Image Segmentation with Shape Priors:
Explicit versus Implicit Representations

Daniel Cremers

Department of Computer Science

TU München, Germany

July 31, 2010

1 Introduction

1.1 Image Analysis and Prior Knowledge

Image segmentation is among the most studied problems in image understanding and computer vision. The goal of image segmentation is to partition the image plane into a set of meaningful regions. Here *meaningful* typically refers to a semantic partitioning where the computed regions correspond to individual objects in the observed scene. Unfortunately, generic purely low-level segmentation algorithms often do not provide the desired segmentation results, because the traditional low level assumptions like intensity or texture homogeneity and strong edge contrast are not sufficient to separate objects in a scene.

To overcome these limitations researchers have proposed to impose prior knowledge into low-level segmentation methods. In the following, we will review methods which allow to impose knowledge about the *shape* of objects of interest into segmentation processes.

In the literature there exist various definitions of the term *shape*, from the very broad notion of shape of Kendall [54] and Bookstein [5] where shape is whatever remains of an object when similiarity transformations are factored out (i.e. a geometrically normalized version of a gray value image) to more specific notions of shape referring to the geometric

outline of an object in 2D or 3D. In this work, we will adopt the latter view and refer to an object's silhouette or boundary as its shape. Intentionally we will leave the exact mathematical definition until later, as different representations of geometry actually imply different definitions of the term *shape*.

One can distinguish various kinds of shape knowledge:

- low-level shape priors which typically simply favor shorter boundary length, i.e. curves with shorter boundary have lower shape energy. where boundary length can be measured in various ways [53, 4, 69, 49, 6],

- mid-level shape priors which favor for example thin and elongated structures, thereby facilitating the segmentation of roads in satellite imagery or of blood vessles in medical imagery [70, 78, 44], and

- high-level shape priors which favor similarity to previously observed shapes, such as hand shapes [50, 21, 36], silhouettes of humans [29, 26] or medical organs like the heart, the prostate, the lungs or the cerebellum [83, 99, 81, 58].

There exists a wealth of works on shape priors for image segmentation. It is beyond the scope of this article to provide a complete overview of existing work. Instead we will present a range of representative works – with many of the examples taken from the author's own work – discuss their advantages and shortcomings. Some of these works are formulated in a probabilistic setting where the challenge is to infer the most likely shape given an image and a set of training shapes. Typically the segmentation is formulated as an optimization problem.

One can distinguish two important challenges:

1. The modeling challenge: How do we formalize distances between shapes? What probability distributions do we impose? What energies should we minimize?

2. The algorithmic challenge: How do we minimize the arising cost function? Are the computed solutions globally optimal? If they are not globally optimal how sensitive are solutions with respect to the initialization?

1.2 Explicit versus Implicit Shape Representation

	Spatially continuous	Spatially discrete	
Explicit	polygons [21, 102], splines [53, 3, 36]	edgel labeling & dyn. progr. [84, 74, 1, 87, 89]	hybrid repres. & LP relaxation [90]
Implicit	level set methods [39, 72], convex relaxation [11, 31]	graph cut methods [49, 6]	

Figure 1: Shapes can be represented explicitly or implicitly, in a spatially continuous or a spatially discrete setting. More recently, researchers have adopted hybrid representations [90], where objects are represented both in terms of their interior (implicitly) and in terms of their boundary (explicitly).

A central question in the modeling of shape similarity is that of how to represent a shape. Typically one can distinguish between *explicit* and *implicit* representations. In the former case, the boundary of the shape is represented explicitly – in a spatially continuous setting this could be a polygon or a spline curve. In a spatially discrete setting this could be a set of edgles (edge elements) forming a regular grid. Alternatively, shapes can be represented implicitly in the sense that one labels all points in space as being part of the interior or the exterior of the object. In the spatially continuous setting the optimization of such implicit shape representations is solved by means of partial differential equations. Among the most popular representatives are the level set method [39, 72] or alternative convex relaxation techniques [11]. In the spatially discrete setting implicit representations have become popular through the graph cut methods [49, 7]. More recently, researchers have also advocated hybrid representations where objects are represented both explicitly and implicitly [90]. Table 1 provides an overview of a few representative works on image segmentation based on explicit and implicit representations of shape.

Figure 2 shows examples of shape representations using an explicit parametric representation by spline curves (spline control points are marked as black boxes), implicit representations by a signed distance function or a binary indicator function and an explicit discrete
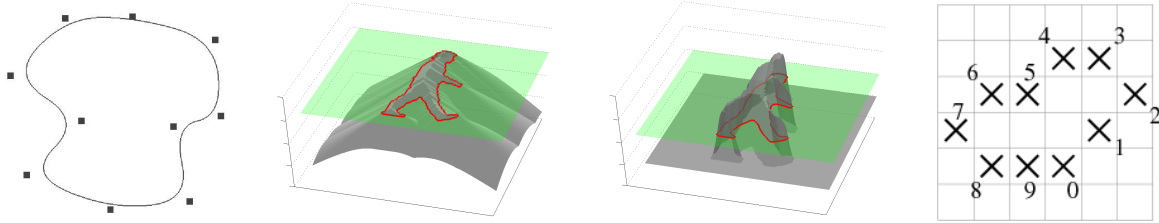
Figure 2: Examples of shape representations by means of a parametric spline curve (1st image), a signed distance function (2nd image), a binary indicator function (3rd image), and an explicit discrete representation (4th image).

representation (4th image).

As we shall see in the following, the choice of shape representation has important consequences on the class of objects that can be modeled, the type of energy that can be minimized and the optimality guarantees that can be obtained. Among the goals of this article is to put in contrast various shape representations and discuss their advantages and limitations. In general one observes that:

- Implicit representations are easily generalized to shapes in arbitrary dimension. Respective algorithms (level set methods, graph cuts, convex relaxation techniques) straightforwardly extend to three or more dimensions. Instead, the extension of explicit shape representations to higher dimensions is by no means straight forward: The notion of arc-length parameterization of curves does not extend to surfaces. Moreover, the discrete polynomial-time shortest-path algorithms [1, 85, 89] which allow to optimally identify pairwise correspondence of points on either shape do not directly extend to minimal-surface algorithms.

- Implicit representations are easily generalized to arbitrary shape topology. Since the implicit representation merely relies on a labeling of space (as being inside or outside the object), the topology of the shape is not constrained. Both level set and graph cut algorithms can therefore easily handle objects of arbitrary topology. Instead, for spatially continuous parametric curves, modeling the transition from a single closed curve to a multiply connected object boundary requires sophisticated splitting and merging techniques [61, 65, 60, 38]. Similarly, discrete polynomial-time algorithms are
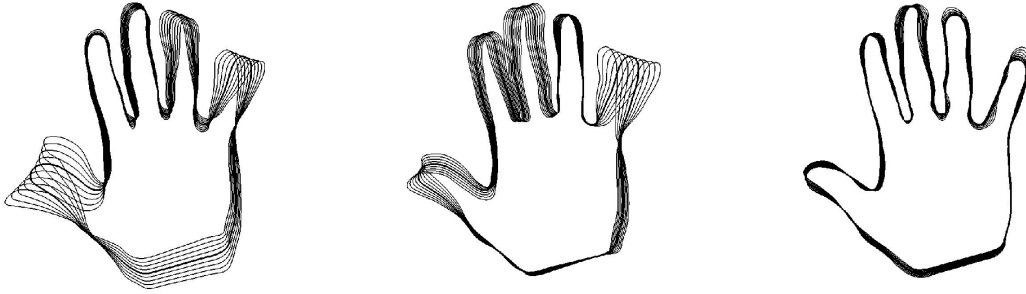
Figure 3: The linear interpolation of spline-based curves (shown here along the first three eigenmodes of a shape distribution) gives rise to a families of intermediate shapes.

typically constrained to finding open [1, 23, 20] or closed curves [86, 89].

- Explicit boundary representations allow to capture the notion of *point correspondence* [47, 85, 89]. The correspondence between points on either of two shapes and the underlying correspondence of semantic parts is of central importance to human notions of shape similarity. The determination of optimal point correspondences, however, is an important combinatorial challenge, especially in higher dimensions.

- For explicit representations the modeling of shape similarity is often more straight forward and intuitive. For example, for two shapes parameterized as spline curves, the linear interpolation of these shapes also gives rise to a spline curve and often captures the human intuition of an *intermediate shape*. Instead, the linear interpolation of implicit representations is generally not straight-forward: Convex combintations of binary-valued functions are no longer binary-valued. And convex combinations of signed distance functions are generally no longer a signed distance function. Figures 3 show examples of a linear interpolations of spline curves and a linear interpolations of signed distance functions. Note that the linear interpolation of signed distance functions may give rise to intermediate silhouettes of varying topology.

In the following, we will give an overview over some of the developments in the domain of shape priors for image segmentation. In Section 2, we will review a formulation of image segmentation by means of Bayesian inference which allows the fusion of input data and shape knowledge in a single energy minimization framework. In Section 3, we will discuss a

5

framework to impose statistical shape priors in a spatially continuous parametric representation. In Section 4, we discuss methods to impose statistical shape priors in level set based image segmentation. In Section 5, we discuss statistical models which allow to represent the temporal evolution of shapes and can serve as dynamical priors for image sequence segmentation. And lastly, in Section 6, we will present recent developments to impose elastic shape priors in a manner which allows to compute globally optimal shape-consistent segmentations in polynomial time.
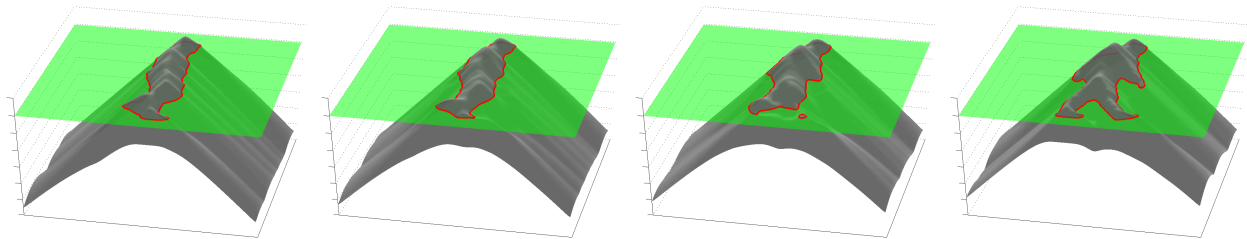

Figure 4: This figure shows the linear interpolation of the signed distance functions associated with two human silhouettes. The interpolation gives rise to intermediate shapes and allows changes of the shape topology. Yet, the linear combination of two signed distance functions is generally no longer a signed distance function.

2 Image Segmentation via Bayesian Inference

Over the last decades Bayesian inference has become an established paradigm to tackle data analysis problems – see [105, 30] for example. Given an input image $I : \Omega \to \mathbb{R}$ on a domain $\Omega \subset \mathbb{R}^2$, a segmentation $\mathcal{C}$ of the image plane Ω can be computed by maximizing the posterior probability

$$\mathcal{P}(\mathcal{C} \,|\, I) = \frac{\mathcal{P}(I \,|\, \mathcal{C}) \, \mathcal{P}(\mathcal{C})}{\mathcal{P}(I)} \tag{1}$$

where $\mathcal{P}(I \,|\, \mathcal{C})$ denotes the data likelihood for a given segmentation $\mathcal{C}$ and $\mathcal{P}(\mathcal{C})$ denotes the prior probability which allows to impose knowledge about which segmentations are *a priori* more or less likely.

Maximizing the posterior distribution can be performed equivalently by minimizing the

negative logarithm of (1) which gives rise to an energy or cost function of the form

$$E(\mathcal{C}) = E_{data}(\mathcal{C}) + E_{shape}(\mathcal{C}), \tag{2}$$

where $E_{data}(\mathcal{C}) = -\log \mathcal{P}(I \,|\, \mathcal{C})$ and $E_{shape}(\mathcal{C}) = -\log \mathcal{P}(\mathcal{C})$ are typically referred to as *data fidelity term* and *regularizer* or *shape prior*. By maximizing the posterior one aims at computing the most likely solution given data and prior. Of course there exist alternative strategies of either computing solutions corresponding to the mean of the distribution rather than its mode, or of retaining the entire posterior distribution in order to propagate multiple hypotheses over time, as done for example in the context of particle filtering [3].

Over the years various data terms have been proposed. In the following, we will simply use a piecewise-constant approximation of the input intensity I [69]:

$$E_{data}(\mathcal{C}) = \sum_{i=1}^{k} \int_{\Omega_i} \big(I(x) - \mu_i\big)^2 dx, \tag{3}$$

where the regions $\Omega_1, \ldots, \Omega_k$ are pairwise disjoint regions separated by the boundary $\mathcal{C}$ and μ_i denotes the average of I over the region Ω_i:

$$\mu_i = \frac{1}{|\Omega_i|} \int_{\Omega_i} I(x) \, dx. \tag{4}$$

More sophisticated data terms based on color likelihoods [57, 8, 103] or texture likelihoods [2, 30] are conceivable.

A glance into the literature indicates that the most prominent image segmentation methods rely on a rather simple geometric shape prior E_{shape} which energetically favors shapes with shorter boundary length [53, 4, 69], a penalizer which – in a spatially discrete setting – dates back at least as far as the Ising model for ferromagnetism [52]. There are several reasons for the popularity of length constraints in image segmentation. Firstly, solid objects in our world indeed tend to be spatially compact. Secondly, such length constraints are mathematically well-studied. They give rise to well-behaved models and algorithms – mean curvature motion in a continuous setting and low-order Markov random fields and submodular cost functions in the discrete setting.

Nevertheless, the preference for a shorter boundary is clearly a very simplistic shape prior. In many applications the user may have a more specific knowledge about what kinds

of shapes are likely to arise in a given segmentation task. For example, in biology one may want to segment cells that all have a rather specific size and shape. In medical imaging one may want to segment organs that all have a rather unique shape – up to a certain variability – and preserve a specific spatial relationship with respect to other organs. In satellite imagery one may be most interested in segmenting thin and elongated roads, or in the analysis of traffic scenes from a driving vehicle, the predominant objects may be cars and pedestrians. In the following sections, we will discuss ways to impose such *higher-level* shape knowledge into image segmentation methods.

3   Statistical Shape Priors for Parametric Shape Representations

Among the most straight-forward ways to represent a shape is to model its outline as a parametric curve. An example is a simple closed spline curve $\mathcal{C} \in \mathcal{C}^k(\mathbb{S}^1, \Omega)$ of the form:

$$\mathcal{C}(s) = \sum_{i=1}^{n} p_i \, B_i(s), \tag{5}$$

where $p_i \in \mathbb{R}^2$ denote a set of spline control points and B_i a set of spline basis functions of degree k [66, 19, 43, 36]. In the special case of linear basis functions, we simply have a polygonal shape, used for example in [102]. With increasing number of control points, we obtain a more and more detailed shape representation – see Figure 5. It shows one of the nice properties of parametric shape representations: The representation is quite *compact* in the sense that very detailed silhouettes can be represented by a few real-valued variables.

Given a family of m shapes each represented by a spline curve of a fixed number of n control points, we can think of these training shapes as a set $\{z_1, \ldots, z_m\}$ of control point vectors

$$z_i = (p_{i1}, \ldots, p_{in}) \in \mathbb{R}^{2n}, \tag{6}$$

where we assume that all control point vectors are normalized with respect to translation, rotation and scaling [41].

With this contour representation, the image segmentation problem boils down to computing an optimal spline control point vector $z \in \mathbb{R}^{2n}$ for a given image. The segmentation

process can be constrained to familiar shapes by imposing a statistical shape prior computed from the set of training shapes.
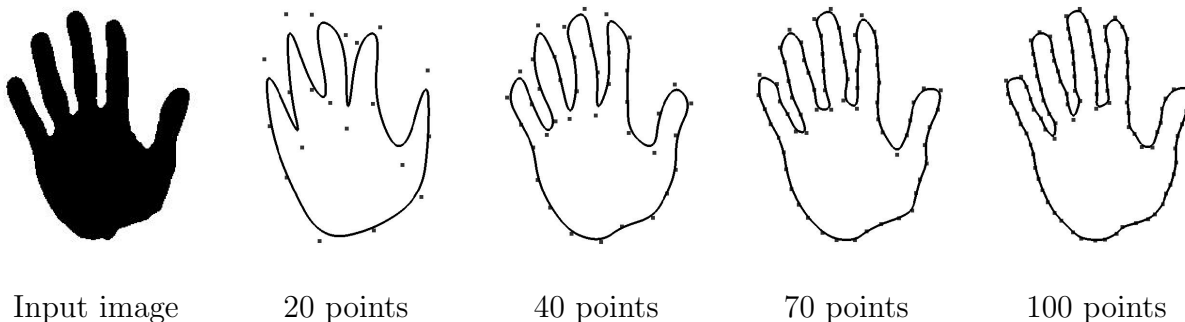

| Input image | 20 points | 40 points | 70 points | 100 points |

Figure 5: Spline representation of a hand shape (left) with increasing resolution.

3.1   Linear Gaussian Shape Priors

Among the most popular shape prior is based on the assumption that the training shapes are Gaussian distributed – see for example [21, 55, 36]. There are several reasons for the popularity of Gaussian distributions. Firstly, according to the central limit theorem the average of a large number of i.i.d. random variables is approximately Gaussian distributed – so if the observed variations of shape were created by independent processes then one could expect the overall distribution to be approximately Gaussian. Secondly, the Gaussian distribution can be seen as a second-order approximation of the true distribution. And thirdly, the Gaussian distribution gives rise to a convex quadratic cost function that allows for easy minimization.

In practice, the number of training shapes m is often much smaller than the number of dimensions $2n$. Therefore, the estimated covariance matrix Σ is degenerate with many zero eigenvalues and thus not invertible. As introduced in [36] a regularized covariance matrix is given by:

$$\Sigma_\perp = \Sigma \,+\, \lambda_\perp \left(I - V\,V^t\right), \tag{7}$$

where V is the matrix of eigenvectors of Σ. In this way, we replace all zero eigenvalues of the sample covariance matrix Σ by a constant $\lambda_\perp \in [0, \lambda_r]$, where λ_r denotes the smallest

non-zero eigenvalue of Σ.[1] In [68] it was shown that $\lambda_\perp$ can be computed from the true covariance matrix by minimizing the Kullback-Leibler divergence between the exact and the approximated distribution. Yet, since we do not have the exact covariance matrix but merely a *sample* covariance matrix, the reasoning for determining $\lambda_\perp$ suggested in [68] is not justified.

The Gaussian shape prior is then given by:

$$\mathcal{P}(z) = \frac{1}{|2\pi\Sigma_\perp|^{1/2}} \exp\left(-\frac{1}{2}(z-\bar{z})^t \Sigma_\perp^{-1}(z-\bar{z})\right),\tag{8}$$

where $\bar{z}$ denotes the mean control point vector.

Based on the Gaussian shape prior, we can define a shape energy that is invariant to similarity transformations (translation, rotation and scaling) by:

$$E_{shape}(z) = -\log\mathcal{P}(\hat{z}),\tag{9}$$

where $\hat{z}$ is the shape vector upon similarity alignment with respect to the training shapes:

$$\hat{z} = \frac{R(z-z_0)}{|R(z-z_0)|},\tag{10}$$

where the optimal translation z_0 and rotation R can be written as functions of z [36]. As a consequence, we can minimize the overall energy

$$E(z) = E_{data}(C(z)) + E_{shape}(z)\tag{11}$$

using gradient descent in z. For details on the numerical minimization we refer to [36, 25].

Figure 6 shows several intermediate steps in a gradient descent evolution on the energy (2) combining the piecewise constant intensity model (3) with a Gaussian shape prior constructed from a set of sample hand shapes. Note how the similarity-invariant shape prior (9) constrains the evolving contour to hand-like shapes without constraining its translation, rotation or scaling.

Figure 7 shows the gradient descent evolution with the same shape prior for an input image of a partially occluded hand. Here the missing part of the silhouette is recovered

---

[1]Note that the inverse $\Sigma_\perp^{-1}$ of the regularized covariance matrix defined in (7) fundamentally differs from the pseudoinverse, the former scaling components in degenerate directions by $\lambda_\perp^{-1}$ while the latter scales them by 0.
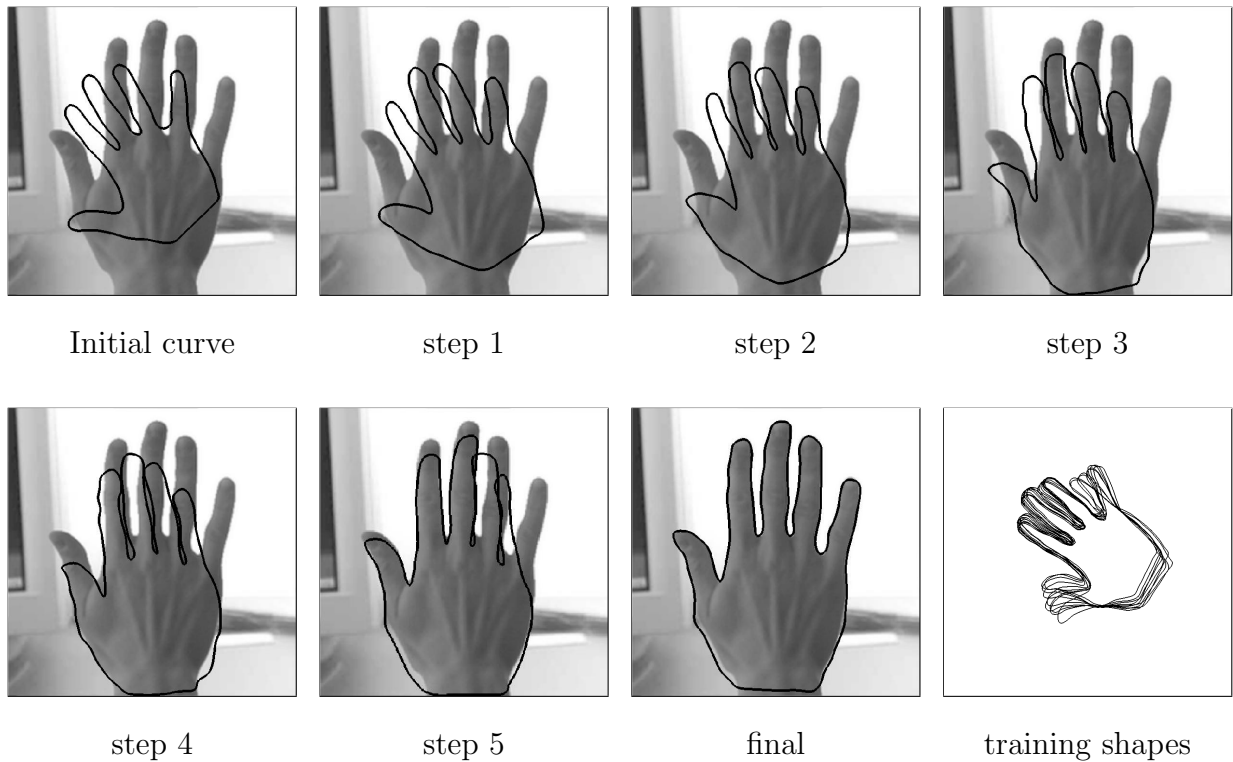
Figure 6: Evolution of a parametric spline curve during gradient descent on the energy (2) combining the piecewise constant intensity model (3) with a Gaussian shape prior constructed from a set of sample hand shapes (lower right). Note that the shape prior is by construction invariant to similiarity transformations. As a consequence, the contour easily undergoes translation, rotation and scaling as these do not affect the energy.

Figure 7: Gradient descent evolution of a parametric curve from initial to final with similarity invariant shape prior. The statistical shape prior permits a reconstruction of the hand silhouette in places where it is occluded.

through the statistical shape prior. These evolutions demonstrate that the curve converges to the correct segmentation over rather large spatial distance, an aspect which is characteristic
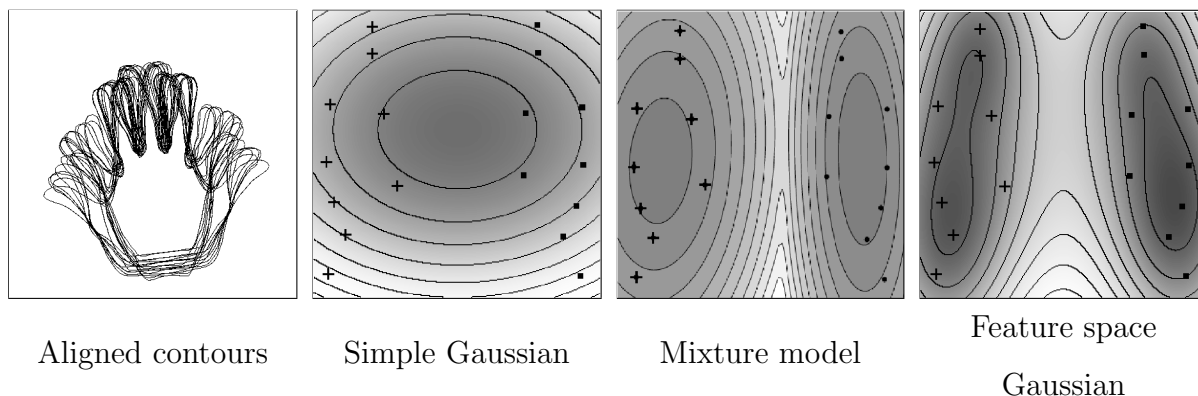
for region-based cost functions like (3).

| Aligned contours | Simple Gaussian | Mixture model | Feature space Gaussian |

Figure 8: Model comparison. Density estimates for a set of left ($\bullet$) and right ($+$) hands, projected onto the first two principal components. **From left to right:** Aligned contours, simple Gaussian, mixture of Gaussians, Gaussian in feature space (13). In contrast to the mixture model, the Gaussian in feature space does not require an iterative (sometimes suboptimal) fitting procedure.

3.2   Nonlinear Statistical Shape Priors

The shape prior (9) was based on the assumption that the training shapes are Gaussian distributed. For collections of real-world shapes this is generally not the case. For example, the various silhouettes of a rigid $3D$ object obviously form a three-dimensional manifold (given that there are only three degrees of freedom in the observation process). Similarly, the various silhouettes of a walking person essentially correspond to a one-dimensional manifold (up to small fluctuations). Furthermore, the manifold of shapes representing deformable objects like human persons are typically very low-dimensional, given that the observed $3D$ structure only has a small number of joints.

Rather than learning the underlying low-dimensional representation (using principal surfaces or other manifold learning techniques), we can simply estimate arbitrary shape distributions by reverting to nonlinear density estimators – *nonlinear* in the sense that the permissible shapes are not simply given by a weighted sum of eigenmodes. Classical approaches for estimating nonlinear distributions are the Gaussian mixture model or the Parzen-Rosenblatt

kernel density estimator – see Section 4.

An alternative technique is to adapt recent kernel learning methods to the problem of density estimation [28]. To this end, we approximate the training shapes by a Gaussian distribution, not in the input space but rather upon transformation $\psi : \mathbb{R}^{2n} \to Y$ to some generally higher-dimensional *feature space* Y:

$$\mathcal{P}_\psi(z) \propto \exp\left(-\frac{1}{2} (\psi(z) - \psi_0)^t \, \Sigma_\psi^{-1} \, (\psi(z) - \psi_0) \right). \tag{12}$$

As before we can define the corresponding shape energy as

$$E(z) = -\log \mathcal{P}_\psi(\hat{z}), \tag{13}$$

with $\hat{z}$ being the similarity-normalized shape given in (10). Here ψ_0 and Σ_ψ denote the mean and covariance matrix computed for the transformed shapes:

$$\psi_0 = \frac{1}{m} \sum_{i=1}^m \psi(z_i), \quad \Sigma_\psi = \frac{1}{m} \sum_{i=1}^m \left(\psi(z_i) - \psi_0 \right) \left(\psi(z_i) - \psi_0 \right)^\top, \tag{14}$$

where Σ_ψ is again regularized as in (7).

As shown in [28], the energy $E(z)$ in (13) can be evaluated without explicitly specifying the nonlinear transformation ψ. It suffices to define the corresponding Mercer kernel [67, 24]:

$$k(x, y) := \langle \psi(x), \psi(y) \rangle, \quad \forall x, y \in \mathbb{R}^{2n}, \tag{15}$$

representing the scalar product of pairs of transformed points $\psi(x)$ and $\psi(y)$. In the following we simply chose a Gaussian kernel function of width σ:

$$k(x, y) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{||x - y||^2}{2\sigma^2} \right). \tag{16}$$

It was shown in [28], that the resulting energy can be seen as a generalization of the classical Parzen-Rosenblatt estimators. In particular, the Gaussian distribution in feature space Y is fundamentally different from the previously presented Gaussian distribution in the input space $\mathbb{R}^{2n}$. Figure 8 shows the level lines of constant shape energy computed from a set of left and right hand silhouettes, displayed in a projection onto the first two eigenmodes of the distribution. While the linear Gaussian model gives rise to elliptical level lines, the Gaussian mixture and the nonlinear Gaussian allow for more general non-elliptical level

| Initial contour | No prior | With prior |

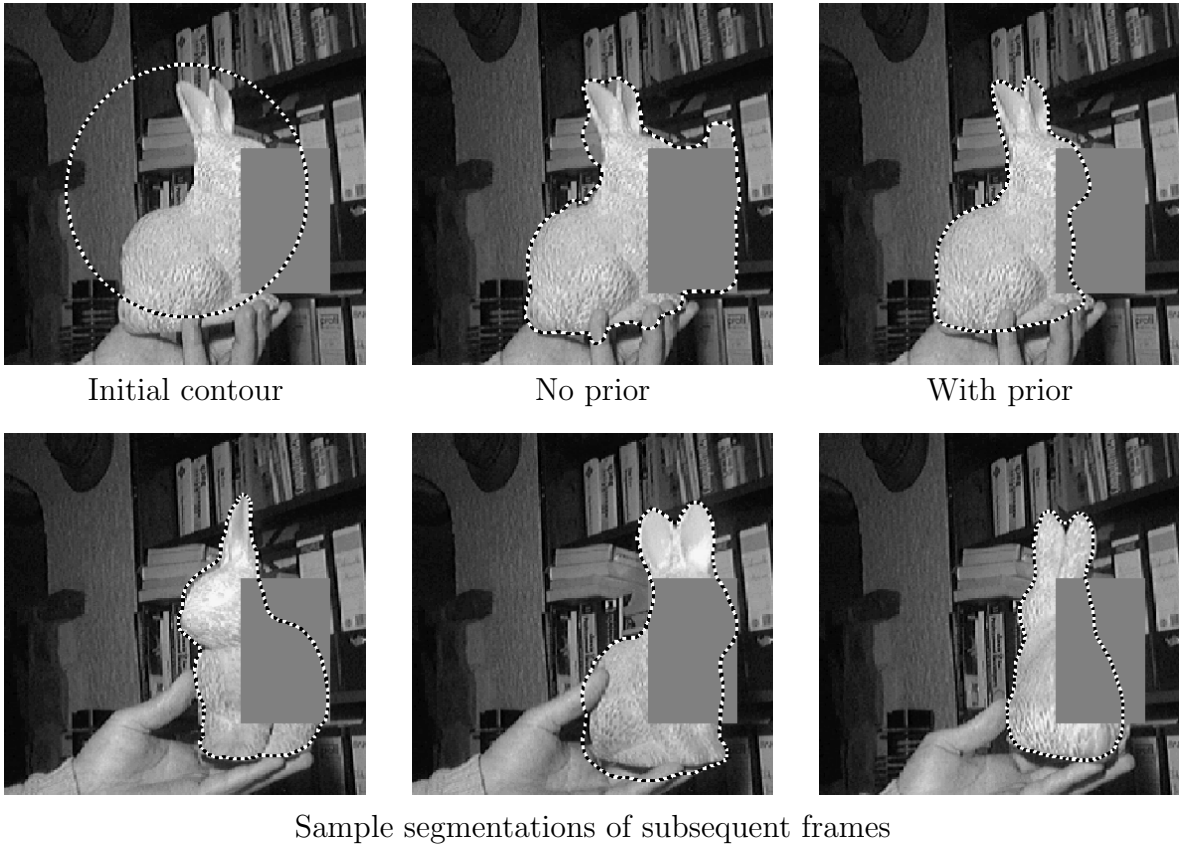Sample segmentations of subsequent frames

Figure 9: Tracking a familiar object over a long image sequence with a nonlinear statistical shape prior. A single shape prior constructed from a set of sample silhouettes allows the emergence of a multitude of familiar shapes, permitting the segmentation process to cope with background clutter and partial occlusions.

lines. In contrast to the mixture model, however, the nonlinear Gaussian does not require an interative parameter estimation process, not does it require or assume a specific number of Gaussians.

Figure 9 shows screenshots of contours computed for an image sequence by gradient descent on the energy (11) with the nonlinear shape energy (13) computed from a set of 100 training silhouettes. Throughout the entire sequence, the object of interest was occluded by an artificially introduced rectangle. Again, the shape prior allows to cope with spurious background clutter and to restore the missing parts of the object's silhouette. Two–dimensional projections of the training data and evolving contour onto the first principal components, shown in Figure 10, demonstrate how the nonlinear shape energy constrains the evolving
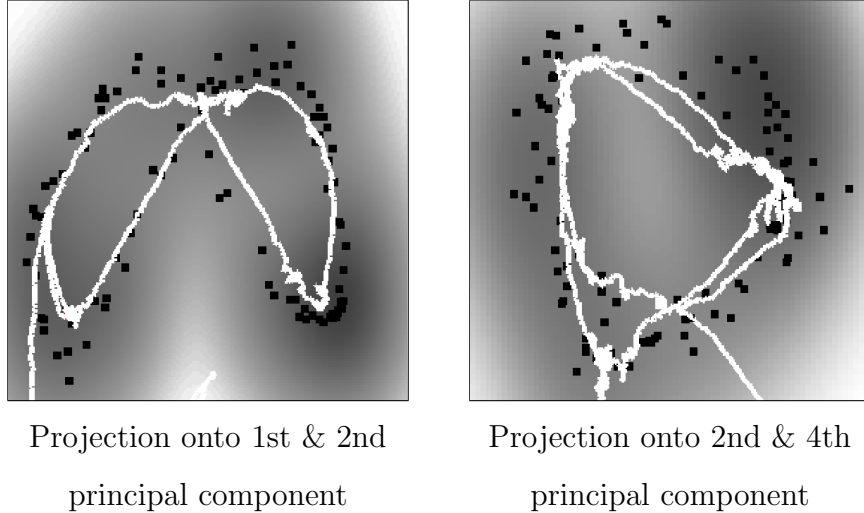
Projection onto 1st & 2nd   Projection onto 2nd & 4th

principal component   principal component

Figure 10: Tracking sequence from Figure 9 visualized. Training data (●), estimated energy density (shaded) and the contour evolution (white curve) in appropriate 2D projections. The evolving contour – see Figures 9 – is constrained to the domains of low energy induced by the training data.

shape to remain close to the training shapes.

4   Statistical Priors for Level Set Representations

Parametric representations of shape as those presented above have numerous favorable properties, in particular, they allow to represent rather complex shapes with few a parameters, resulting in low memory requirements and low computation time. Nevertheless, the explicit representation of shape has several drawbacks:

- The representation of explicit shapes typically depends on a specific choice of representation. To factor out this dependency in the representation and in respective algorithms gives rise to computationally challenging problems. Determining point correspondences, for example, becomes particularly difficult for shapes in higher dimensions (surfaces in $3D$ for example).

- In particular, the evolution of explicit shape representations requires sophisticated numerical regridding procedures to assure an equidistant spacing of control points and

prevent control point overlap.

- Parametric representations are difficult to adapt to varying topology of the represented shape. Numerically topology changes require sophisticated splitting and remerging procedures.

- A number of recent publications [49, 11, 59] indicate that in contrast to explicit shape representations, the implicit representation of shape allows to compute *globally optimal* solutions to shape inference for large classes of commonly used energy functionals.

A mathematical representation of shape which is independent of parameterization was pioneered in the analysis of random shapes by Fréchet [45] and in the school of mathematical morphology founded by Matheron and Serra [64, 94]. The level set method [39, 72] provides a means of propagating contours $\mathcal{C}$ (independent of parameterization) by evolving associated embedding functions ϕ via partial differential equations – see Figure 11 for a visualization of the level set function associated with a human silhouette. It has been adapted to segment images based on numerous low-level criteria such as edge consistency [63, 10, 56], intensity homogeneity [13, 101], texture information [73, 80, 51, 9] and motion information [33].

In this section, we will give a brief insight into shape modeling and shape priors for implicit level set representations. Parts of the following text were adopted from [34, 81, 35].

4.1   Shape Distances for Level Sets

The first step in deriving a shape prior is to define a distance or dissimilarity measure for two shapes encoded by the level set functions ϕ_1 and ϕ_2. We shall briefly discuss three solutions to this problem. In order to guarantee a unique correspondence between a given shape and its embedding function ϕ, we will in the following assume that ϕ is a *signed distance function*, i.e. $\phi > 0$ inside the shape, $\phi < 0$ outside and $|\nabla \phi| = 1$ almost everywhere. A method to project a given embedding function onto the space of signed distance functions was introduced in [98].

Given two shapes encoded by their signed distance functions ϕ_1 and ϕ_2, a simple measure
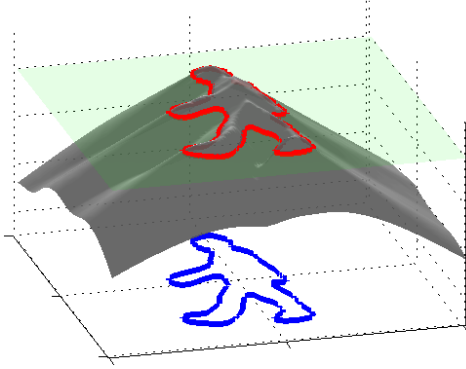
Figure 11: The level set method is based on representing shapes implicitly as the zero level set of a higher-dimensional embedding function.

of their dissimilarity is given by their L$_2$-distance in Ω [62]:

$$\int_\Omega (\phi_1 - \phi_2)^2 \; dx. \tag{17}$$

This measure has the drawback that it depends on the domain of integration Ω. The shape dissimilarity will generally grow if the image domain is increased – even if the relative position of the two shapes remains the same. Various remedies to this problem have been proposed. We refer to [32] for a detailed discussion.

An alternative dissimilarity measure between two implicitly represented shapes represented by the embedding functions ϕ_1 and ϕ_2 is given by the area of the symmetric difference [12, 77, 15]:

$$d^2(\phi_1, \phi_2) = \int_\Omega \Big(H\phi_1(x) - H\phi_2(x) \Big)^2 dx. \tag{18}$$

In the present work, we will define the distance between two shapes based on this measure, because it has several favorable properties. Beyond being independent of the image size Ω, measure (18) defines a distance on the set of shapes: it is non-negative, symmetric and fulfills the triangle inequality. Moreover, it is more consistent with the philosophy of the level set method in that it only depends on the *sign* of the embedding function. In practice, this means that one does not need to constrain the two level set functions to the space of signed distance functions. It can be shown [15] that L$^\infty$ and $W^{1,2}$ norms on the signed distance functions induce equivalent topologies as the metric (18).

Since the distance (18) is not differentiable, we will in practice consider an approximation of the Heaviside function H by a smooth (differentiable) version H_ϵ. Moreover, we will only consider gradients of energies with respect to the L_2–norm on the level set function, because they are easy to compute and because variations in the signed distance function correspond to respective variations of the implicitly represented curve. In general, however, these do not coincide with so-called *shape gradients* – see [46] for a recent work on this topic.

4.2   Invariance by Intrinsic Alignment

One can make use of the shape distance (18) in a segmentation process by adding it as a shape prior $E_{shape}(\phi) = d^2(\phi, \phi_0)$ in a weighted sum to the data term, which we will assume to be the two-phase version of (3) introduced in [14]:

$$E_{data}(\phi) = \int_\Omega (I - u_+)^2 H\phi(x)dx \,+\, \int_\Omega (I - u_-)^2 \big(1 - H\phi(x)\big)dx \,+\, \nu \int_\Omega |\nabla H\phi|dx, \qquad (19)$$

Minimizing the total energy

$$E_{total}(\phi) = E_{data}(\phi) \,+\, \alpha\, E_{shape}(\phi) = E_{data}(\phi) \,+\, \alpha\, d^2(\phi, \phi_0) \qquad (20)$$

with a weight $\alpha > 0$, induces an additional driving term which aims at maximizing the similarity of the evolving shape with a given template shape encoded by the function ϕ_0.

By construction this shape prior is not invariant with respect to certain transformations such as translation, rotation and scaling of the shape represented by ϕ.

A common approach to introduce invariance (c.f. [17, 82, 35]) is to enhance the prior by a set of explicit parameters to account for translation by μ, rotation by an angle θ and scaling by σ of the shape:

$$d^2(\phi, \phi_0, \mu, \theta, \sigma) = \int_\Omega \Big(H\big(\phi\big(\sigma R_\theta(x - \mu)\big)\big) - H\phi_0(x)\Big)^2 dx. \qquad (21)$$

This approach to estimate the appropriate transformation parameters has several drawbacks:

- Optimization of the shape energy (21) is done by local gradient descent. In particular, this implies that one needs to determine an appropriate time step for each parameter, chosen so as to guarantee stability of resulting evolution. In numerical experiments, we found that balancing these parameters requires a careful tuning process.

- The optimization of μ, θ, σ and ϕ is done simultaneously. In practice, however, it is unclear how to alternate between the updates of the respective parameters. How often should one iterate one or the other gradient descent equation? In experiments, we found that the final solution depends on the selected scheme of optimization.

- The optimal values for the transformation parameters will depend on the embedding function ϕ. An accurate shape gradient should therefore take into account this dependency. In other words, the gradient of (21) with respect to ϕ should take into account how the optimal transformation parameters $\mu(\phi)$, $\sigma(\phi)$ and $\theta(\phi)$ vary with ϕ.

Inspired by the normalization for explicit representations introducing in (10), we can elimiate these difficulties associated with the local optimization of explicit transformation parameters by introducing an intrinsic registration process. We will detail this for the cases of translation and scaling. Extensions to rotation and other transformations are conceivable but will not be pursued here.

4.2.1 Translation invariance by intrinsic alignment

Assume that the template shape represented by ϕ_0 is aligned with respect to the shape's centroid. Then we define a shape energy by:

$$E_{shape}(\phi) = d^2(\phi, \phi_0) = \int_\Omega \Big(H\phi(x + \mu_\phi) - H\phi_0(x) \Big)^2 dx, \tag{22}$$

where the function ϕ is evaluated in coordinates relative to its center of gravity μ_ϕ given by:

$$\mu_\phi = \int x\, h\phi\, dx, \quad \text{with } h\phi \equiv \frac{H\phi}{\int_\Omega H\phi\, dx}. \tag{23}$$

This intrinsic alignment guarantees that the distance (22) is invariant to the location of the shape ϕ. In contrast to the shape energy (21), we no longer need to iteratively update an estimate of the location of the object of interest.

4.2.2 Translation and scale invariance via alignment

Given a template shape (represented by ϕ_0) which is normalized with respect to translation and scaling, one can extend the above approach to a shape energy which is invariant to

19

translation and scaling:

$$E_{shape}(\phi) = d^2(\phi, \phi_0) = \int_\Omega \Big(H\phi(\sigma_\phi\, x + \mu_\phi) - H\phi_0(x) \Big)^2 dx, \qquad (24)$$

where the level set function ϕ is evaluated in coordinates relative to its center of gravity μ_ϕ and in units given by its intrinsic scale σ_ϕ defined as:

$$\sigma_\phi = \left(\int (x - \mu)^2\, h\phi\, dx \right)^{\frac{1}{2}}, \quad \text{where } h\phi = \frac{H\phi}{\int_\Omega H\phi\, dx}. \qquad (25)$$

In the following, we will show that functional (24) is invariant with respect to translation and scaling of the shape represented by ϕ. Let ϕ be a level set function representing a shape which is centered and normalized such that $\mu_\phi = 0$ and $\sigma_\phi = 1$. Let $\tilde{\phi}$ be an (arbitrary) level set function encoding the same shape after scaling by $\sigma \in \mathbb{R}$ and shifting by $\mu \in \mathbb{R}^2$:

$$H\tilde{\phi}(x) = H\phi\left(\frac{x - \mu}{\sigma} \right).$$

Indeed, center and intrinsic scale of the transformed shape are given by:

$$\mu_{\tilde{\phi}} = \frac{\int x H\tilde{\phi}dx}{\int H\tilde{\phi}dx} = \frac{\int x H\phi\left(\frac{x-\mu}{\sigma}\right) dx}{\int H\phi\left(\frac{x-\mu}{\sigma}\right) dx} = \frac{\int (\sigma x' + \mu) H\phi(x')\, \sigma dx'}{\int H\phi(x')\, \sigma dx'} = \sigma\mu_\phi + \mu = \mu,$$

$$\sigma_{\tilde{\phi}} = \left(\frac{\int (x - \mu_{\tilde{\phi}})^2 H\tilde{\phi}dx}{\int H\tilde{\phi}dx} \right)^{\frac{1}{2}} = \left(\frac{\int (x - \mu)^2 H\phi\left(\frac{x-\mu}{\sigma}\right) dx}{\int H\phi\left(\frac{x-\mu}{\sigma}\right) dx} \right)^{\frac{1}{2}} = \left(\frac{\int (\sigma x')^2 H\phi(x')dx'}{\int H\phi(x')dx'} \right)^{\frac{1}{2}} = \sigma.$$

The shape energy (24) evaluated for $\tilde{\phi}$ is given by:

$$E_{shape}(\tilde{\phi}) = \int_\Omega \Big(H\tilde{\phi}(\sigma_{\tilde{\phi}} x + \mu_{\tilde{\phi}}) - H\phi_0(x) \Big)^2 dx = \int_\Omega \Big(H\tilde{\phi}(\sigma\, x + \mu) - H\phi_0(x) \Big)^2 dx$$

$$= \int_\Omega \Big(H\phi(x) - H\phi_0(x) \Big)^2 dx \; = \; E_{shape}(\phi)$$

Therefore, the above shape dissimilarity measure is invariant with respect to translation and scaling.

Note, however, that while this analytical solution guarantees an invariant shape distance, the transformation parameters μ_ϕ and σ_ϕ are not necessarily the ones which minimize the shape distance (21). Extensions of this approach to a larger class of invariance are conceivable. For example, one could generate invariance with respect to rotation by rotational alignment with respect to the (oriented) principal axis of the shape encoded by ϕ. We will not pursue this here.

4.3 Kernel Density Estimation in the Level Set Domain

In the previous sections, we have introduced a translation and scale invariant shape energy and demonstrated its effect on the reconstruction of a corrupted version of a single familiar silhouette the pose of which was unknown. In many practical problems, however, we do not have the exact silhouette of the object of interest. There may be several reasons for this:

- The object of interest may be three-dimensional. Rather than trying to reconstruct the three dimensional object (which generally requires multiple images and the estimation of correspondence), one may learn the two dimensional appearance from a set of sample views. A meaningful shape dissimilarity measure should then measure the dissimilarity with respect to this set of projections – see the example in Figure 9.

- The object of interest may be one object out of a class of similar objects (the class of cars or the class of tree leafs). Given a limited number of training shapes sampled from the class, a useful shape energy should provide the dissimilarity of a particular silhouette with respect to this class.

- Even a single object, observed from a single viewpoint, may exhibit strong shape deformation – the deformation of a gesticulating hand or the deformation which a human silhouette undergoes while walking. In the following, we will assume that one can merely generate a set of stills corresponding to various (randomly sampled) views of the object of interest for different deformations – see Figure 12. In the following, we will demonstrate that – without constructing a dynamical model of the walking process – one can exploit this set of sample views in order to improve the segmentation of a walking person.

In the above cases, the construction of appropriate shape dissimilarity measures amounts to a problem of density estimation. In the case of explicitly represented boundaries, this has been addressed by modeling the space of familiar shapes by linear subspaces (PCA) [21] and the related Gaussian distribution [36], by mixture models [22] or nonlinear (multi-modal) representations via simple models in appropriate feature spaces [27, 28].

For level set based shape representations, it was suggested [62, 100, 83] to fit a linear sub-space to the sampled signed distance functions. Alternatively, it was suggested to represent familiar shapes by the level set function encoding the mean shape and a (spatially independent) Gaussian fluctuation at each image location [82]. These approaches were shown to capture some shape variability. Yet, they exhibit two limitations: Firstly, they rely on the assumption of a Gaussian distribution which is not well suited to approximate shape distributions encoding more complex shape variation. Secondly, they work under the assumption that shapes are represented by signed distance functions. Yet, the space of signed distance functions is not a linear space. Therefore, in general, neither the mean nor the linear combination of a set of signed distance functions will correspond to a signed distance function.

In the following, we will propose an alternative approach to generate a statistical shape dissimilarity measure for level set based shape representations. It is based on classical methods of (so-called non-parametric) kernel density estimation and overcomes the above limitations.

Given a set of training shapes $\{\phi_i\}_{i=1...N}$ – such as those shown in Figure 12 – we define a probability density on the space of signed distance functions by integrating the shape distances (22) or (24) in a Parzen-Rosenblatt kernel density estimator [79, 75]:

$$\mathcal{P}(\phi) \propto \frac{1}{N} \sum_{i=1}^{N} \exp\left(-\frac{1}{2\sigma^2} d^2(H\phi, H\phi_i)\right). \tag{26}$$

The kernel density estimator is among the theoretically most studied density estimation methods. It was shown (under fairly mild assumptions) to converge to the true distribution in the limit of infinite samples (and $\sigma \to 0$), the asymptotic convergence rate was studied for different choices of kernel functions.

It should be pointed out that the theory of classical nonparametric density estimation was developed for the case of finite-dimensional data. It is beyond the scope of this work to develop a general theory of probability distributions and density estimation on infinite-dimensional spaces (including issues of integrability and measurable sets). For a general formalism to model probability densities on infinite-dimensional spaces, we refer the reader to the theory of Gaussian processes [76]. In our case, an extension to infinite-dimensional

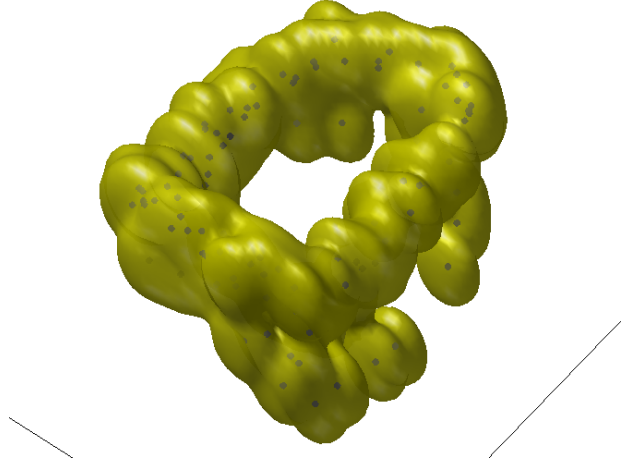Selected sample shapes from a set of a walking silhouettes.

Figure 12: Density estimated using a kernel density estimator for a projection of 100 silhouettes of a walking person (see above) onto the first three principal components.

objects such as level set surfaces $\phi : \Omega \to \mathbb{R}$ could be tackled by considering discrete (finite-dimensional) approximations $\{\phi_{ij} \in \mathbb{R}\}_{i=1,...,N, j=1,...,M}$ of these surfaces at increasing levels of spatial resolution and studying the limit of infinitesimal grid size (i.e. $N, M \to \infty$). Alternatively, given a finite number of samples, one can apply classical density estimation techniques efficiently in the finite-dimensional subspace spanned by the training data [81].

Similarly respective metrics on the space of curves give rise to different kinds of gradient descent flows. Recently researchers have developed rather sophisticated metrics to favor smooth transformations or rigid body motions. We refer the reader to [16, 97] for promising advances in this direction. In the following we will typically limit ourselves to L_2 gradients.

There exist extensive studies on how to optimally choose the kernel width σ, based on asymptotic expansions such as the parametric method [37], heuristic estimates [104, 95] or maximum likelihood optimization by cross validation [42, 18]. We refer to [40, 96] for a detailed discussion. For this work, we simply fix σ^2 to be the mean squared nearest-neighbor

distance:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \min_{j \neq i} d^2(H\phi_i, H\phi_j). \tag{27}$$

The intuition behind this choice is that the width of the Gaussians is chosen such that on the average the next training shape is within one standard deviation.

Reverting to kernel density estimation resolves the drawbacks of existing approaches to shape models for level set segmentation discussed above. In particular:

- The silhouettes of a rigid 3D object or a deformable object with few degrees of freedom can be expected to form fairly low-dimensional manifolds. The kernel density estimator can capture these without imposing the restrictive assumption of a Gaussian distribution. Figure 12, shows a 3D approximation of our method: We simply projected the embedding functions of 100 silhouettes of a walking person onto the first three eigenmodes of the distribution. The projected silhouette data and the kernel density estimate computed in the 3D subspace indicate that the underlying distribution is not Gaussian. The estimated distribution (indicated by an isosurface) shows a closed loop which stems from the fact that the silhouettes were drawn from an essentially periodic process.

- Kernel density estimators were shown to converge to the true distribution in the limit of infinite (independent and identically distributed) training samples [40, 96]. In the context of shape representations, this implies that our approach is capable of accurately representing arbitrarily complex shape deformations.

- By not imposing a linear subspace, we circumvent the problem that the space of shapes (and signed distance functions) is not a linear space. In other words: Kernel density estimation allows to estimate distributions on non-linear (curved) manifolds. In the limit of infinite samples and kernel width σ going to zero, the estimated distribution is more and more constrained to the manifold defined by the shapes.

24

4.4 Gradient Descent Evolution for the Kernel Density Estimator

In the following, we will detail how the statistical distribution (26) can be used to enhance level set based segmentation process. As for the case of parametric curves, we formulate level set segmentation as a problem of Bayesian inference, where the segmentation is obtained by maximizing the conditional probability

$$\mathcal{P}(\phi \,|\, I) = \frac{\mathcal{P}(I \,|\, \phi)\,\mathcal{P}(\phi)}{\mathcal{P}(I)}, \tag{28}$$

with respect to the level set function ϕ, given the input image I. For a given image, this is equivalent to minimizing the negative log-likelihood which is given by a sum of two energies:

$$E(\phi) = E_{data}(\phi) \;+\; E_{shape}(\phi), \tag{29}$$

with

$$E_{shape}(\phi) = -\log \mathcal{P}(\phi), \tag{30}$$

Minimizing the energy (29) generates a segmentation process which simultaneously aims at maximizing intensity homogeneity in the separated phases and a similarity of the evolving shape with respect to all the training shapes encoded through the statistical estimator (26).

Gradient descent with respect to the embedding function amounts to the evolution:

$$\frac{\partial \phi}{\partial t} = -\frac{1}{\alpha}\frac{\partial E_{data}}{\partial \phi} - \frac{\partial E_{shape}}{\partial \phi}, \tag{31}$$

where the knowledge-driven component is given by:

$$\frac{\partial E_{shape}}{\partial \phi} = \frac{\sum \alpha_i \frac{\partial}{\partial \phi} d^2(H\phi, H\phi_i)}{2\sigma^2 \sum \alpha_i}, \tag{32}$$

which simply induces a force in direction of each training shape ϕ_i weighted by the factor:

$$\alpha_i = \exp\left(-\frac{1}{2\sigma^2} d^2(H\phi, H\phi_i)\right), \tag{33}$$

that decays exponentially with the distance from the training shape ϕ_i.

4.5 Nonlinear Shape Priors for Tracking a Walking Person

In the following we apply the above shape prior to the segmentation of a partially occluded walking person. To this end, a sequence of a walking figure was partially occluded by
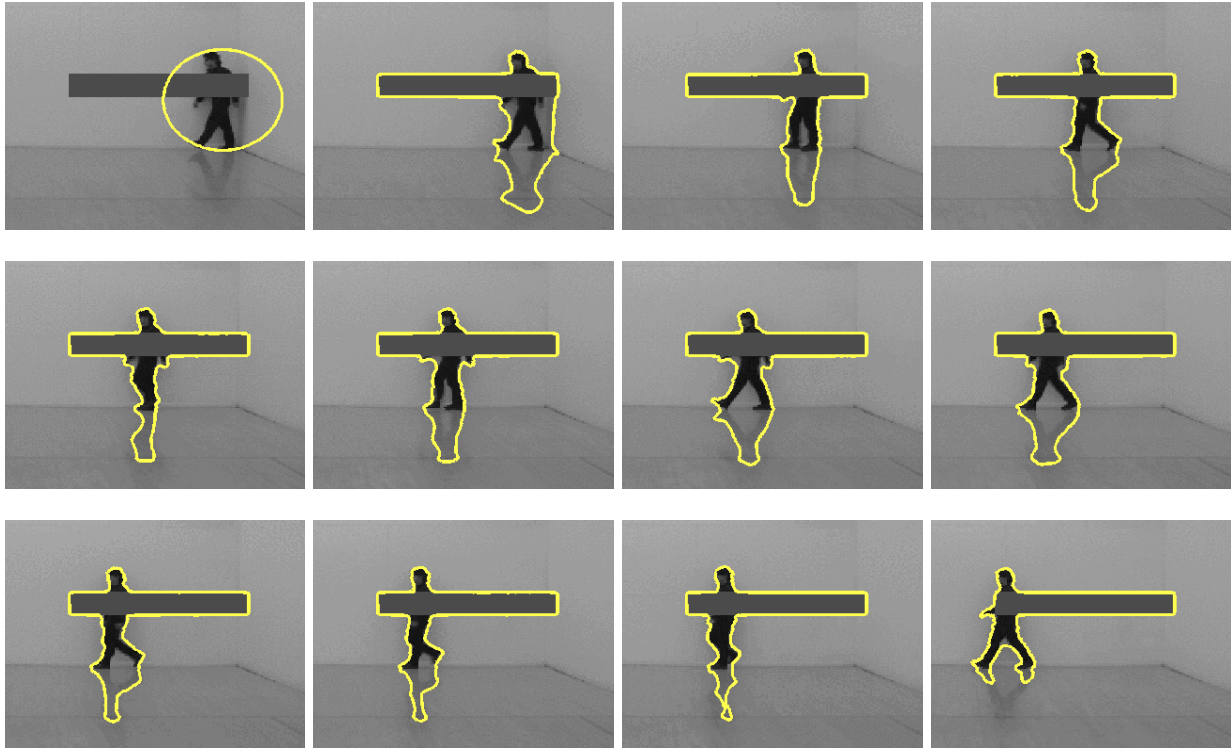
25

Figure 13: Purely intensity-based segmentation. Various frames show the segmentation of a partially occluded walking person generated by minimizing the Chan-Vese energy (19). The walking person cannot be separated from the occlusion and darker areas of the background such as the person's shadow.

an artificial bar. Subsequently we minimized energy (19), segmenting each frame of the sequence using the previous segmentation as initialization. Figure 13 shows that this purely image-driven segmentation scheme is not capable of separating the object of interest from the occluding bar and similarly shaded background regions such as the object's shadow on the floor.

In a second experiment, we manually binarized the images corresponding to the first half of the original sequence (frames 1 through 42) and aligned them to their respective center of gravity to obtain a set of training shape – see Figure 12. Then we ran the segmentation process (31) with the shape prior (26). Apart from adding the shape prior we kept the other parameters constant for comparability.

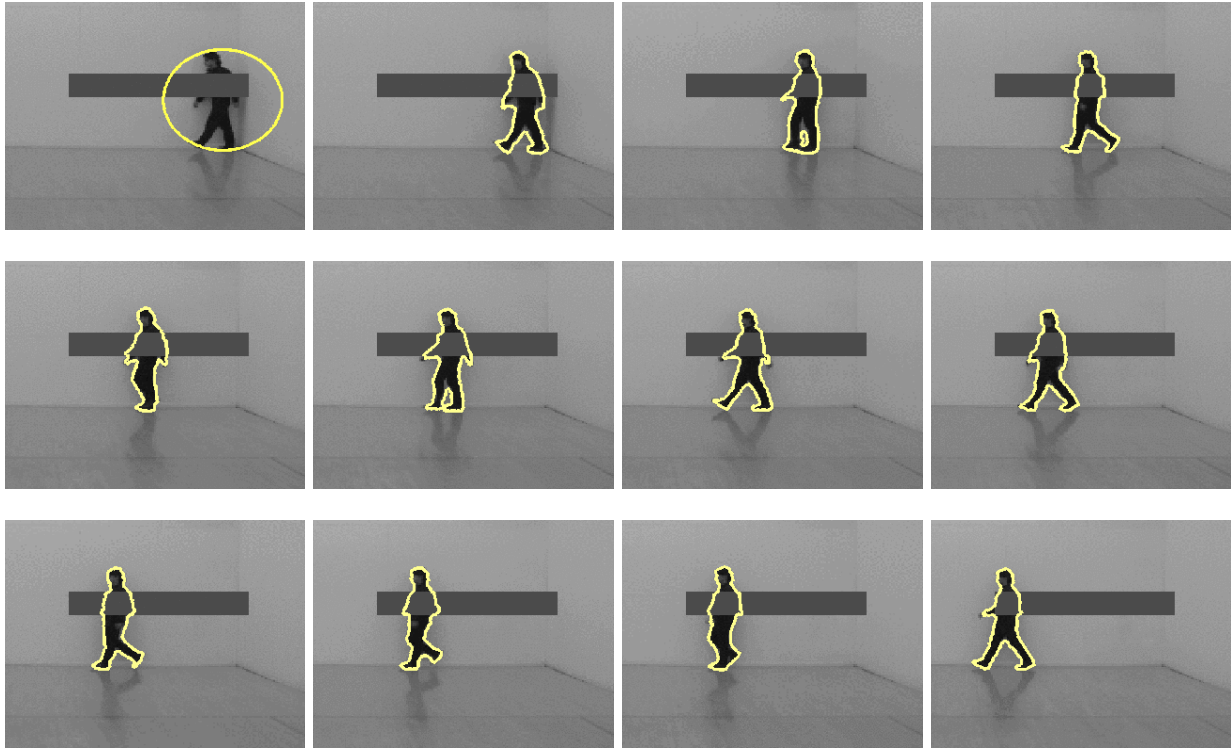Figure 14 shows several frames from this knowledge-driven segmentation. A comparison

Figure 14: Segmentation with nonparametric invariant shape prior. Segmentation generated by minimizing energy (29) combining intensity information with the shape prior (26). For every frame in the sequence, the gradient descent equation was iterated (with fixed parameters), using the previous segmentation as initialization. The shape prior permits to separate the walking person from the occlusion and darker areas of the background such as the shadow. The shapes in the second half of the sequence were not part of the training set.

to the corresponding frames in Figure 13 demonstrates several properties:

- The shape prior permits to accurately reconstruct an entire set of fairly different shapes. Since the shape prior is defined on the level set function ϕ – rather than on the boundary C (cf. [17]) – it can easily handle changing topology.

- The shape prior is invariant to translation such that the object silhouette can be reconstructed in arbitrary locations of the image.

- The statistical nature of the prior allows to also reconstruct silhouettes which were not part of the training set – corresponding to the second half of the images shown (beyond frame 42).

27

5 Dynamical Shape Priors for Implicit Shapes

5.1 Capturing the Temporal Evolution of Shape

In the above works, statistically learned shape information was shown to cope for missing or misleading information in the input images due to noise, clutter and occlusion. The shape priors were developed to segment objects of familiar shape in a given image. Although they can be applied to tracking objects in image sequences, they are not well-suited for this task, because they neglect the *temporal coherence of silhouettes* which characterizes many deforming shapes.

When tracking a deformable object over time, clearly not all shapes are equally likely at a given time instance. Regularly sampled images of a walking person, for example, exhibit a typical pattern of consecutive silhouettes. Similarly, the projections of a rigid 3D object rotating at constant speed are generally not independent samples from a statistical shape distribution. Instead, the resulting set of silhouettes can be expected to contain strong temporal correlations.

In the following, we will present temporal statistical shape models for implicitly represented shapes that were first introduced in [26]. In particular, the shape probability at a given time depends on the shapes observed at previous time steps. The integration of such dynamical shape models into the segmentation process can be elegantly formulated within a Bayesian framework for level set based image sequence segmentation. The resulting optimization by gradient descent induces an evolution of the level set function which is driven both by the intensity information of the current image as well as by a dynamical shape prior which relies on the segmentations obtained on the preceding frames. Experimental evaluation demonstrates that the resulting segmentations are not only similar to previously learned shapes, but they are also consistent with the temporal correlations estimated from sample sequences. The resulting segmentation process can cope with large amounts of noise and occlusion because it exploits prior knowledge about *temporal* shape consistency and because it aggregates information from the input images over time (rather than treating each image independently).

5.2 Level Set Based Tracking via Bayesian Inference

Statistical models can be estimated more reliably if the dimensionality of the model and the data are low. We will therefore cast the Bayesian inference in a low-dimensional formulation within the subspace spanned by the largest principal eigenmodes of a set of sample shapes. We exploit the training sequence in a twofold way: Firstly, it serves to define a low-dimensional subspace in which to perform estimation. And secondly, within this subspace we use it to learn dynamical models for implicit shapes. For static shape priors this concept was already used in [81].

Let $\{\phi_1, \ldots, \phi_N\}$ be a temporal sequence of training shapes.[2] Let ϕ_0 denote the mean shape and $\psi_1, \ldots, \psi_n$ the n largest eigenmodes with $n \ll N$. We will then approximate each training shape as:

$$\phi_i(x) = \phi_0(x) + \sum_{j=1}^{n} \alpha_{ij}\, \psi_j(x), \tag{34}$$

where

$$\alpha_{ij} = \langle \phi_i - \phi_0, \psi_j \rangle \equiv \int (\phi_i - \phi_0)\, \psi_j\, dx. \tag{35}$$

Such PCA based representations of level set functions have been successfully applied for the construction of statistical shape priors in [62, 100, 83, 81]. In the following, we will denote the vector of the first n eigenmodes as $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_n)$. Each sample shape ϕ_i is therefore approximated by the n-dimensional shape vector $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{in})$. Similarly, an arbitrary shape ϕ can be approximated by a shape vector of the form

$$\boldsymbol{\alpha}_\phi = \langle \phi - \phi_0, \boldsymbol{\psi} \rangle. \tag{36}$$

In addition to the deformation parameters $\boldsymbol{\alpha}$, we introduce transformation parameters θ, and we introduce the notation:

$$\phi_{\boldsymbol{\alpha},\theta}(x) = \phi_0(T_\theta x) + \boldsymbol{\alpha}^\top \boldsymbol{\psi}(T_\theta x), \tag{37}$$

---

[2]We assume that all training shapes ϕ_i are signed distance functions. Yet an arbitrary linear combination of eigenmodes will in general not generate a signed distance function. While the discussed statistical shape models favor shapes which are close to the training shapes (and therefore close to the set of signed distance functions), not all shapes sampled in the considered subspace will correspond to signed distance functions.

to denote the embedding function of a shape generated with deformation parameters $\boldsymbol{\alpha}$ and transformed with parameters θ. The transformations T_θ can be translation, rotation and scaling (depending on the application).

With this notation, the goal of image sequence segmentation within this subspace can be stated as follows: Given consecutive images $I_t : \Omega \to \mathbb{R}$ from an image sequence, and given the segmentations $\hat{\boldsymbol{\alpha}}_{1:t-1}$ and transformations $\hat{\theta}_{1:t-1}$ obtained on the previous images $I_{1:t-1}$, compute the most likely deformation $\hat{\boldsymbol{\alpha}}_t$ and transformation $\hat{\theta}_t$ by maximizing the conditional probability

$$\mathcal{P}(\boldsymbol{\alpha}_t, \theta_t \,|\, I_t, \, \hat{\boldsymbol{\alpha}}_{1:t-1}, \, \hat{\theta}_{1:t-1}) = \frac{\mathcal{P}(I_t \,|\, \boldsymbol{\alpha}_t, \theta_t) \, \mathcal{P}(\boldsymbol{\alpha}_t, \theta_t \,|\, \hat{\boldsymbol{\alpha}}_{1:t-1}, \hat{\theta}_{1:t-1})}{\mathcal{P}(I_t \,|\, \hat{\boldsymbol{\alpha}}_{1:t-1}, \hat{\theta}_{1:t-1})}. \tag{38}$$

The key challenge, addressed in the following, is to model the conditional probability

$$\mathcal{P}(\boldsymbol{\alpha}_t, \theta_t \,|\, \hat{\boldsymbol{\alpha}}_{1:t-1}, \hat{\theta}_{1:t-1}), \tag{39}$$

which constitutes the probability for observing a particular shape $\boldsymbol{\alpha}_t$ and a particular transformation θ_t at time t, conditioned on the parameter estimates for shape and transformation obtained on previous images.

5.3   Linear Dynamical Models for Implicit Shapes

For realistic deformable objects one can expect the deformation parameters $\boldsymbol{\alpha}_t$ and the transformation parameters θ_t to be tightly coupled. Yet, we want to learn dynamical shape models which are invariant to the absolute translation, rotation etc. To this end, we can make use of the fact that the transformations form a group which implies that the transformation θ_t at time t can be obtained from the previous transformation θ_{t-1} by applying an incremental transformation $\triangle\theta_t$: $T_{\theta_t} x = T_{\triangle\theta_t} T_{\theta_{t-1}} x$. Instead of learning models of the absolute transformation θ_t, we can simply learn models of the update transformations $\triangle\theta_t$ (e.g. the changes in translation and rotation). By construction, such models are invariant with respect to the global pose or location of the modeled shape.

To jointly model transformation and deformation, we simply obtain for each training shape in the learning sequence the deformation parameters $\boldsymbol{\alpha}_i$ and the transformation

changes $\triangle \theta_i$, and define the *extended shape vector*

$$\boldsymbol{\beta}_t := \begin{pmatrix} \boldsymbol{\alpha}_t \\ \triangle \theta_t \end{pmatrix}. \tag{40}$$

We will then impose a linear dynamical model of order k to approximate the temporal evolution of the extended shape vector:

$$\mathcal{P}(\boldsymbol{\beta}_t \,|\, \hat{\boldsymbol{\beta}}_{1:t-1}) \propto \exp\left(-\frac{1}{2}\,\boldsymbol{v}^\top \Sigma^{-1} \boldsymbol{v}\right), \tag{41}$$

where

$$\boldsymbol{v} \equiv \boldsymbol{\beta}_t - \boldsymbol{\mu} - A_1 \hat{\boldsymbol{\beta}}_{t-1} - A_2 \hat{\boldsymbol{\beta}}_{t-2} \ldots - A_k \hat{\boldsymbol{\beta}}_{t-k} \tag{42}$$

Various methods have been proposed in the literature to estimate the model parameters given by the mean $\boldsymbol{\mu}$ and the transition and noise matrices $A_1, \ldots, A_k, \Sigma$. We applied a stepwise least squares algorithm proposed in [71]. Using dynamical models up to an order of 8, we found that according to Schwarz's Bayesian criterion [92], our training sequences were best approximated by an autoregressive model of second order ($k = 2$).

Figure 15 shows a sequence of statistically synthesized embedding functions and the induced contours given by the zero level line of the respective surfaces – for easier visualization, the transformational degrees are neglected. In particular, the implicit representation allows to synthesize shapes of varying topology. The silhouette on the bottom left of Figure 15, for example, consists of two contours.

5.4 Variational Segmentation with Dynamical Shape Priors

Given an image I_t from an image sequence and given a set of previously segmented shapes with shape parameters $\hat{\boldsymbol{\alpha}}_{1:t-1}$ and transformation parameters $\hat{\theta}_{1:t-1}$, the goal of tracking is to maximize the conditional probability (38) with respect to shape $\boldsymbol{\alpha}_t$ and transformation θ_t. This can be performed by minimizing its negative logarithm, which is – up to a constant – given by an energy of the form:

$$E(\boldsymbol{\alpha}_t, \theta_t) = E_{data}(\boldsymbol{\alpha}_t, \theta_t) + E_{shape}(\boldsymbol{\alpha}_t, \theta_t). \tag{43}$$

For the data term we use the model in (3) with independent intensity variances:

$$E_{data}(\boldsymbol{\alpha}_t, \theta_t) = \int \left(\frac{(I_t - \mu_1)^2}{2\sigma_1^2} + \log \sigma_1\right) H\phi_{\boldsymbol{\alpha}_t, \theta_t} + \left(\frac{(I_t - \mu_2)^2}{2\sigma_2^2} + \log \sigma_2\right)\left(1 - H\phi_{\boldsymbol{\alpha}_t, \theta_t}\right) dx. \tag{44}$$
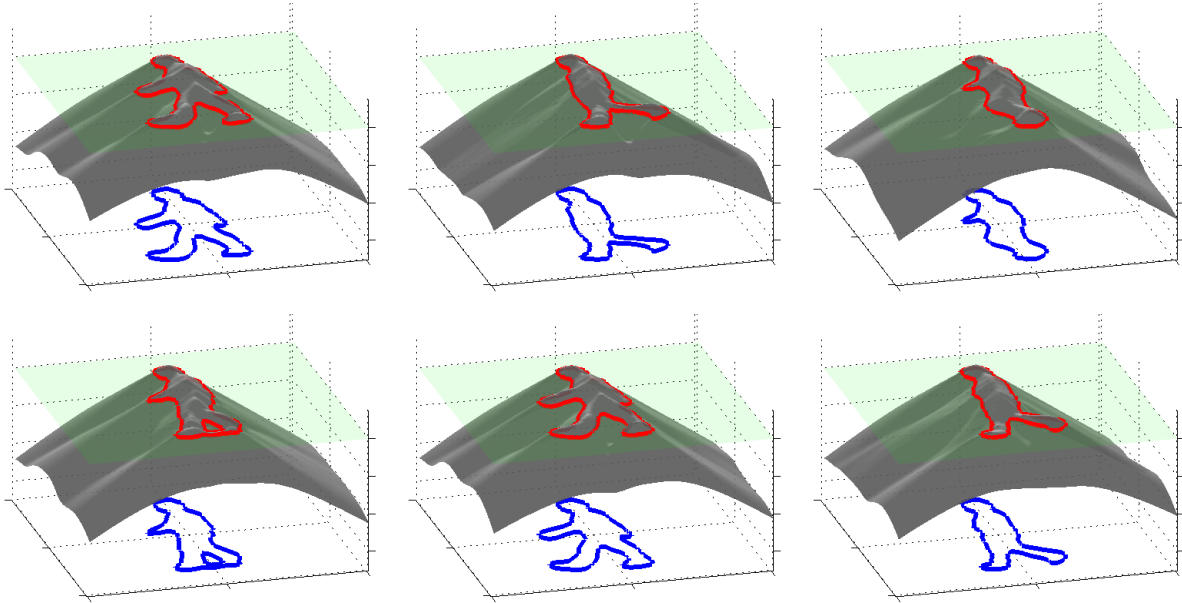
Figure 15: Synthesis of implicit dynamical shapes. Statistically generated embedding surfaces obtained by sampling from a second order autoregressive model, and the contours given by the zero level lines of the synthesized surfaces. The implicit representation allows the embedded contour to change topology (bottom left image).

Using the autoregressive model (41), the shape energy is given by:

$$E_{shape}(\boldsymbol{\alpha}_t, \theta_t) = \frac{1}{2}\, \boldsymbol{v}^\top \Sigma^{-1}\, \boldsymbol{v} \tag{45}$$

with $\boldsymbol{v}$ defined in (42).

The total energy (43) is easily minimized by gradient descent. For details we refer to [26].

Figure 16 shows images from a sequence that was degraded by increasing amounts of noise.

Figure 17 shows segmentation results obtained by minimizing (43) as presented above. Despite prominent amounts of noise, the segmentation process provides reliable segmentations where human observers fail.

Figure 18 shows the segmentation of an image sequence showing a walking person that was corrupted by noise and an occlusion which completely covers the walking person for several frames. The dynamical shape prior allows for reliable segmentations despite noise

25% noise   50% noise   75% noise   90% noise
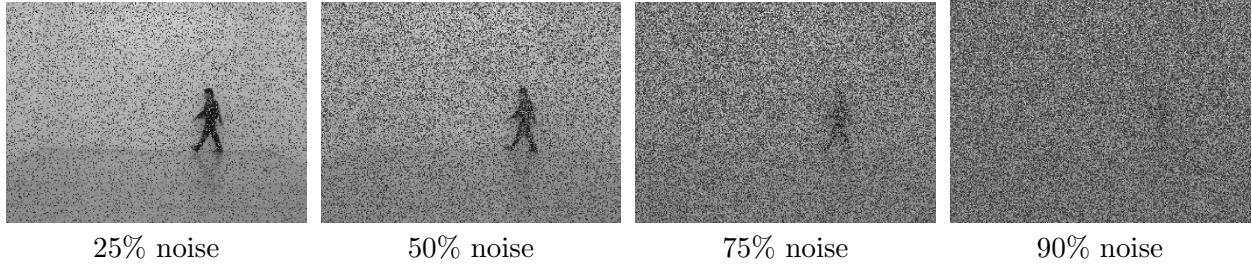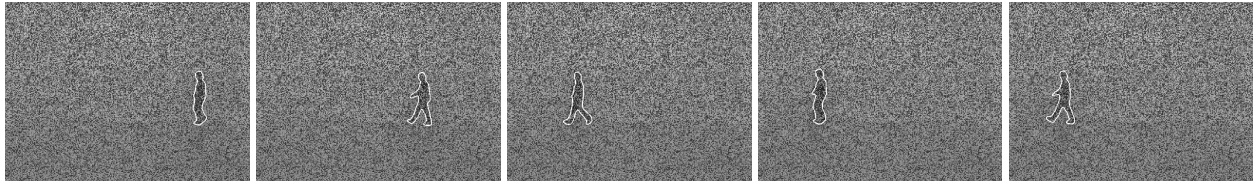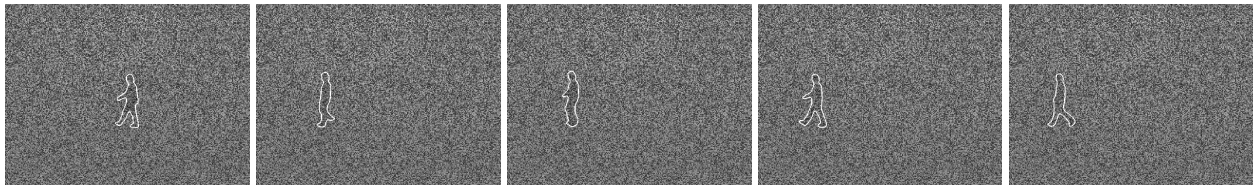
Figure 16: Images from a sequence with increasing amount of noise.

Segmentation results for 75% noise.

Segmentation results for 90% noise.

Figure 17: Variational image sequence segmentation with a dynamical shape prior for various amounts of noise. 90% noise means that nine out of ten intensity values were replaced by a random intensity from a uniform distribution. The statistically learned dynamical model allows for reliable segmentation results despite prominent amounts of noise.

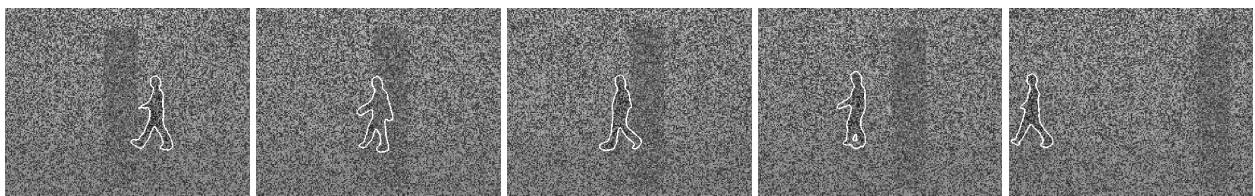and occlusion. For more details and quantitative evaluations we refer to [26].

Figure 18: Tracking in the presence of occlusion. The dynamical shape prior allows to reliably segment the walking person despite noise and occlusion.

6 Parametric Representations Revisited: Combinatorial Solutions for Segmentation with Shape Priors

In previous sections we saw that shape priors allow to improve the segmentation and tracking of familiar deformable objects, biasing the segmentation process to favor familiar shapes or even familiar shape evolution. Unfortunately these approaches are based on locally minimizing respective energies via gradient descent. Since these energies are generally non-convex, respective solutions are bound to be locally optimal only. As a consequence, they depend on an initialization and are likely to be suboptimal in practice. One exception based on implicit shape representations as binary indicator functions and convex relaxation techniques was proposed in [31]. Yet, the linear interpolation of shapes represented by binary indicator functions does not give rise to plausible intermediate shapes such that respective algorithms require a large number of training shapes.
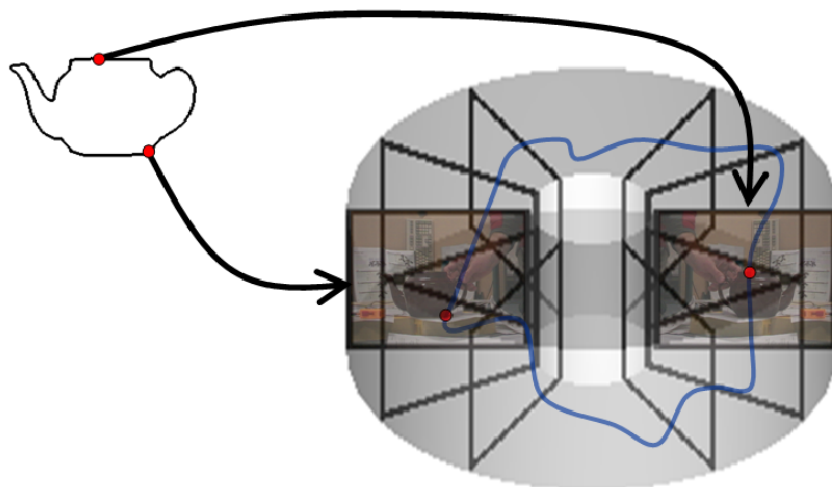
Figure 19: A polynomial-time solution for matching shapes to images: Matching a template curve $\mathcal{C} : \mathbb{S}^1 \to \mathbb{R}^2$ (left) to the image plane $\Omega \subset \mathbb{R}^2$ is equivalent to computing an orientation-preserving cyclic path $\Gamma : \mathbb{S}^1 \to \Omega \times \mathbb{S}^1$ (blue curve) in the product space spanned by the image domain and the template domain. The latter problem can be solved in polynomial time – see [89] for details.

Moreover, while implicit representations like the level set method circumvent the problem of computing correspondences between points on either of two shapes, it is well-known that the aspect of point correspondences plays a vital role in human notions of shape similarity. For matching planar shapes there is abundant literature on how to solve the arising correspondence problem in polynomial time using dynamic programming techniques [48, 93, 85].

Similar concepts of dynamic programming can be employed to localize deformed template curves in images. Coughlan et al. [23] detected open boundaries by shortest path algorithms in higher-dimensional graphs. And Felzenszwalb et al. used dynamic programming in chordal graphs to localize shapes, albeit not on a pixel level.

Polynomial-time solutions for localizing deformable closed template curves in images using minimum ratio cycles or shortest circular paths were proposed in [89], with a further generalization presented in [88]. There the problem of determining a segmentation of an image $I : \Omega \to \mathbb{R}$ that is elastically similar to an observed template $cc : \mathbb{S}^1 \to \mathbb{R}^2$ by computing minimum ratio cycles

$$\Gamma : \mathbb{S}^1 \to \Omega \times \mathbb{S}^1 \tag{46}$$

in the product space spanned by the image domain Ω and template domain $\mathbb{S}^1$. See Figure 19 for a schematic visualization. All points along this circular path provide a pair of corresponding template point and image pixel. In this manner, the matching of template points to image pixels is equivalent to the estimation of orientation-preserving cyclic paths, which can be solved in polynomial time using dynamic programming techniques such as ratio cycles [86] or shortest circular paths [91].

Figure 20 shows an example result obtained with this approach: The algorithm determines a deformed version (right) of a template curve (left) in an image (center) in globally optimal manner. An initialization is no longer required and the best conceivable solution is determined in polynomial time.

Figure 21 shows further examples of tracking objects: Over long sequences of hundreds of frames the objects of interest are tracked reliably – despite low contrast, camera shake, bad visibility and illumination changes. For further details we refer to [89].

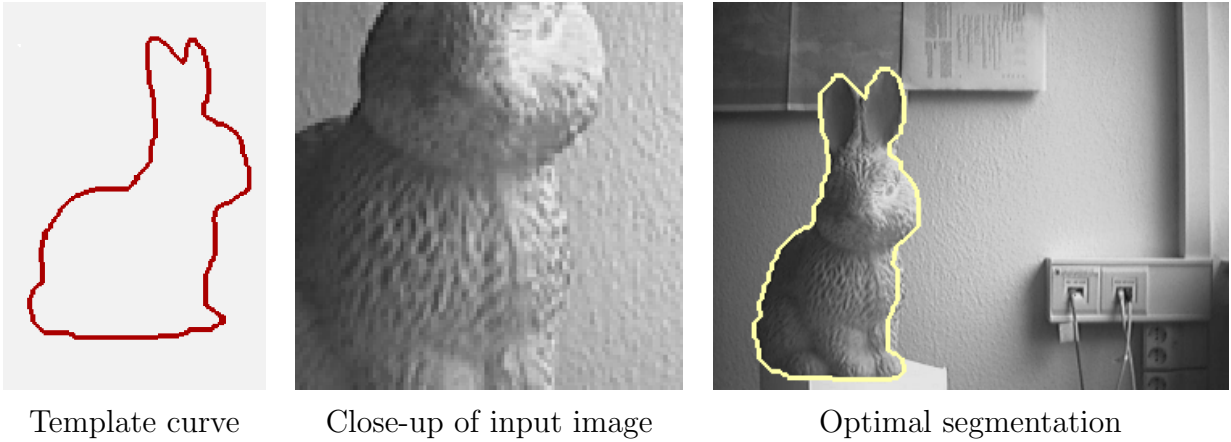| Template curve | Close-up of input image | Optimal segmentation |

Figure 20: Segmentation with a single template: Despite significant deformation and translation, the initial template curve (red) is accurately matched to the low-contrast input image. The globally optimal correspondence between template points and image pixels is computed in polynomial time by dynamic programming techniques [89].

| frame 1 | frame 10 | frame 80 | frame 110 |

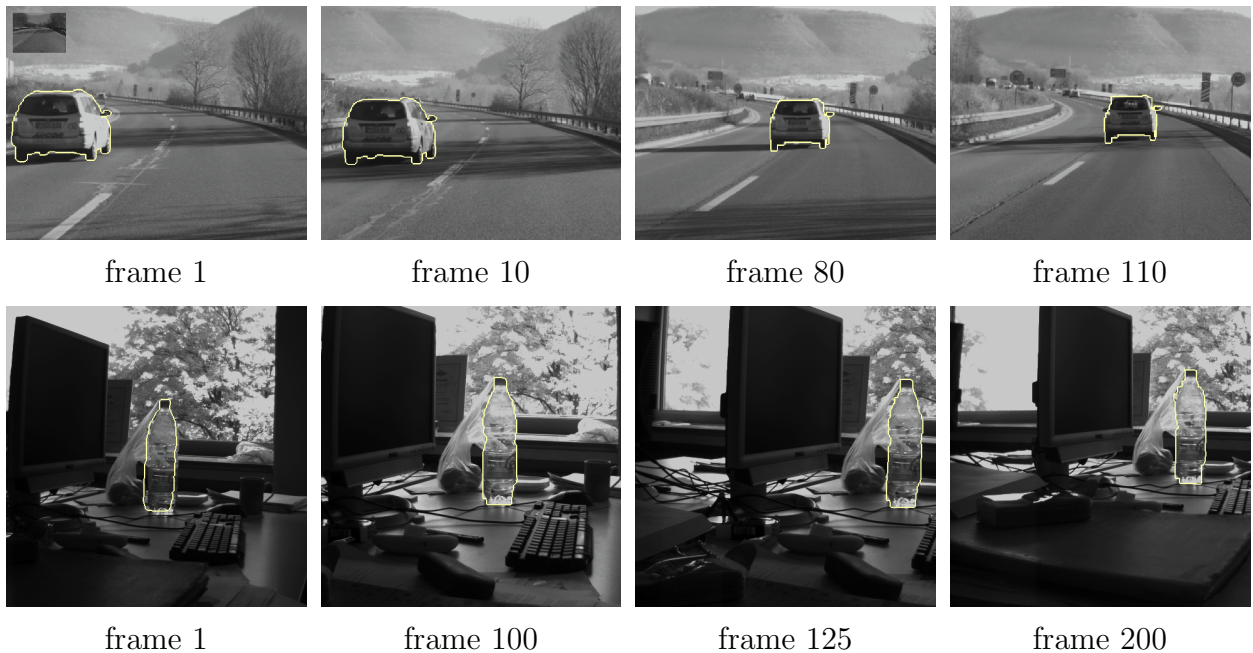| frame 1 | frame 100 | frame 125 | frame 200 |

Figure 21: Tracking of various objects in challenging real-world sequences. [89]. Despite bad visibility, camera shake and substantial lighting changes, the polynomial-time algorithm allows to reliably track objects over hundreds of frames. Image data taken from [89].

7 Conclusion

In the previous sections, we have discussed various ways to impose statistical shape priors into image segmentation methods. We have made several observations:

- By imposing statistically learnt shape information one can generate segmentation processes which favor the emergence of familiar shapes – where familiarity is based on one or several training shapes.

- Statistical shape information can be elegantly combined with the input image data in the framework of Bayesian maximum aposteriori estimation. Maximizing the posterior distribution is equivalent to minimizing a sum of two energies representing the data term and the shape prior. A further generalization allows to impose dynamical shape priors so as to favor familiar deformations of shape in image sequence segmentation.

- While linear Gaussian shape priors are quite popular, the silhouettes of typical objects in our environment are generally not Gaussian distributed. In contrast to linear Gaussian priors, nonlinear statistical shape priors based on Parzen-Rosenblatt kernel density estimators or based on Gaussian distributions in appropriate feature spaces [28] allow to encode a large variety of rather distinct shapes in a single shape energy.

- Shapes can be represented explicitly (as points on the object's boundary or surface) or implicitly (as the indicator function of the interior of the object). They can be represented in a spatially discrete or a spatially continuous setting.

- The choice of shape representation has important consequences regarding the question which optimization algorithms are employed and whether respective energies can be minimized locally or globally. Moreover, different shape representations give rise to different notions of shape similarity and shape interpolation. As a result, there is no single ideal representation of shape. Ultimately one may favor hybrid representations such as the one proposed in [90]. It combines explicit and implicit representations allowing cost functions which represent properties of both the object's interior and its boundary. Subsequent LP relaxation provides minimizers of bounded optimality.

References

[1] A.A. Amini, T.E. Weymouth, and R.C. Jain. Using dynamic programming for solving variational problems in vision. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 12(9):855 – 867, September 1990.

[2] S. P. Awate, T. Tasdizen, and R. T. Whitaker. Unsupervised texture segmentation with nonparametric neighborhood statistics. In *European Conference on Computer Vision (ECCV)*, pages 494–507, Graz, Austria, May 2006. Springer.

[3] A. Blake and M. Isard. *Active Contours.* Springer, London, 1998.

[4] A. Blake and A. Zisserman. *Visual Reconstruction.* MIT Press, 1987.

[5] F. L. Bookstein. *The Measurement of Biological Shape and Shape Change*, volume 24 of *Lect. Notes in Biomath.* Springer, New York, 1978.

[6] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *IEEE Int. Conf. on Computer Vision*, pages 26–33, Nice, 2003.

[7] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 26(9):1124–1137, 2004.

[8] T. Brox, M. Rousson, R. Deriche, and J. Weickert. Unsupervised segmentation incorporating colour, texture, and motion. In N. Petkov and M. A. Westenberg, editors, *Computer Analysis of Images and Patterns*, volume 2756 of *LNCS*, pages 353–360, Groningen, The Netherlands, August 2003. Springer.

[9] T. Brox and J. Weickert. A TV flow based local scale measure for texture discrimination. In T. Pajdla and V. Hlavac, editors, *European Conf. on Computer Vision*, volume 3022 of *LNCS*, pages 578–590, Prague, 2004. Springer.

[10] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. IEEE Intl. Conf. on Comp. Vis.*, pages 694–699, Boston, USA, 1995.

[11] T. Chan, S. Esedoḡlu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006.

[12] T. Chan and W. Zhu. Level set based shape prior segmentation. Technical Report 03-66, Computational Applied Mathematics, UCLA, Los Angeles, 2003.

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

[14] T.F. Chan and L.A. Vese. A level set algorithm for minimizing the Mumford–Shah functional in image processing. In *IEEE Workshop on Variational and Level Set Methods*, pages 161–168, Vancouver, CA, 2001.

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

[16] G. Charpiat, O. Faugeras, J.-P. Pons, and R. Keriven. Generalized gradients: Priors on minimization flows. *Int. J. of Computer Vision*, 73(3):325–344, 2007.

[17] Y. Chen, H. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. S. Gopinath, R. W. Briggs, and E. Geiser. Using shape priors in geometric active contours in a variational framework. *Int. J. of Computer Vision*, 50(3):315–328, 2002.

[18] Y. S. Chow, S. Geman, and L. D. Wu. Consistent cross-validated density estimation. *Annals of Statistics*, 11:25–38, 1983.

[19] R. Cipolla and A. Blake. The dynamic analysis of apparent contours. In *IEEE Int. Conf. on Computer Vision*, pages 616–625. Springer, 1990.

[20] L. Cohen and R. Kimmel. Global minimum for active contour models: A minimal path approach. *Int. J. of Computer Vision*, 24(1):57–78, August 1997.

[21] T. F. Cootes, C. J. Taylor, D. M. Cooper, and J. Graham. Active shape models – their training and application. *Comp. Vision Image Underst.*, 61(1):38–59, 1995.

[22] T.F. Cootes and C.J. Taylor. A mixture model for representing shape variation. *Image and Vision Computing*, 17(8):567–574, 1999.

[23] J. Coughlan, A. Yuille, C. English, and D. Snow. Efficient deformable template detection and localization without user initialization. *Computer Vision and Image Understanding*, 78(3):303–319, 2000.

[24] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 1. Interscience Publishers, Inc., New York, 1953.

[25] D. Cremers. *Statistical Shape Knowledge in Variational Image Segmentation*. PhD thesis, Department of Mathematics and Computer Science, University of Mannheim, Germany, 2002.

[26] D. Cremers. Dynamical statistical shape priors for level set based tracking. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 28(8):1262–1273, August 2006.

[27] D. Cremers, T. Kohlberger, and C. Schnörr. Nonlinear shape statistics in Mumford–Shah based segmentation. In A. Heyden et al., editors, *Europ. Conf. on Comp. Vis.*, volume 2351 of *LNCS*, pages 93–108, Copenhagen, May 2002. Springer.

[28] D. Cremers, T. Kohlberger, and C. Schnörr. Shape statistics in kernel space for variational image segmentation. *Pattern Recognition*, 36(9):1929–1943, 2003.

[29] D. Cremers, S. J. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *Int. J. of Computer Vision*, 69(3):335–351, 2006.

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

[31] D. Cremers, F. R. Schmidt, and F. Barthel. 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, June 2008.

[32] D. Cremers and S. Soatto. A pseudo-distance for shape priors in level set segmentation. In N. Paragios, editor, *IEEE 2nd Int. Workshop on Variational, Geometric and Level Set Methods*, pages 169–176, Nice, 2003.

[33] D. Cremers and S. Soatto. Motion Competition: A variational framework for piecewise parametric motion segmentation. *Int. J. of Computer Vision*, 62(3):249–265, May 2005.

[34] D. Cremers, N. Sochen, and C. Schnörr. Multiphase dynamic labeling for variational recognition-driven image segmentation. In T. Pajdla and V. Hlavac, editors, *European Conf. on Computer Vision*, volume 3024 of *LNCS*, pages 74–86. Springer, 2004.

[35] D. Cremers, N. Sochen, and C. Schnörr. A multiphase dynamic labeling model for variational recognition-driven image segmentation. *Int. J. of Computer Vision*, 66(1):67–81, 2006.

[36] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion Snakes: Introducing statistical shape knowledge into the Mumford–Shah functional. *Int. J. of Computer Vision*, 50(3):295–313, 2002.

[37] P. Deheuvels. Estimation non paramétrique de la densité par histogrammes généralisés. *Revue de Statistique Appliquée*, 25:5–42, 1977.

[38] H. Delingette and J. Montagnat. New algorithms for controlling active contours shape and topology. In D. Vernon, editor, *Proc. of the Europ. Conf. on Comp. Vis.*, volume 1843 of *LNCS*, pages 381–395. Springer, 2000.

[39] A. Dervieux and F. Thomasset. A finite element method for the simulation of Raleigh-Taylor instability. *Springer Lect. Notes in Math.*, 771:145–158, 1979.

[40] L. Devroye and L. Györfi. *Nonparametric Density Estimation. The L1 View.* John Wiley, New York, 1985.

[41] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis.* Wiley, Chichester, 1998.

[42] R. P. W. Duin. On the choice of smoothing parameters for Parzen estimators of probability density functions. *IEEE Trans. on Computers*, 25:1175–1179, 1976.

[43] G. Farin. *Curves and Surfaces for Computer–Aided Geometric Design.* Academic Press, San Diego, 1997.

[44] E. Franchini, S. Morigi, and F. Sgallari. Segmentation of 3d tubular structures by a pde-based anisotropic diffusion model. In *Intl. Conf. on Scale Space and Variational Methods*, volume 5567 of *LNCS*, pages 75–86. Springer, 2009.

[45] M.. Fréchet. Les courbes aléatoires. *Bull. Inst. Internat. Stat.*, 38:499–504, 1961.

[46] K. Fundana, N. C. Overgaard, and A. Heyden. Variational segmentation of image sequences using region-based active contours and deformable shape priors. *Int. J. of Computer Vision*, 80(3):289–299, 2008.

[47] Y. Gdalyahu and D. Weinshall. Flexible syntactic matching of curves and its application to automatic hierarchical classication of silhouettes. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 21(12):1312–1328, 1999.

[48] D. Geiger, A. Gupta, L.A. Costa, and J. Vlontzos. Dynamic programming for detecting, tracking and matching deformable contours. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 17(3):294–302, 1995.

[49] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum *a posteriori* estimation for binary images. *J. Roy. Statist. Soc., Ser. B.*, 51(2):271–279, 1989.

[50] U. Grenander, Y. Chow, and D. M. Keenan. *Hands: A Pattern Theoretic Study of Biological Shapes.* Springer, New York, 1991.

[51] M. Heiler and C. Schnörr. Natural image statistics for natural image segmentation. In *IEEE Int. Conf. on Computer Vision*, pages 1259–1266, 2003.

[52] E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 23:253–258, 1925.

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

[54] D. G. Kendall. The diffusion of shape. *Advances in Applied Probability*, 9:428–430, 1977.

[55] C. Kervrann and F. Heitz. Statistical deformable model-based segmentation of image motion. *IEEE Trans. on Image Processing*, 8:583–588, 1999.

[56] S. Kichenassamy, A. Kumar, P. J. Olver, A. Tannenbaum, and A. J. Yezzi. Gradient flows and geometric active contour models. In *IEEE Int. Conf. on Computer Vision*, pages 810–815, 1995.

[57] J. Kim, J. W. Fisher, A. Yezzi, M. Cetin, and A. Willsky. Nonparametric methods for image segmentation using information theory and curve evolution. In *Int. Conf. on Image Processing*, volume 3, pages 797–800, 2002.

[58] T. Kohlberger, D. Cremers, M. Rousson, and R. Ramaraj. 4d shape priors for level set segmentation of the left myocardium in SPECT sequences. In *Medical Image Computing and Computer Assisted Intervention*, volume 4190 of *LNCS*, pages 92–100, October 2006.

[59] K. Kolev, M. Klodt, T. Brox, and D. Cremers. Continuous global optimization in multview 3d reconstruction. *International Journal of Computer Vision*, 2009.

[60] J.-O. Lachaud and A. Montanvert. Deformable meshes with automated topology changes for coarse-to-fine three-dimensional surface extraction. *Medical Image Analysis*, 3(2):187–207, 1999.

[61] F. Leitner and P. Cinquin. Complex topology 3d objects segmentation. In *SPIE Conf. on Advances in Intelligent Robotics Systems*, volume 1609, Boston, November 1991.

[62] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Int. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 316–323, Hilton Head Island, SC, 2000.

[63] R. Malladi, J. A. Sethian, and B. C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 17(2):158–175, 1995.

[64] G. Matheron. *Random Sets and Integral Geometry*. Wiley & Sons, 1975.

[65] T. McInerney and D. Terzopoulos. Topologically adaptable snakes. In *Proc. 5th Int. Conf. on Computer Vision*, pages 840–845, Los Alamitos, California, June 20–23 1995. IEEE Comp. Soc. Press.

[66] S. Menet, P. Saint-Marc, and G. Medioni. B–snakes: implementation and application to stereo. In *Proc. DARPA Image Underst. Workshop*, pages 720–726, April 6-8 1990.

[67] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philos. Trans. Roy. Soc. London, A*, 209:415–446, 1909.

[68] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 19(7):696–710, 1997.

[69] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.

[70] D. Nain, A. Yezzi, and G. Turk. Vessel segmentation using a shape driven flow. In *MICCAI*, pages 51–59, 2003.

[71] A. Neumaier and T. Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM T. on Mathematical Software*, 27(1):27–57, 2001.

[72] S. J. Osher and J. A. Sethian. Fronts propagation with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations. *J. of Comp. Phys.*, 79:12–49, 1988.

[73] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *Int. J. of Computer Vision*, 46(3):223–247, 2002.

[74] P. Parent and S. W. Zucker. Trace inference, curvature consistency, and curve detection. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 11(8):823–839, 1989.

[75] E. Parzen. On the estimation of a probability density function and the mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.

[76] C.-E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[77] T. Riklin-Raviv, N. Kiryati, and N. Sochen. Unlevel sets: Geometry and prior-based segmentation. In T. Pajdla and V. Hlavac, editors, *European Conf. on Computer Vision*, volume 3024 of *LNCS*, pages 50–61, Prague, 2004. Springer.

[78] M. Rochery, I. Jermyn, and J. Zerubia. Higher order active contours. *Int. J. of Computer Vision*, 2006. To appear.

[79] F. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–837, 1956.

[80] M. Rousson, T. Brox, and R. Deriche. Active unsupervised texture segmentation on a diffusion based feature space. In *Proc. IEEE Conf. on Comp. Vision Patt. Recog.* , pages 699–704, Madison, WI, 2003.

[81] M. Rousson and D. Cremers. Efficient kernel density estimation of shape and intensity priors for level set segmentation. In *MICCAI*, volume 1, pages 757–764, 2005.

[82] M. Rousson and N. Paragios. Shape priors for level set representations. In A. Heyden et al., editors, *Europ. Conf. on Comp. Vis.*, volume 2351 of *LNCS*, pages 78–92. Springer, 2002.

[83] M. Rousson, N. Paragios, and R. Deriche. Implicit active shape models for 3d segmentation in MRI imaging. In *MICCAI*, volume 2217 of *LNCS*, pages 209–216. Springer, 2004.

[84] A. Rosenfeld S. W. Zucker, R. A. Hummel. An application of relaxation labeling to line and curve enhancement. *IEEE Trans. on Computers*, 26(4):394–403, 1977.

[85] F. R. Schmidt, Dirk Farin, and D. Cremers. Fast matching of planar shapes in sub-cubic runtime. In *IEEE Int. Conf. on Computer Vision*, Rio de Janeiro, October 2007.

[86] T. Schoenemann and D. Cremers. Globally optimal image segmentation with an elastic shape prior. In *IEEE Int. Conf. on Computer Vision*, Rio de Janeiro, Brasil, October 2007.

[87] T. Schoenemann and D. Cremers. Introducing curvature into globally optimal image segmentation: Minimum ratio cycles on product graphs. In *IEEE Int. Conf. on Computer Vision*, Rio de Janeiro, October 2007.

[88] T. Schoenemann and D. Cremers. Matching non-rigidly deformable shapes across images: A globally optimal solution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, Alaska, June 2008.

[89] T. Schoenemann and D. Cremers. A combinatorial solution for model-based image segmentation and real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

[90] T. Schoenemann, F. Kahl, and D. Cremers. Curvature regularity for region-based image segmentation and inpainting: A linear programming relaxation. In *IEEE Int. Conf. on Computer Vision*, Kyoto, 2009.

[91] T. Schoenemann, F. R. Schmidt, and D. Cremers. Image segmentation with elastic shape priors via global geodesics in product spaces. In *British Machine Vision Conference*, Leeds, UK, September 2008.

[92] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6:461–464, 1978.

[93] T. Sebastian, P. Klein, and B. Kimia. On aligning curves. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 25(1):116–125, 2003.

[94] J. Serra. *Image Analysis and Mathematical Morophology.* Academic Press, London, 1982.

[95] B. W. Silverman. Choosing the window width when estimating a density. *Biometrika*, 65:1–11, 1978.

[96] B. W. Silverman. *Density estimation for statistics and data analysis.* Chapman and Hall, London, 1992.

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

[98] M. Sussman, Smereka P., and S. J. Osher. A level set approach for computing solutions to incompressible twophase flow. *J. of Comp. Phys.*, 94:146–159, 1994.

[99] A. Tsai, W. Wells, S. K. Warfield, and A. Willsky. Level set methods in an EM framework for shape classification and estimation. In *MICCAI*, 2004.

[100] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, and A. Willsky. Model–based curve evolution technique for image segmentation. In *Comp. Vision Patt. Recog.*, pages 463–468, Kauai, Hawaii, 2001.

[101] A. Tsai, A. J. Yezzi, and A. S. Willsky. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Trans. on Image Processing*, 10(8):1169–1186, 2001.

[102] G. Unal, H. Krim, and A. Y. Yezzi. Information-theoretic active polygons for unsupervised texture segmentation. *Int. J. of Computer Vision*, May 2005.

[103] M. Unger, T. Pock, D. Cremers, and H. Bischof. TVSeg - interactive total variation based image segmentation. In *British Machine Vision Conference (BMVC)*, Leeds, UK, September 2008.

[104] T. J. Wagner. Nonparametric estimates of probability densities. *IEEE Trans. on Inform. Theory*, 21:438–440, 1975.

[105] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 18(9):884–900, 1996.