Pattern Recognition Letters 43 (2014) 3–16

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Differential geometric representations and algorithms for some pattern recognition and computer vision problems

Ruonan Li [a,*], Pavan Turaga [b], Anuj Srivastava [c], Rama Chellappa [d]

[a] Harvard School of Engineering and Applied Sciences, Cambridge, MA 02138, United States
[b] Schools of Arts, Media, Engineering and Electrical Engineering, Arizona State University, Tempe, AZ 85281, United States
[c] Department of Statistics, Florida State University, Tallahassee, FL 32306, United States
[d] Department of Electrical and Computer Engineering, and the Center for Automation Research, UMIACS, University of Maryland, College Park, MD 20742, United States

ARTICLE INFO

Article history:
Available online 16 October 2013

Communicated by H. Bunke

Keywords:
Manifold
Hashing
Activity recognition
Video alignment
Shape analysis

ABSTRACT

Scene geometry, imaging laws, as well as computational mechanisms generate mathematical constraints on both raw data and computed features. In many cases, these constraints place the patterns on geometrically well-defined spaces, described as manifolds. In such cases, we argue that exploiting the geometry of these manifolds is important to our understanding of the objects and semantics in the imagery. Statistics and algorithms accounting for manifolds also yield improved performance in many vision applications. We justify these arguments by presenting our recent research efforts based on manifold theory for addressing a variety of pattern recognition and computer vision problems, including hashing on manifolds for efficient search, statistical modeling on Grassmann/Stiefel manifolds for activity recognition, discriminative learning for group motion recognition, stochastic optimization for spatio-temporal alignment, and shape matching. We also discuss the manifolds of re-parameterizations and elastic shapes, as well as applications of manifolds to face recognition and unsupervised adaptation of classification model from one domain to another.

1. A differential-geometric view for patterns

Patterns, especially those arising from various modalities of imagery, usually encode the interactions between scene geometry and physical phenomena such as illumination and motion. On the one hand, when these scenes are imaged using cameras, the observed patterns obey certain mathematical constraints that are induced by the underlying physical constraints. It is well known that images of a convex object under all possible illumination conditions lie on the so called illumination-cone (Georghiades et al., 1998). Images taken under a stereo-pair are constrained by the epipolar geometry of the cameras (Hartley and Zisserman, 2004). Similarly, the 3D pose of the human head is parameterized by three angles hence, under constant illumination and expression, the observed face of a human under different viewing directions lies on a three-dimensional manifold. On the other hand, when informative patterns are extracted and processed from the raw imagery, additional mathematical constraints are frequently imposed or incurred during the computation, so that the patterns do not occupy the full linear Euclidean space but only reside in a subset that can be analytically defined. Many classifiers, for example, require a normalization of input features so that they have zero mean and unit length. As a result, the set of normalized features x's reside on the unit hypersphere centered at the origin, represented as $x^Tx = 1$.

In a particular application, if these constraints are well-understood, such as those studied in epipolar geometry and illumination modeling, then one can design accurate modeling and inference techniques using those knowledge. In many applications, such as those discussed in later sections including texture classification, human motion recognition, group activity analysis, and video alignment, these constraints have a special form, often enabling a geometric interpretation.

An immediate question, however, is why one should involve these constraints that require new mathematical tools from differential geometry. One may wonder why we cannot use linear methods and multi-variate statistics in the classical Euclidean space, with perhaps some loss of accuracy. A simple answer often quoted is that, extrinsically embedding points into a larger vector space increases the dimensionality of the data, as opposed to working with intrinsic co-ordinates. In addition, there are several reasons why an intrinsic analysis is much desired when considering statistical analysis. Statistical quantities, such as the mean of a sample set, in general are meaningful if they follow the geometric constraints of the data itself. For example, the algebraic mean of a set of points on a sphere, in general will not lie on the sphere. One can extrinsically compute the algebraic mean, and then project it back to the

* Corresponding author. Tel.: +1 (617) 495 4478.
E-mail address: ruonanli@seas.harvard.edu (R. Li).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.09.019

sphere, as a solution. However, the choice of projection operator will result in different estimates of the mean. Similarly, how does one define a meaningful equivalent of 'covariance' for a set of points on the sphere? Another interesting issue arises when one is faced with an optimization problem whose domain of optimization is not $\mathbb{R}^n$, but some geometrically well-defined subset, say, a sphere. If one were to extend the domain of the function to the larger extrinsic vector space, and find the direction of steepest descent in the ambient space, one may obtain a solution that does not lie on the sphere. One is again confronted with the question of how to project the solution back to the sphere, which again raises questions of the dependence of the solution to the choice of projection operation.

The theory of differential geometry provides principled solutions to problems such as the ones described above. We hope to convince the reader in the following sections that geometric interpretation of the patterns in a number of computer vision application leads to improved characterization of the data distributions and improved empirical performances of algorithmic solutions, by showing a series of examples that we recently explored. First, we briefly go through the mathematical basics of differential geometry and manifolds in Section 2. Our examples span a wide spectrum of approaches, including hashing for efficient search over a large database (Section 3), Grassmann/Stiefel manifolds for describing linear dynamic systems (Section 4), discriminative incremental learning (Section 5), stochastic optimization (Section 6), as well as a non-parametric Fisher-Rao metric and its extensions to shape analysis (Section 7). We also introduce manifold-based facial analysis and domain adaptation techniques (Section 8) before we conclude the paper in Section 9.

2. Preliminaries

In this section, we briefly recapitulate the mathematical preliminaries needed to design various computational algorithms such as statistical modeling, discriminative models, and nearest neighbor search, etc., on non-Euclidean manifolds. The discussion below is meant to be a utilitarian collection of concepts and resources that the reader may draw upon to create effective algorithmic structures for learning problems in these non-Euclidean spaces, and is by no means comprehensive or complete. For a comprehensive introduction to differential geometry and manifolds, the reader is referred to systematic treatment such as Lee (2002), Spivak (1999), Boothby (1975), and Kreyszig (1991).

2.1. Basic definitions

A topological space $\mathcal{M}$ is called a manifold if it is *locally Euclidean*. In more formal terms, for each point $p \in \mathcal{M}$, there exists an open neighborhood U of p and a mapping $\phi : U \to \mathbb{R}^n$ such that $\phi(U)$ is open in $\mathbb{R}^n$ and $\phi : U \to \phi(U)$ is a diffeomorphism. The pair (U, ϕ) is called a *coordinate chart* for the points that fall in U. Let $\mathcal{M}$ be an n-dimensional manifold and, for a point $p \in \mathcal{M}$, consider a differentiable curve $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ such that $\gamma(0) = p$. The velocity $\dot{\gamma}(0)$ denotes the velocity of γ at p. This vector has the same dimension as the manifold and is an example of a tangent vector to $\mathcal{M}$ at p. The set of all such tangent vectors is called the tangent space to $\mathcal{M}$ at p. Even though the manifold $\mathcal{M}$ may be nonlinear, the tangent space $T_p(\mathcal{M})$ is always linear and of the same dimension as the manifold $\mathcal{M}$. Please see Fig. 1(a) for an illustration.

2.2. Metrics via geodesic distances

The task of measuring distances on a manifold is accomplished using a Riemannian metric. A Riemannian metric on a differentiable manifold $\mathcal{M}$ is a map $\langle \cdot, \cdot \rangle_p : T_p(\mathcal{M}) \times T_p(\mathcal{M}) \to \mathbb{R}$ that smoothly associates to each point $p \in \mathcal{M}$ a symmetric, bilinear, positive definite form on the tangent space $T_p(\mathcal{M})$, so that $\forall v_i, v_2 \in T_p(\mathcal{M})$ the inner product between them is given by $\langle v_1, v_2 \rangle_p$. As shown in Fig. 1(b), given the Riemannian metric that characterizes the inner products in an infinitesimal local neighborhood of each point $p \in \mathcal{M}$, it becomes possible to define lengths of paths and the geodesic paths on a manifold. Let $\alpha : [0, 1] \to M, \alpha(0) = p_1, \alpha(1) = p_2$ be a differential path, and we can assign it a path length according to

$$L[\alpha] = \int_0^1 \sqrt{\langle \dot{\alpha}(t), \dot{\alpha}(t) \rangle_{\alpha(t)}} dt, \tag{1}$$

where $\langle \cdot, \cdot \rangle_p$ denotes the given Riemannian metric at point p. A minimizer of L is called a *geodesic* path:

$$\alpha^* = \underset{\alpha[0,1] \to M, \alpha(0) = p_1, \alpha(1) = p_2}{\operatorname{argmin}} L[\alpha]. \tag{2}$$

To be precise, a global minimizer is called a *minimal* geodesic while a local minimizer is termed geodesic. A common way to find a geodesic is to minimize an energy function E:

$$E[\alpha] = \int_0^1 \langle \dot{\alpha}(t), \dot{\alpha}(t) \rangle_{\alpha(t)} dt. \tag{3}$$

$E[\alpha]$ differs from $L[\alpha]$ in that its integrand is square of the one in $L[\alpha]$ and one can show that they have the same minima. The problem now shifts to finding minima of $E[\alpha]$.

One way to solve this problem is to use the calculus of variation and derive the Euler–Lagrange equation that the minimizer (geodesic) satisfies. To express this equation in local coordinates, let, for any $p \in \mathcal{M}$, $\{E_{i,p}, i = 1, \ldots, m\}$ denote a basis of the tangent space $T_p(\mathcal{M})$. Using the local coordinates $(x_1, x_2, \ldots, x_m)$ for p, we can express the Riemannian metric in the local coordinates using an $m \times m$ matrix $g_{ij} = \langle E_i, E_j \rangle_p$. Using the Christoffel symbols given by

$$\Gamma^k_{i,j} = \frac{1}{2} \sum_s g^{ks} \left(\frac{\partial g_{si}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_s} + \frac{\partial g_{js}}{\partial x_i} \right) \tag{4}$$

where g^{ks} denotes the (k, s)th-elements of the inverse of g, the Euler–Lagrange or the *geodesic* equation can be written as

$$\frac{d^2 x_k}{dt^2} + \Gamma^k_{i,j} \frac{dx_i}{dt} \frac{dx_j}{dt} = 0. \tag{5}$$

Although this equation provides a necessary condition for the geodesic to satisfy, it is usually difficult to solve. Also note that though the Christoffel symbols are written in the same notation as tensors with index notation, they are not tensors because they do not transform like tensors under a change of coordinates (See, for example, p. 141 of Kreyszig (1991)).

In case where $\mathcal{M}$ is embedded inside a vector space V and the Riemannian metric on $\mathcal{M}$ is the one inherited from V, there is a more direct solution to finding the geodesics. This approach, called *path straightening* (Klassen and Srivastava, 2006), initializes the search for α^* using an arbitrary path and then iteratively updates it using the gradient of E. In this approach, the gradient of E is available in an analytical form, as a tangent vector field along the current path α that can be used to efficiently update α.

2.3. Exponential maps and its inverse

Many statistical modeling procedures are performed on tangent spaces of manifolds. An example in devising a probability density function on a manifold is to map points on the manifold to the tangent space, and estimate a probability density function in the tangent-space using traditional vector-space techniques, then map the estimated probability density function back to the