International Journal of Computer Vision 69(3), 335–351, 2006

DOI: 10.1007/s11263-006-7533-5

VIDEO ONLINE

Kernel Density Estimation and Intrinsic Alignment for Shape Priors in Level Set Segmentation

DANIEL CREMERS
Department of Computer Science, University of Bonn, Germany

STANLEY J. OSHER
Department of Mathematics, University of California at Los Angeles

STEFANO SOATTO
Department of Computer Science, University of California at Los Angeles

Received September 25, 2005; Revised January 25, 2006; Accepted January 31, 2006

First online version published in May, 2006

Abstract. In this paper, we make two contributions to the field of level set based image segmentation. Firstly, we propose shape dissimilarity measures on the space of level set functions which are analytically invariant under the action of certain transformation groups. The invariance is obtained by an intrinsic registration of the evolving level set function. In contrast to existing approaches to invariance in the level set framework, this closed-form solution removes the need to iteratively optimize explicit pose parameters. The resulting shape gradient is more accurate in that it takes into account the effect of boundary variation on the object's pose.

Secondly, based on these invariant shape dissimilarity measures, we propose a statistical shape prior which allows to accurately encode multiple fairly distinct training shapes. This prior constitutes an extension of kernel density estimators to the level set domain. In contrast to the commonly employed Gaussian distribution, such nonparametric density estimators are suited to model aribtrary distributions.

We demonstrate the advantages of this multi-modal shape prior applied to the segmentation and tracking of a partially occluded walking person in a video sequence, and on the segmentation of the left ventricle in cardiac ultrasound images. We give quantitative results on segmentation accuracy and on the dependency of segmentation results on the number of training shapes.

Keywords: image segmentation, shape priors, level set methods, Bayesian inference, alignment, kernel density estimation

1. Introduction

When interpreting a visual scene, it can be advantageous to exploit high-level knowledge about expected objects in order to disambiguate the low-level intensity or color information of a given image. Much research effort has been devoted to integrating prior knowledge into machine-vision algorithms, in particular in the context of image segmentation. In this work, we focus on prior knowledge about the shape of objects of interest.

1.1. *Explicit versus Implicit Shape Representations*

There exist various definitions of the term *shape* in the literature. Kendall (1977) for example defines shape as all the geometrical information that remains when location, scale and rotational effects are filtered out from a collection of point coordinates. For recognition purposes there is no default invariance group. While invariance to affine transformations may be useful for recognizing 3D objects in 2D views, invariance to rotation may be undesirable in other applications: Certain pairs of letters such as "*p*" and "*d*" are identical up to rotation, yet they should not be identified by a character recognition system. In this work, we therefore denote as shape the closed contours defined implicitly as the zero level set of some embedding function $\phi : \mathbb{R}^2 \to \mathbb{R}$. Moreover, we will introduce shape dissimilarity measures for implicit contours which are by construction invariant under the action of certain transformation groups.

Most research on statistical shape modeling has been devoted to explicit contour representations. We refer to the book by Dryden and Mardia (1998) for an overview. The concept of considering shapes as points on an infinite dimensional manifold, representing shape deformations as the action of Lie groups on this manifold and computing statistics on the space of diffeomorphisms was pioneered by Grenander (1976) and Grenander et al. (1991). Some more recent advances were done by Trouvé (1998) and Younes (1998) and by Klassen et al. (2004). Applications of explicit statistical shape models in image segmentation were among others proposed in Grenander et al. (1991), Cootes et al. (1995), Cremers et al. (2002), Mio et al. (2004), Cremers et al. (2003).

A mathematical representation of shape which is independent of parameterization was pioneered in the analysis of random shapes by Fréchet (1961) and in the school of mathematical morphology founded by Matheron (1975). Osher and Sethian introduced the level set method (Osher and Sethian, 1988; Osher and Fedkiw, 2002; Osher and Paragios, 2003) as a means of propagating contours (independent of parameterization) by evolving associated embedding functions via partial differential equations. For a precursor containing some of these ideas we refer to the work of Dervieux and Thomasset (1979).

While we do not claim that implicit boundary representations are superior to explicit ones, we want to clarify why we believe the statistical modeling of implicitly represented shapes to be worth investigating. Explicit parameterizations allow for compact representations of shapes in terms of a few landmarks or control points. They allow to easily define correspondence of parts and the notion of contour shrinking and stretching (cf. Basri et al. (1988) and Gdalyahu and Weinshall (1999)). Yet, factoring out the reparameterization group and identifying an initial point correspondence (when matching shapes) are numerically involved processes (Klassen et al., 2004), especially when generalizing to higher dimensions (surface matching). Moreover, extensions of explicit representations to model multiply-connected objects are not straightforward. Finally, the notion of point-wise correspondence can be introduced into implicit boundary representations as well (Pons et al., 2003). In this work, we therefore adopt the implicit shape representation given by the level set framework.

1.2. Prior Shape Knowledge in Level Set Segmentation

Among variational approaches, the level set method (Osher and Sethian, 1988) has become a popular framework for image segmentation. It has been adapted to segment images based on numerous low-level criteria such as edge consistency (Malladi et al., 1995; Caselles et al., 1995; Kichenassamy et al., 1995), intensity homogeneity (Chan and Vese, 2001; Tsai et al., 2001), texture information (Paragios and Deriche, 2002; Rousson et al., 2003; Heiler and Schnörr, 2003; Brox and Weickert, 2004) and motion information (Cremers and Soatto, 2005).

More recently, it was proposed to integrate prior knowledge about the shape of expected objects into the level set framework. Leventon et al. (2000) suggested to represent a set of training shapes by their signed distance function sampled on a regular grid (of fixed dimension) and to apply principal component analysis (PCA) to this set of training vectors. Subsequently they enhanced a geodesic active contours segmentation process (Caselles et al., 1995; Kichenassamy et al., 1995) by adding a term to the evolution equation which draws the level set function toward the function which is most probable according to the learnt distribution. Tsai et al. (2001) also performed PCA to obtain a set of eigenmodes and subsequently reformulated the segmentation process to directly optimize the parameters associated with the first few deformation modes. Chen et al. (2002) proposed to impose prior knowledge onto the segmenting contour extracted after each iteration of the level set function. While this approach allows to introduce shape information into the segmentation process, it is not entirely in the spirit of the level set scheme since the shape prior acts on the contour and is therefore not capable of modeling topological changes. Rousson and Paragios (2002) and Rousson et al. (2004) impose shape information into the the variational formulation of the level set scheme, either by a model of local (spatially independent) Gaussian fluctuations around a mean level set function or by global deformation modes along the lines of Tsai et al. (2001). An excellent study regarding the equivalence of the topologies induced by three different shape metrics and meaningful extensions of the concepts of sample mean and covariance can be found in the work of Charpiat et al. (2005). More recently, level set formulations were proposed which allow to impose dynamical shape priors (Cremers, 2006) and concepts of tracking (Rathi et al., 2005), to apply shape knowledge selectively in certain image regions (Cremers, Sochen and Schnörr, 2003; Chan and Zhu, 2003), or to impose multiple competing shape priors so as to simultaneously reconstruct several independent objects in a given image sequence (Cremers et al., 2006).

The above approaches allow to improve the level set based segmentation of corrupted images of familiar objects. Yet, existing methods to impose statistical shape information on the evolving embedding function suffer from three limitations:

- The existing statistical models are based on the assumption that the training shapes are distributed according to a Gaussian distribution. As shown in Cremers et al. (2003), this assumption is rather limiting when it comes to modeling more complex shape distributions such as the various silhouettes of a 3D object. Figure 4 shows a set of sample shapes and Figure 5 the density estimated for a 3D projection of the corresponding embedding functions: clearly the projected data do not form a Gaussian distribution. Moreover, as shown in Charpiat et al. (2005), notions such as the *empirical mean shape* of a set of shapes are not always uniquely defined.
- They commonly work under the assumption that shapes are represented by signed distance functions (cf. Leventon et al. (2000) and Rousson et al. (2004)). Yet, for a set of training shapes encoded by their signed distance function, neither the mean level set function nor the linear combination of eigenmodes will in general correspond to a signed distance function, since the space of signed distance functions is not a linear space.[1]
- Invariance of the shape prior with respect to certain transformations is introduced by adding a set of explicit parameters and numerically optimizing the segmentation functional by gradient descent (Chen et al., 2002; Rousson and Paragios, 2002; Yezzi and Soatto, 2003). This iterative optimization not only requires a delicate tuning of associated gradient descent time step sizes (in order to guarantee a stable evolution). It is also not clear in what order and how frequently one is to alternate between the various gradient descent evolutions. In particular, we found in experiments that the order of updating the different pose parameters and the level set function affects the resulting segmentation process.

1.3. Contributions

In this paper, we are building up on the above developments and propose two contributions in order to overcome the discussed limitations:

- We introduce invariance of the shape prior to certain transformations by an intrinsic registration of the evolving level set function. By evaluating the evolving level set function not in global coordinates, but in coordinates of an intrinsic reference frame attached to the evolving surface, we obtain shape distances which are by construction invariant. Such a closed-form solution removes the need to iteratively update local estimates of explicit pose parameters. Moreover, we will argue that this approach is more accurate because the resulting shape gradient contains an additional term which accounts for the effect of boundary variation on the pose of the evolving shape.
- We propose a statistical shape prior by introducing the concept of kernel density estimation (Rosenblatt, 1956; Parzen, 1962) to the domain of level set based shape representations. In contrast to existing approaches of shape priors in level set segmentation (which are based on the assumption of a Gaussian distribution), this prior allows to well approximate arbitrary distributions of shapes. Moreover, our formulation does not require the embedding function to be a signed distance function. Numerical results demonstrate our method applied to the segmentation of a partially occluded walking person.

While numerical results are only computed in two dimensions, the methods developed in this paper can be applied in higher dimensions as well.

The organization of the paper is as follows: In Section 2, we briefly review the level set scheme for the two-phase Mumford-Shah functional (Chan and Vese, 2001; Tsai et al., 2001). In Section 3, we review and discuss dissimilarity measures for two shapes represented by level set functions. In Section 4, we review existing approaches to model pose invariance and introduce a solution to induce invariance by intrinsic alignment. In Section 5, we detail the computation of the Euler-Lagrange equations associated with the proposed invariant shape dissimilarity measures. We demonstrate the invariance properties and the effect of the additionally emerging terms in the shape gradient on the segmentation of a human silhouette. In Section 6, we introduce a novel (multi-modal) statistical shape prior by extending the concept of non-parametric kernel density estimation to the domain of level set based shape representations. In Section 7, we formulate level set segmentation as a problem of Bayesian inference in order to integrate the proposed shape distribution as a prior on the level set function. In Section 8, we provide qualitative and quantitative results of applying the nonparametric shape prior to the segmentation of a partially occluded walking person in a video sequence and to the segmentation of the left ventricle in cardiac ultrasound images. Preliminary results of this work were presented at a conference (Cremers et al., 2004).

2. Level Set Segmentation

Originally introduced in the community of computational physics as a means of propagating interfaces (Osher and Sethian, 1988), the level set method has become a popular framework for image segmentation (Malladi et al., 1995; Caselles et al., 1995; Kichenassamy et al., 1995). The central idea is to implicitly represent a contour C in the image plane $\Omega \subset \mathbb{R}^2$ as the zero-level of an embedding function $\phi : \Omega \to \mathbb{R}$:

$$C = \{x \in \Omega \mid \phi(x) = 0\} \tag{1}$$

Rather than directly evolving the contour C, one evolves the level set function ϕ. The two main advantages are that firstly one does not need to deal with control or marker points (and respective regridding schemes to prevent overlapping). And secondly, the embedded contour is free to undergo topological changes such as splitting and merging which makes it well-suited for the segmentation of multiple or multiply-connected objects.

In the present paper, we use a level set formulation of the piecewise constant Mumford-Shah functional (Mumford and Shah, 1989; Tsai et al., 2001; Chan and Vese, 2001). In particular, a two-phase segmentation of an image $I : \Omega \to \mathbb{R}$ can be generated by minimizing the functional (Chan and Vese, 2001):

$$\begin{aligned} E_{cv}(\phi) = & \int_\Omega (I - u_+)^2 H\phi(x)dx \\ & + \int_\Omega (I - u_-)^2 \big(1 - H\phi(x)\big)dx \\ & + \nu \int_\Omega |\nabla H\phi| dx, \end{aligned} \tag{2}$$

with respect to the embedding function ϕ. Here $H\phi \equiv H(\phi)$ denotes the Heaviside step function and u_+ and u_- represent the mean intensity in the two regions where ϕ is positive or negative, respectively. For related computations based on the use of the Heaviside function, we refer to Zhao et al. (1996). While the first two terms in (2) aim at minimizing the gray value variance in the separated phases, the last term enforces a minimal length of the separating boundary. Gradient descent with respect to ϕ amounts to the evolution equation:

$$\begin{aligned} \frac{\partial\phi}{\partial t} = -\frac{\partial E_{cv}}{\partial\phi} = \delta_\epsilon(\phi)\Bigg[& \nu \,\mathrm{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - (I - u_+)^2 \\ & + (I - u_-)^2 \Bigg]. \end{aligned} \tag{3}$$

Chan and Vese (2001) propose a smooth approximation δ_ϵ of the delta function which allows the detection of interior boundaries.

In the corresponding Bayesian interpretation, the length constraint given by the last term in (2) corresponds to a prior probability which induces the segmentation scheme to favor contours of minimal length. But what if we have more informative prior knowledge about the shape of expected objects? Building up on recent advances (Leventon et al., 2000; Tsai et al., 2001; Chen et al., 2002; Rousson and Paragios, 2002; Cremers, Sochen and Schnörr, 2003; Cremers and Soatto, 2003; Charpiat et al., 2005; Cremers et al., 2006; Chan and Zhu, 2003) and on classical methods of non-parametric density estimation (Rosenblatt, 1956; Parzen, 1962), we will in the following construct a shape prior which statistically approximates an arbitrary distribution of training shapes (without making the restrictive assumption of a Gaussian distribution).

3. 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 Sussman et al. (1994).

Given two shapes encoded by their signed distance functions ϕ_1 and ϕ_2, a simple measure of their dissimilarity is given by their L_2-distance in Ω:

$$\int_\Omega (\phi_1 - \phi_2)^2 \, dx. \tag{4}$$

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.

In Rousson and Paragios (2002), it was proposed to constrain the integral to the domain where ϕ_1 is positive:

$$d_1^2(\phi_1, \phi_2) = \int_\Omega (\phi_1 - \phi_2)^2 \; H\phi_1(x)\, dx, \tag{5}$$

where $H\phi$ again denotes the Heaviside step function. As shown in Cremers and Soatto (2003), this measure can be further improved by normalizing with respect to the area where ϕ_1 is positive and by symmetrizing with respect to the exchange of ϕ_1 and ϕ_2. The resulting dissimilarity measure,

$$d_{1s}^2(\phi_1, \phi_2) = \int_\Omega (\phi_1 - \phi_2)^2 \, \frac{h\phi_1 + h\phi_2}{2} \, dx, \quad \text{with} \quad h\phi \equiv \frac{H\phi}{\int H\phi \, dx}, \tag{6}$$

constitutes a pseudo-distance on the space of signed distance functions (Cremers and Soatto, 2003).

Although the requirement of symmetry may appear to be a theoretical formality, such symmetry considerations can have very relevant practical implications. In particular, asymmetric measures of the form 5 do not allow to impose prior shape information outside the evolving shape (i.e. in areas where $\phi_1 < 0$). Figure 1 shows an example of two circles which only differ by the fact that the second shape has a spike. The measure (5) gives the same distance between the two shapes, no matter how long the spike is, because it only takes into account shape discrepancy inside the first shape. In contrast, the symmetric variant (6) also takes into account shape discrepancies within the second shape. It gives a more informative measure of the shape dissimilarity and therefore allows for more powerful shape priors.

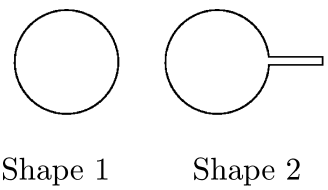

Figure 1. A shape comparison for which the asymmetric shape dissimilarity measures (5) and (7) fail (see text).

Alternatively (Bresson et al., 2003) one can constrain the integration in (4) to the contour C_1 represented by ϕ_1 (i.e. to the area where $\phi = 0$):

$$d_2^2(\phi_1, \phi_2) = \oint_{C_1} \phi_2^2 \, dC_1 = \int_\Omega \phi_2^2(x)\, \delta(\phi_1) |\nabla\phi_1| \, dx. \tag{7}$$

Due to the definition of the signed distance function, this measure corresponds to the distance of the closest point on the contour C_2 (given by $|\phi_2|$) integrated over the entire contour C_1. As with Eq. (5), this measure suffers from not being symmetric. The measure in (7) for example will only take into account points of contour C_2 which are sufficiently close to contour C_1, distant (and possibly disconnected) components of C_2 will be ignored. A symmetric variant of (7) is given by:

$$\begin{aligned} d_{2s}^2(\phi_1, \phi_2) &= \oint_{C_1} \phi_2^2 \, dC_1 + \oint_{C_2} \phi_1^2 \, dC_2 \\ &= \int_\Omega \phi_2^2(x)\, |\nabla H\phi_1| + \phi_1^2(x)\, |\nabla H\phi_2| \, dx. \end{aligned}$$

Further normalization with respect to the contour length is conceivable.

A third variant to compute the dissimilarity of two shapes represented by their embedding functions ϕ_1 and ϕ_2 is to compute the area of the set symmetric difference, as was proposed in Chan and Zhu (2003), Riklin-Raviv et al. (2004), Charpiat et al. (2005):

$$d^2(\phi_1, \phi_2) = \int_\Omega (H\phi_1(x) - H\phi_2(x))^2 \, dx. \quad (8)$$

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 (8) defines a distance: 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 (Charpiat et al., 2005) that L^∞ and $W^{1,2}$ norms on the signed distance functions induce equivalent topologies as the metric (8). Since the distance (8) is not differentiable, we will in practice consider an approximation of the Heaviside function H by a smooth (differentiable) version H_ϵ.

4. Invariance by Intrinsic Alignment

One can make use of the shape distance (8) 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 is in our case the Chan-Vese functional (2). Minimizing the total energy

$$\begin{aligned} E_{total}(\phi) &= E_{cv}(\phi) + \alpha \, E_{shape}(\phi) \\ &= E_{cv}(\phi) + \alpha \, d^2(\phi, \phi_0) \end{aligned} \quad (9)$$

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.

4.1. *Iterative Optimization of Explicit Parameters*

A common approach to introduce invariance (c.f. Chen et al. (2002), Rousson and Paragios (2002), Cremers et al. (2006)) 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:

$$\begin{aligned} d^2(\phi \, , \phi_0, \mu, \theta, \sigma) \\ = \int_\Omega (H(\phi(\sigma R_\theta(x - \mu))) - H\phi_0(x))^2 \, dx. \end{aligned} \quad (10)$$

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

- Optimization of the shape energy (10) 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 (10) with respect to ϕ should take into account how the optimal transformation parameters $\mu(\phi)$, $\sigma(\phi)$ and $\theta(\phi)$ vary with ϕ.

In order to eliminate these difficulties associated with the local optimization of explicit transformation parameters, we will in the following present an alternative approach to obtain invariance. We will show that invariance can be achieved analytically by 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. *Translation Invariance by Intrinsic Alignment*

Assume that the template shape represented by ϕ_0 is aligned with respect to its center of gravity. Then we define a shape energy by:

$$\begin{aligned} E_{shape}(\phi) &= d^2(\phi, \phi_0) \\ &= \int_\Omega (H\phi(x + \mu_\phi) - H\phi_0(x))^2 \, dx, \end{aligned} \quad (11)$$

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

$$\mu_\phi = \int x\, h\phi\, dx, \quad \text{with } h\phi \equiv \frac{H\phi}{\int_\Omega H\phi\, dx}. \quad (12)$$

This intrinsic alignment guarantees that the distance (11) is invariant to the location of the shape ϕ. In contrast to the shape energy (10), we no longer need to iteratively update an estimate of the location of the object of interest.

4.3. *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 translation and scaling:

$$\begin{aligned} E_{shape}(\phi) &= d^2(\phi, \phi_0) \\ &= \int_\Omega (H\phi(\sigma_\phi\, x + \mu_\phi) - H\phi_0(x))^2\, dx, \end{aligned} \quad (13)$$

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:

$$\sigma_\phi = \left(\int (x-\mu)^2\, h\phi\, dx \right)^{\frac{1}{2}}, \quad \text{where } h\phi = \frac{H\phi}{\int_\Omega H\phi\, dx}. \quad (14)$$

In the following, we will show that functional (13) 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:

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

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

The shape energy (13) evaluated for $\tilde{\phi}$ is given by:

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

Therefore, the proposed 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 (10). 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 in the present work. For explicit contour representations, an analogous intrinsic alignment with respect to similarity transformation was proposed in Cremers et al. (2002).

5. Euler-Lagrange Equations for Nested Functions

The two energies (11) and (13) derive their invariance from the fact that ϕ is evaluated in coordinates relative to its own location and scale. In a knowledge-driven segmentation process, one can maximize the similarity of the evolving shape encoded by ϕ and the template shape ϕ_0 by locally minimizing one of the two shape energies. For the sake of differentiability, we will approximate the Heaviside functions H by smoothed approximations H_δ as was done in the work of Chan and Vese (2001)—see Eq. (3).

The associated shape gradient is particularly interesting since the energies (11) and (13) exhibit a nested dependence on ϕ via the moments μ_ϕ and σ_ϕ. In the following, we will detail the computation of the corresponding Gâteaux derivatives for the two invariant energies introduced above.

5.1. Shape Derivative of the Translation Invariant Distance

The gradient of energy (11) with respect to ϕ in direction of an arbitrary deviation $\tilde{\phi}$ is given by the Gâteaux derivative:

$$\begin{aligned}\left.\frac{\partial E}{\partial \phi}\right|_{\tilde{\phi}} &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}(E(\phi+\epsilon\tilde{\phi})-E(\phi)),\\ &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_{\Omega}(H(\phi+\epsilon\tilde{\phi})(x+\mu_{\phi+\epsilon\tilde{\phi}})-H\phi_0(x))^2\\ &\quad -(H(\phi)(x+\mu_{\phi})-H\phi_0(x))^2dx. \end{aligned}\tag{15}$$

With the short-hand notation $\delta\phi \equiv \delta(\phi)$, the effect of the shape variation on the center of gravity is given by:

$$\begin{aligned}\mu_{\phi+\epsilon\tilde{\phi}} &= \int xh(\phi+\epsilon\tilde{\phi})dx = \frac{\int x(H\phi+\epsilon\tilde{\phi}\,\delta\phi)\,dx}{\int(H\phi+\epsilon\tilde{\phi}\,\delta\phi)\,dx}\\ &= \mu_\phi + \frac{\epsilon}{\int H\phi dx}\int (x-\mu_\phi)\tilde{\phi}\,\delta\phi\,dx + \mathcal{O}(\epsilon^2),\end{aligned}\tag{16}$$

Inserting (16) into (15) and further linearization in ϵ leads to a directional shape derivative of the form:

$$\begin{aligned}\left.\frac{\partial E}{\partial \phi}\right|_{\tilde{\phi}} = \; & 2\int (H\phi(\bar{x})-H\phi_0(x))\delta\phi(\bar{x})\\ &\left[\tilde{\phi}(\bar{x})+\nabla\phi(\bar{x})\frac{1}{\int H\phi dx'}\right.\\ &\left.\times\int(x'-\mu_\phi)\tilde{\phi}(x')\delta\phi(x')\,dx'\right]dx,\end{aligned}\tag{17}$$

where $\bar{x} = x+\mu_\phi$ denotes the coordinates upon centering. We therefore deduce that the shape gradient for the translation-invariant energy (11) is given by:

$$\begin{aligned}\frac{\partial E}{\partial \phi} &= 2\,\delta\phi(x)\bigg(\Big(H\phi(x)-H\phi_0(x-\mu_\phi)\Big)+\frac{(x-\mu_\phi)^\top}{\int H\phi\,dx}\\ &\quad\times\int\Big(H\phi(x')-H\phi_0(x'-\mu_\phi)\Big)\delta\phi(x')\nabla\phi(x')\,dx'\bigg).\end{aligned}\tag{18}$$

Let us make several remarks in order to illuminate this result:

- As for the image-driven flow in (3), the entire expression in (18) is weighted by the δ-function which stems from the fact that the function E in (11) only depends on $H\phi$.
- In a gradient descent evolution, the first of the two terms in (18) will draw $H\phi$ to the template $H\phi_0$, transported to the local coordinate frame associated with ϕ.
- The second term in (18) results from the ϕ-dependency of μ_ϕ in (11). It compensates for shape deformations which merely lead to a translation of the center of gravity μ_ϕ. Not surprisingly, this second term contains an integral over the entire domain because the center of gravity is an integral quantity. Figure 2 demonstrates that when applied as a shape prior in a segmentation process, this additional term tends to facilitate the translation of the evolving shape. While the boundary evolution represented in the top row was obtained using the first term of gradient (18) only, the contour flow shown in the bottom row exploits the full shape gradient. The additional term speeds up the convergence (cf. the segmentations obtained after 140 iterations) and generates a more accurate segmentation.

5.2. Shape derivative of the Translation and Scale Invariant Distance

The above computation of a translation invariant shape gradient can be extended to the functional (13). An infinitesimal variation of the level set function ϕ in direction $\tilde{\phi}$ affects the scale σ_ϕ defined in (14) as follows:

$$\begin{aligned}\sigma_{\phi+\epsilon\tilde{\phi}} &= \left(\int(x-\mu_{\phi+\epsilon\tilde{\phi}})^2 h(\phi+\epsilon\tilde{\phi})dx\right)^{\frac{1}{2}}\\ &= \sigma_\phi + \frac{\epsilon}{2\,\sigma_\phi\int H\phi\,dx}\\ &\quad\times\int\Big((x-\mu_\phi)^2-\sigma_\phi^2\Big)\tilde{\phi}\,\delta\phi\,dx+\mathcal{O}(\epsilon^2).\end{aligned}$$

This expression is inserted into the definition (15) of the shape gradient for the shape energy (13). Further linearization in ϵ gives a directional shape derivative of the form:

$$\begin{aligned}\left.\frac{\partial E}{\partial \phi}\right|_{\tilde{\phi}} &= 2\int(H\phi(\bar{x})-H\phi_0(x))\delta\phi(\bar{x})\\ &\quad\times\bigg(\tilde{\phi}(\bar{x})+\nabla\phi(\bar{x})\bigg[\frac{1}{\int H\phi\,dx'}\end{aligned}\tag{19}$$

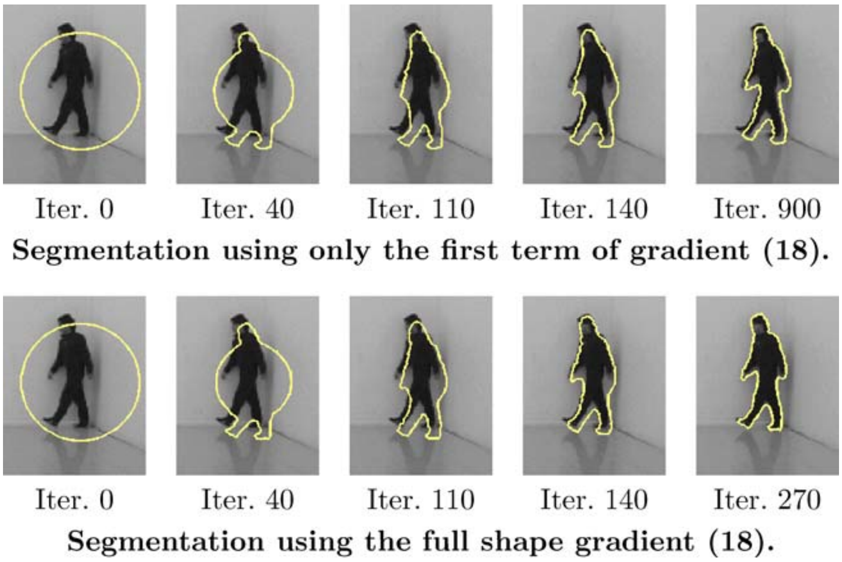

Figure 2. Effect of the additional term in the shape gradient. Segmentation of a human silhouette obtained by minimizing (9), a weighted sum of the data term (2) and a translation-invariant shape prior of the form (11) encoding the given silhouette. The **top row** is obtained by merely using the first term of the shape gradient in (18): The contour does not converge to the desired solution. In contrast, the **bottom row** is obtained by using the full shape gradient, including the second term which is due to the ϕ-dependence of the descriptor μ_ϕ in (11). For the specific choice of parameters (kept constant for the two experiments), including the additional term both speeds up the convergence (cf. the results after 140 iterations) and improves the segmentation.

$$\times \int (x' - \mu_\phi)\, \tilde{\phi}\, \delta\phi\, dx' + \frac{x}{2\sigma_\phi \int H\phi\, dx'}$$
$$\times \int \left((x' - \mu_\phi)^2 - \sigma_\phi^2\right) \tilde{\phi}\, \delta\phi\, dx' \Big]\Big) dx,$$

where $\bar{x} = \sigma_\phi x + \mu_\phi$. Since this directional shape gradient corresponds to a projection of the full shape gradient onto the respective direction $\tilde{\phi}$, i.e.

$$\left.\frac{\partial E}{\partial \phi}\right|_{\tilde{\phi}} = \int \tilde{\phi}(x)\, \frac{\partial E}{\partial \phi}(x)\, dx,$$

we need to rearrange the integrals in order to read off the shape gradient:

$$\begin{aligned}\frac{\partial E}{\partial \phi} &= \frac{2}{\sigma_\phi}\, \delta\phi(x) \bigg(\frac{1}{\sigma_\phi}(H\phi(x) - H\phi_0(Tx)) + \frac{(Tx)^\top}{\int H\phi\, dx} \\ &\quad \times \int (H\phi(x') - H\phi_0(Tx'))\delta\phi(x')\nabla\phi(x')\, dx' \\ &\quad + \frac{(Tx)^2 - 1}{2\int H\phi\, dx} \int (H\phi(x') - H\phi_0(Tx')) \\ &\quad \times \delta\phi(x')(Tx')^\top \nabla\phi(x')\, dx' \bigg),\end{aligned} \tag{20}$$

where $Tx \equiv \frac{x - \mu_\phi}{\sigma_\phi}$ denotes the transformation into the local coordinate frame associated with ϕ. The three terms in the shape gradient (20) can be interpreted as follows:

- The first term draws the contour toward the boundary of the familiar shape represented by ϕ_0, transported to the intrinsic coordinate frame of the evolving function ϕ.
- The second term results from the ϕ-dependency of μ_ϕ. It compensates for deformations which merely result in a shift of the center of gravity.
- The third term stems from the ϕ-dependency of σ_ϕ. Analogous to the second term, it compensates for variations of ϕ which merely lead to changes in the scale σ_ϕ.

To demonstrate the scale-invariant property of the shape energy (13), we applied the segmentation scheme to an image of a partially occluded human silhouette, observed at three different scales. Figure 3 shows the contour evolutions generated by minimizing the total energy (9) with the translation and scale invariant shape energy (13), where ϕ_0 is the level set function associated with a normalized (centered and rescaled) version

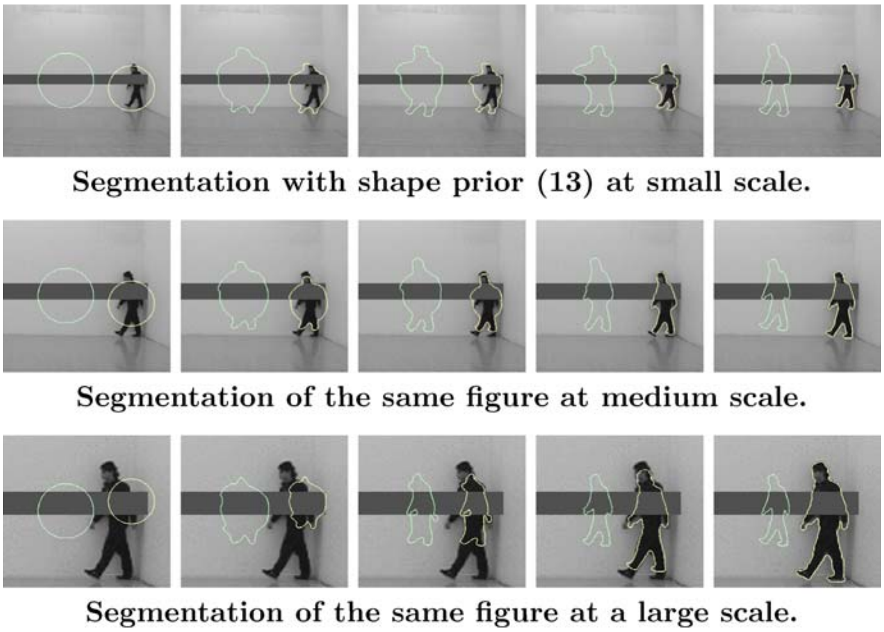

Figure 3. Invariance with respect to scaling and translation. Segmentation of a partially occluded human silhouette obtained by minimizing (9), a weighted sum of the data term (2) and the invariant shape energy (13) encoding the given silhouette. For the three experiments, we kept all involved parameters constant. Due to the analytic invariance of the shape energy to translation and scaling, there is no need to numerically optimize explicit pose parameters in order to reconstruct the object of interest at arbitrary scale and location.

Figure 4. Selected sample shapes from a set of a walking silhouettes.

of the silhouette of interest. The results demonstrate that for the same (fixed) set of parameters, the shape prior enables the reconstruction of the familiar silhouette at arbitrary location and scale. For a visualization of the intrinsic alignment process, we also plotted the evolving contour in the normalized coordinate frame (left). In these normalized coordinates the contour converges to essentially the same solution in all three cases.

6. 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. We refer to Cremers et al. (2003) for such an example.
- The object of interest may be one object out of a class of similar objects (the class of cars or the class of tree leaves). Given a limited number of training shapes sampled from the class, a useful shape energy should

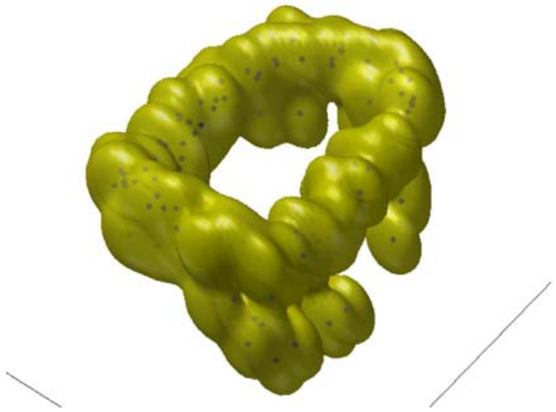

Figure 5. Density estimated for a 3D projection of 100 silhouettes (see Figure 4).

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 many cases, possibly because the camera frame rate is low compared to the speed of the moving hand or person, one is not able to extract a model of the temporal succession of silhouettes. In this paper, 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: Figure 4, shows such sample views for the case of a walking person. In the following, we will demonstrate that—without being able to construct 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) (Cootes et al., 1995) and the related Gaussian distribution (Cremers et al., 2002), by mixture models (Cootes and Taylor, 1999) or nonlinear (multi-modal) representations via simple models in appropriate feature spaces (Cremers et al., 2003).

For level set based shape representations, it was suggested (Leventon et al., 2000; Tsai et al., 2001; Rousson et al., 2004) 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 (Rousson and Paragios, 2002). 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 $\{\phi_i\}_{i=1\ldots N}$ – such as those shown in Figure 4—we define a probability density on the space of signed distance functions by integrating the shape distances (11) or (13) in a Parzen-Rosenblatt kernel density estimator (Rosenblatt, 1956; Parzen, 1962):

$$\mathcal{P}(\phi) \propto \frac{1}{N} \sum_{i=1}^{N} \exp\left(-\frac{1}{2\sigma^2}\, d^2(H\phi, H\phi_i)\right). \quad (21)$$

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. For a general formalism to model probability densities on infinite-dimensional spaces, we refer the reader to the theory of Gaussian processes (Rasmussen and Williams, 2006). In our case, an extension to infinite-dimensional objects such as level set surfaces $\phi : \Omega \to \mathbb{R}$ could be tackled by considering discrete (finite-dimensional) approximations $\{\phi_{ij} \in \mathbb{R}\}_{i=1,\ldots,N,j=1,\ldots,M}$ 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 (Rousson and Cremers, 2005).

There exist extensive studies on how to optimally choose the kernel width σ, based on asymptotic expansions such as the parametric method (Deheuvels, 1977), heuristic estimates (Wagner, 1975; Silverman, 1978) or maximum likelihood optimization by cross validation (Duin, 1976; Chow et al., 1983). We refer to Devroye and Györfi (1985); Silverman (1992) for a detailed discussion. For this work, we simply fix σ^2 to be the mean squared nearest-neighbor distance:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \min_{j \neq i} d^2(H\phi_i, H\phi_j). \qquad (22)$$

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 5, 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 (Devroye and Györfi, 1985; Silverman, 1992). 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.

7. Knowledge-driven Segmentation

In the following, we will detail how the statistical distribution (21) can be used to enhance level set based segmentation process. To this end, we formulate level set segmentation as a problem of Bayesian inference, where the segmentation is obtained by maximizing the conditional probability

$$\mathcal{P}(\phi \mid I) = \frac{\mathcal{P}(I \mid \phi)\, \mathcal{P}(\phi)}{\mathcal{P}(I)}, \qquad (23)$$

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[2]:

$$E(\phi) = \frac{1}{\alpha} E_{cv}(\phi) \;+\; E_{shape}(\phi), \qquad (24)$$

with a positive weighting factor α and the shape energy

$$E_{shape}(\phi) = -\log \mathcal{P}(\phi), \qquad (25)$$

Minimizing the energy (24) 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 (21).

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

$$\frac{\partial \phi}{\partial t} = -\frac{1}{\alpha} \frac{\partial E_{cv}}{\partial \phi} - \frac{\partial E_{shape}}{\partial \phi}, \qquad (26)$$

with the image-driven component of the flow given in (3) and the knowledge-driven component is given by:

$$\frac{\partial E_{shape}}{\partial \phi} = \frac{\sum \alpha_i \frac{\partial}{\partial \phi} d^2(H\phi, H\phi_i)}{2\sigma^2 \sum \alpha_i}, \qquad (27)$$

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

$$\alpha_i = \exp\left(-\frac{1}{2\sigma^2}\, d^2(H\phi, H\phi_i)\right), \qquad (28)$$

which decays exponentially with the distance from the training shape ϕ_i. The invariant shape gradient

$\frac{\partial}{\partial\phi} d^2(H\phi, H\phi_i)$ is given by the expression (18) or (20), respectively.

8. Experimental Results

8.1. Tracking a Walking Person

In the following we apply the proposed shape prior to the segmentation of a partially occluded walking person. To this end, a sequence of a dark figure walking in a (fairly bright) squash court was recorded. We subsequently introduced a partial occlusion and ran an intensity segmentation by iterating the evolution (3) 100 times for each frame (using the previous result as initialization). For a similar application of the Chan-Vese functional (without statistical shape priors), we refer to Moelich and Chan (2003). Figure 6 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 4. Then we ran the segmentation process (26) with the shape prior (21). Apart from adding the shape prior we kept the other parameters constant for comparability.

Figure 7 shows several frames from this knowledge-driven segmentation. A comparison to the corresponding frames in Figure 6 demonstrates several properties of our contribution:

– 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. Chen et al. (2002)—it can easily reproduce the topological changes present in the training set.
– The shape prior is invariant to translation such that the object silhouette can be reconstructed in arbitrary locations of the image. All training shapes are centered at the origin, and the shape energy depends merely on an intrinsically aligned version of the evolving level set function.
– 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).

8.2. Dependency on the Number of Training Shapes

In the experiment of Figure 7, we segmented 85 frames from a walking sequence using a statistical prior constructed on the first 42 shapes. While the number of

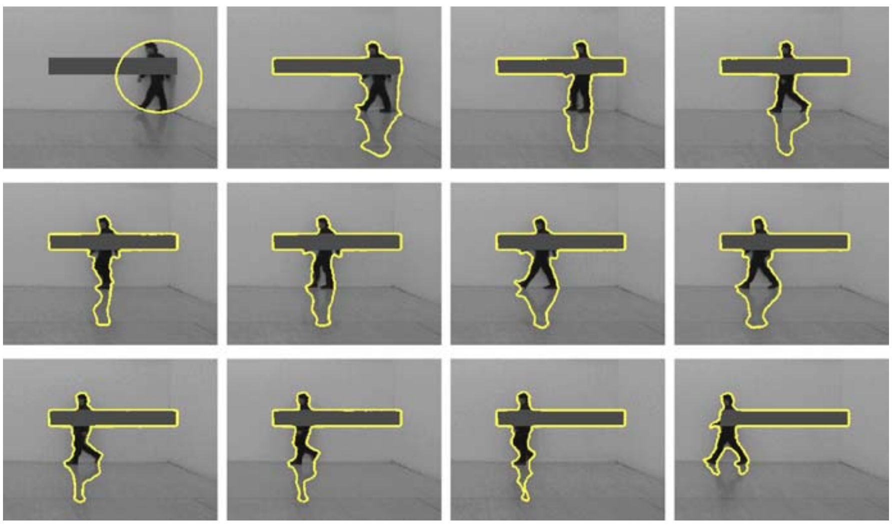

Figure 6. Purely intensity-based segmentation. Various frames show the segmentation of a partially occluded walking person generated by minimizing the Chan-Vese energy (2). The walking person cannot be separated from the occlusion and darker areas of the background such as the person's shadow.

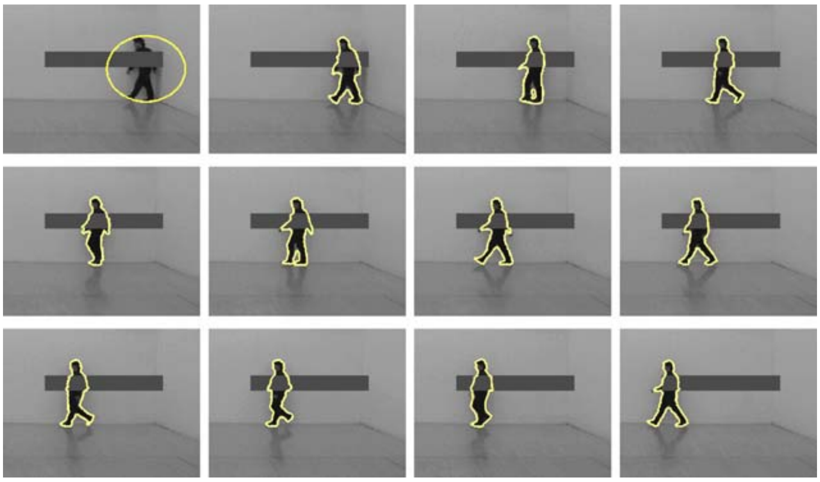

Figure 7. Segmentation with nonparametric invariant shape prior. Segmentation generated by minimizing energy (24) combining intensity information with the statistical shape prior (21). For every frame in the sequence, the gradient descent equation was iterated (with fixed parameters), using the previous segmentation as initialization. Comparison with the respective frames in Fig. 7 shows that the multi-modal 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.

training shapes affects the run time linearly, it is not obvious how it affects the segmentation process. To assess this question numerically, we defined a measure of segmentation accuracy given by the relative error ϵ between the estimated segmentation $H\phi$ and the true segmentation $H\phi_0$:

$$\epsilon = \frac{\int (H\phi - H\phi_0)^2 \, dx}{\int H\phi_0 \, dx}. \tag{29}$$

When reducing the number of training shapes, the average distance between neighboring shapes, and hence our estimate of the kernel width σ in (22) will increase. As a consequence, the estimated shape model becomes less accurate and the segmentation accuracy will degrade. Both of these effects are demonstrated in Figure 8. It shows the estimated kernel width σ and the reconstruction error ϵ averaged over several segmentations of the sequence in Figure 7, using priors constructed from random subsets of the initial 42 training shapes.

8.3. *Segmentation of Cardiac Ultrasound Images*

As a second example, we applied the proposed method to the segmentation of the left ventricle in cardiac ultrasound images. This is a challenging problem, because the object of interest differs very little from the sur-

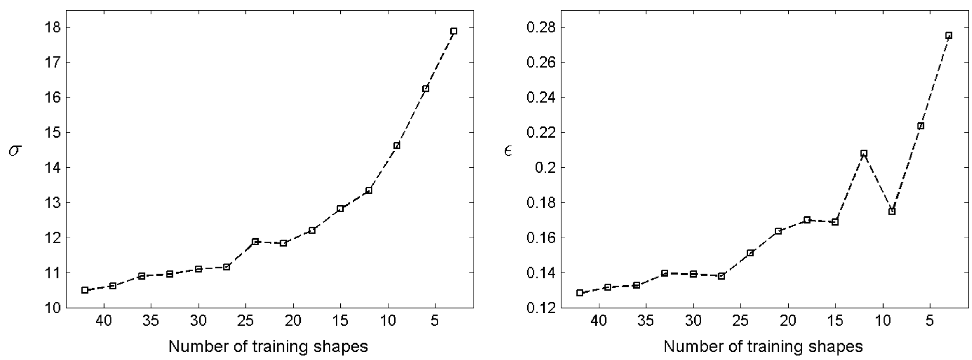

Figure 8. Kernel width σ and segmentation error ϵ versus training set size.

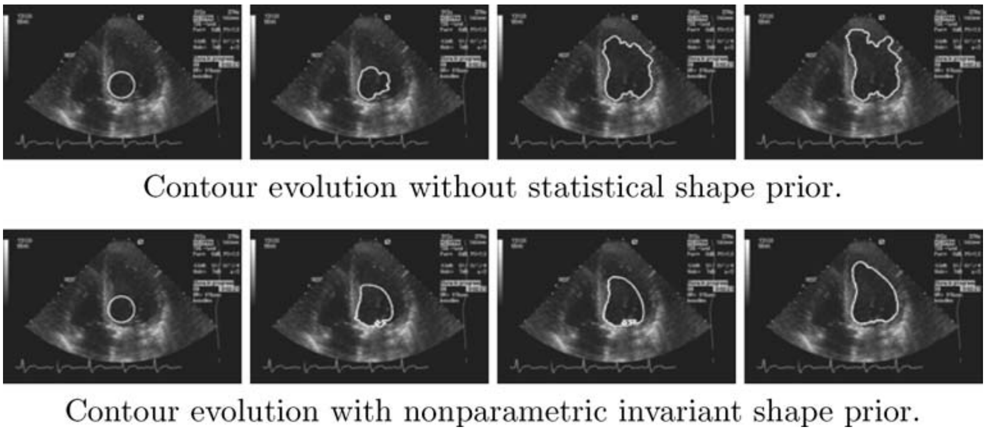

Figure 9. Segmentation of the left ventricle in cardiac ultrasound images. The top row shows a purely data-driven segmentation with the Chan-Vese functional (2). The bottom row shows the boundary evolution obtained with a nonparametric invariant shape prior of the form (21) constructed from a set of 21 aligned manually segmented training shapes. The use of the shape prior speeds up the computation and reduces the segmentation error from 25% (top right) to 9% (bottom right).

rounding background. Figure 9, top row, shows that even by merely imposing a length constraint, one does not obtain an accurate segmentation of the structure of interest. By including a nonparametric shape prior constructed from 21 training images, the segmentation can be drastically improved. Moreover, the segmentation converged after 140 iterations (rather than 900 in the absence of a statistical shape prior). The computation time was approximately 30 seconds (in MATLAB).

9. Conclusion

We proposed solutions to open problems regarding statistical shape priors for level set based segmentation schemes. In particular, we made two contributions:

Firstly, we combined concepts of non-parametric density estimation with level set based shape representations in order to create a statistical shape prior for level set segmentation which can accurately represent arbitrary shape distributions. In contrast to existing approaches, we do not rely on the restrictive assumptions of a Gaussian distribution and can therefore encode fairly distinct shapes. Moreover, by reverting to a non-parametric density estimation technique, we are able to model shape distributions on curved manifolds, thereby circumventing the problem that the space of signed distance functions is not a linear space.

Secondly, we proposed an analytic solution to generate invariance of the shape prior with respect to translation and scaling of the objects of interest. By evaluating the evolving level set function in local coordinates defined relative to its current center of gravity and in units relative to its current scale, our method no longer requires the numerical and iterative optimization of explicit pose parameters. This removes the need to select appropriate gradient time steps and to define a meaningful alternation process for the descent equations. Due to the intrinsic alignment, an additional term emerges in the Euler-Lagrange equations which takes into account the dependency of the transformation parameters on the level set function.

In numerical experiments, we showed that the additional term in the shape gradient both improves and speeds up the convergence of the contour to the desired solution. We showed that the scale invariant prior allows to reconstruct familiar silhouettes at arbitrary scale. Finally, we applied the statistical shape prior to the segmentation and tracking of a partially occluded walking person and to the segmentation of the left ventricle in cardiac ultrasound images. In particular, we demonstrated that the proposed multi-modal shape prior permits to reconstruct fairly distinct silhouettes in arbitrary locations and at arbitrary scales. We numerically validated that a reduction in the number of training shapes leads to a decay in the segmentation accuracy.

Acknowledgments

We thank Alessandro Bissacco and Payam Saisan for providing the image sequence data used for our experiments. We thank David Mumford, Laurent Younes and Mikaël Rousson for fruitful discussions, and Timothée Bailloeul and Gareth Funka-Lea for helpful comments on the manuscript. We thank Marie-Pierre Jolly for providing images and ground truth for the ultrasound sequence. DC and SS were supported by ONR N00014-02-1-0720/N00014-03-1-0850 and AFOSR F49620-03-1-0095/E-16-V91-G2. DC was supported by the German Science Foundation (DFG), grant # CR205/1-1. SO was supported by an NSF IIS-0326388-01, "ITR: Intelligent Deformable Models", Agreement # F5552-01.

Notes

1. One can define a mean shape by back-projection onto the space of signed distance functions
2. In the Bayesian terminology, the length constraint in the Chan-Vese functional (2) should be associated with the shape energy as a (geometric) prior favoring shapes of minimal boundary. However, for notational simplicity, we will only refer to the statistical component as a *shape energy*.

References

Basri, R., Costa, L., Geiger, D., and Jacobs D. 1998. Determining the similarity of deformable shapes. *Vision Research*, 38:2365–2385.

Bresson, X., Vandergheynst, P., and Thirau, J.P. 2003. A priori information in image segmentation: Energy functional based on shape statistical model and image information. In *Proc. IEEE Int. Conf. Image Processing*, pp. 425–428,

Brox, T. and Weickert, J. 2004. A TV flow based local scale measure for texture discrimination. In (eds.), T. Pajdla and V. Hlavac *European Conf. on Computer Vision*, volume 3022 of *Lect. Not. Comp. Sci.*, Prague, Springer.

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

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

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

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

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

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

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

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

Cremers, D. 2006. Dynamical statistical shape priors for level set based tracking. *IEEE Trans. on Patt. Anal. and Mach. Intell.*. To appear.

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

Cremers, D., Osher, S., and Soatto, S. 2004. Kernel density estimation and intrinsic alignment for knowledge-driven segmentation: Teaching level sets to walk. In *Pattern Recogn.*, volume 3175 of *Lect. Not. Comp. Sci.*, pp. 36–44. Springer, C. E. Rasmussen (ed.).

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

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

Cremers, D., Sochen, N., and Schnörr, C. 2003. Towards recognition-based variational segmentation using shape priors and dynamic labeling. In, *Int. Conf. on Scale Space Theories in Computer Vision*, volume 2695 of *Lect. Not. Comp. Sci.*, L. Griffith, editor, pp. 388–400, Isle of Skye, Springer.

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

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

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

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

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

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

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

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

Gdalyahu, Y. and Weinshall, D. 1999. 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.

Grenander, U. 1976. *Lectures in Pattern Theory*. Springer, Berlin.

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

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

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

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

Klassen, E., Srivastava, A., Mio, W., and Joshi, S.H. 2004. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 26(3):372–383.

Leventon, M., Grimson, W., and Faugeras, O. 2000. Statistical shape influence in geodesic active contours. In *CVPR*, *Hilton Head Island*, SC. 1:316–323.

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

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

Mio, W., Srivastava, A., and Liu, X. 2004. Learning and Bayesian shape extraction for object recognition. In *European Conf. on Computer Vision*, volume 3024 of *Lect. Not. Comp. Sci.*, Prague, Springer, pp. 62–73.

Moelich, M. and Chan, T. 2003. Tracking objects with the Chan-Vese algorithm. Technical Report 03-14, Computational Applied Mathematics, UCLA, Los Angeles.

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

Osher, S.J. and Fedkiw, R.P. 2002. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, New York.

Osher, S.J. and Paragios, N. 2003. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer, Telos.

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

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

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

Pons, J.-P., Hermosillo, G., Keriven, R., and Faugeras, O. 2003. How to deal with point correspondences and tangential velocities in the level set framework. In *IEEE Int. Conf. on Computer Vision*, pages 894–899.

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

Rathi, Y., Vaswani, N., Tannenbaum, A., and Yezzi, A. 2005. Particle filtering for geometric active contours and application to tracking deforming objects. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, volume 2, pp. 2–9.

Riklin-Raviv, T., Kiryati, N. and Sochen, N. 2004. Unlevel sets: Geometry and prior-based segmentation. In T. Pajdla and V. Hlavac, editors, *European Conf. on Computer Vision*, volume 3024 of *Lect. Not. Comp. Sci.*, pp. 50–61, Prague. Springer.

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

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

Rousson, M. and Cremers, D. 2005. Efficient kernel density estimation of shape and intensity priors for level set segmentation. In *Intl. Conf. on Medical Image Computing and Comp. Ass. Intervention (MICCAI)*, 1:757–764.

Rousson, M. and Paragios, N. 2002. Shape priors for level set representations. In A. Heyden et al., editors, *Europ. Conf. on Comp. Vis.*, volume 2351 of *Lect. Not. Comp. Sci.*, pages 78–92. Springer.

Rousson, M., Paragios, N., and Deriche, R. 2004. Implicit active shape models for 3d segmentation in MRI imaging. In *Intl. Conf. on Medical Image Computing and Comp. Ass. Intervention (MICCAI)*, volume 2217 of *Lect. Not. Comp. Sci.*, Springer, pp. 209–216

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

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

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

Trouvé, A. 1998. Diffeomorphisms, groups and pattern matching in image analysis. *Int. J. of Computer Vision*, 28(3):213–21.

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

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

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

Yezzi, A. and Soatto, S. 2003. Deformotion: Deforming motions and shape averages. *Int. J. of Computer Vision*, 53(2):153–167.

Younes, L. 1998. Computable elastic distances between shapes. *SIAM J. Appl. Math.*, 58(2):565–586.

Zhao, H.-K., Chan, T., Merriman, B. and Osher, S. 1996. A variational level set approach to multiphase motion. *J. of Comp. Phys.*, 127:179–195.