Graz University of Technology

Institute for Computer Graphics and Vision

Master Thesis

Statistical Shape Models with Signed Distance Functions

Marc Steiner

Graz, Austria, May 2013

Thesis supervisor

Univ.-Prof. DI. Dr. Horst Bischof

Instructor

DI. Dr. Martin Urschler

Abstract

Statistical shape models have gained an increasing interest in the Computer Vision community over the past few decades. Awareness of an objects feasible shape variations provides insight to structural features and has shown to improve image segmentation significantly. Further applications that benefit from shape knowledge include image based analysis, classification and tracking. An essential element of shape analysis is the choice of a proper shape representation. Implicit representations such as the signed distance function (SDF) advantageously provide independence of correspondence, parametrisation and topology. Another crucial aspect concerns the selection of a shape space. Several approaches proposed in the literature revert to the simplified assumption of a linear shape space. However, in general this assumption is invalid, resulting in poor modelling of complex variations such as bending. In order to tackle complex shape deformations, a recent trend is to consider shapes populating a non-linear space. The work at hand details a recently proposed manifold approach for non-linear shape modelling on SDFs. We provide in-depth comparison to linear models and show how the manifold model can be incorporated to image segmentation.

Keywords: Statistical Shape Models, Level Set Representation, Manifold Learning, Image Segmentation

Contents

List of Figures

Chapter 1

Introduction

Contents

1.1 Motivation

Shape is a trivial visual property, mostly perceived subliminal in our everydays life. But let's take a closer look. As Light reaches our eye it was modulated differently by the objects in a certain scene. This physical process equips the scene's image with texture, motion, edges and shape, thereby allowing us to distinguish objects. However, these features are not sufficient for a human to identify an object. During our developmental phase we learn to associate observed feature combinations to an object. Hence we have the remarkable ability to classify e.g. a never seen dog breed as a dog or a letter in an unknown font as such. The latter example illustrates vividly that shape provides a powerful visual clue.
Inspired by our capability of object recognition, shape analysis, i.e. the study of a group of shapes was developed. Originating from biological considerations, an early record of shape analysis dates back to Galileo Galilei (1638) who studied bone shapes of various animals. Shape analysis tries to find relations between shape and some physical property. For

example: how is the femoral bone shape related to an animals weight? Likewise analysis allows to understand how shape is affected by pathology or evolution. Considering several humanoid skull shapes (Figure 1.1), properties such as brain size or historical period can be inferred.

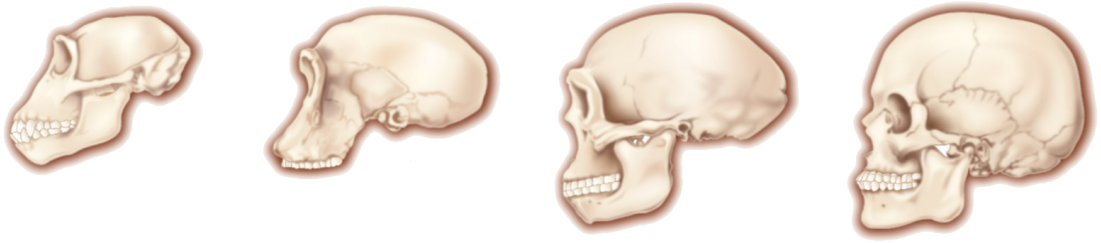


Figure 1.1: A set of humanoid skull shapes relevant in anthropological studies. From left to right: chimpanzee, Homo habilis, Homo erectus, Homo sapiens. Adapted from [48].

Once knowledge of discriminative correlations is established, we can make shape based predictions [27]. Shape analysis is of interest to many applications covering medicine, bionics, industrial quality assurance and augmented reality to name a few. This work explores statistical approaches to extract information from a set of shapes, so to learn a *statistical shape model (SSM)*.

1.2 Modelling Shape

We now give a brief glance at SSM creation and utilisation, see Fig. 1.3 for reference. SSMs rely on a collection of training samples that provide the shape variations to be captured. Variations may be caused by several sources. For example, when moving the viewpoint, a non-symmetrical 3D object produces various contour projections to the image plane. Nonrigid objects introduce additional modes of variability, being of interest in medical and biometric applications. One might also be interested in a model that covers shapes produced by different objects, e.g. for anatomical inter-subject analysis. In practice the mentioned variations often occur combined. In order to extract variations from the training set, alignment to a common reference is required such that corresponding parts overlap. After alignment, shape variations emerge by the distribution of corresponding salient contour points. For instance, when modelling hand shapes, the positions of thumb-tips will occupy a characteristic region. On a larger scale, all contour points of the thumb and eventually the whole hand contour will show specific distributions restricted to act-able

hand gestures. The intention of a SSM is to capture these distributions. Moreover, the model should be able to simplify the shape variations. Consider bare thumb tilt motion. Modelling this variation by the definition of a displacement field for each contour point would be a rather cumbersome approach. Instead we would like to project each individual deformation to a feature parameter. Another desired SSM quality is separation of features, i.e. there should be one parameter solely responsible for thumb tilt while another one controls index finger length. Both, simplification and separation can be achieved by means of feature extraction techniques, being the core component of SSMs. In essence, a SSM is composed of the set of features extracted from the training data as well as a rule to synthesise a shape from a combination of feature values.

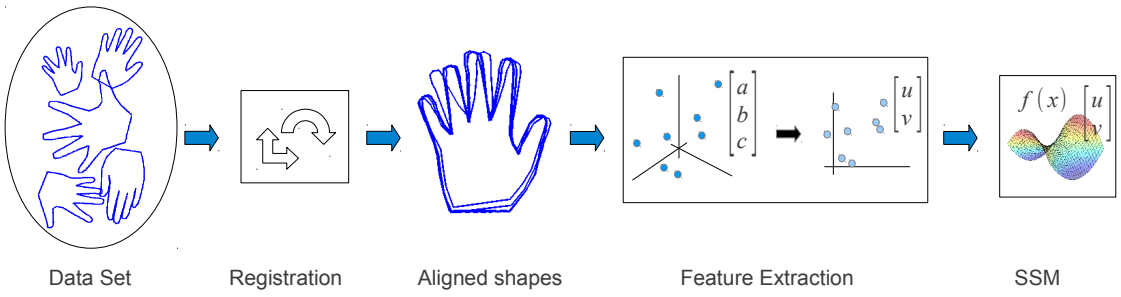


Figure 1.2: Principle of SSM creation.

We now turn to the utilisation of the model. When fed a new shape, the SSM should be able to reproduce that shape by variation of its internal feature vector. The output is the models interpretation of the new shape, relevant for shape restoration and robust image segmentation. Furthermore, the SSM outputs the tuned feature vector which can be used for classification, recognition and comparison.

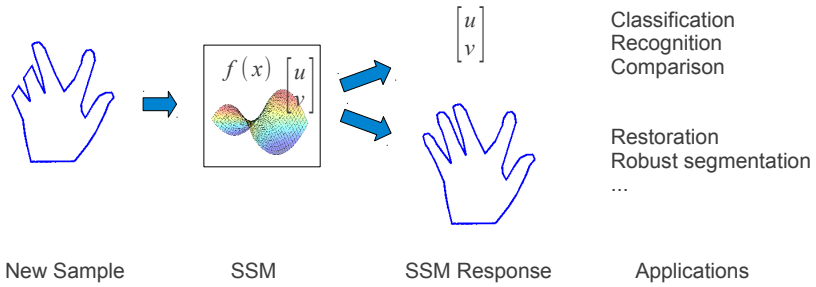


Figure 1.3: Principle of SSM utilisation.

1.3 Shape Model based Image Segmentation

Image segmentation - the task of finding a partition of an image into meaningful regions - is of vital interest in many applications. Advances in medical imaging confront radiologists with a vast amount of data. Segmentation combined with shape analysis can aid radiologists in pathology detection, measuring tissue volumes and treatment planning. Also, reliable segmentation is a big step towards image understanding, allowing machines to derive semantic interpretations from visual information. Depending on the image features used, segmentation techniques may be split roughly in edge and region based approaches. However, in complex images these features may be corrupted by noise, cluttered background and partial occlusion thus rendering segmentation an ill-posed problem.

A well known practice to alleviate this inconsistency is to apply regularisation methods, thereby imposing a restriction to the space of feasible solutions. In order to do so, some kind of a priori knowledge about the object in question has to be incorporated to the segmentation process. This can be achieved by augmenting the segmentation with an SSM. Considering Fig. 1.4, the objective is to separate the partially occluded bunny from its cluttered background. Observing the segmentation result relying solely on low level information (Fig. 1.4a), the contour includes unrelated regions. In contrast, providing shape information to the task yields a sound segmentation result (Fig. 1.4b).

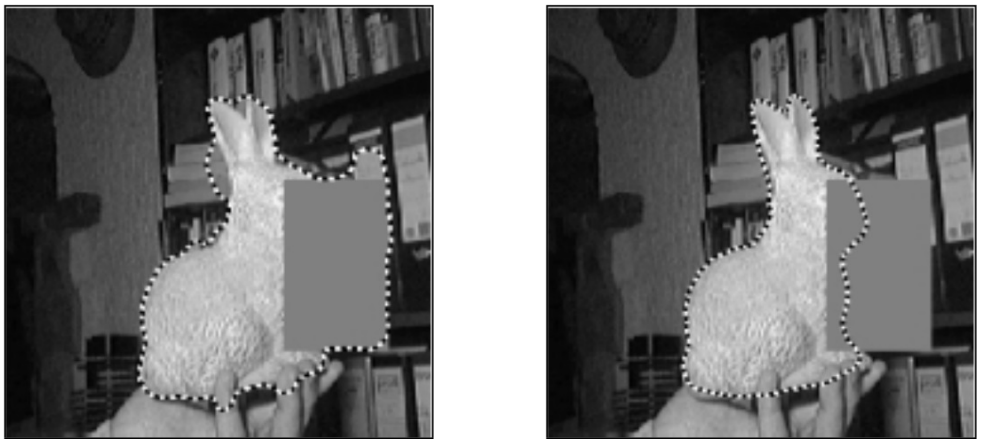


Figure 1.4: The low-level based segmentation result (l) is improved significantly by using a shape prior (r). Reproduced from [20].

1.4 Purpose of this Thesis

The work at hand explores tools to build and apply SSMs for the application of image segmentation. Emphasis is put on implicit shape representation and the requirement to capture non-linear shape variations. We review shape models proposed in the literature and provide detailed insight to a recently introduced manifold framework. Moreover we present extensive evaluation as well as comparison to a linear approache on various data-sets.

1.5 Data Sets

There are several demands for a representative training-set. Depending on the shape variability to be captured by the model, the set should consist of an adequate amount of samples. Furthermore, "extreme" shape instances are desired to cover a wide range of variability. Ideally, the samples are distributed uniformly, i.e. the extent of variation between similar samples should not differ too much. Throughout this work several training sets are used for illustration and evaluation. Figure 1.5 depicts representative examples and Table 1.1 summarises their properties.

	Subject	**#samples**	**Description**
Cup	single rigid	132	Pure pose variation (tilting and rotation) of a coffee cup
Eclipse	single nonrigid	26	Synthetic set: solid disk carved out by a smaller disk
Walker	single nonrigid	30	Cycle of a walking person
FGnet Face [32]	single nonrigid	5000	Video frames of a talking persons face
IMM Faces [51]	multiple nonrigid	240	Pose and articulation variance of 40 individuals
Cause07 [75]	multiple nonrigid	33	3D cerebral structure of left caudate nuclei
SLiver07 [34]	multiple nonrigid	20	3D human liver

Table 1.1: Summary of data-sets used.

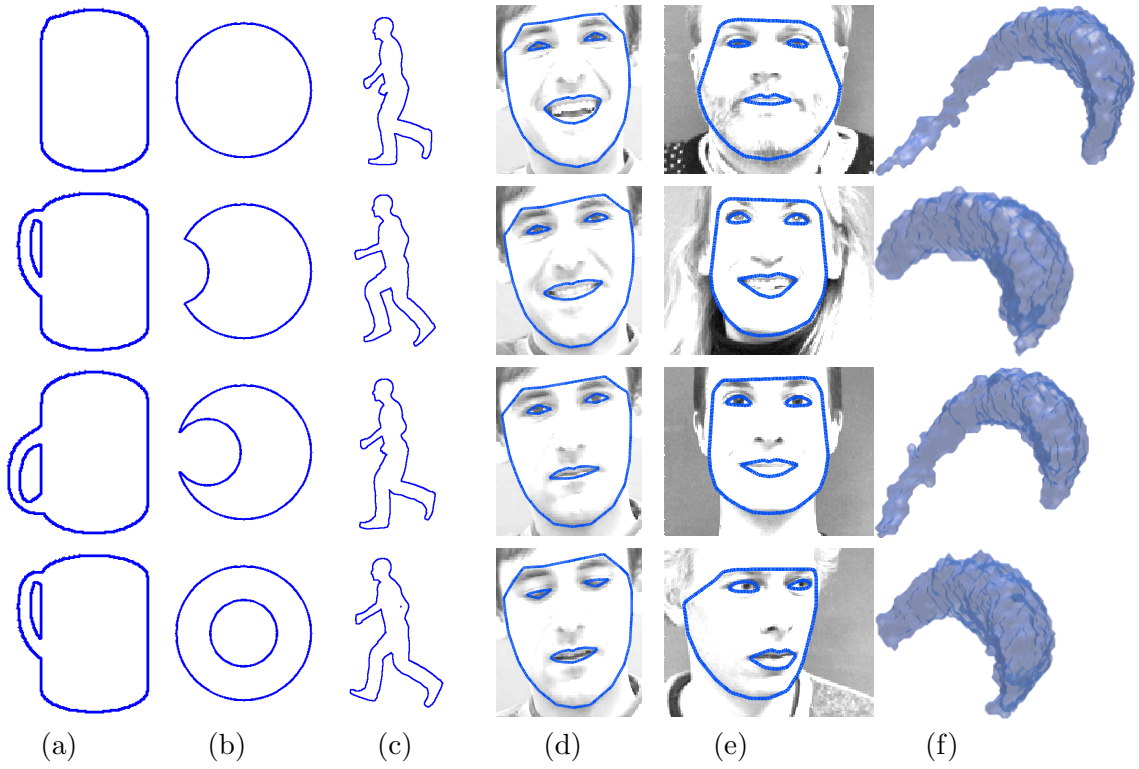


Figure 1.5: Instances of the data-sets used. (a) Cup, (b) Eclipse, (c) Walker, (d) FGnet talking face, (e) IMM faces, (f) caudate nuclei

1.6 Outline

This thesis is organised as follows. Chapter 2 provides an overview of shape representations and introduces essential tools for shape modelling. We first review explicit representations in Section 2.1 followed by implicit ones with a focus on SDFs in Section 2.2.

Section 2.3 is dedicated to dissimilarity metrics for shape comparison. Approaches to rigid shape registration are discussed in Section 2.4.

Feature extraction is the subject of Chapter 3 where key methods for dimensionality reduction are reviewed so to map shape to a lower dimensional representation. We first investigate the linear methods Principle Component Analysis and Multidimensional Scaling in Section 3.3. Non-linear methods including Isomap, Local Linear Embedding, Laplacian Eigenmaps, Kernel-PCA and Diffusion Maps are discussed under consideration of applicability to shape models in Section 3.4.

Chapter 4 summarises approaches for SSMs proposed in the literature. We review several combinations of local, global, linear and non-linear efforts to learn a set of shapes. Active

Shape Models [17] and a similar level set based approach [46] are presented in Section 4.1 and Section 4.2. We investigate in Chord Length Models [5, 15, 44] in Section 4.4 and in density equipped reduced spaces [20, 56] in Section 4.7. Local statistics on implicit representations [58, 59] are reviewed in Section 4.5. Section 4.6 introduces the idea of statistics on deformation fields [11, 12].

In Chapter 5 we present an approach to non-linear shape modelling as proposed by Etyngier and Thorstensen et al. [30, 31, 69, 70] in detail. Training samples are assumed to be sampled from a manifold which is learned by means of non-linear feature extraction. We discuss approaches to warp one shape towards another in Section 5.2 and address the maps between shape space and feature space in Section 5.3 and Section 5.4.

Chapter 6 presents results of the manifold approach while an comprehensive comparison to linear methods is made. We also show how the SSM can be integrated in the task of image segmentation and present promising results.

Finally, Chapter 7 concludes with a review of the insights gained and gives prospects for possible enhancements and future work.

Chapter 2

Shape

Contents

Psychological considerations of shape were established by Wertheimer, Koffka and Köhler in 1928. They founded the theory of *Gestalt Psychology*, which explores the impact of shape on a human observer. In this work we consider a pure geometrical definition, specifically the one given by Kendall [38]:

> *"Shape is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object."*

Indeed, when gathering multiple instances of an objects shape, one has to deal with varying capture conditions. In Section 2.4 we will analyse methods for eliminating those disturbances. For the moment being, let $\mathcal{S}$ be a set of shapes generated by one or several similar objects with above mentioned effects removed. Then the set of all possible shapes spans the *shape space* $\mathbb{S}$. In practice, $\mathcal{S}$ is a finite set of q instances $\{S_i\}$ sampled from shape space. We will refer to $\mathcal{S}$ as training set. The remaining geometrical information, the shape itself, may be expressed as a subset $\mathcal{C}$ of the image domain Ω. We consider $\mathcal{C}$ a set of closed contours in 2D and a set of closed surfaces in 3D. Representations for

such outlines are discussed in the sequel. Although the following excursion focuses on 2D shapes, all described techniques apply to 3D shapes as well.

2.1 Explicit Shape Representation

To start with, we investigate selected explicit shape representations, which use some kind of parametric description.

2.1.1 Control points

Perhaps the most intuitive way to describe a shape is to sub-sample its contour. Then the coordinates of these points $\mathbf{s} := (x_1, y_1, x_2, y_2, ...x_l, y_l)^T$ define the shape. A closed contour is realised by interconnecting control points by piecewise polynomial or linear functions. When building a statistical model, points require correspondence throughout the training set. However, correspondence can not always be established for shapes revealing strong variation. For example, consider the pot handle in Fig. 1.5 which is occluded in certain poses. Additionally, it often requires an expert to label the training set by hand which is, in particular for 3D shapes, a tedious task. Missing feature correspondences may be recovered by the Iterative Closest Point algorithm or the Shape Context descriptor [3]. Noteworthy, these methods do not guarantee to find correct correspondences and are applicable only for well-articulated shapes. Another drawback of the control point representation concerns its lack of handling topological changes such as splitting and merging. Although extensions have been proposed that allow for such operations (Delingette and Montagnat [26]), the notion of correspondence then becomes somewhat diffuse. Also, contour intersection and high shape curvature introduce problems. On the other hand, control points are favoured for their low dimensional description and the resulting low computational needs. Seminal works on control point based shape analysis include the Active Contour formulation by Kass et al. [37] and Active Shape Models by Cootes et al. [17].

2.1.2 Fourier Descriptors

In [64], Staib and Duncan employ elliptical harmonics to represent shape. Considering a point moving along the contour, its projection to the Cartesian axes can be written as a Fourier series expansion. For example, the contour x-component can be decomposed by

$$x(t) = a_0 + \sum_{n=1}^{\infty} a_n cos\frac{2n\pi t}{|\mathcal{C}|} + b_n sin\frac{2n\pi t}{|\mathcal{C}|}, \tag{2.1}$$

and alike for further dimensions. Here $|\mathcal{C}|$ is the contour perimeter and t specifies the evolution along the contour. Taking into account the leading N harmonics, any shape can be approximated up to arbitrary precision. Hence increasing N allows for more detail and curvature. The principle of a three harmonics descriptor is illustrated in Fig. 2.1. Each ellipse point is parametrised by $(N+1)$ coefficients $\{a_i\}$ for each dimension. These coefficients define axis lengths, orientation and phase angles. For any $n > 1$, an ellipse point defines either the center of its prior ellipse or an approximated shape point for $n = N$. Fourier descriptors were used in [57] and more recently in [56].

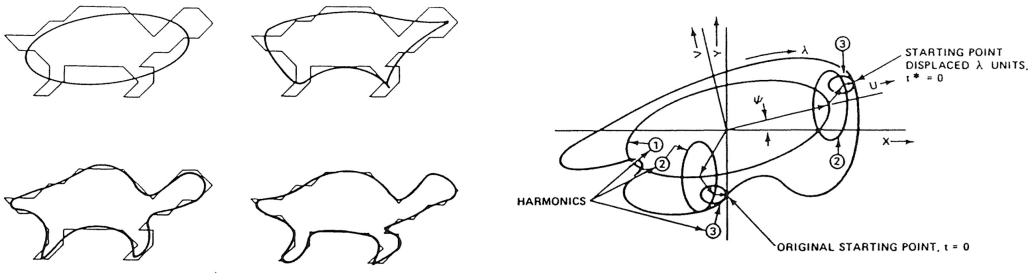


Figure 2.1: Elliptical Fourier shape descriptor: (l) Shape approximation using $N = 1$, 4, 8 and 12 harmonics, (r) Geometric interpretation. Figures are reproduced from [41].

2.2 Implicit Shape Representation

Another prominent regime of shape representations is constituted by implicit approaches. These allow a parameter free description without the requirement for correspondence among a set of shapes.

2.2.1 Currents

Vaillant and Glaunés [73] represent shape by a mathematical construct termed currents. The idea is to compute the flux produced by a set of probing vector fields through the contour as in Equation (2.2) for the 2D case.

$$S(\omega) = \int_{\mathcal{C}} \omega(\mathbf{x})^t \tau(\mathbf{x}) d\mathcal{C} \tag{2.2}$$

Here ω denotes the probing vector field and τ is the tangent at contour point $\mathbf{x}$, see Fig. 2.2 for illustrations. For the probing vector fields the authors propose a square integrable

function convolved several times by a smoothing kernel. In essence, this operation embeds a shape into a linear vector space, where the mean and covariance of a population of currents can be computed. However, deforming a shape in the framework of currents comes at high computational cost. Durleman [29] alleviate this problem by using matching pursuit to find a sparse representation for currents.

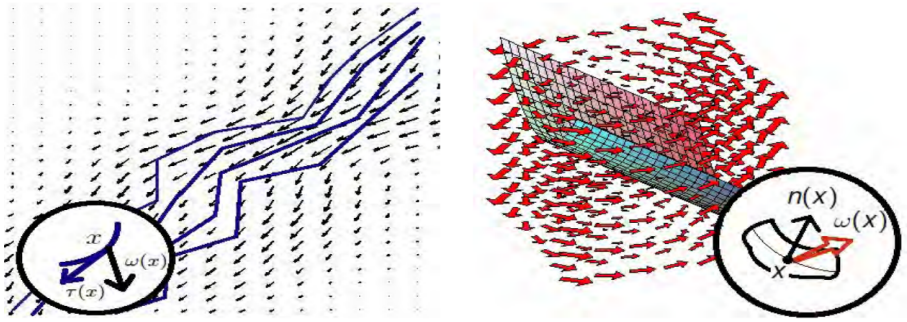


Figure 2.2: Currents integrate vector fields. 2D shape (l), a bundle of lines. 3D shape (r). The probing vector fields are indicated by arrows. Figures are reproduced from [28]

2.2.2 Characteristic Function

One natural way to describe shape implicitly is to define it as the subset of Ω that is enclosed by the contour according to

$$\chi(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \Omega_{in} \\ 0, & \mathbf{x} \in \Omega_{out} \end{cases} \tag{2.3}$$

The characteristic function (also Heaviside step-function) was used for example in [21, 72]. Suppose we introduce a plane $z = \frac{1}{2}$ to the characteristic function. Then the intersection of both functions results in the shape contour $\mathcal{C}$. This concept is developed further in the next section.

2.2.3 Signed Distance Functions

In [52] Osher and Sethian popularised the *Level Set Framework* for physical processes involving propagating fronts. The front or interface is embedded by a higher dimensional function as the set of iso-values of some magnitude (isophote). Caselles [8] carried this idea to applications in Computer Vision by the Geodesic Active Contour formulation. Ever

since, the Level Set method has become a popular approach in Vision applications, as it represents shape in a nonparametric manner and handles topological changes naturally. One heavily considered embedding function is the *Signed Distance Function* (SDF) defined by

$$\phi(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in \mathcal{C} \\ -d_\Omega(\mathbf{x}, \mathcal{C}), & \mathbf{x} \in \Omega_{in} \\ +d_\Omega(\mathbf{x}, \mathcal{C}), & \mathbf{x} \in \Omega_{out} \end{cases} \tag{2.4}$$

To every point $\mathbf{x}$ in Ω, the shortest distance $d_\Omega(\mathbf{x}, \mathcal{C})$ to the contour is assigned. Function values are negative inside the shape Ω_{in} and positive in the outside region Ω_{out}. The contour is then defined implicitly as the zero level set of its SDF $\mathcal{C} : \{\phi(\mathbf{x}) = 0\}$. This distance map has some particularly advantageous properties. The contour unit normal and the curvature are given by

$$\mathbf{n} = \nabla\phi \qquad \text{and} \qquad \kappa = div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) = \Delta\phi \tag{2.5}$$

respectively. Figure 2.3 illustrates a SDF embedding a face contour. Change of shape is

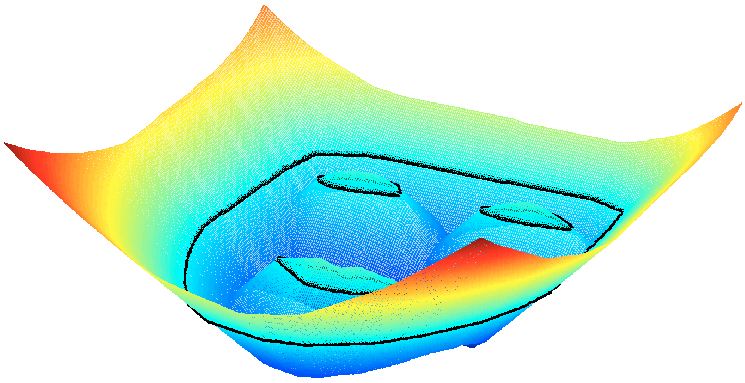


Figure 2.3: Illustration of the signed distance function embedding a face contour.

achieved by deforming the associated SDF. A contour $\mathcal{C}$ moving in the normal direction is coupled to its distance map by the level set equation

$$\frac{\partial\phi}{\partial t} + \nabla\phi \cdot \frac{\partial\mathcal{C}}{\partial t} = 0. \tag{2.6}$$

Note that this formulation alleviates the problem of point correspondences, which can be casted as finding similar SDF values in the normal direction.

2.2.4 Alternative Implicit Shape Representations

Hong et al. [35] depart from the SDF idea and define shape on the image domain by so called *Integral Kernels*. Their concept resembles shape context in a continuous formulation. Briefly, the characteristic shape function χ is convolved by a kernel K_σ, e.g. a Gaussian. A distinction is made for locations being inside respectively outside the shape, see Equation (2.7). Thereby a value of closeness to the contour in a local neighbourhood is assigned to each spatial location $\mathbf{x} \in \Omega$.

$$\phi_{IK}(\chi) = \chi(K_\sigma * (1 - \chi)) + (1 - \chi)(K_\sigma * \chi) \tag{2.7}$$

The authors show the robustness of Integral Kernels to noise and apply it to the correspondence problem. Figure 2.4(m) shows the integral representation of a face shape resulting from a Gaussian with $\sigma = 16$.
Another ansatz is to encode a set of aligned shapes by the probability of being inside the shape, that is $p : \Omega \mapsto [0, 1]$. Cremers et al. exploit this approach in [23]. Pohl and coworkers [55] pursue probability maps by introducing the *Log Odds* defined by $log(p) - log(1-p)$ to shape analysis. See Fig. 2.4(r) for illustration. In contrast to SDFs, probability maps form a vector space that allows for linear feature extraction.

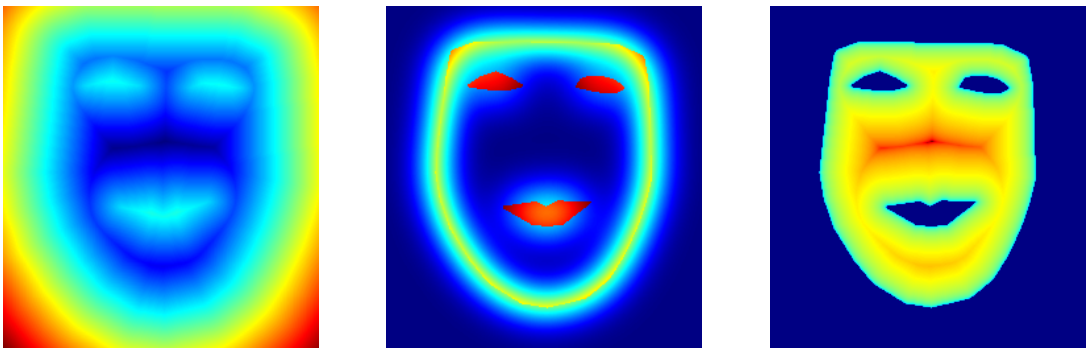


Figure 2.4: Heat map illustration of implicit representations defined on Ω: SDF (l), Integral Kernels (m) and Logodds (r).

2.3 Shape Similarity

To quantify similarity between shapes, an appropriate metric is required. Mathematically speaking, we want to define a mapping $d : \mathbb{S} \times \mathbb{S} \mapsto \mathbb{R}$. One such measure is the Procrustes distance, given by the sum of squared differences (SSD) of corresponding point distances,

see also Section 2.4.1. In this section we outline further measures proposed in the literature. Expressing the difference between complex signals such as shape by a single scalar is ambiguous. Nevertheless, it provides a coarse comparative indicator.

2.3.1 Hausdorff Distance

Given two shapes S_1 and S_2, the Hausdorff distance can be expressed in terms of their SDFs by

$$d_H(S_1, S_2) = sup\left(|\phi_1| - |\phi_2|\right) \tag{2.8}$$

That is the farthest distance found from either contour to the closest point of the other contour. Since the Hausdorff metric is not differentiable, Charpiat et al. propose a smooth approximation in [10], allowing to introduce d_H into gradient based optimisation.

2.3.2 Area of Symmetric Difference

This distance measures the non-overlapping areas of two shapes represented as characteristic functions according to

$$d_{SD}(S_1, S_2)^2 = \int_\Omega \left(\chi_1(\mathbf{x}) - \chi_2(\mathbf{x})\right)^2 \, d\Omega. \tag{2.9}$$

It was advocated by Cremers et al. [21], where an extension for translation invariance is proposed. The metric is symmetric i.e. $d_{SD}(S_1, S_2) = d_{SD}(S_2, S_1)$ and does not depend on the domain size. To obtain an intuitive measure of dissimilarity required later on, we introduce normalisation on two levels of coarseness:

$$d_{SD}^{\Omega_{in}}(S_F, S_M) = \frac{d_{SD}(S_F, S_M)}{\Omega_{in}} \quad \text{with} \;\; \Omega_{in} = \int_\Omega \chi_F \, d\Omega \tag{2.10}$$

$$d_{SD}^{\mathcal{C}}(S_F, S_M) = \frac{d_{SD}(S_F, S_M)}{|\mathcal{C}_\mathcal{F}|} \tag{2.11}$$

Either measure returns 0 for identical shapes. While $d_{SD}^{\Omega_{in}}(\cdot,\cdot) = 1$ indicates a non-overlapping area equal to the area inside of S_F, on the finer scale $d_{SD}^{\mathcal{C}}(\cdot,\cdot) = 1$ is reached for a non-overlapping area equal to the fixed shape's contour length. For an impression of different values of $d_{SD}^{\mathcal{C}}(\cdot,\cdot)$, see Fig. 2.5.

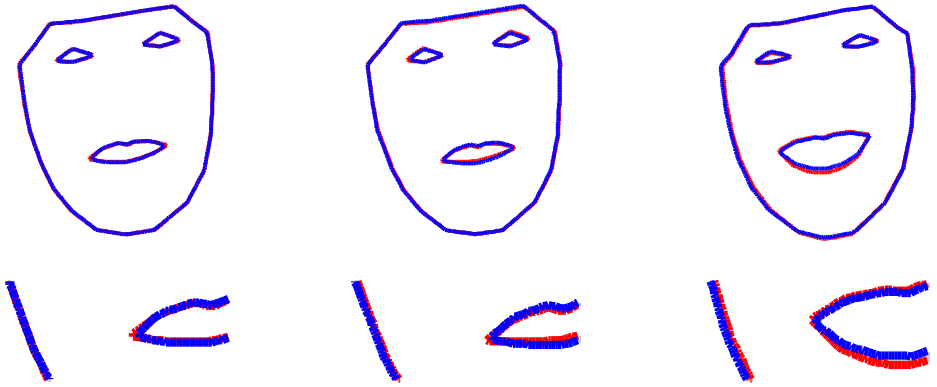


Figure 2.5: Examples of different errors $d_{SD}^{\mathcal{C}}(\cdot,\cdot)$. From left to right: 0.25, 0.50 and 0.75. Bottom row: details of above figures.

2.3.3 Distances on SDFs

Another way to express shape dissimilarity is to equip the space of SDFs with a certain norm, e.g. the L^2 norm according to

$$d_{L^2}(S_1, S_2)^2 = \int_\Omega |\phi_1 - \phi_2|^2 d\Omega. \tag{2.12}$$

Note that this norm depends on the domain size. As noted by Charpiat [11], SDFs and their gradients $\nabla\phi$ are square integrable on Ω. Hence they belong to the Sobolev space $W^{1,2}$ and we can define the Sobolev distance by

$$d_{W^{1,2}}(S_1, S_2)^2 = \int_\Omega |\phi_1 - \phi_2|^2 + |\nabla\phi_1 - \nabla\phi_2|^2 d\Omega. \tag{2.13}$$

This norm is also promoted in [65], for it imposes regularity on active contour energies and can reduce the order of involved PDEs. For a comprehensive evaluation of shape metrics, the reader is referred to [11].

2.4 Shape Registration

We now address the projection of a set of shapes to global transform free shape space. This amounts to eliminate translation, isotropic scale and rotation, as parametrised by a similarity transform. Formally, we search the set of transformations $\{T_i\}$ that align

the training shapes to a fixed reference shape S_F such that a given similarity measure is maximised.

At this point we are not interested in registering shapes non-rigidly by a diffeomorphism to match exactly. Instead the residual variation will provide the information to build a statistical model. Note, that when confronting the model with unknown shapes, e.g. in a segmentation process, these have to be projected to shape space as well, thereby stressing further registrations.

2.4.1 Procrustes Analysis

Given two shapes, each defined by a set of control points along with correspondences, Procrustes Analysis [33] provides means to find the optimal pose registration in a least squares sense. For this purpose we define a shape as $l \times N$ matrix $\mathbf{S}$, with l the dimension of the image domain and N the number of points. Registration is achieved by minimising the Procrustes distance

$$d_{Pr}(\mathbf{S}_M, \mathbf{S}_F)^2 = \frac{1}{N} \sum_{j=1}^{N} \left| \mu \mathbf{R} \mathbf{s}_{M,j} + \mathbf{t} - \mathbf{s}_{F,j} \right|^2 . \tag{2.14}$$

Here $\mathbf{s}_{M,j}$ and $\mathbf{s}_{F,j}$ denote the j^{th} column of $\mathbf{S}_M$ and $\mathbf{S}_R$ respectively. Parameters $\{\mu, \mathbf{R}, \mathbf{t}\}$ constitute the similarity transform in question. Equation (2.14) has a closed form solution which is outlined below. In a first step $\mathbf{S}_M$ is translated to $\mathbf{S}_R$ such that their centres of gravity coincide. The rotation angle is then discovered by Singular Value Decomposition (SVD) on the covariance matrix $\mathbf{U}\mathbf{\Lambda}\mathbf{V}^\mathbf{T} = \mathbf{S}_M^T \mathbf{S}_R$. The so found rotation matrix $\mathbf{R} = \mathbf{V}\mathbf{U}^\mathbf{T}$ allows to compute the scale and finally the translation is adjusted by $\mathbf{t} = \bar{\mathbf{S}}_F - \mu \mathbf{R} \bar{\mathbf{S}}_M$.

To align several shapes, Generalised Procrustes Analysis [6] can be used. Here all shapes are registered successively as described above to a reference shape, initially one in the set. Then the mean of the transformed shapes is computed, serving as reference for the next alignment round. This process is repeated until the mean reaches steady state.

2.4.2 Implicit Shape Registration

Implicit shape representations such as SDFs and characteristic functions are defined on the whole image domain Ω. Therefore their alignment can be performed by techniques established for intensity based image registration. Note that SDFs are robust to slight

misalignment, since nearby values are highly correlated [46]. Tsai and coworkers [71, 72] align a set of shapes simultaneously by minimising a one-shot energy given by

$$E_{SSD_\chi}(\{T\}_{i=1\dots q}) = \sum_{i=1}^{q} \sum_{\substack{j=1 \\ j\neq i}}^{q} \frac{\int_\Omega T_i(\chi_i) - T_j(\chi_j)^2 d\Omega}{\int_\Omega T_i(\chi_i) + T_j(\chi_j)^2 d\Omega}. \tag{2.15}$$

Here T_i is the similarity transform for shape χ_i given by a characteristic function. The metric is a modified version of the symmetric difference, where the denominator counteracts to shrinking. In [53] Paragios et al. propose to register two SDFs ϕ_F, ϕ_M by minimising the SSD

$$E_{SSD_\phi}(T) = \int_\Omega (\mu\ \phi_M(\mathbf{x})\ -\ \phi_F(T(\mathbf{x})))^2\ d\Omega. \tag{2.16}$$

Here μ is the transform's scale and $\mathbf{x}$ a location of the moving SDF. Distance maps are invariant only to translation and rotation [53], hence μ in above equation compensates for scale variance. Huang et al. [36] propose a misalignment metric originating from multimodal medical image registration, Mutual Information (MI). MI establishes a statistical relationship between two images by considering their information content expressed in terms of entropy. In this context, two images are considered aligned, when their mutual information reaches a maximum. Huang et al. employ this metric for the alignment of SDFs, the corresponding functional that is to be minimised is given by

$$E_{MI(\phi)}(T) = -\iint p(a,b)\ \log \frac{p(a,b)}{p(a)\ p(b)}\ da\, db. \tag{2.17}$$

Here a denotes SDF values of the transformed moving image $\phi_M(T(\mathbf{x}))$ and b describes values of the fixed distance map $\phi_F(\mathbf{x})$. Accordingly, $p(a)$ and $p(b)$ represent the marginal probability density functions (PDF). The joint PDF $p(a,b)$ gives the probability of a SDF value having magnitude a in $T(\phi_M)$ and magnitude b in ϕ_F. These densities may be estimated by histograms or Parzen windowing.

In order to compare the aforementioned metrics w.r.t. capture range, we conduct an experiment by transforming an IMM face SDF by various scales and rotations. Then the registration error $d_{SD}^{\Omega_{in}}(S, T(S))$ between transformed SDF and its originator is measured. Since translational displacement can be eliminated by an initial CoG alignment, we do not account for it. For this experiment we revert to the elastix image registration toolbox [39], which provides a rich diversity of adjustable parameters. In all experiments we use the quasi-newton optimiser with a 4 layer image pyramid to avoid local minima. Figure 2.6

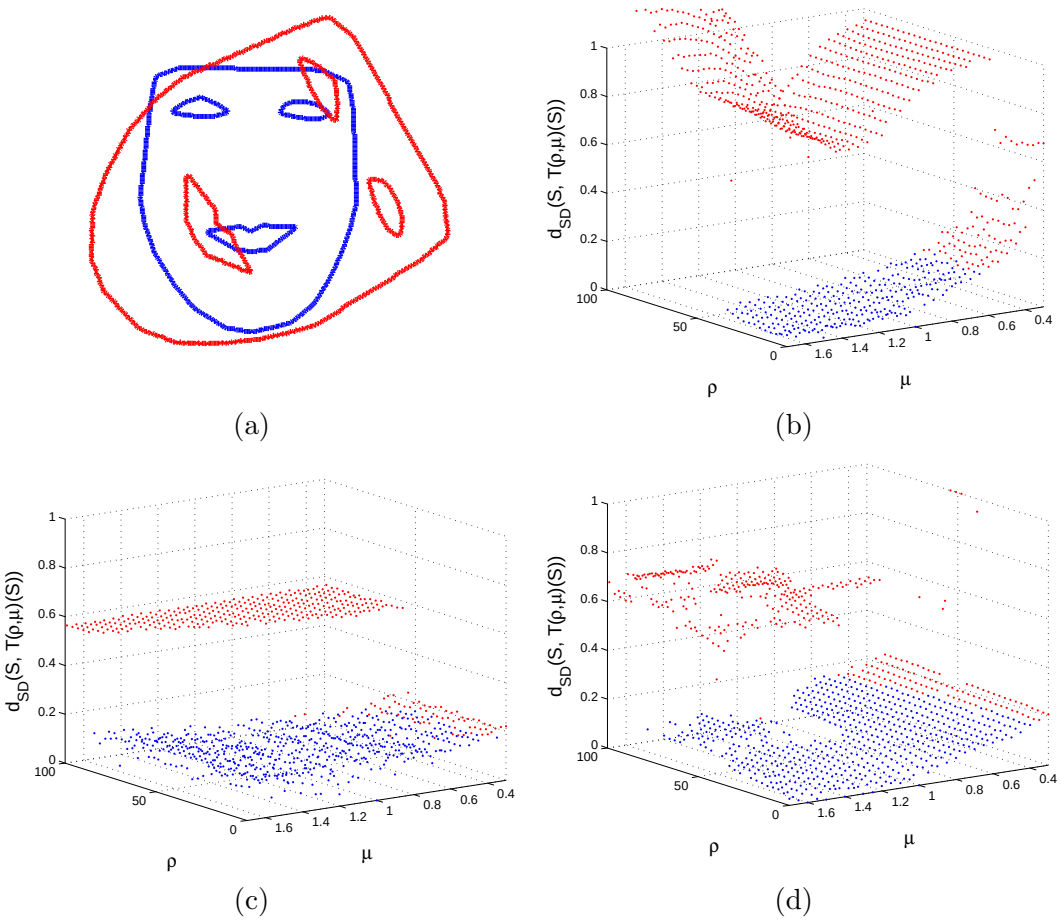


Figure 2.6: Capture range of implicit registration metrics. Evaluation includes various rotation and scale combinations. (a) Exemplary initial transformation for $\rho = 60°, \mu = 1.3$. Registration error $d_{SD}^{\Omega_{in}}(S, T(S))$ for (b) MI_ϕ, (c) SSD_ϕ, (d) SSD_χ. Blue dots indicate an error $d_{SD}^{\mathcal{C}}(S_F, S_M) < 1$.

shows the results. Obviously, MI_ϕ has problems aligning shapes being rotated above 40°, in particular at scales beyond $\mu = 1$. Both, SSD_ϕ and SSD_χ perform well for angles up to $\rho = 60°$ and beyond, depending on scale. Note that for scales below $\mu = 1$ all metrics yield poor precision since the moving SDF is sampled sparsely. Alternative choices for dissimilarity metrics include the sum of absolute differences, cross correlation and joint entropy.

Reconnecting to the application of shape modelling, Fig. 2.7 shows the alignment of multiple shapes to a reference. Having removed global transforms, variations appear as distributions of corresponding contour points.

Figure 2.7: Examples of aligning multiple shapes to a reference with the SSD_ϕ metric: 20 IMM faces (l) and four Cause07 shapes (r).

2.5 Discussion

In this chapter we gave an overview of popular shape representations. We examined explicit representations which require parametrisation as well as point correspondence and can not handle topological changes naturally. Implicit representations alleviate these problems. Especially SDFs are well established representations with advantageous properties defined by the level set framework. Therefore we will focus on SDF hereafter. Furthermore, we investigated in measures for shape similarity. These are of importance for shape registration and for feature extraction techniques discussed in Chapter 3. There seems to be no consensus about the ideal similarity measure in the literature. However, since we decided to use SDFs, we will stick to the L^2 - and Sobolev $W^{1,2}$ norm. Finally, this chapter shed light on techniques for rigid shape registration. Although SDF registration is subject to non-convex optimisation, the cost functions involved turn out to be well-behaved - especially for the SSD_ϕ metric.

Chapter 3

Dimensionality Reduction

Contents

The human sensory perception captures a vast amount of data from the environment. For instance, the bandwidth of a human retina is about 10^5 bit per second. This overwhelming flood of information exceeds the brain's processing capacity and is therefore reduced to the order of 10 bit/s [40]. Obviously, we are equipped with efficient methods to compress information to its essentials. In analogy, many applications in Science and Engineering gather high-dimensional signals for analysis and control. Likewise to the perceptual apparatus, it is desirable to reduce the captured data for further processing. High-dimensional data hides beyond human imagination, for our mind is adapted to a three dimensional world. Nevertheless, algebra allows us to operate in such realms. In this context, the term *Curse of Dimensionality* refers to problems induced by an increasing input space. A finite set of samples populates high-dimensional spaces very sparsely, thereby complicating data analysis. As it turns out, for many applications the intrinsic dimensionality of data is lower due to redundancies and correlations in the measurement. This observation founds the motivation for dimensionality reduction methods to which this chapter is dedicated.

3.1 Definitions

Although dimensionality reduction techniques may be used on arbitrary data, we will emphasise the application on shapes represented as distance maps. Therefore we subsample a SDF ϕ at discrete grid locations. The resulting values are stored in shape vectors $\{\mathbf{s}_i\}_{1\leq i\leq q}$, which can be considered as points in a high-dimensional observation space $\mathbb{R}^N$. Note that even in the limit of an infinite number of shapes gathered from a specific object (that do not vary arbitrarily), $\mathbb{R}^N$ will not be populated exhaustively. The distance map constraint $|\nabla\phi| = 1$ imposes additional narrowing. This gives rise to the assumption that shape space resembles a lower dimensional sub-manifold $\mathcal{M}$. A manifold is a geometric entity of lower dimension than its ambient space. Therefore we aim at finding a mapping (or embedding) Ψ that projects $\{\mathbf{s}_i\} \in \mathcal{M}$ to $\mathbb{R}^n$, where $n << N$ such that most information is retained. We denote the corresponding lower-dimensional feature vectors by $\{\mathbf{y}_i\}_{1\leq i\leq q} \in \mathbb{R}^n$. Both sets of vectors are organised column-wise in matrices $\mathbf{S}$ and $\mathbf{Y}$. Furthermore, let a dissimilarity metric be indicated by $d(\mathbf{s}_i, \mathbf{s}_j)$ and $\delta(\mathbf{y}_i, \mathbf{y}_j)$ measured in observation space and feature space respectively.

3.2 Feature Selection

SDFs are a highly redundant shape representation. For the embedded contour is uniquely defined by close isophotes, it suffices to maintain values in a narrow-band guided by $\mathcal{C}$. Combining the narrow-bands of several aligned shapes, these form a narrow-band overlap map, restricting the area in Ω where the contours are present. This idea is also suggested in [53], its formal description reads

$$\chi_{NB}(\mathbf{x}) = \begin{cases} 1, & min\{|\phi_i(\mathbf{x})|\} \leq \varepsilon_{NB} \\ 0, & min\{|\phi_i(\mathbf{x})|\} > \varepsilon_{NB}, \end{cases} \tag{3.1}$$

with ε_{NB} defining the bands minimal width. The shape vector is then simply reduced to values where $\chi_{NB} = 1$. This allows for a high reduction factor that increases with domain size. Equipped with this insight as an optional preprocessing step, we now turn our attention to unsupervised dimensionality reduction techniques.

3.3 Linear Dimensionality Reduction

Preliminary we consider dimensionality reduction methods that assume the data to reside in a linear subspace embedded in observation space.

3.3.1 Principal Component Analysis (PCA)

PCA was introduced by Pearson [54] and became a popular approach in multivariate statistics. By taking advantage of the belief that several attributes of high dimensional data are correlated linearly, PCA fits a new orthogonal basis to the data. The objective is to maximise the variance along the i^{th} dimension while its axis is orthogonal to all preceding ones. To this end, in a first step the data is centred at the origin by subtracting the mean observation $\bar{\mathbf{s}} = \frac{1}{q}\sum_{i=1}^{q}\mathbf{s}_i$. By stacking the residuals $\mathbf{s}_i - \bar{\mathbf{s}}$ in a $N \times q$ data matrix $\mathbf{X}$ column-wise, the covariance matrix is obtained by

$$\mathbf{C} = \frac{1}{q}\mathbf{X}\mathbf{X}^{\mathbf{T}}. \tag{3.2}$$

Variance maximisation is then achieved by diagonalising $\mathbf{C}$ by eigendecomposition

$$\lambda\mathbf{u} = \mathbf{C}\mathbf{u}. \tag{3.3}$$

The resulting eigenvectors $\{\mathbf{u}_i\}$, coined *principal components*, represent the rotated variance maximising basis while corresponding eigenvalues $\{\lambda_i\}$ constitute the eigenspectrum. The principal components assigned by PCA are in order of decreasing variance which is reflected by a decay in the eigenspectrum. As a result, the tailing dimensions can be discarded thus reducing the dimensionality while most information is retained. To retain 95% of information, the leading n principal components are kept such that $\sum_{i=1}^{n}\lambda_i \geq 0.95\sum_{i=1}^{q-1}\lambda_i$. Finally the input data is projected to the new basis by the set of retained eigenvectors $\mathbf{U}_n$ by

$$\mathbf{y}_i = \Psi(\mathbf{s}_i) = \mathbf{U}_n^T(\mathbf{s}_i - \bar{\mathbf{s}}). \tag{3.4}$$

The very same operation allows to project an unseen shape to feature space. Accordingly a shape may be synthesised by adding a weighted sum of the leading n eigenvectors to the mean

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{U}_n\mathbf{y}. \tag{3.5}$$

To facilitate notation, we refer to $\mathbf{y} = \Psi(\mathbf{s})$ as image of $\mathbf{s}$. By contrast, the shape resulting from the inverse mapping $\mathbf{s} = \Psi^{-1}(\mathbf{y})$ is termed the *pre-image* of $\mathbf{y}$. To keep instances valid, the Gaussian standard deviation constraint

$$-3\sqrt{\lambda_j} \leq \mathbf{Y}(\cdot, j) \leq 3\sqrt{\lambda_j} \tag{3.6}$$

should be considered with $\mathbf{Y}(\cdot, j)$ the j^{th} row of $\mathbf{Y}$. PCA returns an optimal basis transform for multivariate Gaussian distributions, as they are characterised by mean and covariance. Note that natural sources of shape like anatomical structures (faces, organs, etc.) indeed can be assumed Gaussian.

3.3.2 Multidimensional Scaling

In contrast to coordinate based PCA, Multidimensional Scaling (MDS) [18] uses pairwise distances between shapes. These are organised in a matrix $\mathbf{D}$ where $d_{ij} = d(\mathbf{s}_i, \mathbf{s}_j)$. For shapes any similarity metric discussed in Section 2.3 may be chosen. MDS generates a point distribution by optimising a stress function, such that all pairwise distances are preserved as good as possible. Such a mapping is referred to as *isometry*. Classical MDS maps the data in a linear subspace by

$$\min_{\mathbf{y}_1 \cdots \mathbf{y}_q} \sum_{i<j} \left(d(\mathbf{s}_i, \mathbf{s}_j)^2 - \delta(\mathbf{y}_i, \mathbf{y}_j)^2 \right) \tag{3.7}$$

A solution to above equation is obtained by eigendecomposition of the Gram matrix $\mathbf{K}$ given by

$$\mathbf{K} = -\frac{1}{2}\mathbf{HDH} \qquad \text{with} \qquad \mathbf{H} = \mathbf{I} - \frac{1}{q}\mathbf{1}\mathbf{1}^T. \tag{3.8}$$

Here $\mathbf{H}$ is a centring matrix, $\mathbf{I}$ the identity matrix and $\mathbf{1}$ a vector of ones. The decomposition yields eigenvalues $\{\lambda_i\}$ of decreasing magnitude and their eigenvectors $\{\mathbf{u}_i\}$. The embedding to feature space $\mathbb{R}^n$ is then given by

$$\mathbf{Y} = (\sqrt{\lambda_1}\mathbf{u}_1, \ldots, \sqrt{\lambda_n}\mathbf{u}_n)^T \tag{3.9}$$

Evidently Equation (3.7) prefers to conserve large distances, as those contribute most to the sum. To circumvent this issue, Sammon [62] proposed to weight each sum term by $d(\mathbf{s}_i, \mathbf{s}_j)^{-1}$. Minimising the stress is then subject to iterative optimisation.

3.4 Nonlinear Dimensionality Reduction

Let us consider a vivid benchmark manifold, the Swiss roll depicted in Fig. 3.1a. The ultimate ambition is to unfold the data to a rectangle. Apart from the non-Gaussian data distribution, PCA can not handle non-linearity and would project the roll to a 2D spiral. MDS will attempt to preserve inter-winding distances, thus it can not unravel the manifold. Such rigid behaviour gives rise to considerations that respect possible non-linear structures. This insight leads to a variety of nonlinear dimensionality reduction techniques, subject to the remainder of this section. Fig. 3.2 shows dimensional dependencies for some

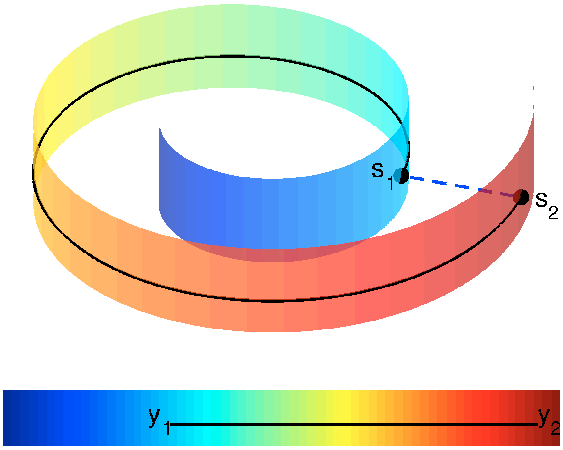


Figure 3.1: The swissroll manifold is shown embedded in $\mathbb{R}^3$. The example illustrates the disparity of geodesic distance along $\mathcal{M}$ (solid line) and the euclidean distance (dashed line) between s_1 and s_2. The desired embedding to $\mathbb{R}^2$ is given by the unravelled roll.

of our shape data-sets. We plot the most prominent principal components y_1 versus y_2 and y_1 versus y_3 for the Eclipse, Cup and FGnet set. Obviously all observed data-sets reveal strong non-linear dependencies, stressing the necessity for nonlinear approaches for our application.

3.4.1 Isometric Feature Mapping (Isomap)

Kernel PCA yields a nonlinear map but does not assume the data to have manifold structure. One method that does so is Isomap proposed by Tennenbaum et al. [67]. Here the basic idea is to approximate the geodesic distance interconnecting two points over the manifold by the sum of Euclidean distances of points in between. Isomap builds an

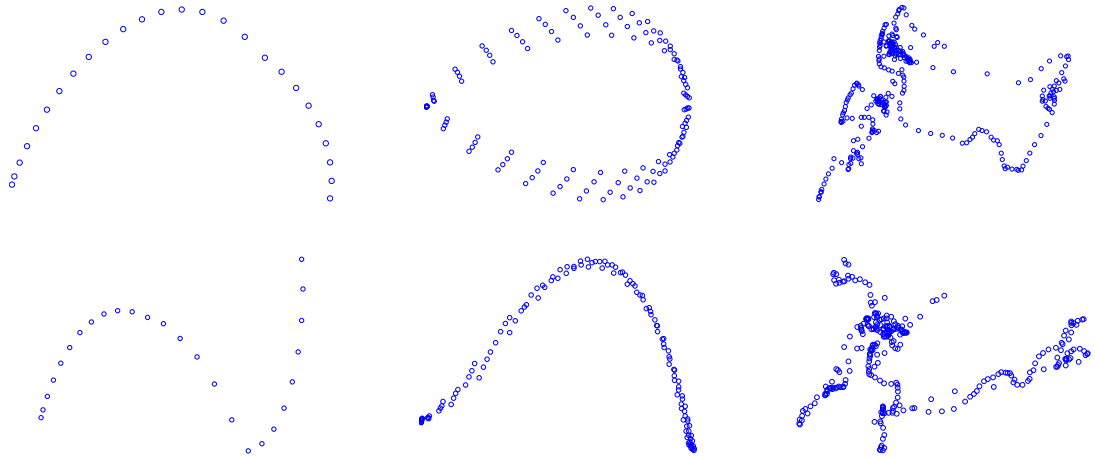


Figure 3.2: Non-linearity of SDF shape data-sets. We plot the principal component y_1 versus y_2 in the first row and y_1 versus y_3 in the second row. (l) Eclipse, (m) Cup, (r) FGnet.

adjacency graph where nodes represent the input samples and links are interconnecting each point with its set of neighbours $\mathcal{N}$. The authors suggest two policies for the notion of neighbourhood: All points within an ϵ-ball centred at the point in question, or alternatively, its k nearest neighbours. The link weights are assigned to input space distances $d(\mathbf{s}_i, \mathbf{s}_j),\ \mathbf{s}_j \in \mathcal{N}_i$, thus the graph is a sparse representation of the MDS distance matrix. From the resulting graph, all-to-all geodesic distances are derived by following the cheapest path through the graph connecting each pair of points, see Fig. 3.1. These approximated geodesic distances are stored in a symmetric $q \times q$ matrix $\mathbf{D}_{\mathcal{M}}$ which is fed to classical MDS as described earlier. The choice of neighbourhood parameters is crucial for a correct embedding since noisy data is likely to reroute the geodesics. Also, Isomap is not able to embed non-convex data correctly. Considering a hole in the Swiss roll band, geodesics will bypass the hole resulting in a distorted embedding [74].

3.4.2 Local Linear Embedding (LLE)

Proposed by Roweis and Saul [61], LLE approximates the data manifold linearly at a small scale. The basic idea is to describe each point solely by a fixed number of neighbours. Like in Isomap, for each point $\mathbf{s}_i$, its nearest neighbours $\mathcal{N}$ are identified. This neighbourhood is assumed to be in the tangent space of $\mathbf{s}_i$. Barycentric weights associating $\mathbf{s}_i$ with its neighbours are then determined and stored in rows of the $q \times q$ weight matrix $\mathbf{W}$. Rows of $\mathbf{W}$ sum up to one and $w_{ij} = 0$ if $\mathbf{s}_j \notin \mathcal{N}$. LLE performs an isometric embedding with

respect to these weights, so each point is reconstructed from its neighbours with the same weights in lower dimensional space. The cost function is therefore given as

$$C(\mathbf{Y}) = \sum_i |\mathbf{y}_i - \sum_j w_{ij}\mathbf{y}_j|^2 \tag{3.10}$$

which can be solved in terms of an eigendecomposition. Exclusion of the trivial solution $\mathbf{Y} = \mathbf{0}$ is achieved by the constraint $\|\mathbf{y}\|^2 = 1$. However, because of this rather simple constraint, LLE is likely to collapse most of the data to a narrow area and may cause undesired manifold rescaling. Also, LLE is known to perform poorly for nonuniform data density [74].

3.4.3 Laplacian Eigenmaps

Another geometrically motivated method termed Laplacian Eigenmaps (LEM) was introduced by Belkin and Niyogi [2]. In fact, LEM is similar to LLE as it computes a map that preserves distances in the local neighbourhood. LEM aims at minimising the weighted sum

$$\sum_{i,j} d(\mathbf{y}_i, \mathbf{y}_j)^2_{L^2} W_{ij}. \tag{3.11}$$

The matrix $\mathbf{W}$ is derived from the adjacency graph obtained from some neighbourhood policy, by weighting links with the Gaussian kernel eq. (3.16). Since short links receive high weight, the minimisation will map similar shapes close to each other. As shown by the authors, above objective is achieved by solving the eigenproblem

$$\mathbf{Lu} = \lambda \mathbf{Du} \tag{3.12}$$

where the degree matrix $\mathbf{D}$ is of diagonal form with entries $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$ and $\mathbf{L} = \mathbf{D} - \mathbf{W}$ being the Laplacian matrix of the graph. It can be shown that in the limit of infinite sample size, the graph Laplacian converges to $\Delta + E$. Here Δ is the Laplace-Beltrami operator, a generalisation of the Laplace operator to curved spaces and E a data density dependent scalar [2]. Omitting the trivial solution $\lambda_0 = 0$, the embedding is given by the eigenvectors corresponding to eigenvalues $\lambda_1, ..., \lambda_m$.

3.4.4 Kernel Principal Component Analysis (KPCA)

Kernel PCA was introduced by Scholkopf and Smola [63]. It generalises standard PCA to a nonlinear method by mapping the input data to a higher dimensional feature space.

This mapping can be achieved implicitly by employing the kernel trick. Briefly, the inner product in feature space is substituted by a kernel function $k : \mathcal{S} \times \mathcal{S} \mapsto \mathbb{R}$

$$k(\mathbf{s}_i, \mathbf{s}_j) = \left\langle \hat{\Psi}(\mathbf{s}_i), \hat{\Psi}(\mathbf{s}_j) \right\rangle, \tag{3.13}$$

mapping the data to a hyper-ellipsoid in feature space. Unlike its linear pendant, KPCA avoids to compute the principal components in feature space. After all we are only interested in the projections of the data onto those vectors. The projection operation can be expressed solely in terms of an inner product, for which we can insert the kernel. Hence there is no need to compute the mappings $\hat{\Psi}$ explicitly. One then proceeds as follows. The kernel is centred in feature space as in Equation (3.8), thereby obtaining the Gram matrix $\mathbf{K}$. By solving the eigenproblem

$$q\lambda\mathbf{u} = \mathbf{K}\mathbf{u} \tag{3.14}$$

the embedding is obtained by

$$\mathbf{y}_i = \left(\mathbf{u}_1 \cdot \mathbf{k}_i, ..., \mathbf{u}_n \cdot \mathbf{k}_i\right)^T, \tag{3.15}$$

with $\mathbf{k}_i$ the i^{th} column of matrix $\mathbf{K}$. This can be understood as performing nonlinear PCA in observation space. An open question concerns the choice of the kernel function. Any symmetric positive-semidefinite kernel may be used. Well-established variants include the Gaussian

$$k(\mathbf{s}_i, \mathbf{s}_j) = \exp\left(-\frac{d(\mathbf{s}_i, \mathbf{s}_j)^2}{2\sigma^2}\right) \tag{3.16}$$

or polynomial functions. For the linear function, we obtain standard PCA. However, such choice is data dependent and non-trivial. Weinberger et al. [76] address this issue by a technique termed *Maximum Variance Unfolding* (MVU). In their approach, the optimal kernel for a given data set is learned by semidefinite programing. MVU explicitly attempts to unfold the manifold by maximising inter point distances [74]. Another variation related to KPCA is *Gaussian Process Latent Variable Model* (GP-LVM) [45], a probabilistic approach where the kernel is applied in feature space.

3.4.5 Diffusion Maps

Another method based solely on a distance measure are Diffusion Maps (DFM), introduced by Coifman, Lafon and coworkers in [14, 43]. The central idea of DFM is to express similarity in terms of a Markov chain that captures the probability of a random walker

to move from one shape to another as in a diffusion process. As before, the data can be thought of as the nodes of a graph interconnected by links. To limit connectivity to a local scale, links are weighted by a rotation invariant kernel $k(s_i, s_j)$. Since the diffusion process is closely related to the heat equation, the authors suggest the Gaussian kernel. Methods such as LEM combine geometry and density information into the embedding. For our application, non-uniform data density can be considered as sampling artifact and should not influence the embedding. As shown in [43], this issue can be resolved by using the normalised kernel

$$\tilde{k}(s_i, s_j) = \frac{k(s_i, s_j)}{q(s_i)q(s_j)} \qquad \text{with} \quad q(s) = \sum_{z \in \mathcal{S}} k(s, z) \tag{3.17}$$

being the degree of node s, indicating its connectivity. Subsequently a Markov chain is obtained by a second normalisation according to

$$p(s_i, s_j) = \frac{\tilde{k}(s_i, s_j)}{\sum_{s_k \in \mathcal{S}} \tilde{k}(s_i, s_k)}. \tag{3.18}$$

The resulting matrix $\mathbf{P}_{i,j} = p(s_i, s_j)$ reveals the probability of transition from shape s_i to shape s_j. Moreover, taking the t^{th} power of $\mathbf{P}$ yields the transition probabilities for t time steps, allowing to define the local neighbourhood on different scales. In the limit of large sample size and narrow kernel width, the entity $\mathbf{I} - \mathbf{P}$ converges to the Laplace-Beltrami operator. For a comprehensive justification of the Laplace-Beltrami being desirable for embeddings, see [2]. DFM aims at an isometry that preserves implicitly the diffusion distance

$$d_{DFM}(s_i, s_j)^2 = \sum_{s_k \in \mathcal{S}} \frac{(p^t(s_i, s_k) - p^t(s_j, s_k))^2}{\mathbf{v}_0(s_k)}. \tag{3.19}$$

With $\mathbf{v}_0$ the top left eigenvector of $\mathbf{P}$. In contrast to the geodesic distance, this measure integrates over all paths of length t, resulting in a robust measure regarding noise. Thus the diffusion distance takes small values for points connected by links with high weight. By Eigendecomposition of matrix $\mathbf{P}$ we get real eigenvalues $\lambda_0 = 1 \geq \lambda_1 \geq \lambda_2 \geq ... \geq 0$ and corresponding right eigenvectors $\{\mathbf{u}_i\}$. Since the trivial eigenvector $\mathbf{u}_0$ is constant, it can be omitted. Hence the diffusion maps embedding is given by

$$\mathbf{Y} = \left(\lambda_1^t \mathbf{u}_1, \lambda_2^t \mathbf{u}_2, ... \lambda_n^t \mathbf{u}_n\right)^T \tag{3.20}$$

As shown in [14], in general few eigenvectors suffice to obtain a good approximation to the diffusion distance.

3.5 Discussion

In this chapter we investigated techniques for dimensionality reduction to extract modes of variability from a set of shapes. To this end most methods employ the eigenvectors of a distance matrix. This reflects the desire to preserve inter sample distances so to find an isometric mapping. For a more comprehensive overview of the diversity of dimensionality reduction methods, the interested reader is referred to van der Maaten et al. [74]. We examined PCA and MDS which are not well suited for non-linear variations occurring in our data sets. We then studied the non-linear approaches Isomap, LLE, Laplacian Eigenmaps, KPCA and DFM. Geodesics based Isomap can not handle non-convexity and is prone to noise. Sparse spectral techniques LLE and Laplacian Eigenmaps suffer from the curse of dimensionality of real-world data and the high dynamic range of eigenvalues resulting in numerical issues [74]. DFM equips the input space with the diffusion distance that has shown to yield a robust mapping w.r.t. outliers [14]. Furthermore, DFM approximates well the input space distances with few eigenvectors while providing a density independent mapping. For these advantageous properties, the remainder of this work will focus on DFM for shape manifold construction.

Chapter 4

Statistical Shape Models

Contents

After investigating tools to represent, handle and reduce shape information we now review key shape models proposed in the literature. A coarse categorisation can be made into linear and non-linear models, considering respective types of shape variations. Another distinction splits global and local approaches. Global shape models, having received most attention, deform the whole contour at once. Local models treat parts of the contour separately.

4.1 Active Shape Models

In their influential work termed *Active Shape Models* Cootes, Taylor and coworkers [17] introduced the point distribution model (PDM). A set of shapes is modelled by application of PCA on a control point based representation. Therefore training shapes are aligned

by generalised procrustes analysis as discussed in Section 2.4.1. The methodology of PCA was already discussed in Section 3.3.1. Using control points amounts to treat their coordinates as input vectors, i.e. $\mathbf{s} := (x_1, y_1, x_2, y_2, ...x_l, y_l)^T$ in 2D. Hence new shapes may be generated by a linear combination of eigenmodes (Equation (3.5)). Figure 4.1a-b illustrates the so obtained most significant modes for the IMM faces data-set. As can be seen, the first mode obtained by $\mathbf{y} = (y_1, 0, ..., 0)^T$ controls mostly the left-right head pose. The second mode alters nodding variation along with prominent facial expression. Active Shape Models performs segmentation by deforming an initial model instance according to image information. The same authors later introduced Active Appearance Models [16]. Here the shape information is augmented by learning the appearance enclosed by the shape. Consistently, the image intensities warped onto a reference shape are subject to PCA.

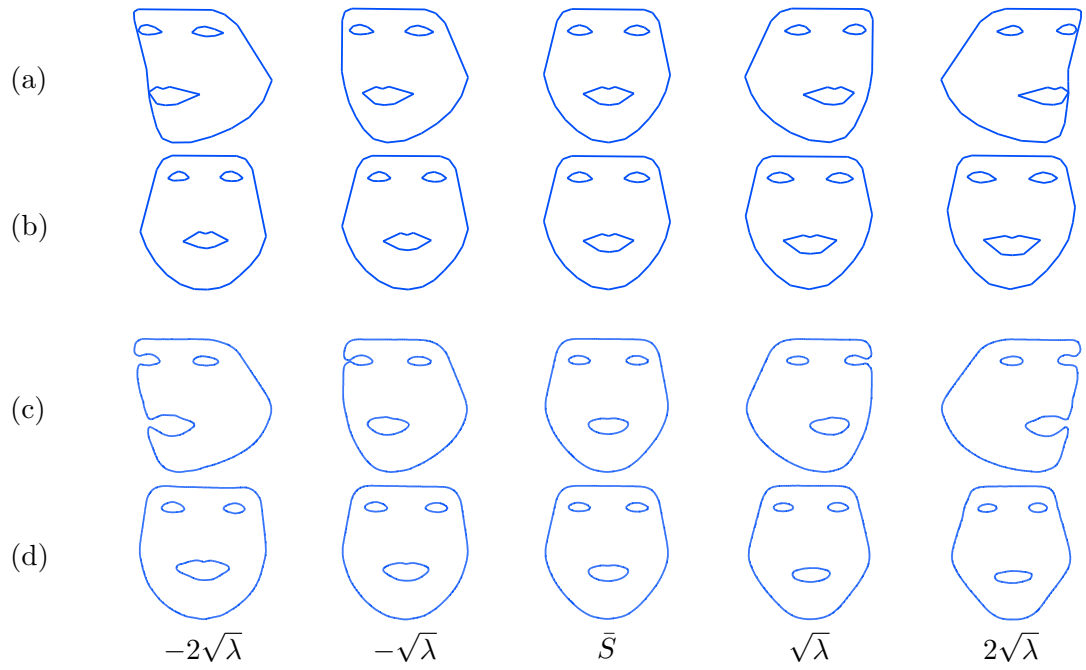


Figure 4.1: Modes of variation: rows (a) and (b) are generated by a PDM, representing the two most significant modes. Rows (c) and (d) show the first two modes of the same training set using Leventon's approach. For this model 120 IMMfaces were trained.

4.2 Liner Model On Distance Maps

Leventon, Grimson and Faugeras introduced shape priors to the level set framework in [46]. Adopting the principle of PDM, linear PCA is applied to SDFs to obtain eigen distance maps. The authors augment the geodesic active contour (GAC) segmentation energy [8] with prior knowledge formulated by a *maximum a posteriori* (MAP) approach stated as

$$P(\mathbf{y}, T \,|\, \phi, \nabla I) = \frac{P(\phi \,|\, \mathbf{y}, T)\; P(\nabla I \,|\, \mathbf{y}, T, \phi)\; P(\mathbf{y})\; P(T)}{P(\phi, \nabla I)}. \tag{4.1}$$

Above equation estimates the probability of a certain shape modelled by coefficient vector $\mathbf{y}$ and its pose T, given a test shape ϕ and image edge response ∇I. Considering a Gaussian distribution on $\mathbf{y}$, the prior term is modelled by

$$P(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{C}_n|}}\; \exp\left(-\frac{1}{2}\mathbf{y}^T \mathbf{C}_n^{-1} \mathbf{y}\right), \tag{4.2}$$

with $\mathbf{C}_n$ the model variance of the n retained eigenmodes. Without interaction of the other terms in Equation (4.1), $P(\mathbf{y})$ reaches its maximum for the mean shape. As noted by the authors, the space of SDFs does not form a vector space, thus linear combinations will not result in correct SDFs. Also, a variation of the SDF and the cause on the contour are not related in a direct manner. Therefore meaningful shape modes are not concentrated in the first components, but are scattered throughout all modes. This requires high n in contrast to the PDM approach [22]. As an example, the leading two modes of the IMM face data-set are shown in Fig. 4.1c-d. As can be seen, this formulation tends to smooth out details. Also, mode deflections within the Gaussian assumption produce unnatural shapes compared to the point distribution model. Nevertheless, well-balanced combinations of these modes (as produced by Equation (3.4) and Equation (3.5)) generate valid instances.

This model inspired various extensions and modifications. For example, in [24] Cremers et al. use the prior energy Equation (4.2) on an explicit representation without projecting to a reduced space. Since this might lead to a covariance matrix of deficit rank, regularisation is applied to compute its inverse. In [19] Cremers applies PCA on distance maps to define dynamic priors. Given a temporal sequence of shape variation, the goal is to find the MAP shape for the next frame. As shown by the author, this approach handles a great deal of noise and may be used to synthesise oscillating variations such as a walkers shape. In [60] Rousson et al. propose to enforce the mean shape to be a SDF. Tsai et al. [71] and Bresson et al. [7] adopted this model with an improved segmentation energy formulation.

4.3 Linear Model on Probability Maps

Cremers and coworkers [23] propose to replace SDFs by probability functions, where shape is defined by $p : \Omega \mapsto [0, 1]$, assigning each pixel the probability $p(\mathbf{x} \in \Omega_{in})$. Since the resulting shape space forms a convex set, mean and covariance can be obtained under the linear assumption. By applying PCA on the probability maps, their eigenmodes are recovered. To restrict pre-images of the model to feasible shapes, an iterative projection scheme is proposed. This scheme alternates two steps until convergence. First the evolving shape is approximated by the model eigenmodes. Then shapes outside the probabilistic shape definition are projected onto shape space by clipping values $p(\mathbf{x}) \notin [0, 1]$.

Probability maps were also used by Pohl et al. [55] in terms of *Log Odds*, the logarithm of the ratio $\frac{p}{1-p}$. Log odds are a generalisation of SDFs with the advantage of embedding shapes to a linear space.

4.4 Chord Length based Model

Another approach is to not consider control point coordinates but inter-point distances, also known as *chord lengths* instead. This idea was proposed by Cootes [15] among others to account for non-linear variations like bending and relative rotation. Here the authors apply PCA to the vectors of chord lengths to obtain a model. As discussed in [17], chord lengths are invariant to rotation and translation and after appropriate normalisation also to scale.

Besbes [5] extended the chord-length approach to a local model. Here the authors use *shape maps* by Langs and Paragios [44], an extension to diffusion maps to eliminate redundant shape information. Shape maps are applied to cluster control points into groups of high correlation. Control points closest to the centres of each cluster are then identified as points of reference. Next a graph is constructed by assigning edges from center points to corresponding cluster points. Finally, for each of the so obtained edges a Gaussian kernel density is learned from training samples. The model probability reads

$$p(\mathbf{x}_1, ...\mathbf{x}_n) \propto \prod_{(i,j)\in R_k} p(\mathbf{x}_i, \mathbf{x}_j)^{\frac{2}{k}} \prod_{(i,j)\in E_k} p(\mathbf{x}_i, \mathbf{x}_j)^{\frac{1}{k}} \tag{4.3}$$

with the first product over inter-center probabilities and the second one over center to non-center points. In order to model shape instances from a random point cloud, the

authors maximise Equation (4.3). This is achieved by iteratively applied constrained perturbations, see Fig. 4.2.

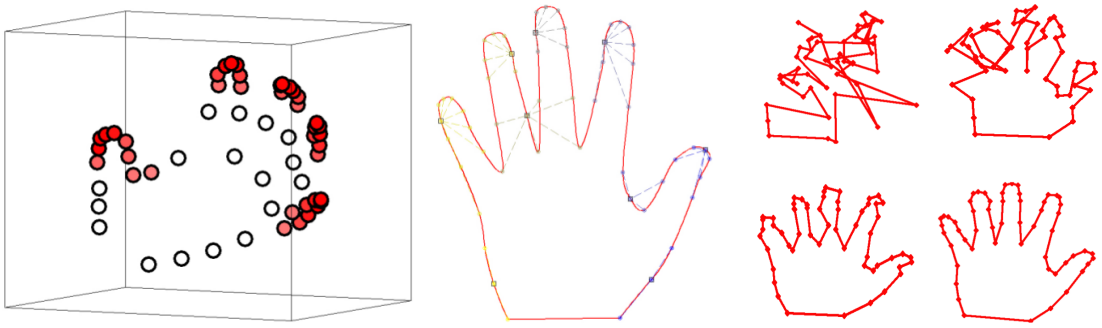


Figure 4.2: Local Chord length model. Clusters in first three shape maps dimensions with colour coded density (l). Control points assigned to their centres depicted by squares (m). Deformation of random points towards the mean shape (r). Reproduced from [5].

4.5 Local Statistics on SDFs

The methods mentioned above impose the SSM globally, i.e. uniformly distributed along the contour. In contrast, Rousson and Paragios [58, 59] pursue a local approach. From a set of aligned SDFs, the objective is to extract the Gaussian probability density distributions at the pixel level. This boils down to estimate a mean shape ϕ_m along with a confidence map σ_m that best describes the training set. Hence for each pixel $\mathbf{x}$, the goal is to find the distribution

$$P(\phi(\mathbf{x})) = \frac{1}{\sqrt{2\pi}\sigma_m(\mathbf{x})}\, exp\left(-\frac{(\phi(\mathbf{x}) - \phi_m(\mathbf{x}))^2}{2\sigma_m^2(\mathbf{x})}\right), \tag{4.4}$$

while constraining ϕ_m to a SDF. Translating above equation to a MAP formulation, one obtains the energy functional

$$E(\phi_m(\mathbf{x}), \sigma_m^2(\mathbf{x})) = -\log \sum_{i=1}^{p} P(\phi(\mathbf{x})) = \frac{1}{2}\sum_{i=1}^{p}\left[\log(\sigma_m^2(\mathbf{x})) + \frac{(\phi_i(\mathbf{x}) - \phi_m(\mathbf{x})^2)}{\sigma_m^2(\mathbf{x})}\right] \tag{4.5}$$

which is subject to minimisation. Extending this formulation to the whole image domain, eventually we get

$$E(\phi_m, \sigma_m^2) = \alpha \sum_{i=1}^{p} \int_\Omega \left[\log(\sigma_m^2) + \frac{(\phi_i - \phi_m)^2}{\sigma_m^2} \right] d\Omega \; + \int_\Omega \psi(\nabla \sigma_m^2) \; d\Omega. \tag{4.6}$$

The added last term acts as variance regularisation, causing σ_m to vary smoothly. Figure 4.3 illustrates an example of a mean shape and its associated confidence map for two horse shapes. Bright regions in the latter indicate low confidence which is propagated to domain borders. In practice ϕ_m is computed as the empirical mean $\phi_m = \frac{1}{p} \sum_{i=1}^{p} \phi_i$ followed by a projection to the space of SDFs. Based on this result, the confidence map is obtained by gradient descent. Opposite to the global model from Section 4.2, local statistics allow to enforce the prior locally varying. Thus in regions of low prior confidence the segmentation energy is given more weight.

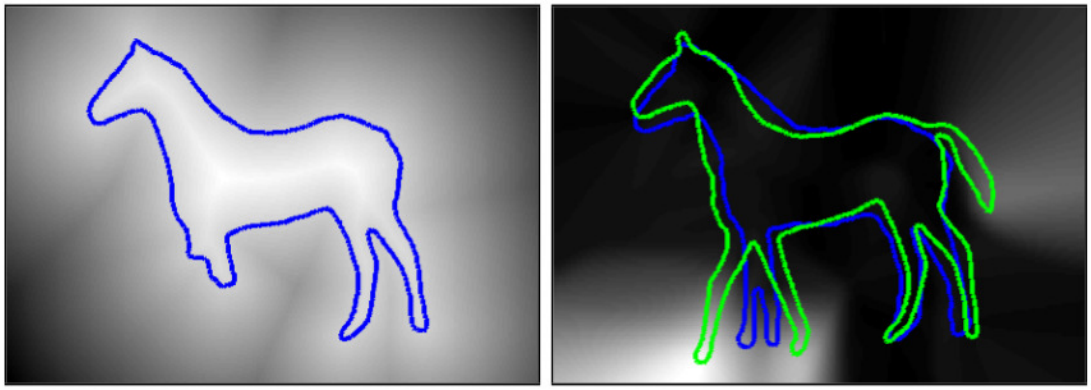


Figure 4.3: Mean shape resulting from two horse shapes along with its zero level (l). The confidence map and the two training samples are pictured (r). Reproduced from [59].

4.6 Statistics on Deformation Fields

In [11, 12] Charpiat et al. propose to apply statistics based on shape distances, thereby exploiting the advantage of independence regarding representation. The authors devise a control point based approach with a continuous approximation of the Hausdorff metric as well as a SDF approach using the Sobolev $W^{1,2}$ norm. Minimising the distance $d(S_i, S_j)$ between two shapes defines a normal deformation field β on the contour that warps one shape towards another. While the methods in Section 4.1, 4.2 assume a linear shape space,

here the idea is to linearise the space of normal deformations around a mean. Therefore, the mean of a set of shapes is defined as the Karcher mean

$$\bar{S} = \arg\min_{S} \sum_{i=1}^{q} d(S, S_i)^2. \tag{4.7}$$

Based on $\bar{S}$, there exists a set $\{\beta_i\}_{i=1\ldots q}$ of deformation fields that warp $\bar{S}$ to S_i. Hence a covariance matrix can be defined on those deformations and PCA yields their ordered modes of variance. This can be interpreted as moving in the tangent space of the deformation space, attached to $\bar{S}$. Figure 4.4 illustrates the so obtained mean and modes resulting from 12 starfish shapes. In practice above mean is computed by gradient descent which does not guarantee to find the global minimum. Hence the mean becomes less meaningful for large data-sets as more local minima emerge [11].

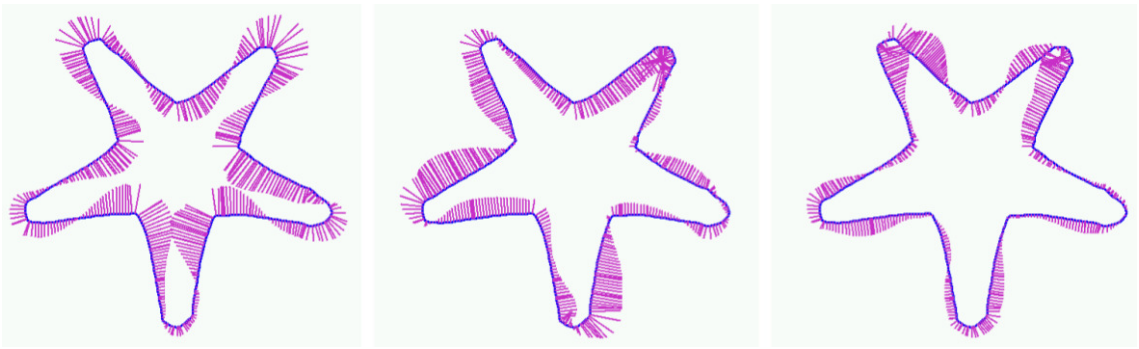


Figure 4.4: Charpiat's Model learnt from 12 starfish shapes. The mean is depicted by the contour along with the leading three modes of deformations. Lines indicate the transport of deformation fields caused by single mode variation. Reprinted from [12].

4.7 Density equipped Feature Space

In an extension to [24], Cremers, Kohlberger and Schnörr [20] build a nonlinear shape model by combining the ideas of KPCA and Parzen density estimation. Control points are used to parametrise shape. Assuming a Gaussian data distribution in KPCA feature space, non-Gaussian density estimates are obtained in input space. An unknown shape is then attracted to regions of high sample density by minimising the energy

$$E(S) = \Psi(S)^T \, \mathbf{C}^{-1} \, \Psi(S). \tag{4.8}$$

With $\Psi(S)$ being the mapping to feature space. To cope with non-invertible covariance matrices $\mathbf{C}$, a regularisation is introduced which effectively splits the energy in a distance in feature space and a distance from feature space. The authors integrate above prior into the Mumford-Shah segmentation.

Pricasariu and Reid [56] follow a similar idea. They describe shape by Fourier Descriptors as discussed in Section 2.1.2. It is assumed that the Fourier harmonics of the training set lie on a non-linear manifold. A model is derived by embedding the harmonics of the training shapes to reduced space by Gaussian Process Latent Variable Model (GP-LVM) [45]. Here the Gaussian process is defined by a mean and covariance function. LVM is a non-linear dimensionality reduction technique which requires constraints to make it an isometry. GP-LVM equips the reduced space with a variance map indicating locations likely to generate valid shapes. Figure 4.5 exemplifies this concept by a heat-map. An

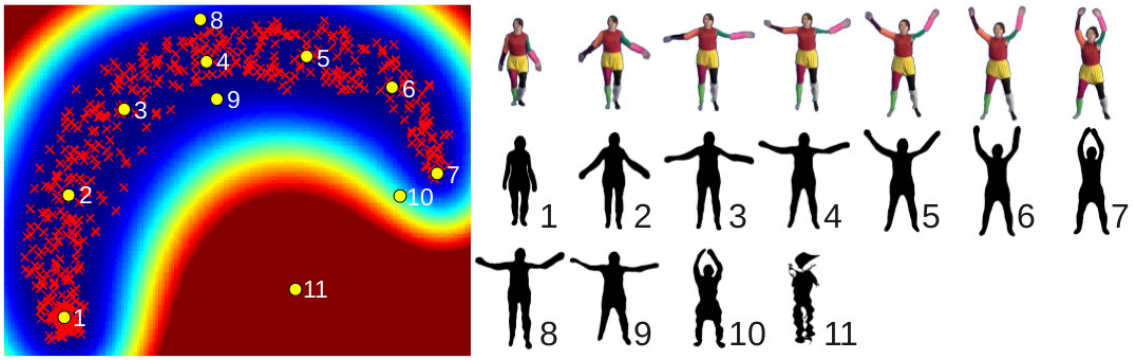


Figure 4.5: GP-LVM shape model. 2D feature space, the heat map indicates regions of low variance / high probability in cold colours and unlikely locations in hot colours (l). Crosses depict embeddings of training shapes. Labelled points generate the correspondingly labelled shapes (r). Figure is reprinted from [56].

unknown shape is projected to feature space by a maximum likelihood approach. Prior knowledge is imposed to a segmentation process by attracting the evolving contour towards regions of low variance. Since the image driven shape is represented by a level set function and the prior as Fourier descriptor, both are unified to a control point description by sub-sampling. The authors show that even strong articulation can be recovered from a 2D feature space.

4.8 Discussion

In this chapter we reviewed approaches to SSMs based on various shape representations and feature extraction techniques. As discussed, linear feature extraction combined with control points can not model complex variations such as bending. Plugging in SDFs as done by Leventon et al. [46] relaxes this constraint since SDFs are not directly related to specific contour points. Although this approach does not operate in the space of SDFs, it has shown to work well in practice [22]. Chord length models alleviate the liner restriction but do not provide the advantages of SDFs. Having no internal parametrisation, the approach in Section 4.5 is more of an segmentation regularisation term than a SSM. The linear deformation space of Section 4.6 is an interesting idea to model nonlinear variations. One step further is to utilise a non-linear feature extraction technique as in Section 4.7 that seem most promising for our purpose. Being mostly related to the techniques in Sections 4.6 and 4.7, the next chapter details a manifold approach within the space of SDF representations.

Chapter 5

Shape Manifolds

Contents

Hitherto the principles for shape representation and SSMs were established. Our goal is to find a model meeting concerns with strong nonlinear shape variations using the advantageous implicit shape representation by signed distance functions (SDFs). In this chapter we detail the model proposed by Etyngier and Thorstensen [31, 69] which follows the assumption of shapes being sampled from a low-dimensional manifold. Indeed, shape variation is often caused by few parameters. Consider for example the multiple 2D views produced by a rigid 3D object, where shape is determined by only two spherical coordinates. The basic idea is then to map the shape manifold comprised by training samples to low dimensional Euclidean space in order to recover the independent modes of variation. Within this approach three central requirements emerge:

- We need a way to estimate the manifold continuously between samples,
- define a projection operator that attracts a new, potentially noisy shape onto the manifold and
- define a map from feature space back to shape space.

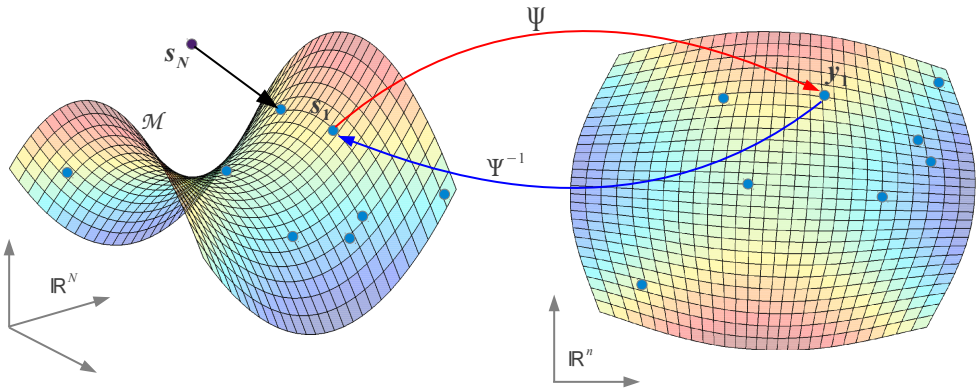


Figure 5.1: Illustration of the manifold framework. The shape manifold $\mathcal{M}$ is indicated by a saddle surface embedded in $\mathbb{R}^{N=3}$, a few samples are shown as points (l). Embedding to $\mathbb{R}^{n=2}$ (r). Embedding $\mathbf{y} = \Psi(\mathbf{s})$ of a sample $\mathbf{s}$ and its inverse, the pre-image mapping $\mathbf{s} = \Psi^{-1}(\mathbf{y})$ and the projection of a noisy shape $\mathbf{s}_N$ are illustrated.

These concepts are illustrated in Fig. 5.1. In this chapter we explore approaches addressing these issues, allowing for the formulation of a proper nonlinear shape model.

5.1 Discovery of the Shape Manifold

As proposed in [31, 69] we recover the shape manifold indicated by q samples by a map to a lower dimensional Euclidean space $\Psi : \mathcal{S} \to \mathbb{R}^n$. To project the samples to shape space, we first align the training set by registration with the $SSD(\phi)$ metric, see Section 2.4.2. For the input space distance we employ the Sobolev $W^{1,2}$ norm defined in Section 2.3.3 as it includes information on the SDF gradient. Bear in mind that the presented framework can be applied with any differentiable distance. With the Gaussian kernel, the initial Diffusion kernel reads

$$w(\phi_i, \phi_j) = \exp\left(-\frac{d_{W^{1,2}}(\phi_i, \phi_j)^2}{2\sigma^2}\right). \tag{5.1}$$

Appropriate normalisation and eigendecompositon as discussed in Section 3.4.5 yields the embedding

$$\Psi : \phi_i \in \mathbb{R}^N \mapsto \mathbf{y}_i \in \mathbb{R}^n \qquad \forall \phi_i \in \mathcal{S}. \tag{5.2}$$

Examples of embeddings for $n = 2$ are given in Fig. 5.2. Note how DFM clusters shapes with distinct head poses. Based on the embedding, designing a head pose classificator is

straight forward. The IMM faces form a parabolic structure while the nuclei caudate are distributed more uniformly with one distinct outlier at the top.

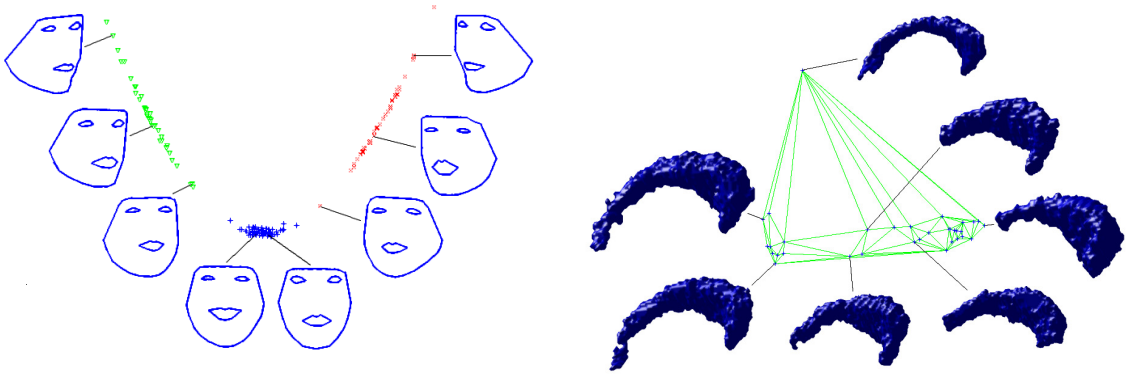


Figure 5.2: Embedding of 120 IMM faces to $\mathbb{R}^2$ (l), embedding of 33 3D shapes of caudate nuclei (r). The illustrated shapes reference to their location on the manifold.

The intrinsic manifold dimension, the reduced space dimension n can be interpreted as the number of modes retained. It may be known beforehand (e.g. it is one for the eclipse shapes) or be inferred from the eigenspectrum of the Diffusion kernel. Hereafter we set n as a parameter, usually low values of $n \leq 4$ suffice even for complex data sets. The parameter $\sigma > 0$ in Equation (5.1) is of crucial importance as it defines the local connectivity among samples and therefore determines the obtained manifold geometry. High σ transfers most eigen-power to the leading eigenvalues thereby capturing most variation in few reduced dimensions but degrades the ability to capture nonlinear structures. On the other hand, choosing σ too small leads to a disconnected graph. In this work we set σ to the median of nearest neighbour distances, ensuring a connected graph. This heuristic was proposed throughout the literature [1, 31]. Since Diffusion Maps provides an isometry w.r.t. the diffusion distance, the local neighbourhood of shapes can be inferred in feature space.
In [31], a Delaunay triangulation in reduced space is performed, partitioning feature space into n-simplices, each having $n + 1$ samples as vertices. The union of those simplices defines the shape manifold $\mathcal{M}$.

5.2 Shape Warping

Given a finite set of training shapes sampling the manifold, we want to move continuously along its structure, i.e. between samples. To do so, a basic requirement is to warp a

shape S continuously towards a target shape S_T. In this work we tackle shape warping by minimisation of a dissimilarity energy $E : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ comparing S and S_T, Hence we first need to establish the notion of the gradient ∇E with respect to a shape.

5.2.1 Calculus of Variations

In mathematics, Calculus of Variations deals with finding extrema of functionals mapping from the domain of functions to the real axis. Then one is interested in finding functions S for which the functional $E(S)$ becomes stationary. In this context the Gâteaux variation generalises the directional derivative according to

$$\delta E(S, \beta) = \lim_{\epsilon \to 0} \frac{E(S + \epsilon\beta, S_T) - E(S, S_T)}{\epsilon} = \frac{\partial}{\partial \epsilon} E(S + \epsilon\beta)\big|_{\epsilon=0} \tag{5.3}$$

Here δE is called *first variation* and β is a small but arbitrary test function. For S represented by a contour, β is a normal deformation field, while for S being a SDF we have $\beta : \Omega \to \mathbb{R}$. Equation (5.3) relates an infinitesimal change of deformation to the change of energy. In this sense $\delta E(S, \beta)$ is the derivative of E with respect to S in "direction" β. Let us consider the generic functional

$$E(S(x)) = \int_\Omega f(x, S(x), S'(x))dx \qquad \text{with} \qquad S' = \frac{dS}{dx}. \tag{5.4}$$

Then it can be shown that for δE to vanish, i.e. the functional reaching an extrema, the Euler-Lagrange Equation (5.5) has to hold.

$$\nabla_S E = \frac{\partial E}{\partial S} - \frac{d}{dx}\frac{\partial f}{\partial S'} \stackrel{!}{=} 0 \tag{5.5}$$

5.2.2 Variational Shape Warping

Note that unlike usual variational problems we are aware of the energy's minimum. It is reached for $S = S_T$. We are rather interested in shapes between defined ones. An open question concerns the choice of the dissimilarity energy. Candidates are continuous shape distances discussed in Sec. 2.3. For example, inserting the Sobolev energy

$$E(S, S_T) = d_{W^{1,2}}(S, S_T)^2 = \int_\Omega |\phi - \phi_T|^2 + |\nabla\phi - \nabla\phi_T|^2 d\Omega. \tag{5.6}$$

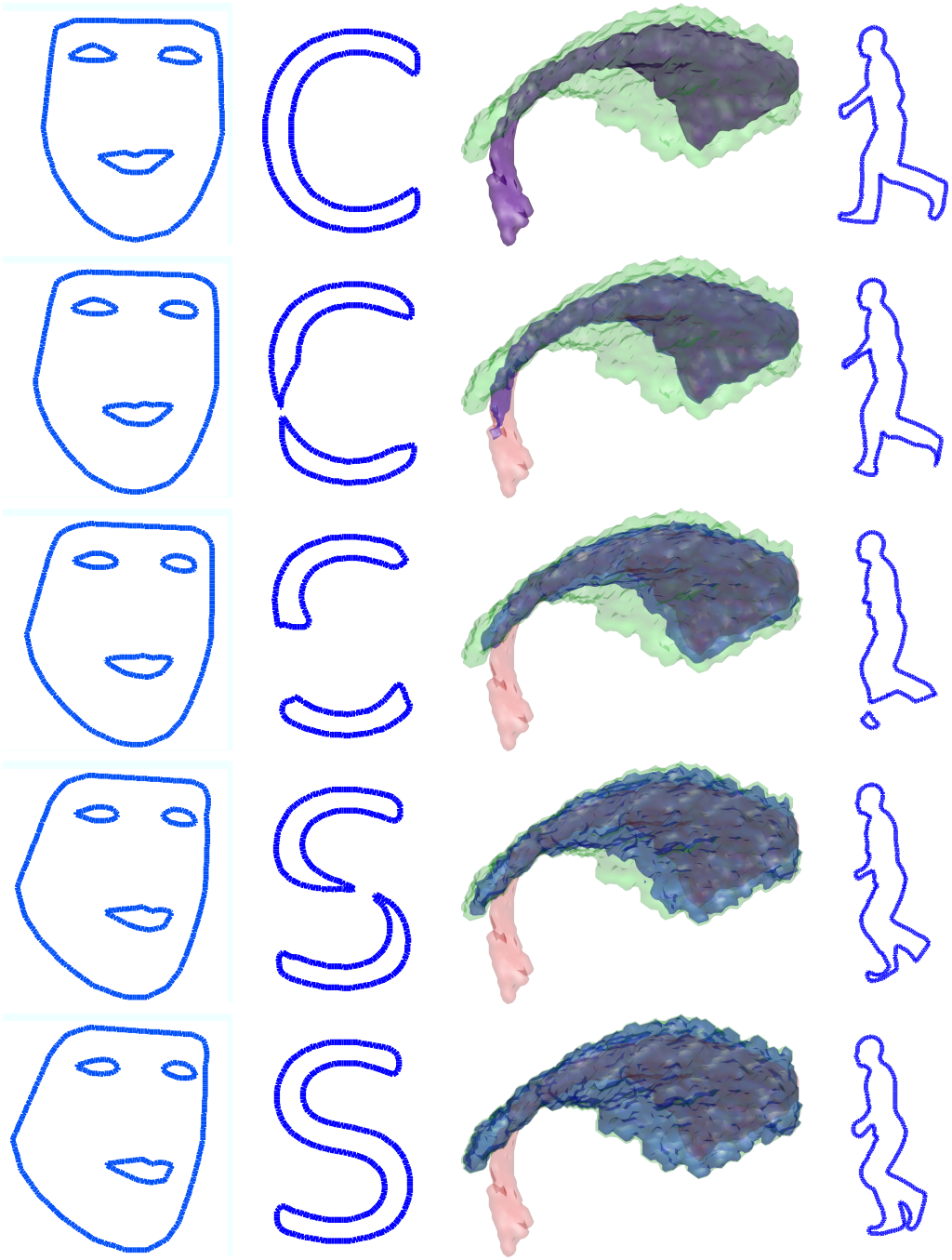


Figure 5.3: Intermediate shapes obtained by warping the topmost shape towards the bottom one by minimising the Sobolev $W^{1,2}$ norm.

to the Euler-Lagrange Equation (5.5), we get

$$\nabla_\phi E(\phi) = (\phi - \phi_T) - \operatorname{div}(\nabla(\phi - \phi_T)) \tag{5.7}$$

Evolving ϕ towards ϕ_T by gradient descent is then achieved by solving

$$\frac{\partial \phi}{\partial t} = -\nabla_\phi E(\phi) \tag{5.8}$$

Note that this approach also provides a solution for the correspondence problem when contour points are tracked during evolution. For an exhaustive analysis of deformation variants, the reader is referred to [11]. Examples of intermediate shapes obtained by the Sobolev warp are shown in Figure 5.3. As can be seen, the warp produces a natural transition between a frontal and rotated face. When warping the letter C to letter S, common parts at top and bottom remain while the middle part splits at first, then reconnects. Transition between 3D nuclei shapes also produces plausible intermediate shapes. However, attempting to warp between shapes having barely overlap of corresponding parts, yields poor results (Fig. 5.3d). This issue will be addressed by incorporating manifold knowledge such that the warp bypasses intermediate shapes of the training set.

5.2.3 Distance Map Creation and Reinitialisation

The warp presented in the previous section deforms shapes without consideration of the SDF condition $|\nabla\phi| = 1 \quad \forall \mathbf{x} \in \Omega$. This issue also occurs when an SDF is evolved by a segmentation energy. To counteract this behaviour, Lie et al. [47] propose a regularisation term for active contour evolutions by incorporating the SDF error into the energy. However, their method is restricted to a narrowband. Starting with an initial function embedding a zero level, several algorithms were proposed to compute the associated full domain SDF. One way is to evolve the PDE (partial differential equation) of Sussman et al. [66] until steady state:

$$\frac{d\phi}{dt} = \operatorname{sign}(\phi_0)(1 - |\nabla\phi|). \tag{5.9}$$

With the sign function taking values -1 inside and 1 outside the shape. During evolution values of the opposite region act as boundary conditions. While this method is intended to preserve the zero level, in practice it tends to smooth away high curvature due to discretisation errors. The algorithm yields highly accurate SDFs coming at high computational cost, especially for ϕ_0 being far from a SDF.

Another approach, termed *fast marching*, starts with an initial band of SDF values adjacent to the zero level. Distance information is propagated according to a normal flow of the band, thereby updating the traversed pixels to SDF values. This approach can be implemented efficiently by a binary tree resulting in a cost of $O(n \log n)$ with n being the number of grid points. In its original formulation fast marching is first order accurate.

Maurer et al. [49] construct a SDF by principles of dimensionality reduction and partial Voronoi diagrams. This results in a very efficient algorithm that executes in linear time. We evaluated the time to construct a 256 $\times$ 256 SDF from its characteristic function on

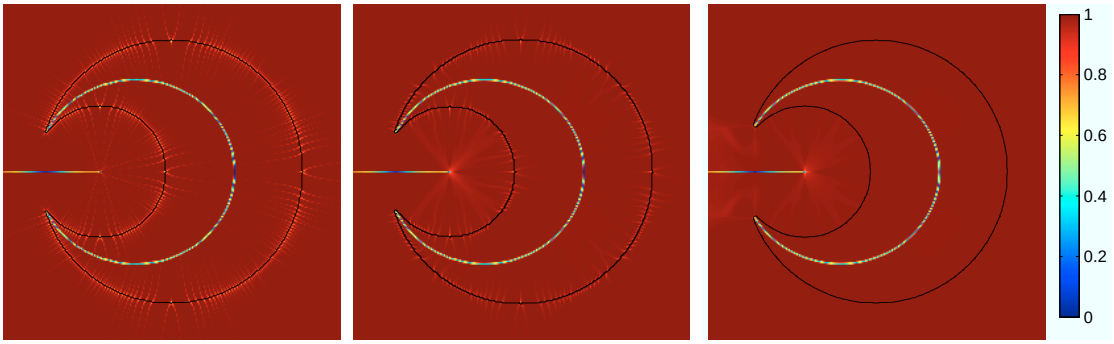


Figure 5.4: The SDF gradient $|\nabla\phi|$ obtained by the algorithms of Maurer (l), fast marching (m) and Sussman (r). The resulting zero level is superimposed as contour. Ridges indicate points of equal distance to at least two contour points.

an Intel® C2D at 3 GHz. Sussman's method requires 9 s for reaching convergence after 175 iterations. Fast marching takes 120 ms, Maurer's algorithm only 10 ms. Figure 5.4 shows the resulting SDF gradient in colour code. Obviously, Sussmans' method produces fewest numerical artifacts at the contour proximity resulting in a smooth zero level. All mentioned methods are applicable for 3D shapes as well. For a 256^3 SDF Maurer's method takes 7.33 m, fast marching finishes after 16 s.

5.2.4 Warp by Velocity Projection

In order to avoid reinitialisation, Chen, Charpiat and Radke [13] propose a warp by *velocity projection* within the space of SDFs. To this end the authors establish the relation between contour gradient $\nabla E_{\mathcal{C}}$ and SDF gradient ∇E_{ϕ} according to Equation (5.10).

$$\nabla_C E(\mathcal{C}(c)) = -\int_{l(c)} \nabla_\phi E(\phi(\mathbf{x}))\, |1 - \kappa(\mathcal{C}(c))\, \phi(\mathbf{x})|\, dr\mathbf{n}_{\mathcal{C}(c)} \tag{5.10}$$

Here $l(c)$ represents the normal projection line through contour point c with $\kappa(\mathcal{C}(c))$ being the local contour curvature. Above equation relates a change of the contour $\nabla E_{\mathcal{C}}$ at contour point $\mathcal{C}(c)$ to the change in the corresponding SDF. Applying a normal deformation vector at $\mathcal{C}(c)$ causes a constant SDF update along the normal projection line. The magnitude of the update is given by integrating the weighted SDF gradient along the projection line. This scheme allows to warp with gradient formulations defined along the contour which are intricate for the EL minimisation. As shown in Fig. 5.5a, the resulting d_{L^2}-warp features a nice fluid-like flow. In contrast, the EL warp would resolve the lower shape followed by spawning it at the top shape.

In practice projection lines are replaced by radial widening rays for integration. Also, occasional SDF reinitialisations are required to correct for accumulating numerical errors. In another example, we warp a flower-like shape, see Fig. 5.5b. At first the flower shrinks to adjust for the radial size while splitting, followed by reconnecting to adjust for finger width. Clearly, this technique is not suitable for our application as we require intermediate contours being of weighted mean nature. We found this warp not straight forward to extend for shapes involving multiple contours like the face shapes. Furthermore, the algorithm's performance does not scale well with domain size.

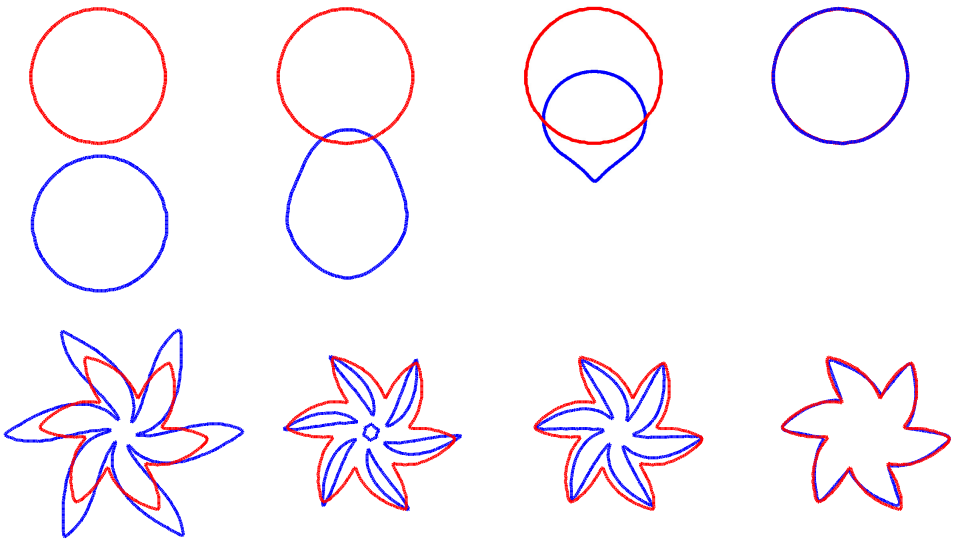


Figure 5.5: Warp by velocity projection: warping the lower circle to the upper one reveals fluid-like behaviour (top). Warping a flower like shape shows unexpected topological changes (bottom).

5.3 The Out-Of-Sample Problem

The embedding $\Psi : \mathcal{S} \mapsto \mathbb{R}^n$ is defined for the training set but not for unknown shapes. Therefore we require a way to map arbitrary shapes e.g. an evolving segmentation shape S to feature space. This is known by the term *Out-of-Sample Problem*, depicted in Fig. 5.1. For linear PCA, this simply amounts to evaluate Equation (3.4). However, kernel based maps require a different approach. A naive solution is to augment the training set with the test shape and then recompute the embedding, This approach strains an eigendecompositon, which is expensive for large data-sets, having complexity $O(q^3)$. The eigendecomposition should be kept as a pre-computation step in the model creation phase. Hence a more efficient alternative is desirable.

5.3.1 The Nyström Extension

Since the embedding coordinates given by Diffusion Maps are constituted by eigenvectors of a transition kernel, it seems a natural idea to extend them for an additional sample. A widely uses technique exploiting this concept is the Nyström extension as detailed subsequently. Since each sample in S_i satisfies

$$\sum_{S_j \in \mathcal{S}} p(S_i, S_j)\ \mathbf{u}_k(S_j) = \lambda_k \mathbf{u}_k(S_i), \tag{5.11}$$

the Nyström extension embeds S according to

$$\tilde{\Psi}_k(S) = \frac{1}{\lambda_k} \sum_{S_j \in \mathcal{S}} p(S_j, S)\ \mathbf{u}_k(S_j). \tag{5.12}$$

In words, the k^{th} coordinate of the embedding $\tilde{\Psi}(S)$ is given by a weighted sum over the elements of the k^{th} eigenvector $\mathbf{u}_k$. Weights are constituted by the transition probability kernel $p(S_j, S)$. Eigenvalues λ_k and eigenvectors are those of the training kernel. This amounts to compute the distances $\{d(S_i, S)\}_{1 \leq i \leq q}$ followed by their integration to the diffusion kernel, having time complexity $O(q)$. Notably, when S differs significantly from shapes in $\mathcal{S}$, the extension maps close to most similar shapes.

As shown in by Bengio et al. [4], this technique can be applied to most kernel methods. Figure 5.6 presents an evaluation of the so obtained extension for the eclipse data-set. We plot the two most significant Diffusion kernel eigenfunctions as solid lines. In a leave-one-out fashion all but one shapes are used to train a model, while the Nyström extension

is applied to the left out shape. Resulting extensions are indicated by crosses. From the systematic variation in the data-set it can be expected that extended values lie between or close to neighbours. As seen in Fig. 5.6 this expectation is met. Note that even samples 1 and 26 in the rim of coordinates are embedded quite accurately. In [43], Lafon and

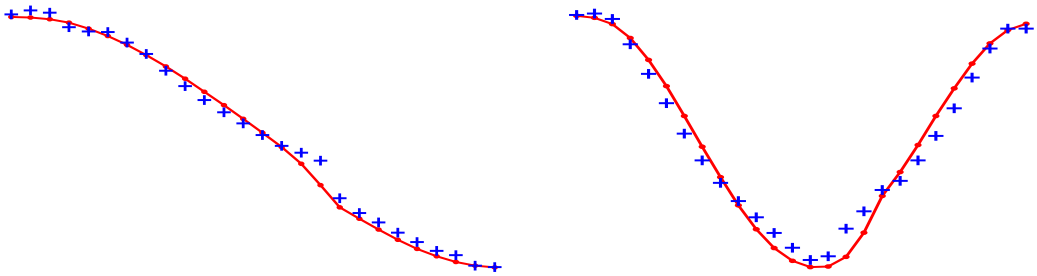


Figure 5.6: Evaluation of the Nyström extension. The approximate first (l) and second (r) eigenfunction are plotted as solid line. Crosses illustrate the values estimated by the Nyström extension.

coworkers suggest a multi-scale extension that finds a separate extension kernel for each diffusion coordinate. This is justified by the fact that each eigenfunction is of different complexity and so is extendable more or less. The resulting algorithm requires a repeated eigendecomposition and the method introduces two additional parameters which were found hard to tune for generic data-sets.

5.3.2 Embeddings outside the Manifold

We now address the scenario when the point $\mathbf{y} = \tilde{\Psi}(S)$ comes to lie outside the convex hull of $\mathbf{Y}$. This might occur when S differs significantly from the learned shapes. Another cause may be an inappropriate choice of the target dimension. We handle this exception by following the suggestion of Etyngier [31] to project $\mathbf{y}$ onto the convex hull $\mathcal{CH}(\mathbf{Y})$. The facets of $\mathcal{CH}(\mathbf{Y})$ resemble $n-1$ dimensional simplices, subsets of hyperplanes in $\mathbb{R}^n$. The orthogonal projection $\tilde{\mathbf{y}}$ of a point $\mathbf{y}$ onto a hyperplane is given by

$$\tilde{\mathbf{y}} = \mathbf{y} - \frac{(\mathbf{y} - \mathbf{y}_i) \cdot \mathbf{n}}{\mathbf{n} \cdot \mathbf{n}} \cdot \mathbf{n}. \tag{5.13}$$

In this Equation, $\mathbf{y}_i$ refers to any point on the hyperplane and $\mathbf{n}$ is its normal vector. We perform a search among the facets to find the hull simplex which includes the projection $\tilde{\mathbf{y}}$. Then we regard those shapes as neighbours, which constitute the vertices of that simplex. Obviously, in such a case we get n neighbours instead of $n+1$ for $\mathbf{y}$ inside $\mathcal{CH}(\mathbf{Y})$. However,

such a projection is not guaranteed to exist. This circumstance is handled by using the $n-1$ nearest neighbours of $\mathbf{y}$ in $\mathbb{R}^n$.

5.4 Pre-Image Problem

Having performed operations in low dimensional feature space, we are interested in mapping back a point $\mathbf{y}$ to shape space. Formally, we search the pre-image $S = \Psi^{-1}(\mathbf{y})$ of $\mathbf{y}$ such that $\Psi(S) = \mathbf{y}$. In the context of PCA any shape can be recovered from its low dimensional counterpart by a linear combination of eigenmodes. However, when embedding data by non-linear kernel methods, more sophisticated approaches are required. As pointed out by Mika et al. [50], in such cases the pre-image problem is ill-posed and an unique solution might not exist. Nevertheless, it is common practice to revert to an approximation of S. One such approximation is given by the shape having minimal embedding distance to $\mathbf{y}$ as stated by Equation (5.14).

$$S = \Psi^{-1}(\mathbf{y}) = \underset{S}{\operatorname{argmin}} \left\| \Psi(S) - \mathbf{y} \right\|^2 \tag{5.14}$$

Likewise, maximisation of the collinearity criterion may be performed as was done by Dambreville et al. [25] and [50]. Kwok and Tsang [42] approach the pre-image problem by relating shape space distances to feature space distances as done by multi dimensional scaling. A comparison of various pre-image approaches is given by Arias et al. [1]. Note that a solution to the pre-image problem allows the projection of an arbitrary shape onto the manifold by $\pi_{\mathcal{M}}(S) = \Psi^{-1}(\Psi(S))$.

5.4.1 Pre-Image by Minimum Embedding Distance

In [31, 69, 70] Etyngier and Thorstenson et al. utilise the manifold assumption to compute the pre-image. Their approach is to interpolate shapes as weighted Karcher mean in the tangent space of the manifold. Recall that $\mathcal{M}$ is parametrised by a set of n-simplices. Given the point in question $\mathbf{y}$, we can easily determine the enclosing simplex $\Delta_{\mathbf{y}}$. The corresponding simplex Δ_S in input space has vertices given by the $(n+1)$ neighbouring shapes $S_i \in \mathcal{N}$ according to the Delaunay triangulation. The simplex Δ_S is employed to parametrise the tangent space of manifold $\mathcal{M}$ at shape S. In the vicinity of Δ_S, each shape is related to the vertices S_i by barycentric weights θ_i with $\sum_{i=1}^{n+1} \theta_i = 1$. The authors

compute the pre-image in Equation (5.14) while constraining S to a Karcher mean

$$S = \underset{S_\theta \in \mathcal{S}}{\operatorname{argmin}} \sum_{S_i \in \mathcal{N}} \theta_i \; d(S_\theta, S_i)^2. \tag{5.15}$$

To this end S is initialised with the simplex shape having the closest embedding to $\mathbf{y}$ and minimisation is performed by gradient descent. For the Euclidean input space a small variation of the barycentrics causes a shape deformation

$$\partial S = \sum_{i=1}^{n+1} \partial\theta_i S_i \tag{5.16}$$

Evaluation of $\|\Psi(S + \partial S) - \mathbf{y}\|^2$ produces a gradient flow drawing the shape towards minimum embedding distance to $\mathbf{y}$. In practice this scheme computes the pre-image in few iterations. The optimisation is performed in an orthogonalized tangent space as the barycentrics are not linearly independent. As shown in [68], this method outperforms the approaches by Arias [1] and Dumbreville [25] in the application of image de-noising. Instead of dividing the manifold by a fixed triangulation, in [69] the n-simplex defined by the nearest neighbours of $\mathbf{y}$ is used for pre-image computation.

5.4.2 Manifold Projection in Shape Space

Etyngier, Ségonne and Keriven [30] propose to compute the projection $\pi_\mathcal{M}(S)$ directly in shape space. Indeed, when provided a new shape e.g. from a segmentation process, the embedding step can be omitted. Having identified the tangent space at S, i.e. the simplex Δ_S, the objective is to find the manifold shape $S_\mathcal{M}$ that minimises the distance to S:

$$\pi_\mathcal{M}(S) = \underset{S_\mathcal{M}}{\operatorname{argmin}} \;\; d(S, S_\mathcal{M}). \tag{5.17}$$

As stated by the authors, above optimisation problem is intricate and involves time consuming second order shape gradients. Therefore an approximation coined the *snail algorithm* is proposed. The algorithm approaches $S_\mathcal{M}$ by alternated warps between two shapes in Δ_S. Etyngier et al. incorporate an adaptive step size that itself is subject to optimisation. However, we will use a slightly modified and thus simplified version that has shown to yield good approximations.

Provided a smooth and convex shape distance $d(\cdot, \cdot)$, we can formulate the procedure as described in Algorithm 1. See also Fig. 5.7 for an illustration of the concept. Note that

when dividing one warp in a fixed number of steps, the step size refines implicitly. In practice few iterations suffice for a good approximation.

Algorithm 1 Approximate the manifold projection $S_{\mathcal{M}}$ of shape S

Require: The neighbouring shapes $\{S_i\}_{i=1\ldots n+1}$ of S forming simplex Δ_S
Ensure: $d(S, S_{\mathcal{M}})$ is minimal
  Construct a cyclic list L of $\{S_i\}$:
    Insert the the shape being most similar to S into L
    Insert remaining shapes ordered by adjacency
  $S_{\mathcal{M}} \leftarrow L_0$
  $S_T \leftarrow L_{k=1}$
  while $\delta d(S, S_{\mathcal{M}}) > \epsilon$ **do**
    while $d(S, S_{\mathcal{M}})$ improves **do**
      Warp $S_{\mathcal{M}}$ towards S_T
    end while
    $S_T \leftarrow L_{k+1}$
  end while

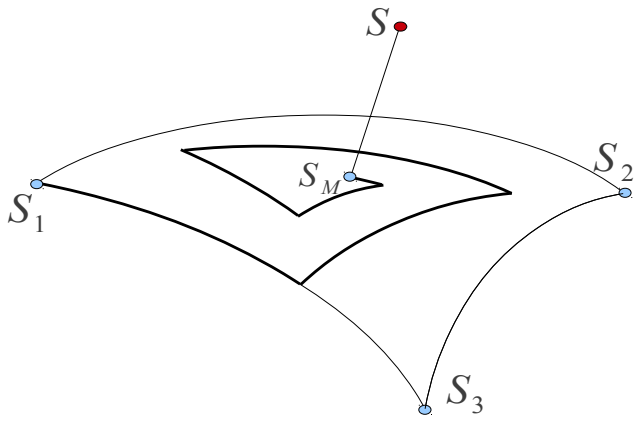


Figure 5.7: Illustration of the snail algorithm for $n = 2$. Shape S is projected to the manifold by alternated warps between simplex shapes S_i.

Chapter 6

Evaluation and Application to Image Segmentation

Contents

The manifold approach (DFM-SDF) discussed in Chapter 5 seems to have promising attributes for a shape model. To confirm our belief, in this chapter we will conduct evaluations with respect to

- **model generalisation**
  how well can the SSM reconstruct unseen shapes,
- **shape disturbance handling**
  how does the model respond to noisy or occluded shapes,
- **training pool size**
  how does the number of training samples influence the results and
- **sample density**
  what impact does the degree of similarity among training samples have.

We also provide comparison to the linear model by Leventon et al. hereafter termed PCA-SDF. Eventually we show how the DFM-SDF model can be incorporated to level set based image segmentation and show results on synthetic and medical data.

6.1 Manifold Explorer

In order to gain a better insight in various combinations of concepts discussed in this thesis, we designed the Manifold Explorer which provides a graphical interface to study manifold creation by Diffusion Maps, the Out-of-sample extension and Pre-image algorithms. The screenshot in Fig. 6.1 shows the 2D manifold of slices from the SLiver07 data-set along with an pre-image.

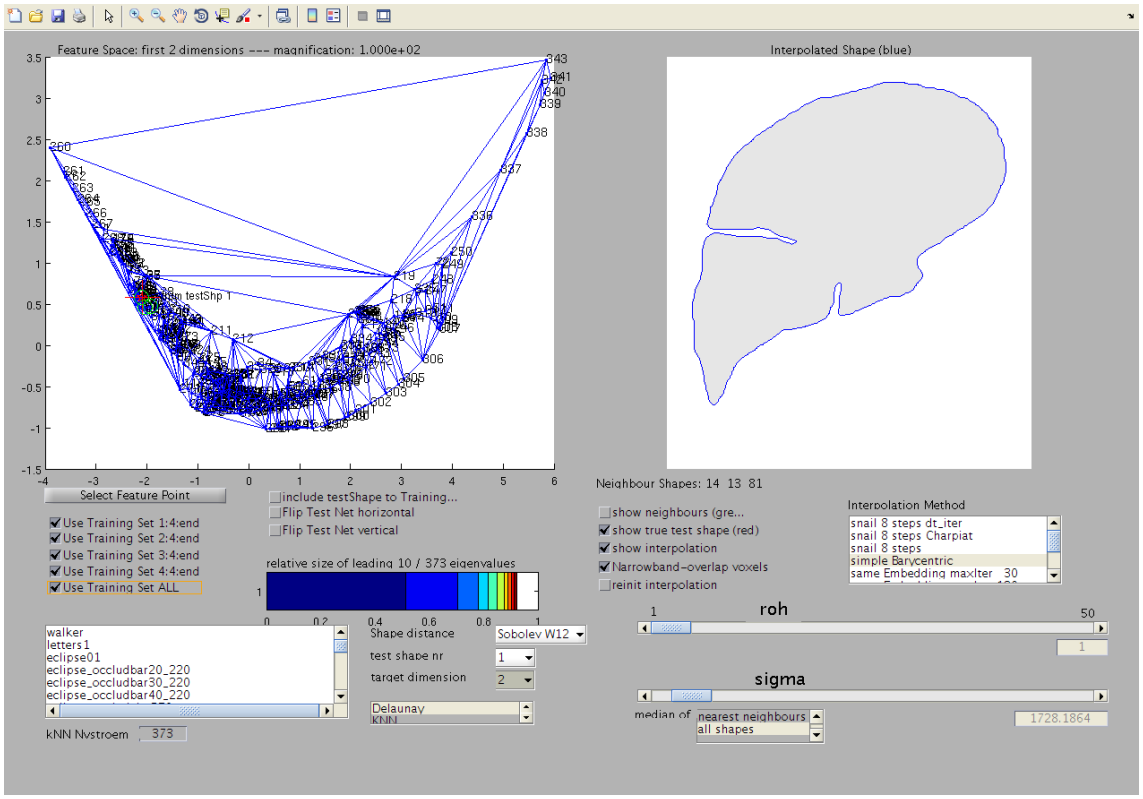


Figure 6.1: The Manifold Explorer was designed to facilitate the concepts in this work.

6.2 Manifold Model Evaluation

We now evaluate the DFM-SDF approach and observe the impact of altering parameters. Therefore a certain amount of shapes is learned followed by observing the model response to various test shapes. To measure the quality of the response we will use the normalised measure $d_{SD}^{\mathcal{C}}(S_{response}, S_{expected})$ defined in Section 2.3.2.

Since the FGNet data-set provides most shapes along with fine-grained sampling, we will use a specific test setup - the *FGNet-Test* - repeatedly with various modifications. The test successively expands the training pool size $q = \{10, 15, 20, ..., 95, 100\}$ by picking equally spaced frames from the data-set. If not mentioned otherwise, each 5^{th} frame is taken. For each q the model is queried for 25 unseen shapes and the reconstruction error $d_{SD}^{\mathcal{C}}$ is evaluated in terms of mean and variance. Recall that a value of $d_{SD}^{\mathcal{C}} = 1$ indicates a non-overlapping area that equals the length of the expected shape contour evaluated on the discrete pixel grid. Note that we use different test shapes for each q, such that test samples are equally spaced between training samples.

6.2.1 Manifold Projection and Tangent Space Policy

First off, we will utilise the FGNet-Test to compare the manifold projection operators *minimum embedding distance* (MED) from Section 5.4.1 and *projection in shape space* (PSS) from Section 5.4.2. Moreover we include different policies on manifold tangent space selection, i.e. how to pick the samples defining the simplices. In [31] a Delaunay triangulation was performed while [69] used the $n+1$ neighbourhood $\mathcal{N}_y$ in feature space. As a third option we add the $n+1$ neighbourhood in shape space $\mathcal{N}_S$. The obtained results for target dimensions $n = 3$ and $n = 4$ are depicted in Fig. 6.2. Obviously, PSS yields way more accurate shapes with less deviations than MED. Setting $n = 3$ we get rather inconsistent results for Delaunay and $\mathcal{N}_y$, see Fig. 6.2l. For non-uniform feature point distributions, Delaunay allows large simplices of training shapes being highly different (see e.g. Fig. 5.2). Thus poor interpolations are obtained in these manifold regions. Since we used a rather low feature space dimension, DFM embeds the shapes quasi-isometrically [31], causing $\mathcal{N}_y \neq \mathcal{N}_S$. Hence $\mathcal{N}_S$ provides better projections. Considering Fig. 6.2r, $n = 4$ is a better choice. Delaunay performs better due to the fact that most test shapes lie outside the manifold and are projected on its convex hull. Also for $\mathcal{N}_y$ the additional dimension contributed a major improvement - $\mathcal{N}_y$ now coincides better with $\mathcal{N}_S$. Overall $\mathcal{N}_S$ cooperates best with MED and PSS.

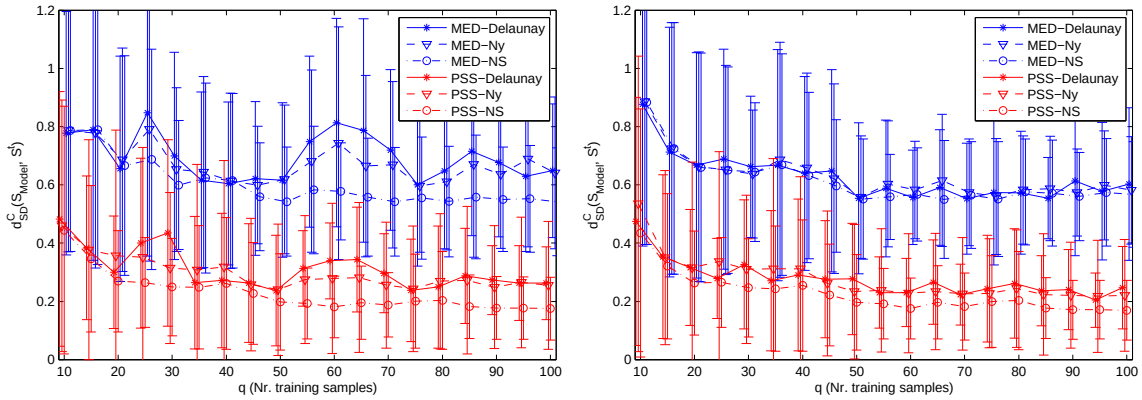


Figure 6.2: FGNet-Test of the projection operators MED and PSS in conjunction with simplex selection policies. Target dimensions $n = 3$ (l) and $n = 4$ (r).

6.2.2 On the number of Training Samples

Since any model is founded by a set of training shapes, the question of the sample pool size arises naturally. Yet there is no universal answer. Major factors regard the variations to be captured by the model. Variation amplitude as well as the number of variation modes contribute to the answer. A principled approach for a given data-set is to find a pool size q such that the reconstruction error falls below an acceptable threshold. Result diagrams of the FGNet-Test - e.g. refFigfig:fgnettestpreimgneigh - show the reconstruction improvement for up to 100 training samples. Apparently, for the variation encoded in the selected FGNet shapes, 50 samples suffice for accurate model responses. A larger pool size generates only marginal better results.

6.2.3 Model Interpretations

This section evaluates DFM-SDF model responses for our data-sets. In a leave-one-out fashion all but one samples are trained in turn and the model interpretation by PSS-$\mathcal{N}_S$ of the omitted shapes is observed. In Fig. 6.3 we show the best and worst responses for each data-set along with error distributions $d_{SD}^{\mathcal{C}}$. The eclipse results are very good with most errors below 0.1. The single poor response is caused by the test shape falling outside the rim of the 1D-manifold. Since the cup data has an sufficient amount of uniformly sampled shapes, responses are quite accurate. Contrary, the walker results are rather poor with most errors being located in the upper error zone. IMM faces score little better with errors heap at lower errors. Apparently the poor results for walker and IMM are caused by to

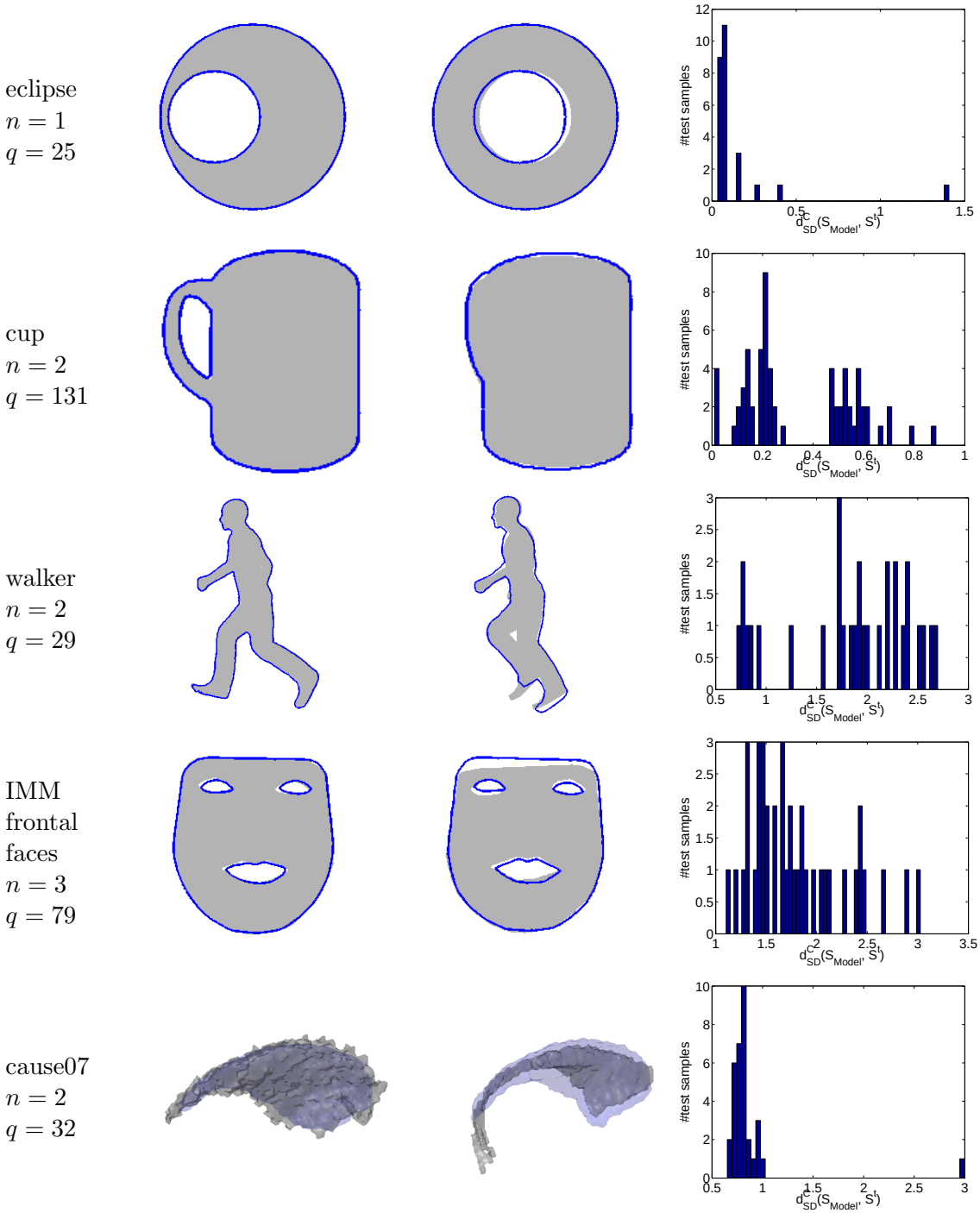


Figure 6.3: DFM-SDF model interpretations. Rows correspond to data-sets, solid shapes indicate the expected result, contours depict the model response. Best (l) and worst (m) responses along with error distributions $d^{\mathcal{C}}_{SD}$ (r).

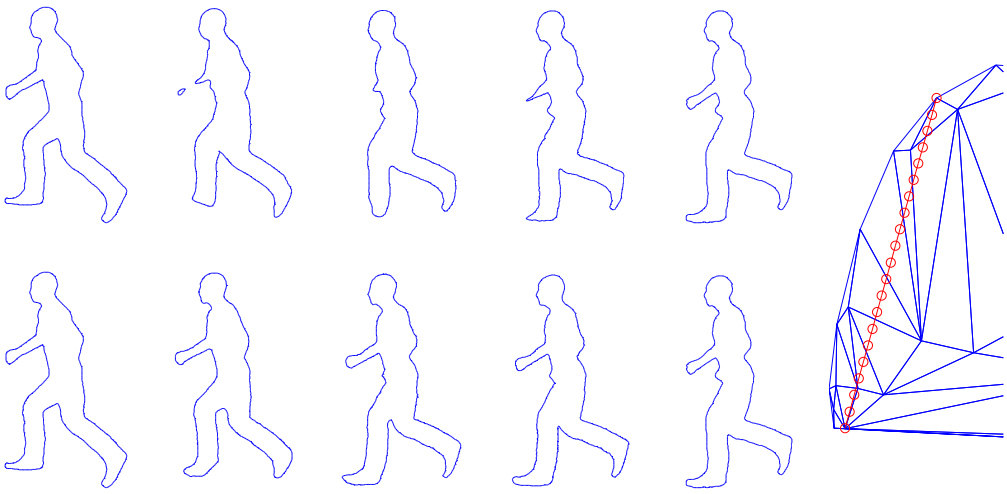


Table 6.1: Evolution of shape. Direct warp of left to rightmost shape (top). Warp by moving along the manifold (bottom). Evolving shape projected to feature space (r).

few training samples available to capture the heavy variations. The cause07 3D shapes perform well except for one distinct sample differing too much from the other ones.

6.2.4 Manifold Warp

Revisiting the defective warp of Section 5.2.2 caused by too different shapes, we now introduce manifold knowledge to the task. Figure 6.1 shows the warp between two shapes from the Runner set. The direct warp by Equation (5.8) (top row) deforms arm and feet in an implausible manner as they dissolve and reemerge. Incorporating the manifold, transition becomes naturally as the evolving shape bypasses intermediate samples (bottom row).

6.3 Comparison To Linear Models

This section provides comparison of DFM-SDF to linear PCA-SDF by Leventon et al. [46]. Since for PCA-SDF shape information is scattered through all principal modes, the eigenspectrum is not representative. Therefore we decided to keep 20% of modes. We also include PDM evaluation for data-sets available in control point representation. Here 95% of information as indicated by the eigenspectrum is retained.

6.3.1 Modes of Variation

To study the capability of manifold modes separating shape features, we construct a 3D manifold from 200 samples of the FGNet data-set. Therefore we perform orthogonal feature space warps through the center of the manifold by MED-$\mathcal{N}_y$. In Fig. 6.4 the manifold is pictured by a wire-frame with traces of the warps pictured as point interconnected lines. Shapes produced by the warps are presented in Fig. 6.2a-c. Apparently mode (a) controls head rotation from frontal to upper left. Mode (b) deflects narrow eyes. All three modes deflect mouth opening.
Linear mode (d) is similar to (a) and most mouth opening is captured by mode (e). Mode (f) deflects head shape alike to mode (c). PCA-SDF encodes the spatial location of variation by its eigen-SDFs. By contrast, distance based DFM-SDF does not distinguish whether a distance difference is caused be a face's eye or its chin. Hence feature separation is handled better by PCA-SDF.

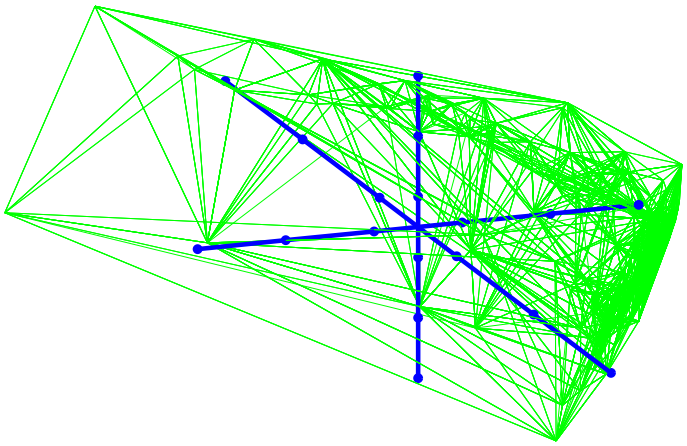


Figure 6.4: The FGNet 3D diffusion maps manifold pictured by a wire-frame along with traces of mode directions a, b and c.

6.3.2 Sample Density

As stated before, the sample density has a major impact on the reconstruction result for DFM-SDF. In other words, for a given training set size the samples should not vary too much. To demonstrate the influence of density we conduct the FGNet-Test with training samples picked from increasingly spaced sequence frames. Fig. 6.5 shows the results for spacings of 1 up to 40 frames for DFM-SDF ($n = 4$) and PCA-SDF ($n = 6$). As expected both methods achieve most accurate results for high density, both approaches

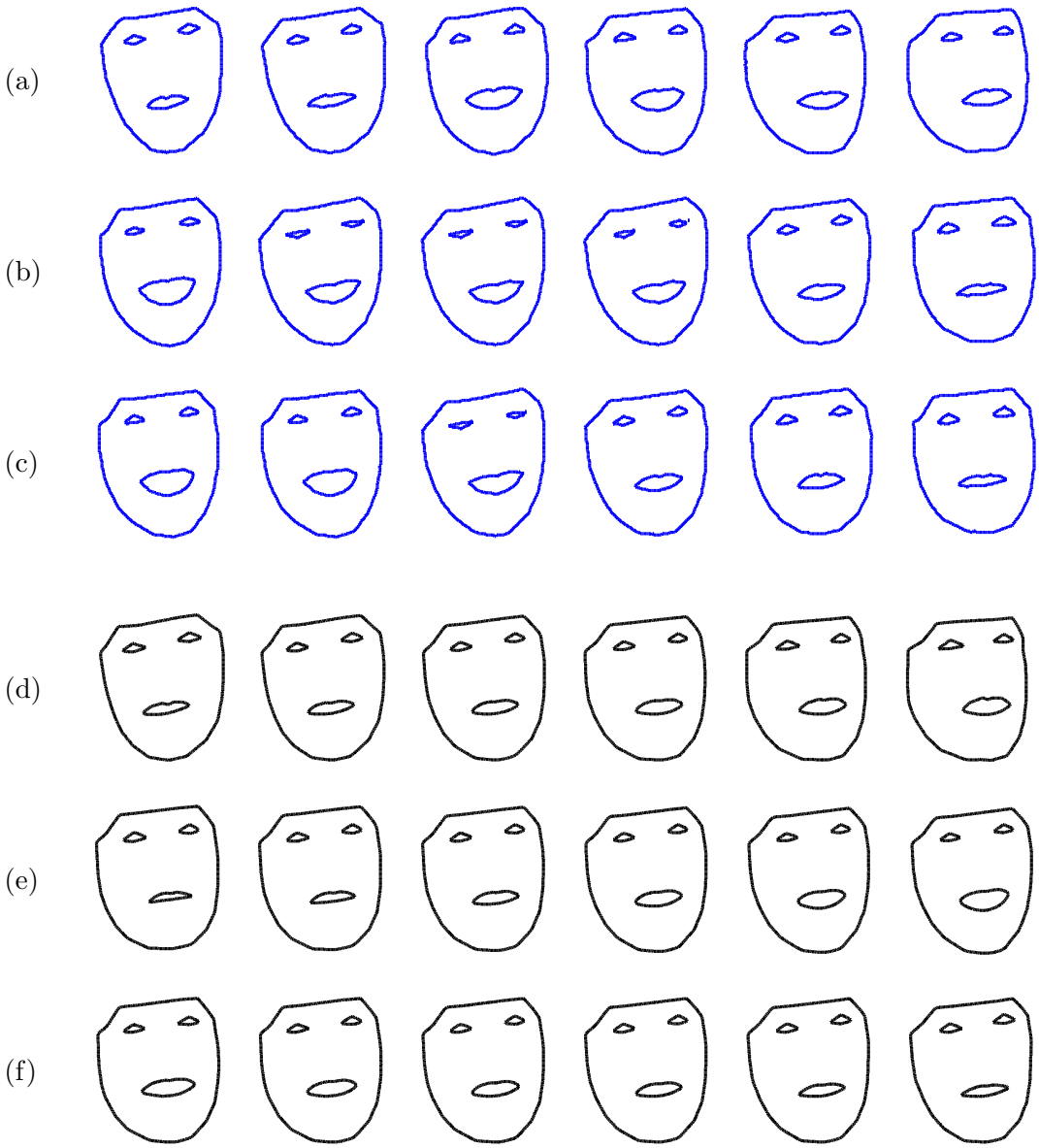


Table 6.2: Comparison of the modes of variation for the FGNet data-set embedded to $\mathbb{R}^3$. Rows (a), (b) and (c) are obtained by warping in orthogonal directions of the diffusion maps manifold, see also Fig. 6.4. For comparison, rows (d), (e) and (f) show the modes of the linear PCA model by deflecting the mean from $-2\sqrt{\lambda}$ to $2\sqrt{\lambda}$.

perform almost equally for $dt = 1$. For lower density ($dt \geq 5$ frames), PCA-SDF clearly outperforms DFM-SDF. The reason being that PCA-SDF takes into account variations of the whole training set, thus allowing better adjust to unseen features. Contrary, DFM-SDF considers only the most similar samples. See also Fig. 2.5 to interpret the reconstruction errors.

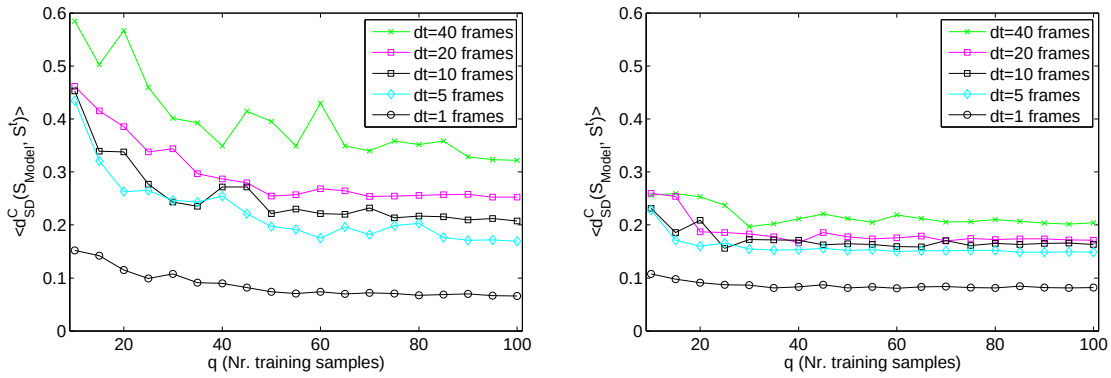
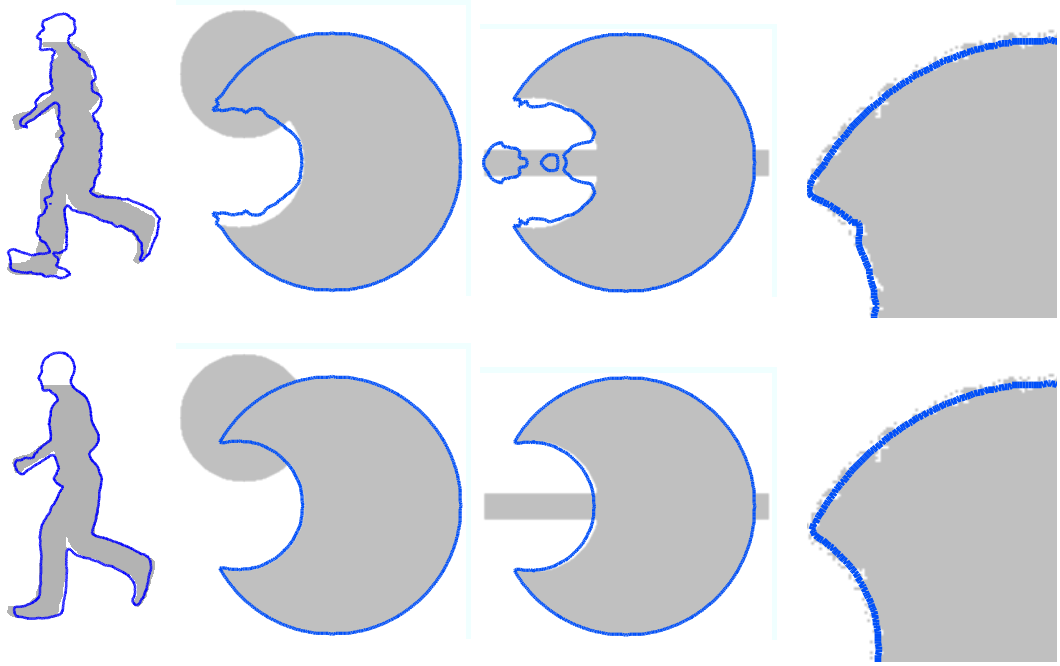


Figure 6.5: FGNet-Test for sample density variation by training increasingly spaced sequence frames. We plot the reconstruction error averaged over 25 test shapes for DFM-SDF (l) and PCA-SDF (r).

6.3.3 Coping with Shape Distortions

A question of great practical interest concerns the tolerance of a shape model regarding occlusion and noise. Hence we perform a test by confronting the model with an corrupted instance and observe the ability to recover the shape. We compare DFM-SDF to PCA-SDF using the walker and eclipse data-set. Artificial occlusion is introduced by removing the head from the walker shape and by an additional disk in the upper left corner / a horizontal bar for eclipse shapes. For the noisy sample we distort the eclipse boundary with 70% salt & pepper noise. See Fig. 6.6 for the queried shapes and obtained results. Evidently, PCA-SDF tries to imitate the corruptions by deflecting modes highly, thereby generating non-plausible shapes. In contrast, DFM-SDF returns accurate reconstructions even for substantial disturbances.

For a more quantitative analysis, we introduce occlusion to the FGNet-Test by removing the mouth sub-contour. The model should be able to reconstruct the shape including the mouth. Figure 6.7l-m exemplify reconstructions for PCA-SDF and DFM-SDF respectively. As can be seen, PCA-SDF avoids forming the mouth and has problems with the eye sub-contours. Note that PCA-SDF deflects modes up to 100 σ, causing heavy con-



(a) (b) (c) (d)

Figure 6.6: Occlusion and noise handling by PCA-SDF (top row) and DFM-SDF (bottom row). The corrupted shape is pictured solid while the contour represents the model response.

tour distortions which we corrected by restricting modes to $\pm 2\sigma$. In contrast, DFM-SDF yields a sound reconstruction. See Figure 6.7-l for the whole evaluation results. Having lower mean and variance in reconstruction errors, DFM-SDF clearly excels in the task of occlusion handling. From this experiments we can infer that the rigidity of DFM-SDF causing inferior results for low sample density in Section 6.3.2 proves beneficial for shape distortions.

6.3.4 Projection Timings

Run times for our MATLAB implementations of pre-image algorithms on an Intel® C2D at 3GHz are presented in Table 6.3. Various parameter combinations are considered: SDF domain sizes, dissimilarity norm, target dimension n and number of snail iterations k. Timings do not include SDF restoration. PSS computes the projections very fast for $k = 6$ and the L_2-norm with timings below those of of PCA. MED is considerably slower since the embedding has to be computed within the gradient descent. Thus the cost for MED increases with sample pool size. The impact of higher target dimension n seems to

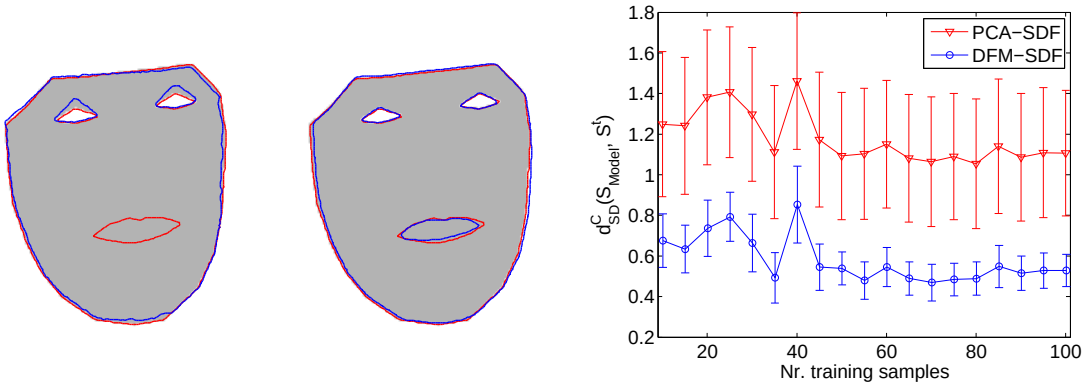


Figure 6.7: FGNet-Test for shape occlusion by removing the mouth. Examples of model responses for mode restricted PCA-SDF (l) and DFM-SDF (m), overall reconstruction errors (r). Solid shape represents the test shape presented to the models, the red contour is the expected reconstruction and the blue contour is the model response.

			128×128		256×256		512×512	
	q	n	L^2	$W^{1,2}$	L^2	$W^{1,2}$	L^2	$W^{1,2}$
PCA	60	30	5		24		148	
		59	12		54		279	
PSS$(k=6)$	60	2	3	15	10	67	222	2280
		3	12	66	38	120	231	2330
MED	30	2	333	1829	1520	5080	5007	29860
		3	718	2132	1879	5830	6592	43000
MED	60	2	710	3885	2743	8180	9470	62900
		3	1080	3726	2640	10050	12750	95700

Table 6.3: Timings in (ms) for pre-image algorithms with various domain sizes and parameters.

decrease for higher domain sizes. Overall, computing the Sobolev $W_{1,2}$ is more expensive since two SDF gradients are required.

6.4 Image Segmentation

Following Etyngier et al.[31], we now demonstrate how the DFM-SDF model can be integrated to the task of level set based image segmentation and show the results.

6.4.1 Geodesic Active Contours

A popular approach for segmentation is the Geodesic Active Contour (GAC) energy proposed by Caselles et al. [8]. In its general formulation, the GAC energy is given by

$$E_{GAC} = \int_0^{|\mathcal{C}|} g(|\nabla I(\mathcal{C})|)ds. \tag{6.1}$$

Here I is the input image and g is a monotonic decreasing function mapping to $(0, 1]$. Hence E_{GAC} weights the Euclidean contour elements ds depending on the covered image gradient. As shown in [8], minimisation of the energy can be achieved by deriving the associated Euler-Lagrange equation. Formulated in the level set framework, the evolution writes

$$\frac{\partial \phi}{\partial t} = g|\nabla \phi|\kappa + \nabla g \cdot \nabla \phi. \tag{6.2}$$

The first term on the right hand side of above equation imposes a smoothness constraint, driving the contour towards a minimal surface. The second term takes into account the input image, slowing down or stopping the evolution at high gradients. An additional normal flow term $\alpha g|\nabla \phi|$ can be incorporated to bypass low gradient regions and increase convergence speed.

6.4.2 Adding the Shape Model

Since the GAC is stated as an energy minimisation, it is a natural idea to add an energy E_{Prior} accounting for the shape prior knowledge. An intuitive approach is to define a force that attracts the evolving SDF ϕ_t towards the shape manifold $\mathcal{M}$ [25, 31]. This amounts to minimise the distance to its projection $\pi_{\mathcal{M}}(\phi_t)$ on the manifold . Another aspect to address is the global transformation between evolving shape ϕ_t and the manifold shapes. Therefore we introduce a similarity transform $T(\phi)$ mapping ϕ_t to the prior shape space. The resulting energy is then given by Equation (6.3) leading to the overall segmentation energy Equation (6.4) where γ is a weighting factor.

$$E_{Prior} = d(T(\phi_t), \pi_{\mathcal{M}}(T(\phi_t)))^2 \tag{6.3}$$

$$E_{Segment} = E_{GAC} + \gamma E_{Prior} \tag{6.4}$$

In practice we alternate GAC and prior iterations with intermediate pose transform updates. To increase convergence speed, we employ an adaptive weight γ and increase the

prior influence when the GAC energy becomes static. In most of the following experiments an initial value of $\gamma = 0.3$ was used.

6.4.3 Results

First we apply the pure GAC energy to a binary image from the walker data-set occluded by vertical bars. As seen in Fig. 6.8-top, the GAC driven shape sticks to edges regardless of the disturbances. However, when augmenting the energy with SSM information, the process overcomes the obstacles and yields a sound segmentation of the walker - see Fig. 6.8-bottom. In another experiment we test the ability to handle topological changes.

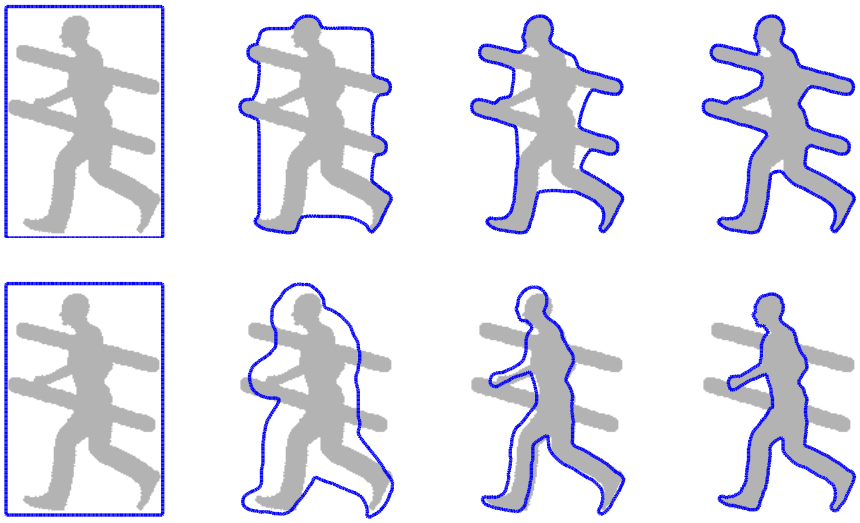
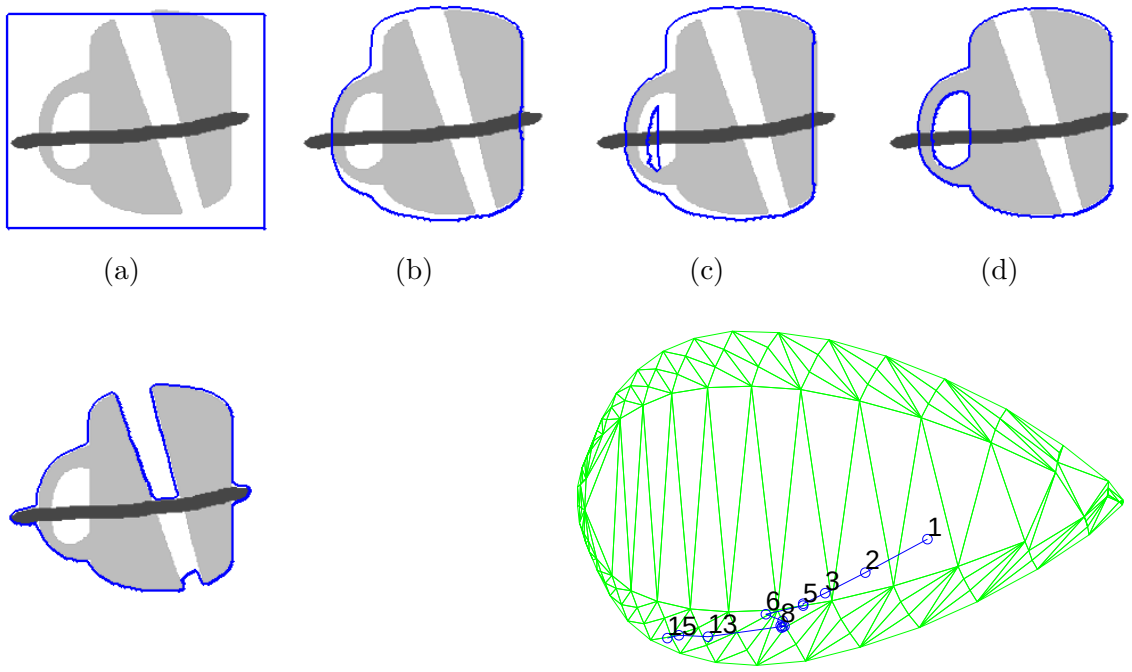


Figure 6.8: Segmentation of an occluded binary image. Shape evolution is from left to right. Pure GAC energy (top). Evolution including the shape manifold (bottom).

Therefore we use a cup shape with distinct handle cavity and add severe distortions. The evolution of the segmenting shape is shown in Fig. 6.9. As expected, the pure GAC energy sticks to the outer cup boundary with the given initialisation. Incorporating the prior term allows to bypass this minimum by gradually developing a cavity until the final result is obtained. Note that the cup data-set could not be as easily modelled and segmented with a point distribution model. Figure 6.9f shows the corresponding shape manifold for $n = 3$ along with projections $\mathbf{y}_t$ of the evolving shape.

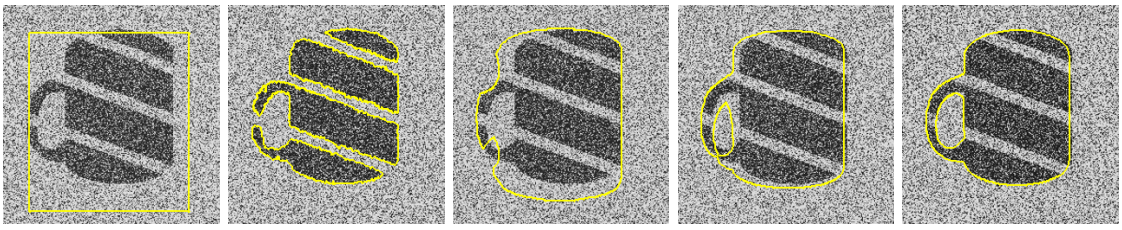
The DFM-SDF model can easily be integrated in other segmentation energies. For example, when facing severe noise distortion, the Chan-Vese energy [9] is a preferable choice.



(a) (b) (c) (d) (e) (f)

Figure 6.9: GAC prior segmentation: initial shape (a), DFM-SDF model guided segmentation for iterations 10, 12 and 15 (b-d), result without prior information (e), projection of the evolving shape on the cup manifold (f).

Replacing E_{GAC} with the Chan-Vese energy in Equation (6.4), we obtain the result shown in Fig. 6.10. As can be seen, the pure Chan-Vese energy en-wraps homogeneous regions but can not account for occlusions. In contrast, the prior augmented energy yiels a sound cup-shaped result while changing topology.

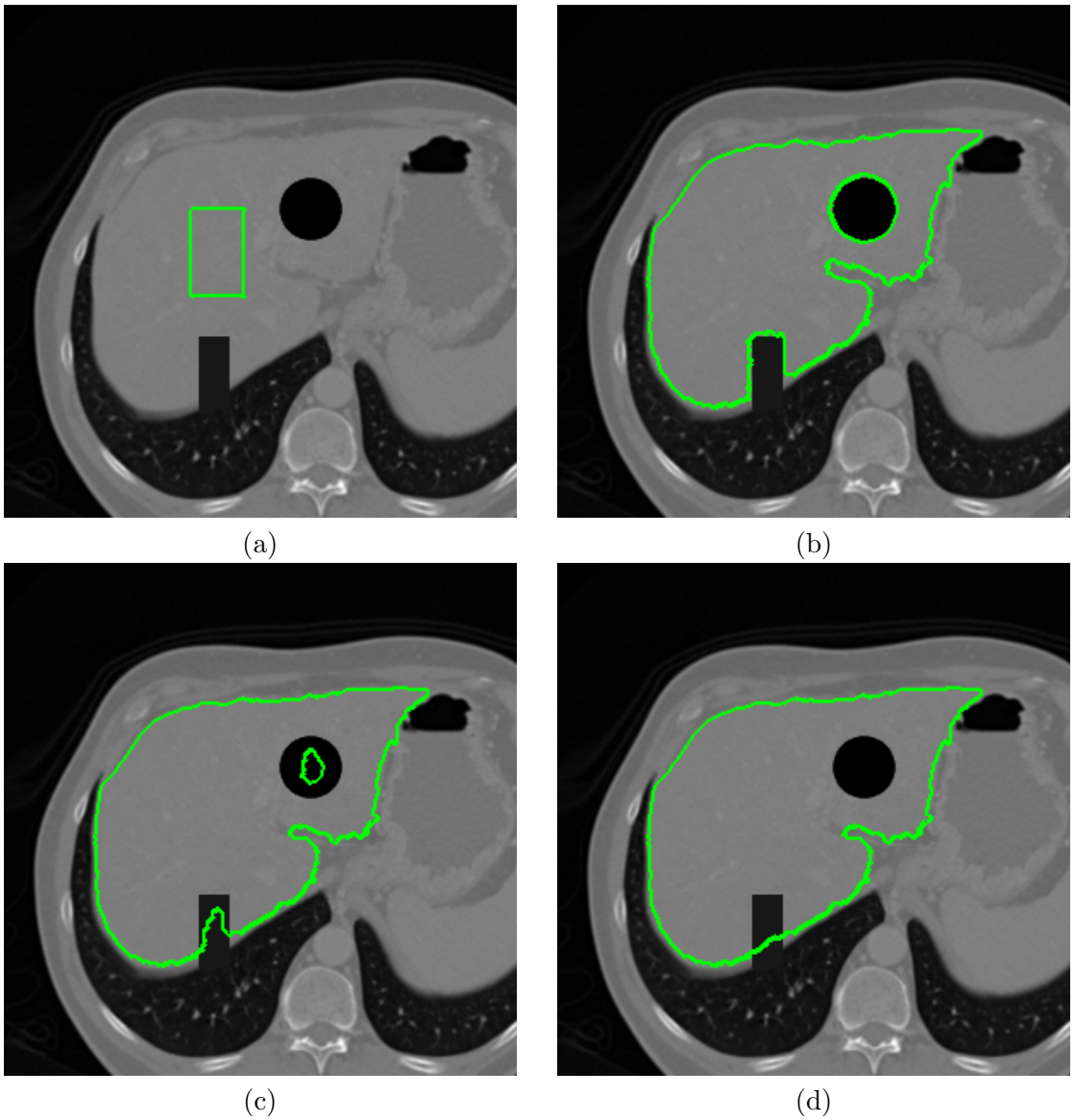


(a) (b) (c) (d) (e)

Figure 6.10: Chan-Vese segmentation: (a) initial contour, (b) result of pure Chan-Vese energy, (c) - (e) prior enhanced segmentation process at iterations 3, 6 and 8.

We now introduce the manifold shape prior to medical image segmentation. Therefore a total of 300 transversal slices of the SLiver 3D data-set [34] is trained to a model with $n = 3$.

As artificial occlusion is imposed to the image, the pure GAC energy can not segment the shape correctly Fig. 6.11b. However, projecting the GAC shape to the manifold, a sound result is obtained - see Fig. 6.11d. Note that to overcome the low tissue contrast, histogram equalisation is performed on the image.



(a) (b)

(c) (d)

Figure 6.11: Liver segmentation: (a) input image and initial contour, (b) result with pure GAC energy. Introducing the shape prior: (c) iteration 5, (d) iteration 15.

Chapter 7

Conclusion and Outlook

Contents

7.1 Conclusion

In this work we investigated an approach for global non-linear SSMs combined with the implicit shape representation by SDFs. SDFs avoid the need for correspondence and advantageously handle topological changes. The lack of correspondence may result in flawed shape warps at non-overlapping regions. However, this shortcoming is compensated by manifold knowledge so to achieve natural transitions between strongly differing shapes. Although complex in nature, the DFM-SDF model has proven to give sound shape interpretations at reasonable computational cost.

As we have seen, the DFM-SDF technique does not generalise as precisely to unseen shapes as PCA-SDF does. The reason being is the rigid global nature considering the most similar shapes only. In this respect the PCA-SDF approach is more flexible as it may deflect some modes - being unintended for this purpose - to compensate for one unseen feature while other modes serve other features. On the other hand, DFM-SDF benefits from this very same rigidity in that disturbances such as noise and occlusion are tackled robustly without attempt of false imitation. We saw that the distances involved in DFM-SDF can not encode the location of variability. By contrast, PCA-SDF uses eigen-SDFs resulting in better feature separation.

In terms of model compactness, PCA-SDF might discard components up to required precision. Alike DFM-SDF could exclude shapes being sampled near by.
DFM-SDF applies very well to image segmentation as inaccurate model responses are caught by low level image information while even severe image distortions are coped by the the model.

7.2 Where to go from here

Several further evaluations and applications are worthy of consideration. The shape manifold might be extended for multiple highly differing objects, being of interest for object classification or prior based segmentation not tied to a specific object. Thus applications such as activity recognition - e.g. hand gesture- or head pose estimation, shape retrieval and animations come to mind. A rather challenging issue to address is better feature separation. While distinct sub-contours such as the faces mouth could be handled independently, separating sub-parts of a closed contour is not straight forward for SDFs. A hybrid DFM-PCA model might be designed with PCA applied to clusters found by DFM, so to combine the strengths of both.
In order to tackle strong global transform deviations, discrete transforms could be incorporated into the manifold. Another possibility is an adaptive learner being able to accumulate shape knowledge on-the-fly. In a related manner, handling very large datasets could be approached by a hirarchical manifold that subjoins finer levels of detail on demand.
Also worthwhile of investigation is model aided 3D reconstruction from multiple views. Since the manifold framework is generic in nature, i.e. not bound to SDF shapes, other shape representations might be incorporated. Moreover the approach could be applied to appearance manifolds in the spirit of Cootes et al. [16].
For feature extraction we investigated in diffusion maps with a Gaussian kernel exclusively. Hence an evaluation of other dimensionality reduction methods with possible data dependent kernels would be interesting.

Bibliography

[1] Arias, P., Randall, G., and Sapiro, G. (2007). Connecting the out-of-sample and pre-image problems in kernel methods. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 0:1–8.

[2] Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6):1373–1396.

[3] Belongie, S. and Malik, J. (2000). Matching with shape contexts. In *Content-based Access of Image and Video Libraries (CBAIVL)*, pages 20–26, Hilton Head, SC.

[4] Bengio, Y., Paiement, J.-F., Vincent, P., Delalleau, O., Roux, N. L., and Ouimet, M. (2004). Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. In *In Advances in Neural Information Processing Systems*, pages 177–184. MIT Press.

[5] Besbes, A. (2010). *Image Segmentation using MRFs and Statistical Shape Modeling*. PhD thesis, Ecole Centrale Paris.

[6] Bookstein, F. L. (1996). Landmark methods for forms without landmarks: Localizing group differences in outline shape. In *Proceedings of the 1996 Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '96)*, MMBIA '96, pages 279–, Washington, DC, USA. IEEE Computer Society.

[7] Bresson, X., Vandergheynst, P., and Thiran, J. (2006). A variational model for object segmentation using boundary Information and Shape Prior Driven by the Mumford-Shah Functional. *International Journal of Computer Vision*, 28(2):145 – 162.

[8] Caselles, V., Kimmel, R., and Sapiro, G. (1997). Geodesic active contours. *International Journal of Computer Vision*, 22:61–79.

[9] Chan, T. F. and Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277.

[10] Charpiat, G., Faugeras, O., and Keriven, R. (2003). Shape metrics, warping and statistics. In *International Conference on Image Processing*, volume 2, pages II–627–630.

[11] Charpiat, G., Faugeras, O., and Keriven, R. (2005). Approximations of shape metrics and application to shape warping and empirical shape statistics. *Foundations of Computational Mathematics*, 5(1):1–58.

[12] Charpiat, G., Keriven, R., and Faugeras, O. (2007). Shape statistics for image segmentation with priors. In *Conference on Computer Vison and Pattern Recognition*.

[13] Chen, S., Charpiat, G., and Radke, R. J. (2010). Converting level set gradients to shape gradients. In *11th European Conference on Computer Vision, ECCV 2010*, Hersonissos, Greece. Springer.

[14] Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Warner, F., and Zucker, S. (2005). Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. In *Proceedings of the National Academy of Sciences*, pages 7426–7431.

[15] Cootes, T. F., Cooper, D. H., Taylor, C. J., and Graham, J. (1992). Trainable method of parametric shape description. *Image Vision Comput.*, 10(5):289–294.

[16] Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23:681–685.

[17] Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models-their training and application. *Computer Vision and Image Understanding*, 61:38–59.

[18] COX, T. and COX, M. (2000). *MULTIDIMENSIONAL SCALING*. Monographs on statistics and applied probability. Chapman & Hall/CRC.

[19] Cremers, D. (2006). Dynamical statistical shape priors for level set-based tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28:1262–1273.

[20] Cremers, D., Kohlberger, T., and Schnörr, C. (2002). Nonlinear shape statistics in mumford-shah based segmentation. In *In European Conference on Computer Vision*, pages 93–108. Springer.

[21] Cremers, D., Osher, S. J., and Soatto, S. (2004). Kernel density estimation and intrinsic alignment for knowledge-driven segmentation: Teaching level sets to walk. In *International Journal of Computer Vision*, pages 36–44. Springer.

[22] Cremers, D., Rousson, M., and Deriche, R. (2007). A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *International Journal of Computer Vision*, 72(2):195–215.

[23] Cremers, D., Schmidt, F. R., and Barthel, F. (2008). Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, Alaska.

[24] Cremers, D., Schnörr, C., and Weickert, J. (2001). Diffusion–snakes: Combining statistical shape knowledge and image information in a variational framework. In *IEEE First Workshop on Variational and Level Set Methods in Computer Vision*, pages 237–244, Vancouver, Canada. IEEE Comp. Soc.

[25] Dambreville, S., Rathi, Y., and Tannen, A. (2006). Shape-based approach to robust image segmentation using kernel pca. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:977–984.

[26] Delingette, H. and Montagnat, J. (2000). New algorithms for controlling active contours shape and topology. In *European Conference on Computer Vision (ECCV'2000), NUMBER 1843 IN LNCS*, pages 381–395. Springer.

[27] Dryden, I. and Mardia, K. (1998). *Statistical shape analysis*. Wiley series in probability and statistics. Wiley, Chichester [u.a.].

[28] Durrleman, S. (2010). *Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution*. Thèse de sciences (phd thesis), Université de Nice-Sophia Antipolis.

[29] Durrleman, S., Pennec, X., Trouvé, A., and Ayache, N. (2008). Sparse approximation of currents for statistics on curves and surfaces. In Metaxas, D., Axel, L., Székely, G., and Fichtinger, G., editors, *Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), Part II*, volume 5242 of *LNCS*, pages 390–398, New-York, USA. Springer. PMID: 18982629.

[30] Etyngier, P. and Keriven, R. (2007). Towards segmentation based on a shape prior manifold. In *In 1st International Conference on Scale Space and Variational Methods in Computer Vision*, pages 895–906.

[31] Etyngier, P., Ségonne, F., and Keriven, R. (2007). Shape priors using manifold learning techniques. pages 1–8.

[32] FGnet (2000). IST-2000-26434 Face and Gesture Recognition Working Group.

[33] Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):285–339.

[34] Heimann, T., van Ginneken, B., Styner, M., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., Bello, F., Binnig, G., Bischof, H., Bornik, A., Cashman, P., Chi, Y., Cordova, A., Dawant, B., Fidrich, M., Furst, J., Furukawa, D., Grenacher, L., Hornegger, J., Kainmueller, D., Kitney, R., Kobatake, H., Lamecker, H., Lange, T., Lee, J., Lennon, B., Li, R., Li, S., Meinzer, H.-P., Nemeth, G., Raicu, D., Rau, A.-M., van Rikxoort, E., Rousson, M., Rusko, L., Saddi, K., Schmidt, G., Seghers, D., Shimizu, A., Slagmolen, P., Sorantin, E., Soza, G., Susomboon, R., Waite, J., Wimmer, A., and Wolf, I. (2009). Comparison and evaluation of methods for liver segmentation from ct datasets. *IEEE Transactions on Medical Imaging*, 28(8):1251–1265.

[35] Hong, B., Prados, E., Soatto, S., and Vese, L. (2006). Shape representation based on integral kernels: Application to image matching and segmentation. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 833–840.

[36] Huang, X., Paragios, N., and D.N., M. (2006). Shape registration in implicit spaces using information theory and free form deformations. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 1303–1318.

[37] Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331.

[38] Kendall, D. G. (1977). The diffusion of shape.

[39] Klein, S., Staring, M., Murphy, K., Viergever, M., and Pluim, J. (2010). elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205.

[40] Koch, K., McLean, J., Segev, R., Freed, M. A., II, M. J. B., Balasubramanian, V., and Sterling, P. (2006). How much the eye tells the brain. *Current Biology*, 16(14):1428 – 1434.

[41] Kuhl, F. P. and Giardina, C. R. (1982). Elliptic fourier features of a closed contour. *Graphical Models /graphical Models and Image Processing /computer Vision, Graphics, and Image Processing*, 18:236–258.

[42] Kwok, J. T. Y. and Tsang, I. W. H. (2004). The pre-image problem in kernel methods. *Neural Networks, IEEE Transactions on*, 15(6):1517–1525.

[43] Lafon, S., Keller, Y., and Coifman, R. R. (2006). Data fusion and multicue data matching by diffusion maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:1784–1797.

[44] Langs, G. and Paragios, N. (2008). Modeling the structure of multivariate manifolds: Shape maps. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8.

[45] Lawrence, N. (2005). Probabilistic non-linear principal component analysis with gaussian process latent variable models. *J. Mach. Learn. Res.*, 6:1783–1816.

[46] Leventon, M. E., Grimson, W. E. L., and Faugeras, O. D. (2000). Statistical shape influence in geodesic active contours. In *CVPR'00*, pages 1316–1323.

[47] Li, C., Xu, C., Gui, C., and Fox, M. D. (2010). Distance regularized level set evolution and its application to image segmentation. *IEEE Transactions on Image Processing*, 19(12):3243–3254.

[48] Maizels, D. (1994). Data for brain volumes by caroll s. *Nature*, 422:849–857.

[49] Maurer, C. R., Qi, R., and Raghavan, V. (2003). A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):265–270.

[50] Mika, S., Schölkopf, B., Smola, A., Müller, K. R., Scholz, M., and Rätsch, G. (1999). Kernel PCA and de-noising in feature spaces. In *Advances in Neural Information Processing Systems 11*, pages 536–542. MIT Press.

[51] Nordstrøm, M. M., M., L., J., S., and Stegmann, M. B. (2004). The imm face database - an annotated dataset of 240 face images. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby.

[52] Osher, S. and Sethian, J. A. (1988). Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79(1):12–49.

[53] Paragios, N., M., R., and V., R. (2002). Matching distance functions: A shape-to-area variational approach for global-to-local registration. In *ECCV (2)*, pages 775–789.

[54] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572.

[55] Pohl, K. M., Fisher, J., Shenton, M., Mccarley, R. W., Eric, W., Grimson, L., Kikinis, R., and Wells, W. M. (2006). Logarithm odds maps for shape representation. In *In Proc. MICCAI, volume II*, pages 955–963.

[56] Prisacariu, V. A. and Reid, I. (2011). Nonlinear shape manifolds as shape priors in level set segmentation and tracking. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*.

[57] Rajpoot, N., Arif, M., and Bhalerao, A. (2007). Unsupervised learning of shape manifolds. In *In Proceedings of the British Machine Vision Conference*.

[58] Rousson, M. and Paragios, N. (2002). Shape priors for level set representations. In *ECCV (2)*, pages 78–92.

[59] Rousson, M. and Paragios, N. (2008). Prior knowledge, level set representations & visual grouping. *Int. J. Comput. Vision*, pages 231–243.

[60] Rousson, M., Paragios, N., and Deriche, R. (2003). Active shape models from a level set perspective. Technical Report RR-4984, INRIA.

[61] Roweis, S. and Saul, L. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326.

[62] Sammon, J. W. (1969). A nonlinear mapping for data structure analysis. *IEEE Trans. Comput.*, 18:401–409.

[63] Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computing*, 10:1299–1319.

[64] Staib, L. and Duncan, J. (1992). Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:1061–1075.

[65] Sundaramoorthi, G., Yezzi, A. J., Mennucci, A., and Sapiro, G. (2009). New possibilities with sobolev active contours. *International Journal of Computer Vision*, 84(2):113–129.

[66] Sussman, M., Smereka, P., and Osher, S. (1994). A level set approach for computing solutions to incompressible two-phase flow. *J. Comput. Phys.*, 114:146–159.

[67] Tenenbaum, J. B., Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323.

[68] Thorstensen, N. (2010). *Manifold learning and applications to shape and image processing*. PhD thesis, Ècole Nationale des Ponts et Chausseès.

[69] Thorstensen, N., Etyngier, P., Ségonne, F., and Keriven, R. (2011). Diffusion maps as a framework for shape modeling. *Comput. Vis. Image Underst.*, 115:520–530.

[70] Thorstensen, N., Segonne, F., and Keriven, R. (2009). Pre-image as karcher mean using diffusion maps:application to shape and image denoising. In *Proc. Second International Conference on Scale Space Methods and Variational Methods in Computer Vision (SSVM) 2009*, LNCS. Springer Verlag. Voss, Norway.

[71] Tsai, A., Yezzi, A., Wells III, W. M., Tempany, C. M., Tucker, D., Fan, A., Grimson, W. E. L., and Willsky, A. (2003). A shape-based approach to the segmentation of medical imagery using level sets. 22(2):137–154.

[72] Tsai, A., Yezzi, J. A., Wells III, W., Tempany, C., Tucker, D., Fan, A., Grimson, W. E., and Willsky, A. (2001). Model-based curve evolution technique for image segmentation. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:463.

[73] Vaillant, M. and GlaunÃ¨s, J. (2005). Surface matching via currents. In *Proceedings of Information Processing in Medical Imaging (IPMI 2005), number 3565 in Lecture Notes in Computer Science*, pages 381–392.

[74] van der Maaten, L., Postma, E., and van den Herik H.J. (2008). Dimensionality reduction: A comparative review.

[75] van Ginneken, B., Heimann, T., and Styner, M. (2007). 3d segmentation in the clinic: A grand challenge. In *3D Segmentation in the Clinic: A Grand Challenge*, pages 7–15.

[76] Weinberger, K. Q., Sha, F., and Saul, L. K. (2004). Learning a kernel matrix for nonlinear dimensionality reduction. In *In Proceedings of the Twenty First International Conference on Machine Learning (ICML-04*, pages 839–846. ACM Press.

Deutsche Fassung:
Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008
Genehmigung des Senates am 1.12.2008

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am ……. ……..
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

……. ……..
date (signature)