vertical motion on three factor manifolds (Grassmannian) since they are orthogonal matrices. The geodesic distance on the product manifold formed by combining the three factor manifolds was used for action classification. The use of tensor decomposition was further demonstrated in conjunction with a tangent bundle on a Grassmann manifold [48]. Since the logarithmic map is a diffeomorphic function, it was applied to charting the manifold totangent spaces. Consequently, the tangent vectors on three tangent spaces associated with appearance, horizontal and vertical motions were exploited for action recognition. The introduction of a tangent bundle facilitates the integration with other classifiers because tangent spaces are Euclidean.

Spatio-temporal alignment is a key step for appearance-based action classification. Veeraraghavan et al. [42] studied the rate-invariant temporal alignment for human activities. In this work, rate variation within a fixed time interval was modeled as a diffeomorphism such that time warping functions can be represented as a temporal affine transformation. Li and Chellappa [43] employed Sequential Importance Sampling (SIS) on a Stiefel manifold for spatio-temporal alignment. The alignment parameters were divided into spatial and temporal spaces viewed as submanifolds. The parameter estimation was considered an optimization problem through SIS.

Another school of thought for action classification is in the use of feature-based methods. Feature-based methods compute a covariance descriptor from a set of image features. The covariance descriptor is a symmetric positive definite matrix which can be viewed as an inner product on the tangent space of Sym_n^+; as a result, covariance descriptors are the natural elements on a Riemannian manifold. Instead of adopting Riemannian manifolds, Guo et al. [45,46] employed the matrix logarithm to map the covariance descriptor to the vector space. A sparse linear representation was then applied to encode the actions. The approximate nearest neighbor search on Riemannian manifolds [44,50] and Grassmann manifolds [50] was also investigated.

3.3. Clustering

Grouping similar patterns into a cluster or computing data centroids is an important process in unsupervised learning. While the concept of averaging in non-Euclidean spaces can be generally characterized as the Karcher mean computation [51] or the rotation group $\mathcal{SO}(3)$ averaging [52], many variants have been derived for myriad applications. We first summarize the clustering methods on matrix manifolds in Table 3.

Since the parameters of rigid motions can be characterized as elements of orthogonal groups, motion estimation can naturally be cast on $\mathcal{SO}(3)$. Govindu [53] exploited this characteristic and iteratively computed the motion average on the Lie-algebra. Because of the closed form expression of the Baker–Campbell–Hausdorff (BCH) formula, the intrinsic averaging was performed on the Lie algebra without exponential and logarithm mappings, resulting in a faster calculation.

Tuzel et al. [54] incorporated Lie groups with the Mean-Shift (MS) algorithm for 3D motion estimation. The motion parameters represented on a Lie group were estimated via mode finding on the sampled distribution. The kernel density function was expressed using an intrinsic distance through the BCH formula. The modes of the underlying distribution were then computed iteratively. Subbarao and Meer [55,56] expressed the MS algorithms on Lie groups and Grassmann manifolds. This method computed the mean shift as weighted tangent vectors on tangent spaces and projected it back to the manifold via the exponential map. On the other hand, Cetingul and Vidal [59] proposed an alternative method for the nonlinear MS algorithms on Stiefel and Grassmann manifolds. These nonlinear MS methods avoided the involvement of tangent spaces; in other words, no exponential map was needed. The kernel density functions were estimated on the manifolds so the modes of a distribution were located intrinsically through iterative optimization.

Gruber and Theis [58] derived a clustering algorithm using the Chordal distance (Projection F-norm) on Grassmann manifolds. The clustering processing was shown to be a linear optimization problem on a convex set with the centroid being pseudo orthogonal; consequently, the optimum clusters can be located at the corners of the convex set.

The use of Riemannian manifolds for clustering has also received attention. Arsigny et al. [62] introduced an efficient way for computing the geometric mean on Sym_n^+ where the space of log-Euclidean was considered. Sirvalingam et al. [60] investigated the use of covariance descriptors for clustering with metric learning. Since the tensor metric on Sym_n^+ is symmetric positive definite, it can be represented as a class of Mahalanobis distances. As such, thetensor metric was determined by minimizing the LogDet divergence under linear constraints. The clustering was performed using the pairwise constrained K means algorithm.

Begelfor and Werman [57] investigated three clustering algorithms on Grassmann manifolds including k-means, mean shift, and average link methods. The authors reported that the clustering algorithms on Grassmann manifolds yielded lower classification errors and increased robustness to the presence of noise when compared with methods operating in Euclidean space. The aspect of open-ended clustering for human actions where the number of clusters was unknown was also investigated using bag-of-feature and product manifold methods in [61]. The product manifold representation was found to be superior, particularly when the activities consisted of gross motions.

Table 3
The use of matrix manifolds for clustering.

Authors & reference	Year	Method	Manifold	Dataset
Govindu [53]	2004	Intrinsic average	Lie group	Proprietary
Tuzel et al. [54]	2005	Mean-shift	Lie group	Proprietary
Subbarao and Meer [55,56]	2006	Mean-shift	Grassmann	Proprietary
Begelfor and Werman [57]	2006	Karcher mean	Grassmann	Proprietary
Gruber and Theis [58]	2006	Closed-form	Grassmann	Proprietary
Cetingul and Vidal [59]	2009	Mean-shift	Stiefel, Grassmann	ETH-80, Hopkins-155
Sirvalingam et al. [60]	2009	Constrained K-means	Sym_n^+	Proprietary
O'Hara et al. [61]	2011	Agglomerate	Grassmann	UCSD-Expression, Cambridge-Gesture, KTH
Turaga et al. [34]	appear	Karcher mean	Grassmann	USF-Figure-Skating

3.4. Visual tracking

In recent years, many visual tracking algorithms have benefited from matrix manifold representations. Table 4 provides a summary of visual tracking on matrix manifolds.

Rigid body tracking can be formulated and parametrized through a set of three dimensional matrices in which the parameter space forms a Lie group. Drummond and Cipolla [63–65] expressed the differentiation of each mode of an affine transformation as a generator of the Lie group. As a consequence, local transformations near the identity formed the basis for the Lie algebra; the exponential map was applied to project the resulting transformation back to the group. Bayro-Corrochano and Ortegón-Aguilar [67] also related the group action as the composition of the transformations. The transformations were then represented as a weighted sum of the generators followed by the exponential map. Particle filtering has also been implemented on the affine group for visual tracking [74–76,79].

Subspace methods are frequently employed for visual tracking. Wang et al. [69] viewed online face tracking as a subspace tracking problem from a Grassmann manifold with the projection matrix varying over time. At each time step, the principal components of appearance were obtained from a Grassmann manifold while a particle filter was used to track the object. The direction of the movement of the subspace was then updated using a Kalman filter and mapped back to the Grassmann manifold using the exponential map. An efficient method for subspace tracking using particle filtering on Grassmann manifolds was studied in [81].

The use of Riemannian geometry [66,68,70–73,77,78,80] for visual tracking has also been studied in recent years. The common theme of these approaches is using covariance descriptors to characterize object appearances where covariance matrices are computed from a set of image features. By imposing a meaningful inner product on the tangent spaces of Sym_n^+, the space of covariance matrices becomes a Riemannian manifold. The representation of a covariance descriptor on a Riemannian manifold has also been applied to pedestrian detection [82–84], head pose detection [85], and stereo matching [86]. The covariance feature sets for face recognition and pedestrian were evaluated in [87].

3.5. Motion grouping and segmentation

Motion estimation is one of the pioneering applications using matrix manifolds because rigid motions can be represented by orthogonal groups. A summary of this work is given in Table 5.

Horn [88] first explicitly considered the orthogonal constraint from the properties of differential geometry when measuring the orientation between cameras. Taylor and Kriegman [89] also estimated the camera rotational error from rotation matrices characterized as points in $\mathcal{SO}(3)$. As such, the optimization applied in these methods was unconstrained in the underlying space. Ma et al. [90,91] formulated the motion recovery problem by using the epipolar constraint $x_2^T \hat{T} R x_1 = 0$ where x_1 and x_2 were the image points, and R and $\hat{T}$ were the essential matrices. From a geometric point of view, $R \in \mathcal{SO}(3)$ and $\hat{T} \in so(3)$ reside on the Lie group and Lie algebra, and the product of these spaces is a product of Stiefel manifolds. The use of a Lie group for trajectory generation from rigid bodies was also investigated in [92] where the optimal trajectory in the ambient space was first constructed followed by a projection on the Euclidean group.

Table 5
The use of matrix manifolds for motion grouping and segmentation.

Authors & reference	Year	Manifold
Horn [88]	1990	Lie group
Taylor and Kriegman [89]	1995	Lie group
Ma et al. [90,91]	1998	Lie group
Belta and Kumar [92]	2002	Lie group
Subbarao and Meer [93]	2006	Grassmann
da Silva and Costeria [94]	2008	Grassmann
Lin et al. [95,96]	2009	Lie group
Li and Chellappa [97]	2010	Lie group
Mégret et al. [98]	2010	Lie group

Analyzing a group of motion in videos gives rise to many visual applications, such as video surveillance and weather modeling. A group of motion known as visual flow can be considered a dynamic flow and parametrized in geometric transforms. Lin et al. [95] employed a 2D affine transformation to model the motion patterns. The flow process was parametrized using the Lie algebra representation such that the intermediate transformations along the geodesic path remained in the subgroup.

Li and Chellappa [97] expressed the group motion as a driving force represented by an affine group. The affine group is essentially a Lie group. To assimilate the nonlinear nature of a Lie group, the driving force was mapped to a Lie algebra so that linear dependence between location and velocity was captured. The temporal sequence was then parametrized in the Cartesian product space between the driving force and the effective area.

The Lie algebraic representation of dynamic flows casts flow modeling as parameter estimation. Instead of matrix parametrization, Lin et al. [96] directly estimated the visual flow by decomposing the image differences based on infinitesimal generators. The infinitesimal generator of each flow was represented as a linear combination of the motion patterns. The authors reported that the frame differencing was more reliable with smooth textures.

Motion grouping has been viewed as subspace segmentation. Subbarao and Meer [93] considered subspace estimation using m-estimators. The parameters of the m-estimators were tracked using a conjugate gradient algorithm on a Grassmann manifold. Newton's

Table 4
The use of matrix manifolds for visual tracking.

Authors & reference	Year	Model	Manifold
Drummond et al. [63–65]	1999	Optimization	Lie group
Porikli et al. [66]	2006	Exhaustive search	Lie group
Bayro-Corrochano and Ortegón-Aguilar [67]	2007	Optimization	Lie group
Tuzel et al. [68]	2008	Multiscale	Lie group
Wang et al. [69]	2008	Kalman Filter + Particle filter	Grassmann
Li et al. [70,71]	2008	Particle filter	Sym_n^+
Wu et al. [72]	2008	Particle filter	Sym_n^+
Palaio and Batisa [73]	2008	Particle filter	Sym_n^+
Kwon et al. [74,75]	2009	Particle filter	Lie group
Porikli and Pan [76]	2009	Particle filter	Lie group
Wang et al. [77,78]	2009	Particle filter	Sym_n^+
Li et al. [79]	2010	Particle filter	Lie group
Ding et al. [80]	2010	Particle filter	Sym_n^+

method was also applied to find the subspace with the maximum number of inliers in Grassmannian for motion grouping [94]. Recently, Mégret et al. [98] reformulated the Lucas-Kanade alignment algorithm on Lie groups using bidirectional composition where both the image and the template were warped incrementally. The compositional increment was parametrized in the Lie algebra.

4. Discussion and summary

When a pattern can be characterized by a state which is a natural element in a particular manifold, there is a connection between the pattern space and the underlying group structure. We can then parametrize the underlying space via algebraic entitiesrather than directly on the patterns. The group element acts as an operator in the space making some attribute in the space invariant. Algebraic characterization of the geometry plays the key role in matrix manifolds.

In computer vision, we relate visual data to a particular matrix manifold through the invariance imposed on the data. In some applications, the characteristics of visual patterns may be described by some transformation whose state transition is governed by a group action. We may then exploit the parameter in the transformation space. Because group elements are closed under a group action, the constraint imposed on the data is automatically maintained.

A Grassmann manifold parametrizes the set of all p-dimensional subspaces in $\mathbb{R}^n$ so that every point in a Grassmann manifold is rotation invariant. That is, rotating the column space of a matrix by multiplying an orthogonal matrixon the right such that the imposed data (a matrix in this case) are invariant to the linear span of the data. On the other hand, the isotropy subgroup of a Stiefel manifold puts a constraint on the order of the matrix. That constraint should be an important consideration when choosing a Stiefel manifold for a particular application.

A metric space may also be geometric invariant. Affine invariance can be imposed to the tensor metric of a Riemannian manifold. In particular, there exists a map $f: Sym_n^+ \rightarrow Sym_n^+$ which is an isometry (a distance preserving map) for symmetric positive definite matrices. From the quotient representation of $Sym_n^+ = GL_n^+/\mathcal{O}(n)$, this gives the invariance to the distance metric, i.e., $dist(U \sum U^T) = dist(\sum)$ where $U \in \mathcal{O}(n)$ and $\sum \in Sym_n^+$.

In summary, geometry may be the most fundamental basis in pattern analysis. Matrix manifolds provide a natural way to characterize some visual objects. There are many emerging computer vision applications on the horizon. It can be advantageous to exploit the underlying geometry of the data. There has been a recent interest in associating matrix manifolds with statistical learning theory [99,100]. Since computer vision is intimately related to machine learning, there may be potential impacts of investigating this area.

Matrix manifolds were developed more than a century ago but are still relatively new in the computer vision community. They are well-defined, rich, and elegant. The strides of matrix manifold representations will likely produce dividends.

Acknowledgments

The author thanks Dr. J. Ross Beveridge and Mr. Stephen O'Hara from Colorado State University for reading the manuscript. The author also thanks the anonymous reviewers for their valuable comments which helped to improve the quality of this manuscript.

References

[1] H. Seung, D. Lee, The manifold ways of perception, Science (2000) 2268–2269.
[2] J. Tenenbaum, V. de Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction, Science (2000) 2319–2323.
[3] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science (2000) 2323–2326.
[4] E. Scholz, The concept of manifold, 1850–1950, in: I.M. James (Ed.), History of Topology, 1999, pp. 25–64.
[5] J. Lee, Introduction to Smooth Manifolds, Springer, 2003.
[6] P. Peterson, Riemannian Geometry, Springer Graduate Texts in Mathematics, 2000.
[7] W. Rossmann, Lie Groups: An Introduction Through Linear Groups, Oxford Graduate Text in Mathematics, 2002.
[8] P.-A. Absil, R. Mahony, R. Sepulchre, Optimization Algorithms on Matrix Manifolds, Princeton University Press, 2008.
[9] Y. Chikuse, Statistics on Special Manifolds, Lecture Notes in Statistics, Springer, New York, 2003.
[10] A. Edelman, R. Arias, S. Smith, The geometry of algorithms with orthogonal constraints, SIAM Journal on Matrix Analysis and Applications 2 (1999) 303–353.
[11] X. Pennec, P. Fillard, N. Ayache, A Riemannian framework for tensor computing, International Journal of Computer Vision 66 (2006) 44–66.
[12] X. Liu, A. Srivastava, K. Gallivan, Optimal linear representations of images for object recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (5) (2004) 662–666.
[13] S. Yan, Z. Tang, Trace quotient problems revisited, European Conference on Computer Vision, Graz, Austria, 2006.
[14] D. Lin, S. Yan, X. Tang, Pursuing informative projection on Grassmann manifold, IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 386–389.
[15] J.-M. Cheng, J.R. Beveridge, B. Draper, M. Kirby, H. Kley, C. Peterson, Illumination face spaces are idiosyncratic, The International Conference on Image Processing, Computer Vision, and Pattern Recognition, 2006, pp. 390–396.
[16] J.-M. Cheng, M. Kirby, H. Kley, C. Peterson, B. Draper, J.R. Beveridge, Recognition of digital images of the human face at ultra low resolution via illumination spaces, Asian Conference on Computer Vision, 2007, pp. 733–743.
[17] J. Hamm, D. Lee, Grassmann discriminant analysis: a unifying view on subspace-based learning, International Conference on Machine Learning, 2008, pp. 376–383.
[18] J. Hamm, D. Lee, Extended Grassmann kernels for subspace-based learning, Advances in Neural Information Processing System, 2008, pp. 601–608.
[19] D.-S. Pham, S. Venkatesh, Robust learning of discriminative projection for multicategory classification on the Stiefel manifold, IEEE Conference on Computer Vision and Pattern Recognition, 2008.
[20] R. Wang, S. Shan, X. Chen, W. Gao, Manifold-manifold distance with application to face recognition based on image set, IEEE Conference on Computer Vision and Pattern Recognition, 2008.
[21] P. Turaga, A. Veeraraghavan, R. Chellappa, Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision, IEEE Conference on Computer Vision and Pattern Recognition, 2008.
[22] R. Tron, R. Vidal, Distributed face recognition via consensus on se(3), Workshop on Omnidirectional Vision, Marseille, France, 2008.
[23] Y. Pang, Y. Yuan, X. Li, Gabor-based region covariance matrices for face recognition, IEEE Transactions on Circuits and Systems for Video Technology 18 (2008) 989–993.
[24] D. Zhao, Z. Lin, X. Tang, Classification via semi-Riemannian spaces, Computer Vision and Pattern Recognition, Anchorage, Alaska, 2008.
[25] Y.M. Lui, J.R. Beveridge, B.A. Draper, M. Kirby, Image-set matching using a geodesic distance and cohort normalization, IEEE International Conference on Automatic Face and Gesture Recognition, Amsterdam, The Netherlands, 2008.
[26] Y.M. Lui, J.R. Beveridge, Grassmann registration manifolds for face recognition, European Conference on Computer Vision, Marseille, France, 2008.
[27] Y.M. Lui, J.R. Beveridge, M. Kirby, Canonical Stiefel quotient and its application to generic face recognition in illumination spaces, IEEE International Conference on Biometrics : Theory, Applications and Systems, Washington, D.C., 2009.
[28] J.R. Beveridge, B.A. Draper, J.-M. Chang, M. Kirby, H. Kley, C. Peterson, Principal angles separate subject illumination spaces in ydb and cmu-pie, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2) (2009) 351–363.
[29] T. Wang, P. Shi, Kernel Grassmannian distances and discriminant analysis for face recognition from image sets, Pattern Recognition Letters 30 (13) (2009) 1161–1165.
[30] D. Yu, J. Lu, J. Yang, Geodesic discriminant analysis on curved Riemannian manifold, International Conference on Fuzzy Systems and Knowledge Discovery, Tianjin, China, 2009.
[31] R. Sirvalingam, D. Boley, V. Morellas, N. Papanikolopoulos, Tensor space coding for region covariances, European Conference on Computer Vision, Crete, Greece, 2010.
[32] S.W. Park, M. Savvides, The multifactor extension of Grassmann manifolds for face recognition, IEEE International Conference on Automatic Face and Gesture Recognition, Santa Barbara, 2011.
[33] M.T. Harandi, C. Sanderson, S. Shirazi, B.C. Lovell, Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching, IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, 2011.
[34] Turaga, P.; Veeraraghavan, A.; Srivastava, A.; Chellappa, R., Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (appear).
[35] O. Yamaguchi, K. Fukui, K. Maeda, Face recognition using temporal image sequence, International Conference on Face and Gesture Recognition, Nara, Japan, 1998, pp. 318–323.
[36] T.-K. Kim, J. Kittler, R. Cipolla, Discriminative learning and recognition of image set classes using canonical correlations, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (6) (2007) 1–14.

[37] P. Turaga, S. Biswas, R. Chellappa, The role of geometry for age estimation, IEEE International conference Acoustics, Speech and Signal Processing, 2010.
[38] S. Taheri, P. Turaga, R. Chellappa, Towards view-invariant expression analysis using analytic shape manifolds, IEEE International Conference on Automatic Face and Gesture Recognition, Santa Barbara, 2011.
[39] A. Veeraraghavan, A.K. Roy-Chowdhury, R. Chellappa, Matching shape sequences in video with applications in human movement analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (2005) 1896–1909.
[40] P. Turaga, A. Veeraraghavan, R. Chellappa, From videos to verbs: mining videos for events using a cascade of dynamical systems, IEEE Conference on Computer Vision and Pattern Recognition, 2007.
[41] P. Turaga, R. Chellappa, Locally time-invariant models of human activities using trajectories on the Grassmannian, IEEE Conference on Computer Vision and Pattern Recognition, 2009.
[42] A. Veeraraghavan, A. Srivastava, A.K. Roy-Chowdhury, R. Chellappa, Rate-invariant recognition of humans and their activities, IEEE Transactions on Image Processing 6 (2009) 1326–1339.
[43] R. Li, R. Chellappa, Aligning spatio-temporal signals on a special manifold, European Conference on Computer Vision, Crete, Greece, 2010.
[44] R. Chaudhry, Y. Ivanov, Fast approximate nearest neighbor methods for non-euclidean manifolds with applications to human activity analysis in videos, European Conference on Computer Vision, Crete, Greece, 2010.
[45] K. Guo, P. Ishwar, J. Konrad, Action recognition in video by sparse representation on covariance manifolds of silhouette tunnels, International Conference on Pattern Recognition, Istanbul, Turkey, 2010.
[46] K. Guo, P. Ishwar, J. Konrad, Action recognition using sparse representation on covariance manifolds of optical flow, International Conference on Advanced Video and Signal-Based Surveillance, Boston, 2010.
[47] Y.M. Lui, J.R. Beveridge, M. Kirby, Action classification on product manifolds, IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, 2010.
[48] Y.M. Lui, J.R. Beveridge, Tangent bundle for human action recognition, IEEE International Conference on Automatic Face and Gesture Recognition, Santa Barbara, 2011.
[49] M.F. Abdelkader, W. Abd-Almageeda, A. Srivastavab, R. Chellappa, Silhouette-based gesture and action recognition via modeling trajectories on Riemannian manifolds, Computer Vision and Image Understanding 115 (3) (2011) 439–455.
[50] P. Turaga, R. Chellappa, Nearest-neighbor algorithms on non-euclidean manifolds for computer vision applications, Indian Conference on Computer Vision, Graphics, and Image Processing, 2011.
[51] H. Karcher, Riemannian center of mass and mollifier smoothing, Communications on Pure and Applied Mathematics 30 (5) (1977) 509–541.
[52] M. Moakher, Means and averaging in the group of rotations, SIAM Journal on Matrix Analysis and Applications 24 (1) (2002) 1–16.
[53] V.M. Govindu, Lie-algebraic averaging for globally consistent motion estimation, i, IEEE Conference on Computer Vision and Pattern Recognition, 2004.
[54] O. Tuzel, R. Subbarao, P. Meer, Simultaneous multiple 3d motion estimation via mode finding on Lie groups, International Conference on Computer Vision, 2005, pp. 18–25.
[55] R. Subbarao, P. Meer, nonlinear mean shift for clustering over analytic manifolds, IEEE Conference on Computer Vision and Pattern Recognition, 2006.
[56] R. Subbarao, P. Meer, Nonlinear mean shift over Riemannian manifolds, International Journal of Computer Vision (2009) 1–20.
[57] E. Begelfor, M. Werman, Affine invariance revisited, IEEE Conference on Computer Vision and Pattern Recognition, New York, 2006.
[58] P. Gruber, F.J. Theis, Grassmann clustering, Proceedings of the European Signal Processing Conference, 2006.
[59] H.E. Cetingul, R. Vidal, Intrinsic mean shift for clustering on Stiefel and Grassmann manifolds, IEEE Conference on Computer Vision and Pattern Recognition, 2009.
[60] R. Sirvalingam, V. Morellas, D. Boley, N. Papanikolopoulos, Metric learning for semi-supervised clustering of region covariance descriptors, IEEE International Conference on Distributed Smart Cameras, 2009.
[61] S. O'Hara, Y.M. Lui, B. Draper, Unsupervised learning of human expressions, gestures, and actions, IEEE International Conference on Automatic Face and Gesture Recognition, Santa Barbara, 2011.
[62] V. Arsigny, P. Fillard, X. Pennec, A. Ayache, Geometric means in a novel vector space structure on symmetric positive-definite matrices, SIAM Journal on Matrix Analysis and Applications 1 (2006) 328–347.
[63] T. Drummond, R. Cipolla, Visual tracking and control using Lie algebras, IEEE Conference on Computer Vision and Pattern Recognition, 1999, pp. 652–657.
[64] T. Drummond, R. Cipolla, Application of Lie algebras to visual serving, International Journal of Computer Vision 1 (2000) 21–41.
[65] T. Drummond, R. Cipolla, Real-time visual tracking of complex structures, IEEE Transactions on Pattern Analysis and Machine Intelligence 7 (2002) 932–946.
[66] F. Porikli, O. Tuzel, P. Meer, Covariance tracking using model update based on Lie algebra, IEEE Conference on Computer Vision and Pattern Recognition, 2006.
[67] E. Bayro-Corrochano, J. Ortegón-Aguilar, Lie algebra approach for tracking and 3 d motion estimation using monocular vision, Image and Vision Computing 25 (6) (2007) 907–921.
[68] O. Tuzel, F. Porikli, P. Meer, Learning on Lie groups for invariant detection and tracking, IEEE Conference on Computer Vision and Pattern Recognition, 2008.
[69] T. Wang, A.G. Backhouse, I.Y. Gu, Online subspace learning on Grassmann manifold for moving object tracking in video, IEEE International on Acoustics, Speech and Signal Processing, Las Vegas, 2008, pp. 969–972.
[70] X. Li, W. Hu, Z. Zhang, X. Zhang, M. Zhu, J. Cheng, Visual tracking via incremental log-euclidean Riemannian subspace learning, IEEE International Conference on Computer Vision, 2008.
[71] X. Li, W. Hu, Z. Zhang, X. Zhang, Robust visual tracking based on an effective appearance model, European Conference on Computer Vision, 2008.
[72] Y. Wu, B. Wu, J. Liu, H. Lu, Probabilistic tracking on Riemannian manifolds, International Conference on Pattern Recognition, 2008.
[73] H. Palaio, J. Batista, Multi-object tracking using an adaptive transition model particle filter with region covariance data association, International Conference on Pattern Recognition, 2008.
[74] J. Kwon, K.M. Lee, F.C. Park, Visual tracking via geometric particle filtering on the affine group with optimal importance functions, IEEE Conference on Computer Vision and Pattern Recognition, 2009.
[75] J. Kwon, F.C. Park, Visual tracking via particle filtering on the affine group, International Journal of Robotics Research 29 (2–3) (2009) 198–217.
[76] F. Porikli, P. Pan, Regressed importance sampling on manifolds for efficient object tracking, IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009.
[77] G. Wang, Y. Liu, H. Shi, Covariance tracking via geometric particle filtering, International Conference on Intelligent Computation Technology and Automation, 2009.
[78] Y. Liu, G. Li, Z. Shi, Covariance tracking via geometric particle filtering, EURASIP Journal on Advances in Signal Processing (2010) 1–9.
[79] M. Li, W. Chen, K. Huang, T. Tan, Visual tracking via incremental self-tuning particle filtering on the affine group, IEEE Conference on Computer Vision and Pattern Recognition, 2010.
[80] X. Ding, L. Xu, X. Wang, G. Lv, Region covariance matrices for object tracking in quasi-monte carlo filter, in: T.-H. Kim (Ed.), Signal Processing and Multimedia, Communications in Computer and Information Science, Springer, Berlin Heidelberg, 2005.
[81] Q. Rentmeesters, P.-A. Absil, P.V. Dooren, K. Gallivan, A. Srivastava, An efficient particle filtering technique on the Grassmann manifold, IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, 2010.
[82] O. Tuzel, F. Porikli, P. Meer, Pedestrian detection via classification on Riemannian manifolds, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (10) (2008) 1713–1727.
[83] S. Paisitkriangkrai, C. Shen, J. Zhang, Fast pedestrian detection using a cascade of boosted covariance features, IEEE Transactions on Circuits and Systems for Video Technology 18 (8) (2008) 1140–1151.
[84] D. Tosato, M. Farenzena, M. Cistani, V. Mutino, A re-evaluation of pedestrian detection on Riemannian manifolds, International Conference on Pattern Recognition, 2010.
[85] D. Tosato, M. Farenzena, M. Spera, V. Murino, M. Cristani, Multi-class classification on Riemannian manifolds for video surveillance, European Conference on Computer Vision, Crete, Greece, 2010.
[86] Q. Gu, J. Zhou, Belief propagation on Riemannian manifold for stereo matching, International Conference on Image Processing, San Diego, 2008, pp. 1788–1791.
[87] P. Cortez-Cargill, C. Undurraga-Rius, D. Mery, A. Soto, Performance evaluation of the covariance descriptor for target detection, International Conference of the Chilean Computer Society, Chile, 2009.
[88] B. Horn, Relative orientation, International Journal of Computer Vision 4 (1990) 59–78.
[89] C. Taylor, D. Kriegman, Structure and motion from line segments in multiple images, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1995) 1021–1032.
[90] Y. Ma, J. Kosecká, S. Sastry, Motion estimation in computer vision: optimization on Stiefel manifolds, IEEE Conference on Decision and Control, Tempa, Florida, 1998, pp. 3751–3756.
[91] Y. Ma, J. Kosecká, S. Sastry, Optimization criteria and geometric algorithms for motion and structure estimation, International Journal of Computer Vision 44 (3) (2001) 219–249.
[92] C. Belta, V. Kumar, Euclidean metrics for motion generation on se(3), Journal of Mechanical Engineering Science Part C (2002) 47–61.
[93] R. Subbarao, P. Meer, Subspace estimation using projection based m-estimators over Grassmann manifolds, European Conference on Computer Vision, 2006.
[94] N.P. da Silva, J.P. Costeria, Subspace segmentation with outliers: a Grassmannian approach to the maximum consensus subspace, IEEE Conference on Computer Vision and Pattern Recognition, 2008.
[95] D. Lin, E. Grimson, J. Fisher, Learning visual flows: a Lie algebraic approach, IEEE Conference on Computer Vision and Pattern Recognition, 2009.
[96] D. Lin, E. Grimson, J. Fisher, Modeling and estimating persistent motion with geometric flows, IEEE Conference on Computer Vision and Pattern Recognition, 2010.
[97] R. Li, R. Chellappa, Group motion segmentation using a spatio-temporal driving force model, IEEE Conference on Computer Vision and Pattern Recognition, 2010.
[98] R. Mégret, J.-B. Authesserre, Y. Berthoumieu, Bidirectional composition on Lie groups for gradient-based image alignment, IEEE Transactions on Image Processing 19 (9) (2010) 2369–2381.
[99] G. Lebanon, Metric learning for text documents, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (4) (2006) 497–508.
[100] S. Watanabe, Algebraic Geometry and Statistical Learning Theory, Cambridge University Press, 2009.