The International Journal of Computer Vision, Kluwer 2005. To appear.

A Multiphase Dynamic Labeling Model for Variational Recognition-driven Image Segmentation

Daniel Cremers
Department of Computer Science
University of California, Los Angeles, USA
http://www.cs.ucla.edu/∼cremers

Nir Sochen
Department of Applied Mathematics
Tel Aviv University, Israel
http://www.math.tau.ac.il/∼sochen

Christoph Schnörr
Department of Mathematics and Computer Science
University of Mannheim, Germany
http://www.cvgpr.uni-mannheim.de

Abstract. We propose a variational framework for the integration of multiple competing shape priors into level set based segmentation schemes. By optimizing an appropriate cost functional with respect to both a level set function and a (vector-valued) labeling function, we jointly generate a segmentation (by the level set function) and a recognition-driven partition of the image domain (by the labeling function) which indicates where to enforce certain shape priors. Our framework fundamentally extends previous work on shape priors in level set segmentation by directly addressing the central question of *where* to apply *which* prior. It allows for the seamless integration of numerous shape priors such that – while segmenting both multiple known and unknown objects – the level set process may selectively use specific shape knowledge for simultaneously enhancing segmentation and recognizing shape.

Keywords: Image segmentation, shape priors, variational methods, level set methods, dynamic labeling

1. Introduction

Image segmentation and object recognition in vision are driven both by low-level cues such as intensities, color or texture properties, and by prior knowledge about objects in our environment. Modeling the interaction between data-driven and model-based processes has become the focus of current research on image segmentation in the field of computer vision. In this work, we consider prior knowledge given by the shapes associated with a set of objects and focus on the problem of how to exploit such knowledge for images containing multiple objects, some of which may be familiar, while others may be unfamiliar.

Following their introduction as a means of front propagation [24][1], level set based contour representations have become a popular framework for image segmentation [2, 22]. They permit to elegantly model topological changes of the implicitly represented boundary, which makes them well suited for segmenting images containing *multiple* objects. Level set segmentation schemes can be formulated to exploit various low level image properties such as edge information [22, 3, 20], intensity homogeneity [5, 33], texture [25, 30, 19, 1] or motion information [13].

In recent years, there has been much effort in trying to integrate prior shape knowledge into level set based segmentation. This was shown to make the segmentation process robust to misleading low-level information caused by noise, background clutter or partial occlusion of an object of interest (cf. [21, 32, 7, 27]).

All of these approaches were designed to segment a *single* known object in a given image. Yet, in general a given image will contain several familiar or unfamiliar objects. A key problem in this context is therefore to ensure that prior knowledge is *selectively* applied at image locations only where image data indicate a familiar object. Conversely, lack of any evidence for the presence of some familiar object should result in a purely data-driven segmentation process.

Clearly, any use of shape priors consistent with the philosophy of the level set method should retain the capacity of the resulting segmentation scheme to deal with multiple independent objects, no matter whether they are familiar or not. One may instead suggest to iteratively apply the segmentation scheme with a different prior at each time and thereby successively segment the respective objects. We believe, however, that such a sequential processing mode will not scale up to large databases of objects and that – even more importantly – the parallel use of *competing priors* is essential for modeling the chicken-egg relationship between segmentation and recognition.

In this paper, we propose a variational framework for image segmentation which allows the integration of multiple competing shape priors into a segmentation process which can simultaneously handle multiple known and unknown objects in a given image. To this end, we propose to introduce a labeling or decision function in order to restrict the effect of given priors to specific domains of the image plane. Learnt shape information is thereby applied in regions where the image data indicates the presence of a familiar object. For a recent variant of the labeling approach, we refer to [4]. During optimization, this labeling function evolves so as to select image regions where given shape models are applied. The resulting process segments scenes containing corrupted

[1] Precursors containing key ideas of the level set method appeared in [17, 18].

versions of known objects in a way which does not affect the correct segmentation of other unfamiliar objects. A smoothness constraint on the labeling function induces the process to distinguish between occlusions (which are close to the familiar object) and separate independent[2] objects (assumed to be sufficiently far from the object of interest).

In this work, the term *shape prior* refers to fixed templates with variable 2D pose and location. However, the proposed framework of selective shape priors could be extended to statistical shape models which would additionally allow certain deformation modes of each template. For promising advances regarding level set based statistical shape representations, we refer to [6].

This paper comprises and extends work which was presented on two conferences [14, 15]. The outline of the paper is as follows: In Section 2, we briefly review the level set formulation of the piecewise constant Mumford-Shah functional proposed in [5]. In Section 3, we augment this variational framework by a labeling function which selectively imposes a single shape prior in certain image regions. In Section 4, we enhance this prior by explicit transformation parameters for pose and location and demonstrate the effect of simultaneous optimization of pose and location in an image for which the exact transformation parameters of the familiar object are unknown. The resulting segmentation process not only selects appropriate regions where to apply the prior, it also selects appropriate pose and translation parameters associated with a given prior. In Section 5, we extend the labeling approach from the case of one known object and background to that of two independent known objects. In Section 6, we introduce the concept of *multiphase dynamic labeling* which allows the generalization of the labeling approach to an arbitrary number of known and unknown objects by means of a vector-valued labeling function. In Section 7, we derive the gradient descent equations which minimize the proposed functional. In subsequent sections we present numerical results to illuminate various properties of our approach: We demonstrate that the segmentation scheme is capable of reconstructing corrupted versions of multiple known objects displayed in a scene containing other unknown objects. The segmentation of multiple partially occluded objects moving independently in image sequences illuminates how the evolution of the labeling or decision functions is driven by the input data. This evolution can be interpreted in the sense that different shape models compete for areas of influence. In the context of mutual occlusion, we show that the segmentation process

[2] In this paper, we assume that objects share the same Gaussian intensity model, by "independent" we mean that their pose and location are independent. Extensions which allow each object to have its own intensity model are conceivable but they are beyond the scope of this paper.

is forced to decide for one or the other shape model. The experimental results demonstrate that our variational framework couples the input intensity data, the shape models and the labeling or decision functions in a recognition-driven segmentation process. We end with a discussion of limitations and open problems.

2. Data-Driven Level Set Segmentation

Level set representations of moving interfaces [24, 17] have become a popular framework for image segmentation. A contour C is represented as the zero level set function $\phi : \Omega \rightarrow I\!R$ on the image domain $\Omega \subset I\!R^2$:

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

During the segmentation process, this contour is propagated implicitly by evolving the embedding function ϕ. In contrast to explicit parameterizations, one avoids the issues of control point regridding. Moreover, the implicitly represented contour can undergo topological changes such as splitting and merging during the evolution of the embedding function. This makes the level set formalism well suited for the segmentation of *multiple* objects. In this work, we revert to a region-based level set scheme introduced by Chan and Vese [5]. However, other data-driven level set schemes could be employed.

In [5] Chan and Vese introduce a level set formulation of the piecewise constant Mumford-Shah functional [23]. In particular, they propose to generate a segmentation of an input image $f(x)$ with two gray value constants μ_1 and μ_2 by minimizing the functional

$$E_{CV}(\{\mu_i\}, \phi) = \int_\Omega (f-\mu_1)^2 H\phi + (f-\mu_2)^2 (1-H\phi) dx + \nu \int_\Omega |\nabla H\phi|, \tag{2}$$

with respect to the scalar variables μ_1 and μ_2 and the embedding level set function ϕ. Here H denotes the Heaviside function

$$H\phi \equiv H(\phi(x)) = \begin{cases} 1, & \phi(x) \geq 0 \\ 0, & \text{else} \end{cases} . \tag{3}$$

The last term in (2) measures the length of the zero-crossing of ϕ.

The Euler-Lagrange equation for this functional is implemented by gradient descent:

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[\nu \ \text{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - (f-\mu_1)^2 + (f-\mu_2)^2 \right], \tag{4}$$

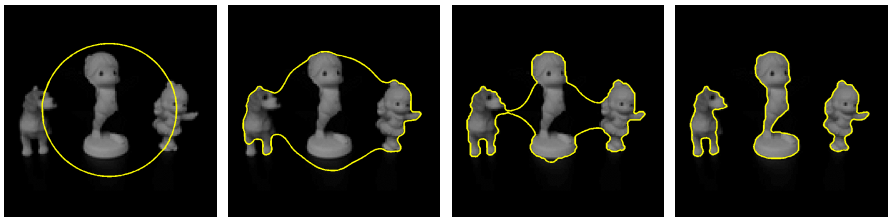


Figure 1. **Purely intensity-based segmentation.** Contour evolution generated by minimizing the Chan-Vese model (2) [5]. The central figure is partially corrupted, i.e. one leg and two arms are missing.

where μ_1 and μ_2 are updated in alternation with the level set evolution to take on the mean gray value of the input image f in the regions defined by $\phi > 0$ and $\phi < 0$, respectively:

$$\mu_1 = \frac{\int f(x) H\phi dx}{\int H\phi dx}, \quad \mu_2 = \frac{\int f(x)(1 - H\phi) dx}{\int (1 - H\phi) dx}. \tag{5}$$

Figure 1 shows a representative contour evolution obtained for an image containing three figures, the middle one being partially corrupted.

3. Selective Shape Priors by Dynamic Labeling

The evolution in Figure 1 demonstrates the well-known fact that the level set based segmentation process can cope with multiple objects in a given scene. However, if the low-level segmentation criterion is violated due to unfavorable lighting conditions, background clutter or partial occlusion of the objects of interest, then the purely image-based segmentation scheme will fail to converge to the desired segmentation (see Figure 8, top row).

To cope with such degraded low-level information, it was proposed to introduce prior shape knowledge into the level set scheme (cf. [21, 32, 27]). The basic idea is to extend the image-based cost functional by a shape energy which favors certain contour formations:

$$E_{total}(\phi) \;=\; E_{CV}(\mu_1, \mu_2, \phi) \;+\; \alpha\, E_{shape}(\phi) \qquad (\alpha > 0). \tag{6}$$

In general, the proposed shape constraints affect the embedding surface ϕ globally (i.e. on the entire domain Ω). In the simplest case, such a prior has the form:

$$E_{shape}(\phi) = \int_\Omega (\phi(x) - \phi_0(x))^2 \, dx, \tag{7}$$

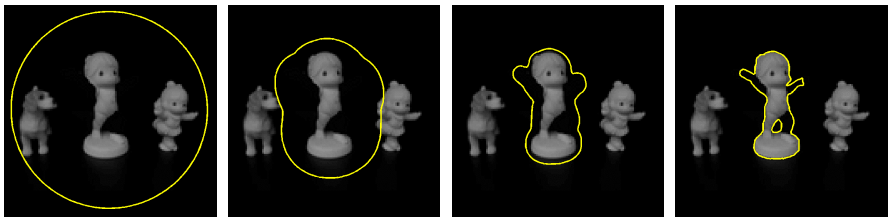


Figure 2. **Global shape prior.** Contour evolution generated by minimizing the total energy (6) with a global shape prior of the form (7) encoding the figure in the center. Due to the global constraint on the embedding function, the familiar object is reconstructed while all unfamiliar structures are suppressed in the final segmentation. The resulting segmentation scheme lost its capacity to deal with multiple independent objects.

where ϕ_0 is the level set function embedding a given training shape (or the mean of a set of training shapes). Uniqueness of the embedding function associated with a given shape is guaranteed by imposing ϕ_0 to be a signed distance function (cf. [21]). For consistency, we also project the segmenting level set function ϕ to the space of distance functions during the optimization – see [31] for details on redistancing.

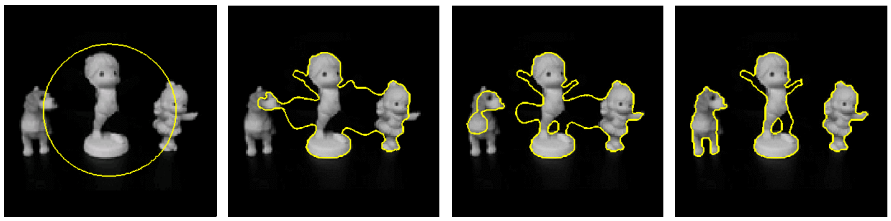
Figure 2 shows several steps in the contour evolution with such a prior, where ϕ_0 is the level set function associated with the middle figure. The shape prior permits to reconstruct the object of interest, yet in the process, all unfamiliar objects are suppressed from the segmentation. The segmentation process with shape prior obviously lost its capacity to handle multiple (independent) objects.

In order to retain this favorable property of the level set method, we proposed in [14] to introduce a labeling or decision function $L : \Omega \rightarrow I\!R$, which indicates the regions of the image where a given prior is to be enforced. During optimization, the labeling evolves dynamically in order to select these regions in a recognition-driven way. The corresponding shape energy is given by:

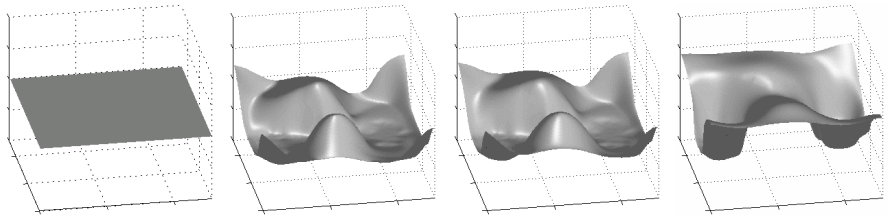
$$E_{shape}(\phi, L) = \int (\phi - \phi_0)^2 (L+1)^2 dx + \int \lambda^2 (L-1)^2 dx + \gamma \int |\nabla L| dx, \quad (8)$$

with two parameters $\lambda, \gamma > 0$. The labeling L enforces the shape prior in those areas of the image where the level set function is similar to the prior (associated with labeling $L = 1$). In particular, for fixed ϕ, minimizing the first two terms in (8) induces the following qualitative behavior of the labeling:

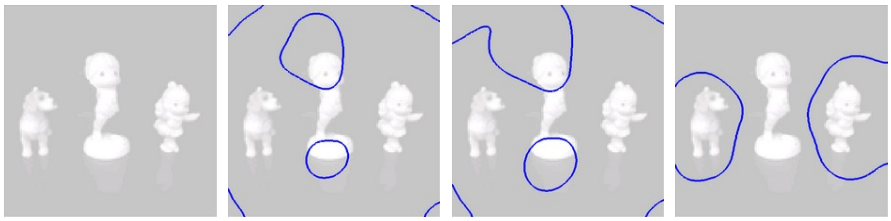
$$\begin{aligned} L &\rightarrow +1, \qquad \text{if } |\phi - \phi_0| < \lambda \\ L &\rightarrow -1, \qquad \text{if } |\phi - \phi_0| > \lambda \end{aligned}$$



Evolution of the segmenting contour.



Simultaneously evolving labeling function.



Zero-crossing of the labeling function.

Figure 3. **Selective shape prior by dynamic labeling.** Contour evolution generated by minimizing the total energy (6) with a selective shape prior of the form (8) encoding the figure in the center. Due to the simultaneous optimization of a labeling function $L(x)$ (middle and bottom row), the shape prior is restricted to act only in selected areas. The familiar shape is reconstructed, while the correct segmentation of separate (unfamiliar) objects remains unaffected. The resulting segmentation scheme thereby retains its capacity to deal with multiple independent objects. In this and all subsequent examples, labeling functions are initialized by $L \equiv 0$.

In addition, the last term in equation (8) imposes a TV regularization on the labeling function stating that neighboring pixels are *a priori* likely to be associated with the same same object (or the background). As is well-known from literature [29], TV regularization favors piecewise constant solutions.[3]

Figure 3, top row, shows the contour evolution generated with the prior (8), where ϕ_0 encodes the middle figure as before. Again the shape prior permits to reconstruct the corrupted figure. In contrast

[3] Note that the TV regulariztion on the labeling L differs from the one we had erroneously reported in the conference versions of this work [14, 15].

to the global prior (7) in Figure 2, however, the process dynamically selects the region where to impose the prior. Consequently, the correct segmentation of the two unknown objects is unaffected by the prior. This selection process is shown in the corresponding evolution of the labeling function in Figure 3, middle row. Its zero crossing (shown in the bottom row) separates the regions associated with the familiar object (where the shape model is enforced) from those associated with the background (where no shape prior is applied).

4. A Similarity-Invariant Formulation

In the above formalism of dynamic labeling, the pose and location of the object of interest is assumed to be known. In a realistic segmentation problem, one generally does not know pose and location of objects. If an object of interest is no longer in the same location as the prior ϕ_0, the labeling approach will fail to generate the desired segmentation. This is demonstrated in Figure 4. While the labeling still separates areas of known objects from areas of unknown objects, the known shape is not reconstructed correctly, since the pose of the prior and that of the object in the image differ.

A possible solution is to introduce a set of pose parameters associated with a given prior ϕ_0 (cf. [27, 7, 12]). The corresponding shape energy

$$\begin{aligned} E_{shape}(\phi, L, s, \theta, h) \;=\; & \int \left(\phi(x) - \frac{1}{s} \phi_0(s R_\theta x + h) \right)^2 (L+1)^2 \, dx \\ & + \int \lambda^2 \, (L-1)^2 \, dx \;+\; \gamma \int |\nabla L| \, dx \end{aligned} \tag{9}$$

is simultaneously optimized with respect to the segmenting level set function ϕ, the labeling function L and transformation parameters, which account for translation h, rotation by an angle θ and scaling s of the template. The division by s guarantees that the resulting shape remains a distance function. Alternatively, one can analytically factor out certain transformation groups by *intrinsic alignment* [11], we will not pursue this alternative here.

Figure 5 shows the resulting segmentation: Again the labeling selects the regions where to apply the given prior, but now the algorithm simultaneously estimates pose and location of the object.

For the sake of simplifying the exposition, we will assume for now, that the initial pose and location of familiar objects is known. Moreover, we will drop the transformation parameters associated with each shape template from the equations, so as to simplify the notation.

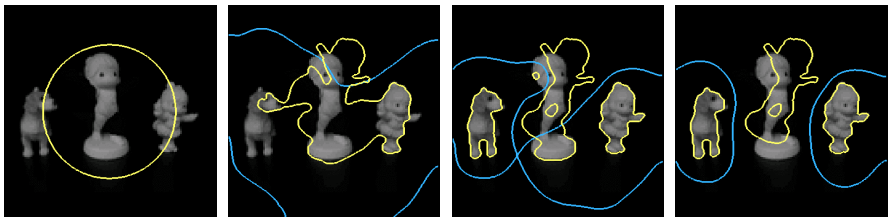


Figure 4. **Missing similarity invariance.** Evolution of contour (yellow) and labeling (light blue) with selective shape prior (8) and a displaced template ϕ_0. Without simultaneous optimization of transformation parameters, the familiar shape is forced to appear in the displaced position.

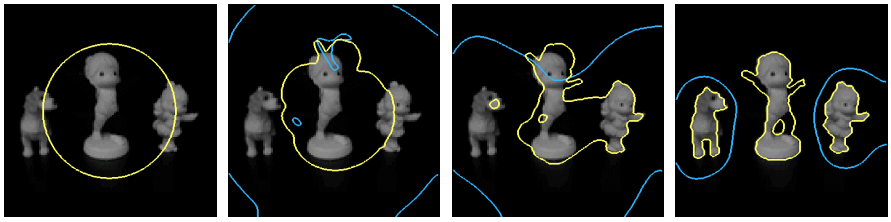


Figure 5. **Similarity Invariance.** By simultaneously optimizing a set of transformation parameters in the shape energy (9), one jointly solves the problems of estimating the area where to impose a prior and pose and location of the respective prior. Note that the estimates of the transformation parameters are gradually improved during the energy minimization.

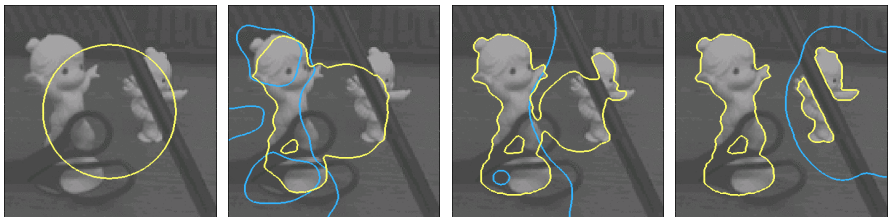
5. Extension to Two Known Objects

A serious limitation of the labeling approach in (8) is that it only allows for a *single* known object (and multiple unknown objects). What if there are several familiar objects in the scene? How can one integrate prior knowledge about multiple shapes such as those given by a database of known objects? Before considering the general case, let us first study the case of *two* known objects.

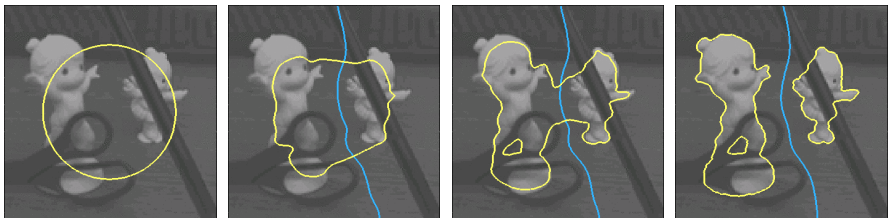
The following modification of (8) allows for two different familiar objects associated with embedding functions ϕ_1 and ϕ_2:

$$E_{shape}(\phi, L) = \frac{1}{\sigma_1^2}\int(\phi-\phi_1)^2\,(L+1)^2 dx + \frac{1}{\sigma_2^2}\int(\phi-\phi_2)^2\,(L-1)^2 dx + \gamma\int|\nabla L|\,dx. \quad (10)$$

The terms associated with the two objects were normalized with respect to the variance of the respective template: $\sigma_i^2 = \int\phi_i^2 dx - (\int\phi_i dx)^2$. The resulting shape prior has therefore merely one (instead



Dynamic Labeling with a single prior and background.



Dynamic Labeling allowing for two competing priors.

Figure 6. **Extension to two priors.** Evolutions of contour (**yellow**) and labeling (**blue**) generated by minimizing energy (6) with a selective prior of the form (8) encoding the left figure (**top**) and with a selective prior of the form (10) encoding both figures (**bottom**). In both cases, the left figure is correctly reconstructed despite prominent occlusions by the scissors. However, while the structure on the right is treated as unfamiliar and thereby segmented based on intensities only (top row), the extension to two priors permits to simultaneously reconstruct both known objects (bottom row).

of two) free parameters. The evolution of the labeling function is now driven by two competing shape priors: each image location will be ascribed to one or the other prior.

Figure 6 shows a comparison: The upper row indicates the contour evolution generated with the shape energy (8), where ϕ_0 encodes the figure on the left. The lower row shows the respective evolution obtained with the shape energy (10), with ϕ_1 and ϕ_2 encoding the left and right figures, respectively. Whereas the object on the right (occluded by a pen) is treated as unknown in the original formulation (upper row), both figures can be reconstructed by simultaneously imposing two competing priors in different domains (lower row).

6. The General Case: Multiphase Dynamic Labeling

The above example showed that the dynamic labeling approach can be transformed to allow for two shape priors rather than a single shape prior and possible background.

Let us now consider the general case of a larger number of known objects and possibly some further independent unknown objects (which should therefore be segmented based on their intensity only). To this end, we introduce a vector-valued labeling function

$$\boldsymbol{L} : \Omega \rightarrow I\!R^n, \qquad \boldsymbol{L}(x) = \Big(L_1(x), \ldots, L_n(x)\Big). \tag{11}$$

We employ the $m = 2^n$ vertices of the polytope $[-1, +1]^n$ to encode m different regions, $L_j \in \{+1, -1\}$, and denote by $\chi_i, i = 1, \ldots, m$ the indicator function for each of these regions. See [34] for a related concept in the context of multi-region segmentation. For example, for $n = 2$, four regions are modeled by the indicator functions:

$$\chi_1(\boldsymbol{L}) = \tfrac{1}{16}(L_1 - 1)^2\,(L_2 - 1)^2, \qquad \chi_2(\boldsymbol{L}) = \tfrac{1}{16}(L_1 + 1)^2\,(L_2 - 1)^2,$$
$$\chi_3(\boldsymbol{L}) = \tfrac{1}{16}(L_1 - 1)^2\,(L_2 + 1)^2, \qquad \chi_4(\boldsymbol{L}) = \tfrac{1}{16}(L_1 + 1)^2\,(L_2 + 1)^2.$$

In the general case of an n-dimensional labeling function, each indicator function will be of the form

$$\chi_i(\boldsymbol{L}) \equiv \chi_{l_1 \ldots l_n}(\boldsymbol{L}) = \frac{1}{4^n} \prod_{j=1}^{n} (L_j + l_j)^2, \quad \text{with } \; l_j \in \{+1, -1\}. \tag{12}$$

With this notation, the extension of the dynamic labeling approach to up to $m = 2^n$ regions can be cast into a cost functional of the form:

$$E_{total}(\phi, \boldsymbol{L}, \mu_1, \mu_2) = E_{CV}(\phi, \mu_1, \mu_2) + \alpha E_{shape}(\phi, \boldsymbol{L}), \tag{13}$$

$$E_{shape} = \sum_{i=1}^{m-1} \int \frac{(\phi - \phi_i)^2}{\sigma_i^2} \chi_i(\boldsymbol{L}) dx + \int \lambda^2 \chi_m(\boldsymbol{L}) dx + \gamma \sum_{i=1}^{m} \int |\nabla L_i| dx.$$

Here, each ϕ_i corresponds to a particular known shape with its variance given by σ_i.

As mentioned before, we have – for better readability – omitted the transformation parameters associated with each template. These can be incorporated by the replacements:

$$\phi_i \;\longrightarrow\; \frac{1}{s_i} \phi_i(s_i R_{\theta_i} x + h_i) \quad \text{ and } \; E_{shape}(\phi, \boldsymbol{L}) \;\longrightarrow\; E_{shape}(\phi, \boldsymbol{L}, \boldsymbol{p}),$$

where $\boldsymbol{p} = (p_1, \ldots, p_m)$ denotes the vector of transformation parameters $p_i = (s_i, \theta_i, h_i)$ associated with each known shape.

7. Energy Minimization

In the previous sections, we have introduced variational formulations of increasing complexity to tackle the problem of multi-object segmentation with shape priors. The corresponding segmentation processes are generated by minimizing these functionals. In this section, we will detail the minimization scheme in order to illuminate how the different components of the proposed cost functionals affect the segmentation process. Minimization of the functional (13) is performed by alternating the update of the mean intensities μ_1 and μ_2 according to (5) with a gradient descent evolution for the level set function ϕ, the labeling functions L_j and the associated pose parameters p_j. The latter will be detailed in the following:

7.1. Evolution of the Segmentation

For fixed labeling, the level set function ϕ evolves according to:

$$\frac{\partial \phi}{\partial t} = -\frac{\partial E_{total}}{\partial \phi} = -\frac{\partial E_{CV}}{\partial \phi} - 2\,\alpha \sum_{i=1}^{m-1} \frac{\phi - \phi_i}{\sigma_i^2} \chi_i(\boldsymbol{L}). \qquad (14)$$

Apart from the image-driven first component given by the Chan-Vese evolution in equation (4), we additionally have a relaxation toward the template ϕ_i in all image locations where $\chi_i > 0$.

7.2. Evolution of the Decision Functions

For fixed level set function ϕ, minimization by gradient descent with respect to the labeling functions L_j corresponds to an evolution of the form:

$$\frac{1}{\alpha}\frac{\partial L_j}{\partial t} = -\sum_{i=1}^{m-1} \frac{(\phi - \phi_i)^2}{\sigma_i^2} \frac{\partial \chi_i}{\partial L_j} - \lambda^2 \frac{\partial \chi_m}{\partial L_j} - \gamma \nabla \left(\frac{\nabla L_j}{|\nabla L_j|} \right), \qquad (15)$$

where the derivatives of the indicator functions χ_i are easily obtained from (12). The first two terms in (15) drive the labeling $\boldsymbol{L}$ to indicate the template ϕ_i which is most similar to the given function ϕ (or alternatively the background). The last term imposes spatial regularity ofthe labeling L_j. This has two effects: Firstly, it induces the labeling to decide for one of the possible templates (or the background), i.e. mixing of templates with label values between $+1$ and -1 are suppressed. Secondly, it enforces the decision regions (regions of constant label) to be "compact", because label flipping is energetically unfavorable. This constraint reflects the assumption that neighboring image locations are likely to belong to the same object.

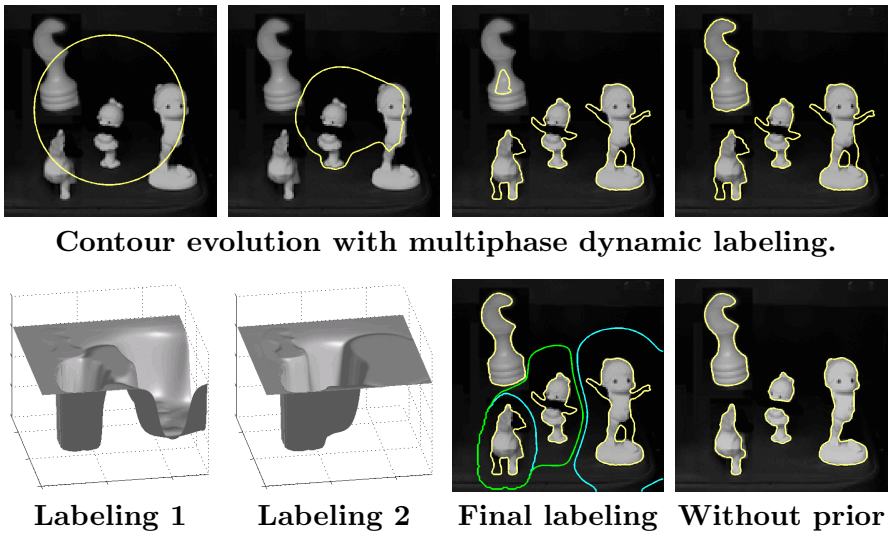


Figure 7. **Coping with several objects by multiphase dynamic labeling.** Contour evolution generated by minimizing energy (6) with a multiphase selective shape prior of the form (13) encoding the three figures on the bottom. The appearance of the known objects is corrupted. Due to the simultaneous optimization of a vector-valued labeling function, several regions associated with each shape prior are selected, in which the given prior is enforced. All familiar shapes are restored while the correct segmentation of separate (unfamiliar) objects remains unaffected. The images on the bottom show the final labeling and – for comparison – the segmentation without prior (right).

Figure 7 shows a contour evolution obtained with the multiphase dynamic labeling model (13) and $n = 2$ labeling functions. The image contains three corrupted objects which are assumed to be familiar and one unfamiliar object (in the top left corner). The top row shows the evolution of the segmenting contour (yellow) superimposed on the input image. The segmentation process with a vector-valued labeling function selects regions corresponding to the different objects in an unsupervised manner and simultaneously applies three competing shape priors which permit to reconstruct the familiar objects. Corresponding 3D plots of the two labeling functions in the bottom rows of Figure 7 show which areas of the image have been associated with which label configuration. For example, the object in the center has been identified by the labeling $\boldsymbol{L} = (+1, -1)$.

7.3. Pose Optimization

For fixed labeling $\boldsymbol{L}$ and level set function ϕ, local optimization of the pose parameters p_i can be implemented by gradient descent. With

$g_i \equiv s_i R_{\theta_i} x + h_i$, the evolution equations for translation h_i, rotation θ_i and scaling s_i associated with each shape model ϕ_i are given by:

$$\frac{\partial h_i}{\partial t} = -\frac{\partial E_{total}}{\partial h_i} = \int \frac{(\phi - \phi_i)}{\sigma_i^2 s_i} \chi_i(\boldsymbol{L}) \nabla \phi_i(g_i) dx, \tag{16}$$

$$\frac{\partial \theta_i}{\partial t} = -\frac{\partial E_{total}}{\partial \theta_i} = \int \frac{(\phi - \phi_i)}{\sigma_i^2} \chi_i(\boldsymbol{L}) \nabla \phi_i(g_i)^\top \frac{\partial R_{\theta_i}}{\partial \theta_i} x dx, \tag{17}$$

$$\frac{\partial s_i}{\partial t} = -\frac{\partial E_{total}}{\partial s_i} = \int \frac{(\phi - \phi_i)}{\sigma_i^2 s_i} \chi_i(\boldsymbol{L}) \left[\nabla \phi_i(g_i)^\top R_{\theta_i} x - \frac{1}{s_i} \phi_i(g_i) \right] dx. \tag{18}$$

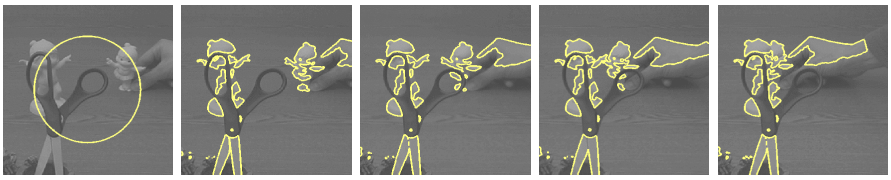
These equations are analogous to the ones derived for a single shape prior (cf. [27, 12]), except that the indicator function $\chi_i(\boldsymbol{L})$ constrains the integrals to the domain of interest associated with shape ϕ_i, i.e. to the area where $\chi_i > 0$.

8. Competition of Shape Models

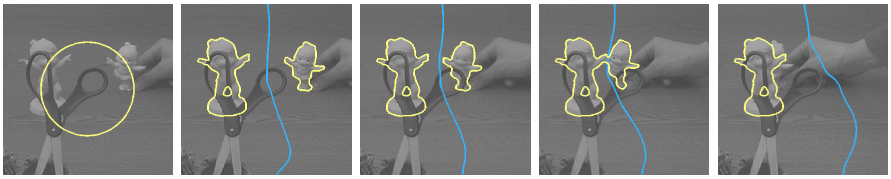
In the presence of multiple shape models, the evolution of the decision boundaries is driven by a competition of the respective shape models. By construction, the energy minimization leads to a partition of the image plane into areas of influence associated with each shape model. Yet what happens if two known objects occlude one another in the same image location?

Figure 8, top row, shows the purely intensity-based segmentation of an image sequence showing two objects one of which is displaced until it is occluded by the other one. We simply iterated the Chan-Vese model until convergence on each frame of the sequence using the segmentation from the previous frame as initialization. These images demonstrate that the objects of interest are clearly not well-defined in terms of intensity homogeneity: They are cut up into pieces according to the brightness constraint.

Figure 8, bottom row, shows segmentation results for the same sequence obtained by adding shape prior of the form (10) to the Chan-Vese functional. Comparison with the purely intensity-based segmentation demonstrates three properties of our approach: Firstly, the integration of two shape priors allows the simultaneous reconstruction of the objects of interest. In particular, the segmentation of background clutter is suppressed by the shape prior. Secondly, the joint optimization of pose parameters allows to keep track of the correct pose of each object. Thirdly, the competition process is such that upon occlusion of one object by the other, one shape model suppresses the other. Due to

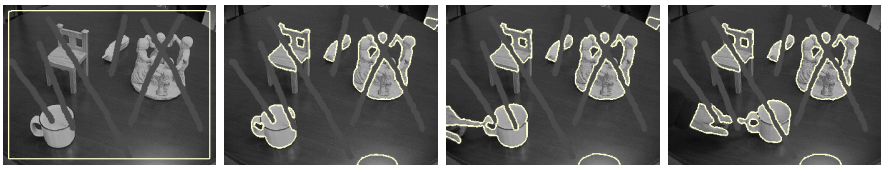


Segmentation by intensity information only.



Segmentation by intensity and two competing shape priors.

Figure 8. **Competing shape priors in the context of mutual occlusion. Top row:** The segmentation of the image sequence with functional (2) shows that the objects of interest cannot be segmented based on the criterion of homogeneous intensity alone. **Bottom row:** Integration of a shape energy of the form (10) allows to simultaneously reconstruct the two objects of interest. Simultaneous pose optimization keeps track of the pose of each object. The labeling or decision function is driven by changes in the intensity data. Upon mutual occlusion, the process is forced to decide for one of the two competing shape models (bottom right).



Intensity-based segmentation.

Computed two-phase reconstruction of the scene.

Figure 9. Intensity-based segmentation of partially occluded moving objects in an image sequence generated by minimizing the functional (2). The segmentation process merely separates bright and dark areas, thereby associating the light reflection on the table with the objects and the legs of the chair and occluded parts with the background. Obviously the brightness criterion does not permit a meaningful reconstruction (lower row) of the objects of interest.

the formulation each location can only be associated with one shape model. In cases of occlusion, the algorithm is therefore forced to decide for the shape model which is favored by the image data.

9. Data-driven Decision Process

In our last experiment, we will demonstrate in which sense our variational approach couples the three levels of input intensity data, shape models and decision functions in a segmentation process. Again, we start with a purely intensity-based segmentation of a sequence containing four objects and artificially introduced occlusion. Figure 9 shows segmentation results obtained with functional (2), using the segmentation of the previous frame as initialization for each of the subsequent frames. The process separates bright and dark areas thereby assigning brighter parts of objects and background (such as reflections on the table and the moving hand in the later part of the sequence) to one phase and darker areas (such as the legs of the chair and the occluded parts of the four objects to the other phase. Again, the intensity information is clearly insufficient to define the objects of interest.

Figure 10 shows the same segmentation process with a multi-phase dynamic labeling allowing for four different shape models. The initial pose of each object is assumed to be known. Energy minimization on the first frame generates a partition of the image plane (by the labeling functions) into four areas of influence associated with each template and a reconstruction of each silhouette according to the shape models. The segmentation of the subsequent frames of the sequence clearly shows that this partition is data-driven: Changes in the input data due to the motion of an object in the sequence affect the decision functions and thereby modify the areas of influence. The decision functions, in turn, will affect where a given shape information is imposed. In this sense, our approach allows to work with shape priors without forcing them onto the data: The data information itself drives the decision process.

10. Dynamic Labeling as a Problem of Bayesian Inference

In the previous sections, we introduced and evaluated the concept of dynamic labeling. By gradually augmenting the complexity of the segmentation task, we derived a variational framework which allows the integration of multiple competing shape priors into an image segmentation process. In this section, we will step back and try to illuminate the proposed framework from a statistical point of view.

The task of image segmentation can be seen as a problem of Bayesian inference: Given the observed image f, we want to infer the most probable partition of the image domain. Obviously this partition should depend on the data. It should also depend on the prior knowledge about expected shape(s). In this work, we restrict our analysis to bi-

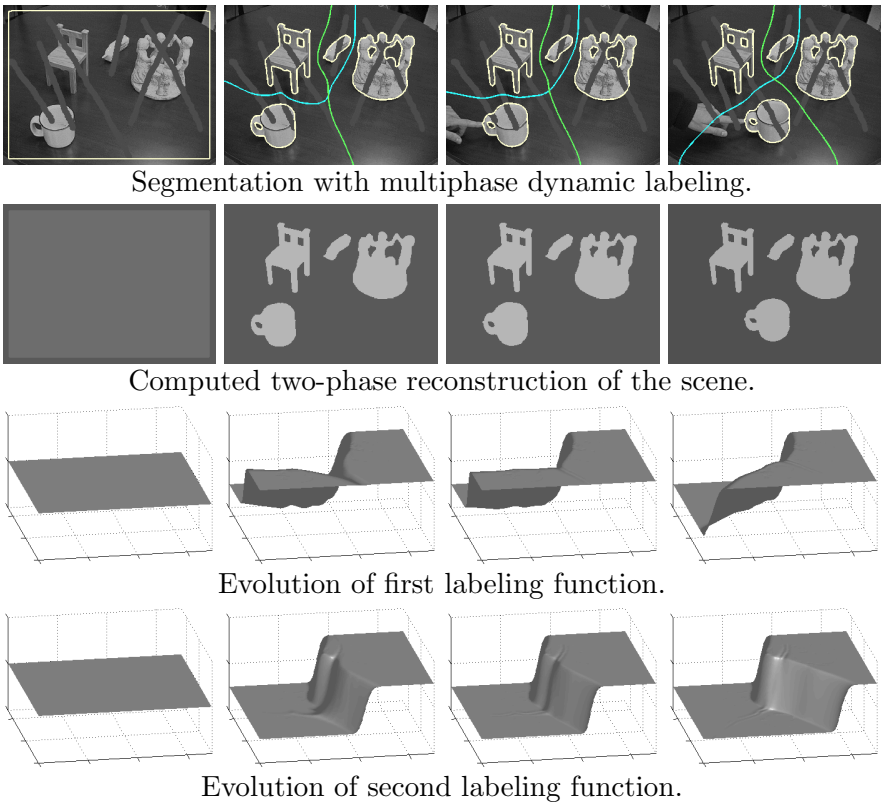


Figure 10. Segmentation of the sequence in Figure 9 with multiphase dynamic labeling allowing for the integration of four shape priors. The two labeling functions (lower rows) identify regions of influence associated with each shape. Changes in the input data (the displacement of the cup) will affect the evolution of these decision functions which in turn will affect the way shape information is integrated into the segmentation scheme.

partitions, where each pixel can be assigned to one of two intensity models. A generalization to multiple models is conceivable. We assume that the gray values in each region are distributed normally with means μ_1 and μ_2 and the same variance for both regions.[4] Following Chan and Vese [5], we use the signed distance (level set) function ϕ to encode the partition. The best partition and intensity models can be estimated by maximizing the a-posteriori probability (MAP) given by:

$$\mathcal{P}(\phi\,;\mu_1,\mu_2\,|\,f) = \frac{\mathcal{P}(f\,|\,\phi\,;\mu_1,\mu_2)\;\mathcal{P}(\phi)}{\mathcal{P}(f)} \tag{19}$$

[4] For the straight-forward extension to differing intensity variances, we refer to the work of Zhu and Yuille [35]. In the case of identical variance considered here, we can simply fix the variance to 1.

where $\mathcal{P}(\phi)$ represents our knowledge and/or expectation about the shape of the object(s) in the image. We assumed a uniform prior for the model intensities μ_i. According to the above assumption of Gaussian intensity distributions, we have:

$$\mathcal{P}(f(x) \,|\, \phi \,; \mu_1, \mu_2) \;\propto\; e^{-(f(x)-\mu(x))^2}, \tag{20}$$

where

$$\mu(x) = \begin{cases} \mu_1, & \text{if } \phi(x) \geq 0 \\ \mu_2, & \text{if } \phi(x) < 0 \end{cases} \tag{21}$$

Assuming that μ_1 and μ_2 are given and that the pixel intensities are independent and identically distributed (i.i.d.) samples, we get

$$\mathcal{P}(f \,|\, \phi \,; \mu_1, \mu_2) \;\propto\; \prod_x e^{-(f-\mu)^2} = e^{-\int \left((f-\mu_1)^2 H\phi + (f-\mu_2)^2 (1-H\phi)\right) dx}$$

Next we need to specify our prior on the segmentation. The prior is composed of two types: syntactic and semantic. The syntactic part penalizes the length of the separating curve and ensures its "compactness" and smoothness. The semantic part penalizes according to the distance of the segmenting shape from a given known shape. In the most simple approach we apply an isotropic Gaussian distribution:

$$\mathcal{P}(\phi) \;\propto\; e^{-\nu \int (\phi - \phi_0)^2} \tag{22}$$

More elaborate approaches include the active shape model [8, 21, 32, 28], in which the modeled shape is constrained to a linear subspace:

$$\phi = \phi_0 + \sum_j \lambda_j \, U_j. \tag{23}$$

Here ϕ_0 is the sample mean and U_j are the eigenvectors of the sample covariance matrix associated with a set of training shapes. For an extensive discussion of the concepts of sample mean and sample covariance for implicit boundary representations, we refer to [6].

For the purpose of clarity we will constrain our analysis to the isotropic Gaussian (i.e. $\lambda_j = 0$) and use one example only as ϕ_0. Extensions to the more general statistical distributions are conceivable. Such statistical shape models include low-dimensional linear subspaces such as the active shape models [8, 21, 32, 28], the (regularized) Gaussian distribution [16], mixtures of Gaussians [9], Gaussian distributions in feature space [10] or distributions inferred by non-parametric density estimation [11].

This can be generalized for the case of dynamic labeling as follows: Assume for simplicity that there is one familiar object and few unfamiliar objects in the image. To correctly segment the image we need

to partition into regions of familiar and unfamiliar objects. To this end, we introduce an indicator function L that separates familiar and not familiar regions. The joint probability of the segmentation and the labeling given the image f is

$$\begin{aligned} \mathcal{P}(\phi, L\,;\mu_1,\mu_2\,|\,f) \;&\propto\; \mathcal{P}(f\,|\,\phi, L\,;\mu_1,\mu_2)\,\mathcal{P}(\phi, L\,;\mu_1,\mu_2) \\ &=\; \mathcal{P}(f\,|\,\phi\,;\mu_1,\mu_2)\,\mathcal{P}(\phi\;|\,L)\,\mathcal{P}(L) \end{aligned} \tag{24}$$

where $\mathcal{P}(f\,|\,\phi\,;\mu_1,\mu_2)$ is given above and

$$\mathcal{P}(\phi\;|\,L) \propto e^{-\int(\phi-\phi_0)^2(L+1)^2} \tag{25}$$

The prior $\mathcal{P}(L)$ is similar in structure to $\mathcal{P}(\phi)$:

$$\mathcal{P}(L) = \mathcal{P}_{syntax}(L)\,\mathcal{P}_{semantics}(L) \propto e^{-\gamma\int|\nabla L|}\,e^{-\int\lambda^2(L-1)^2}, \tag{26}$$

where the semantic prior states that apriori classification as background ($L=+1$) is more likely and that spatially smooth labelings are prefered. Maximization of the conditional probability (24) is indeed equivalent to minimizing functional (6) with shape energy (8).

11. Limitations and Ongoing Research

In the previous sections, we introduced a variational framework which allows to perform segmentation with multiple shape priors. During energy minimization, a vector-valued labeling function identifies areas of influence and pose parameters associated with each shape model. This decision process is driven by a competition of the different shape priors for areas of influence indicated by the image data. Yet there are several open issues which we are currently investigating:

- Each individual prior consists of a fixed silhouette with a set of pose parameters allowing for translation, rotation and scaling. Current effort is focused on extensions of the dynamic labeling approach to also include statistical shape models. For the segmentation of a single object in the level set framework, linear [21, 32, 27] and nonlinear [11] statistical shape models have been proposed.

- The pose of each object is estimated by local optimization of associated pose parameters. In practice, this implies that one needs to either know the exact initial pose of each object (cf. Figures 8 and 10) or have a rough estimate of it (cf. Figure 5). If this is not the case then the present approach will fail to reconstruct a given object. Current effort aims at overcoming this limitation.

- We are investigating generalization of the pose invariance from similarity to perspective transformations as proposed in [26].
- Our formulation is designed to associate each image location with exactly one shape model (or the background). In the case of mutual occlusion of different known objects – as the one shown in Figure 8 – the algorithm therefore decides for one or the other shape model. In certain applications this behavior may not be desirable. Instead one may want to explicitly model occlusions and allow for the simultaneous presence of multiple objects in the same location.
- While the integration of multiple shape priors is modeled as a competition process between different shape models, modeling unknown background objects requires the selection of an additional parameter λ in the shape energy (13) to balance the competition between background region and shape model regions. In practice, most segmentation tasks can be solved with the same value for λ. Yet changes in λ will certainly affect the outcome of the segmentation process, increasing or decreasing the relative size of the identified background region. We are currently investigating means to determine meaningful values for λ in an unsupervised manner.

12. Conclusion

We introduced the framework of multiphase dynamic labeling, which allows to integrate multiple competing shape priors into level set based segmentation schemes. The proposed cost functional is simultaneously optimized with respect to a level set function defining the segmentation, a vector-valued labeling function indicating regions where particular shape priors should be enforced, and a set of pose parameters associated with each prior. Each shape prior is given by a fixed template and respective pose parameters, yet an extension to statistical shape priors (which additionally allow deformation modes associated with each model) is conceivable.

We argued that the proposed mechanism fundamentally generalizes previous approaches to shape priors in level set segmentation. Firstly, it is consistent with the philosophy of level sets because it retains the capacity of the resulting segmentation scheme to cope with multiple independent objects in a given image. Secondly, it addresses the central question of where to apply which shape prior. We showed that the coupled processes of segmentation and recognition-driven selection of areas of influence associated with each object can be derived in the framework of Bayesian inference.

The suggested cost functional couples the three levels of a segmentation process given by the input intensity information, the learnt shape information and the decision functions indicating where to apply certain shape information. As demonstrated in experimental results, the selection of appropriate regions associated with each prior is generated by the dynamic labeling in a recognition-driven manner. In this sense, our work demonstrates how a recognition process can be modeled in a variational segmentation framework.

Acknowledgments

We thank Jianbo Shi for suggesting the occlusion experiment of Figure 8. We also thank the reviewers for helpful comments on the manuscript.

References

1. 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.
2. V. Caselles, F. Catté, T. Coll, and F. Dibos. A geometric model for active contours in image processing. *Numer. Math.*, 66:1–31, 1993.
3. V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. IEEE Intl. Conf. on Comp. Vis.*, pages 694–699, Boston, USA, 1995.
4. T. Chan and W. Zhu. Level set based shape prior segmentation. Technical Report 03-66, Computational Applied Mathematics, UCLA, Los Angeles, 2003.
5. T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.
6. 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*, 2004. To appear.
7. 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. *Intl. J. of Computer Vision*, 50(3):315–328, 2002.
8. T. Cootes, C. Taylor, D. H. Cooper, and J. Graham. Training models of shape from sets of examples. In *Brit. Mach. Vis. Conf.*, pages 9–18, 1992.
9. T.F. Cootes and C.J. Taylor. A mixture model for representing shape variation. *Image and Vision Computing*, 17(8):567–574, 1999.
10. D. Cremers, T. Kohlberger, and C. Schnörr. Shape statistics in kernel space for variational image segmentation. *Pattern Recognition*, 36(9):1929–1943, 2003.
11. D. Cremers, S. J. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for knowledge-driven segmentation: Teaching level sets to walk. In C. E. Rasmussen, editor, *Pattern Recognition*, volume 3175 of *LNCS*, pages 36–44. Springer, 2004.
12. 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.

13. D. Cremers and S. Soatto. Motion Competition: A variational framework for piecewise parametric motion segmentation. *Intl. J. of Computer Vision*, May 2005. To appear.
14. D. Cremers, N. Sochen, and C. Schnörr. Towards recognition-based variational segmentation using shape priors and dynamic labeling. In L. Griffith, editor, *Int. Conf. on Scale Space Theories in Computer Vision*, volume 2695 of *LNCS*, pages 388–400, Isle of Skye, 2003. Springer.
15. 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.
16. D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion Snakes: Introducing statistical shape knowledge into the Mumford–Shah functional. *Intl. J. of Computer Vision*, 50(3):295–313, 2002.
17. 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.
18. A. Dervieux and F. Thomasset. Multifluid incompressible flows by a finite element method. *Lecture Notes in Physics*, 11:158–163, 1981.
19. M. Heiler and C. Schnörr. Natural image statistics for natural image segmentation. In *IEEE Int. Conf. on Comp. Vis.*, pages 1259–1266, 2003.
20. S. Kichenassamy, A. Kumar, P. J. Olver, A. Tannenbaum, and A. J. Yezzi. Gradient flows and geometric active contour models. In *Proc. IEEE Intl. Conf. on Comp. Vis.*, pages 810–815, Boston, USA, 1995.
21. M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proc. Conf. Computer Vis. and Pattern Recog.*, volume 1, pages 316–323, Hilton Head Island, SC, June 13–15, 2000.
22. R. Malladi, J. A. Sethian, and B. C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE PAMI*, 17(2):158–175, 1995.
23. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
24. 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.
25. N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *Intl. J. of Computer Vision*, 46(3):223–247, 2002.
26. 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.
27. M. Rousson and N. Paragios. Shape priors for level set representations. In A. Heyden et al., editors, *Proc. of the Europ. Conf. on Comp. Vis.*, volume 2351 of *LNCS*, pages 78–92, Copenhagen, May 2002. Springer, Berlin.
28. M. Rousson, N. Paragios, and R. Deriche. Implicit active shape models for 3d segmentation in MRI imaging. In *Intl. Conf. on Medical Image Computing and Comp. Ass. Intervention (MICCAI)*, pages 209–216, Rennes/St. Malo, 2004.
29. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
30. C. Sagiv, N. Sochen, and Y. Zeevi. Geodesic active contours applied to texture feature space. In M. Kerckhove, editor, *Intl. Conf. on Scale Space Theories for Comp. Vision*, volume 2106 of *LNCS*, pages 344–352. Springer, 2001.

31. 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.
32. 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.
33. 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.
34. L.A. Vese and T.F. Chan. A multiphase level set framework for image processing using the Mumford–Shah functional. *Intl. J. of Computer Vision*, 50(3):271–293, 2002.
35. S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE PAMI*, 18(9):884–900, 1996.