

# Shape Distances for Binary Image Segmentation

Frank R. Schmidt, Lena Gorelick, Ismail Ben Ayed, Yuri Boykov, and  
Thomas Brox

**Abstract** Shape distances are an important measure to guide the task of shape classification. In this chapter we show that the right choice of shape similarity is also important for the task of image segmentation, even at the absence of any shape prior. To this end, we will study three different shape distances and explore how well they can be used in a trust region framework. In particular, we explore which distance can be easily incorporated into trust region optimization and how well these distances work for theoretical and practical examples.

## 1 Shape Acquisition and Shape Distances

An important task of shape analysis is the acquisition of shapes that we want to analyze. One classical approach is *binary image segmentation* that can be formulated as an energy minimization approach. In other words, we define an energy function  $E : \mathcal{S} \rightarrow \mathbb{R}$  that evaluates how well a certain shape  $S \in \mathcal{S}$  of a *chosen shape space*  $\mathcal{S}$  fits to the observed image and we are interested in the minimizer  $S^* := \operatorname{argmin}_{S \in \mathcal{S}} E(S)$  of the energy  $E$ .

The shape space  $\mathcal{S}$  is usually equipped with a distance  $\operatorname{dist} : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}_0^+$ . The literature is divided on the exact definition of a *distance*. Sometimes, but not always

---

Frank R. Schmidt • Thomas Brox  
Computer Science Department and BIOSS Centre for Biological Signalling Studies, University of  
Freiburg, Georges-Köhler Allee 52, Freiburg, Germany,  
e-mail: schmidt@cs.uni-freiburg.de, brox@cs.uni-freiburg.de

Lena Gorelick • Yuri Boykov  
Computer Science Department, University of Western Ontario, London, Ontario,  
e-mail: lenagorelick@gmail.com, yuri@csd.uwo.ca

Ismail Ben Ayed  
École de Technologie Supérieure, University of Quebec, Montreal, Quebec,  
e-mail: ismail.benayed@etsmtl.ca

it is equated with a metric. In this chapter we call any positive-definite function  $\text{dist}(\cdot, \cdot)$  a distance. Such functions satisfy

$$\text{dist}(A, B) = 0 \Leftrightarrow A = B.$$

In the literature these functions are also referred to as *pre-metrics*. Any shape distance defines a topology of the shape space. In contrast to finite dimensional metric spaces, these topologies are in general not equivalent to one another. In other words, whether a shape  $S \in \mathcal{S}$  is a local minimum of an energy  $E$  depends on the chosen shape distance  $\text{dist}(\cdot, \cdot)$ .

In this chapter, we will explore the influence that a shape distance can have on an image segmentation problem. This influence is only observable if  $E(\cdot)$  cannot be minimized globally. Note that in contrast to other image segmentation applications like [15, 9, 7, 8], we do not use a shape distance in order to enforce a specific shape prior. The only influence that the shape distance has on our optimization task is the definition of a local minimum of the energy  $E$ .

This chapter is organized as follows. In Section 2, we will revisit binary image segmentation that can be solved globally and its extension to the trust region approach [14]. In Section 3, we will present different shape distances and explore if they can be used in a trust region framework. In Section 4, we will show how the chosen shape distance drives the optimization process. Section 5 provides a summary of this chapter.

## 2 Binary Image Segmentation

Binary image segmentation is an important task in computer vision. The goal is to distinguish the object from the background within an image. An image is a mapping  $I: \Omega \rightarrow \mathbb{R}^3$  that assigns to every pixel  $x \in \Omega$  of the  $d$ -dimensional connected image domain  $\Omega \subset \mathbb{R}^d$  a color  $I(x) \in \mathbb{R}^3$ . A binary segmentation can be modelled as a mapping  $u: \Omega \rightarrow \mathbb{B}$  where  $\mathbb{B} = \{0, 1\}$  encodes the object ( $u(x) = 1$ ) and the background ( $u(x) = 0$ ) respectively. A segmentation can also be represented as a subset  $S \subset \Omega$ . The relationship between  $S$  and  $u$  is described via  $x \in S \Leftrightarrow u(x) = 1$ . In the following, we call the binary labeling  $u$  a *segmentation* and the set  $S$  a *shape*.

Given a shape  $S$ , one can apply different image filters to object and background in order to emphasize the object (cf. Fig. 1). In medical image analysis,  $S$  can be used to visualize organs or arteries [23]. Object detection tasks can be addressed better if one works with a segmented object instead of a bounding box [12].



**Fig. 1** If the observed object is easy to distinguish from the background (left image), a per-pixel data term works very well in practice. To remove noise from a threshold solution (central image), an additional length term is used (right image). The resulting energy (1) can be easily optimized via a graph cut [1] or a primal-dual approach [5].

## 2.1 Appearance Models

Classically, the binary image segmentation models the object and the background of an image as a sampling from color distributions  $\text{pdf}_{\text{obj}}$  and  $\text{pdf}_{\text{bg}}$ . Using the notation  $\langle f, g \rangle := \int_{\Omega} f(x) \cdot g(x) dx$ , image segmentation can be cast as minimizing the energy

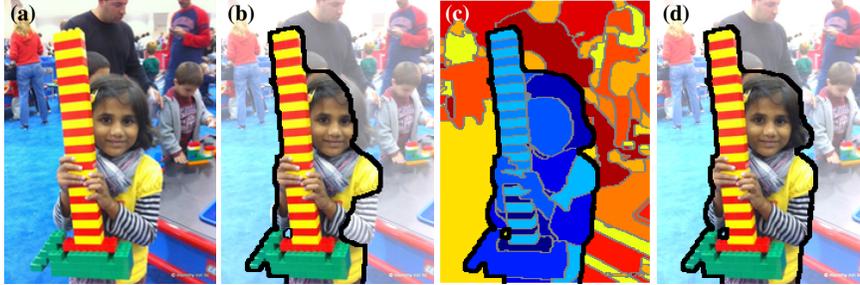
$$E(u) = \langle f, u \rangle \quad \text{with } f(x) = \log \left( \frac{\text{pdf}_{\text{bg}}(x)}{\text{pdf}_{\text{obj}}(x)} \right).$$

While this energy can be easily optimized via a simple thresholding method, the optimal solution exhibits typically a high amount of noise (cf. Fig. 1). Therefore, Mumford and Shah proposed in [18] to add the length of the segmentation's boundary as a regularizing term to the energy, resulting in

$$E(u) = \langle f, u \rangle + \text{len}(\partial S) \quad \text{with } S = \{x \in \Omega | u(x) = 1\}. \quad (1)$$

A discrete formulation approximates the length term via the *Cauchy-Crofton formula* and minimizes the energy via a graph cut approach [1]. A continuous formulation solves the problem via a primal-dual approach that can be efficiently parallelized on GPUs [5]. In the following, we assume that we work in a computer environment where (1) can be easily optimized. Whether the discrete or the continuous formulation is used is not important for the rest of this chapter.

In general,  $f : \Omega \rightarrow \mathbb{R}$  can be an arbitrary integrable function that need not to be derived from color distributions. In the past, different attempts have been made to model the appearance of object and background by using more information than just the color information  $I(x)$  at a pixel  $x$ . Besides using more modalities like depth or infra-red information, it is common to use local features like Fourier features, Gabor features or more general texture features [24, 4]. All these approaches can be seen as an attempt of altering the data term  $f$  in (1). In practice, these approaches improve the segmentation. Nonetheless, these patch-based approaches become less reliable for pixels close to the object's boundary, since the features will then mix object and background information. In the following, we revisit alternative approaches.



**Fig. 2** If the object and the background contain similar appearances (a), the global optimum of (1) does not provide a good segmentation (b). Performing a hierarchical segmentation [10] improves the model of the scene (c) and provides a more accurate binary segmentation (d).

## 2.2 Multiple Models and Holistic Distributions

While the energy (1) can be applied very successfully if the appearance of object and background vary considerably, it struggles if certain appearances appear in both, the object and the background regions. It was therefore suggested by Delong et al. [10] not to use one but multiple distributions for object and background (cf. Fig. 2). The method computes a sub-labeling  $u_0: \Omega \rightarrow \{1, \dots, K\}$  of the image domain  $\Omega$  and a binary labeling  $u_1: \{1, \dots, K\} \rightarrow \mathbb{B}$  of these  $K$  sub-labels. As a result, the method computes simultaneously a superpixel representation  $u_0$  and its binary segmentation  $u_1$ . In conjunction, these two functions induce a binary labeling  $u: \Omega \rightarrow \mathbb{B}$  via  $u(x) = u_1 \circ u_0(x)$ .

While the resulting minimization problem is the instance of an NP hard problem, the approximation that is obtained via  $\alpha$ -expansion [3, 11] proved to be more reliable than the binary segmentation driven by (1). Nonetheless, the optimization process can take a long time and is therefore not fit for fast segmentation tasks.

In order to model the appearance of different colors without the need to find an optimal superpixel representation, we advocate the concept of *holistic histograms*. To this end, let us assume that we have pre-detected  $n$  appearances in an image. An appearance can be based on color, texture or other features. Further, assume that we can decide for every pixel  $x \in \Omega$  whether this appearance is present at  $x$ . This results in  $n$  appearance detectors  $f_i: \Omega \rightarrow \mathbb{B}$ . If we partition an image into disjoint areas of the same color, each  $f_i$  would represent the indicator function of one of these areas. Nonetheless, it is also possible that different  $f_i$  intersect in certain areas. This is for example the case if we have one detector for “blue pixels” and one feature-based detector for the image class “sky”. Given a segmentation  $u$ , we can now compute the following histogram

$$h(u) = (\langle f_1, u \rangle, \dots, \langle f_n, u \rangle) \in \mathbb{R}^n. \quad (2)$$

Note that this histogram cannot be pre-computed on the pixel-level. It depends on the whole segmentation  $u$  and will change during the optimization process. Since there are also detectors that provide only probabilities about the presence of a certain appearance, we can extend the detectors to  $f_i: \Omega \rightarrow \mathbb{R}_0^+$ .

If we want to solve a segmentation task that is scale-invariant, we prefer to work with distributions, *i.e.*, normalized histograms, instead of histograms. Given the appearance detectors  $f_i$  as above, we obtain the *holistic distribution*

$$p(u) = \left( \frac{\langle f_1, u \rangle}{\langle \sum_{j=1}^n f_j, u \rangle}, \dots, \frac{\langle f_n, u \rangle}{\langle \sum_{j=1}^n f_j, u \rangle} \right) \in \mathbb{R}^n. \quad (3)$$

If a prior distribution  $q \in \mathbb{R}^n$  is learned, we would like to use a distribution distance to penalize the deviation of  $p(u)$  from the prior  $q$ . Combining the Bhattacharyya distance between the distributions with a length term results in the energy

$$E(u) = -\log \left( \sum_{i=1}^n \sqrt{\frac{q_i \cdot \langle f_i, u \rangle}{\langle \sum_{j=1}^n f_j, u \rangle}} \right) + \text{len}(u) \quad (4)$$

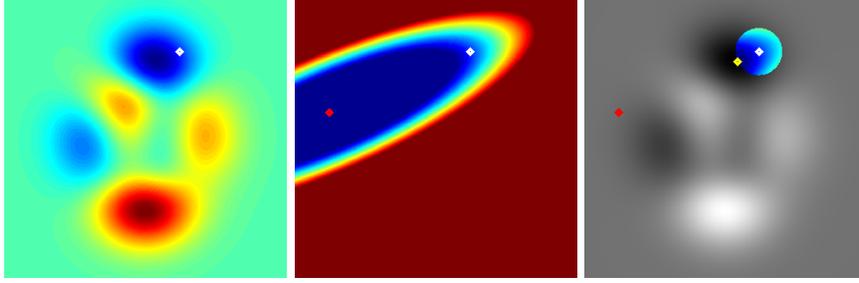
that we want to minimize.

### 2.3 Submodular and Convex Relaxations

Recently, Tang et al. [25] proposed an unsupervised segmentation approach that rewards the  $L^1$ -distance between the object's histogram  $h(u)$  and the background's histogram  $h(1-u)$ . Since this results in the minimization of a submodular energy, it can be solved globally and its solution provides for a much better segmentation than the optimization of (1). Nonetheless, it cannot be used in order to solve (4), which uses distributions instead of histograms.

Nieuwenhuis et al. [19] addressed a problem related to (4). Instead of a binary segmentation they addressed a multi-region segmentation, where ratio constraints for each region are encouraged. They addressed this problem by computing the global optimum of an approximation of the original energy with respect to labelings  $u_i: \Omega \rightarrow [0, 1]$ . Since the *threshold theorem* [6] is not satisfied for the convex function, it cannot be guaranteed that the derived segmentations  $\hat{u}_i: \Omega \rightarrow \mathbb{B}$  is even a local optimum of the approximation.

To guarantee local optimality, Gorelick et al. [14] proposed a method that combines the trust region framework with a class of energies that also includes (4). Since we want to explore the relationship between local optimization methods and shape distances, we will focus on the approximation scheme of [14]. After revisiting it in Section 2.4, in Section 3, we will study shape distances that define different *trust regions* and thus, compute different local minima.



**Fig. 3** *Left:* For a complex energy  $E$  one can use an approximation  $\tilde{E}$  that is exact at a point  $u_0$  (white dot). *Middle:* The global minimizer  $u^*$  (red dot) of  $\tilde{E}$  will in general not improve the value of the original energy  $E$ . *Right:* Trust region approaches trust  $\tilde{E}$  in a small vicinity of  $u_0$  (colored circle). For a sufficient small vicinity, the optimizer  $\hat{u}$  (yellow dot) of (6) improves the energy  $E$ .

## 2.4 Trust Region

*Trust region* methods are used to find a (local) minimum of a function  $E$ . Naturally, these methods are only used if it is difficult to find the global optimum of the energy  $E$ . The idea is to use an approximation  $\tilde{E}$  of  $E$  that is exact at a certain feasible solution  $u_0$ . If the set of all feasible solutions is equipped with a distance function  $\text{dist}(\cdot, \cdot)$ , the trust region approach iteratively solves the *trust region sub-problem*

$$\underset{\text{dist}(u_0, u) < d}{\text{argmin}} \tilde{E}(u). \quad (5)$$

If the solution  $\hat{u}$  of this problem reduces the actual energy considerably, *i.e.*,

$$E(\hat{u}) \leq \alpha E(u_0) \quad \text{with } 0 < \alpha < 1,$$

$\hat{u}$  is accepted as a new approximate solution  $u_0$  (cf. Fig. 3). Otherwise the region in which we trust the approximation is reduced, *i.e.*,  $d$  is multiplied with a factor  $\beta$ ,  $0 < \beta < 1$ . These steps are repeated until the distance  $d$  is small enough.

Since we have to minimize the *trust region sub-problem* (5) globally, we like to use approximations  $\tilde{E}$  that are easy to optimize. If  $E$  is differentiable it can be approximated by a linear Taylor approximation. In the case that the space of feasible solutions is a linear space  $\mathbb{R}^N$  equipped with the canonical metric  $\text{dist}(u_0, u) = \|u_0 - u\|$ , a solution of (5) is

$$\underset{\|u - u_0\| < d}{\text{argmin}} E(u_0) + \langle E'(u_0), u - u_0 \rangle = u_0 - d \cdot \frac{E'(u_0)}{\|E'(u_0)\|}.$$

Therefore, the trust region approach can be understood as a generalization of the *normalized gradient descent* approach. In practice, second order approximations of  $E$  are used [20].

Gorelick et al. [14] used functions  $E$  that can be described as the sum of a differentiable function  $E_1$  and a length term. The approximation  $\tilde{E}$  only uses a linear approximation for  $E_1$ . The length term is not approximated at all, resulting in:

$$E(u) = E_1(u) + \text{len}(u) \quad \tilde{E}(u) = E_1(u_0) + \langle E_1'(u_0), u - u_0 \rangle + \text{len}(u)$$

To solve the trust region sub-problem (5), a Lagrangian formulation<sup>1</sup>

$$\underset{u}{\text{argmin}} \langle E_1'(u_0), u \rangle + \text{len}(u) + \lambda \text{dist}(u_0, u) \quad (6)$$

was used and a reciprocal relationship between the Lagrangian factor  $\lambda$  and the distance  $d$  was exploited. For more details we refer to [14].

*Remark 1.* Note that in contrast to a gradient descent approach, the length term need not to be approximated, since we can optimize energies of the form (1) that also include length terms. If we also approximated the length term, the resulting sub-problem would include a curvature motion as explored in the level set framework [21]. It was shown in [13] that not approximating the length term is beneficial in practice. The resulting method is faster and possesses fewer local minima than the level set approach of [16].

*Remark 2.* The Lagrangian formulation (6) uses the current solution  $u_0$  as a prior. If we want to trust  $\tilde{E}$  in a smaller vicinity,  $\lambda$  is automatically increased and the prior has a stronger influence. This results in a process where the global optimum of (6) is pushed towards  $u_0$  with increasing  $\lambda$ . Note that it is not necessary to *tune the parameter*  $\lambda$  to the application.  $\lambda$  is instead automatically adapted by the trust region framework. This adaptation is driven by the original energy  $E$ .

Since the prior in (6) depends on the distance  $\text{dist}(\cdot, \cdot)$ , we explore in the next section different distance functions for shapes. These distances define different sub-problems (6) and thus different local minima of  $E$ . In order to globally optimize (6), we focus on shape distance functions that are affine in  $u$ . In these cases, the trust region sub-problem is of the form (1), which we can easily optimize.

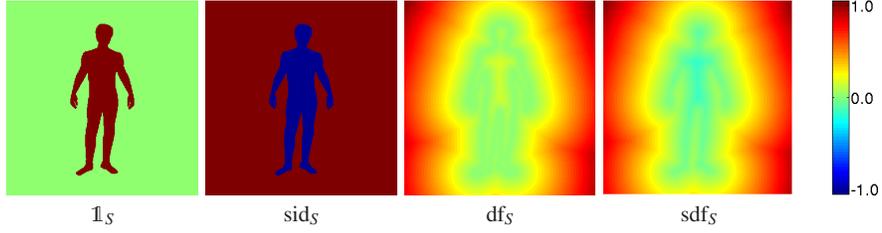
### 3 Shapes and Shape Distances

In order to avoid shapes  $S \subset \Omega$  that can only be created by the set-theoretical *axiom of choice* or sets that are null-sets in the Lebesguean sense, we want to focus on shapes  $S$  that are open sets. Since we will also be interested in the boundary  $\partial S$  of a shape  $S$ , we want to exclude those shapes whose boundaries are empty. With

$$\mathcal{S} := \{S \subset \Omega \mid S \text{ is open and } \partial S \neq \emptyset\} \quad (7)$$

---

<sup>1</sup> Since we are only interested in the minimizer, we removed constant terms from the energy.



**Fig. 4** For a shape  $S \in \mathcal{S}$ , we use different implicit representations, the *indicator function*  $\mathbb{1}_S$ , the *signed indicator function*  $\text{sid}_S$ , the *distance function*  $\text{df}_S$  and the *signed distance function*  $\text{sdf}_S$ .

we denote the set of all those shapes. Since the boundary  $\partial S := \overline{S} \cap \overline{S^c}$  is the intersection of the closure of  $S$  and the closure of its complement  $S^c$ , only the empty set and the whole domain  $\Omega$  are exempted from the shape space  $\mathcal{S}$ . This is a consequence of  $\Omega$  being connected.

In order to equip the shape space  $\mathcal{S}$  with a distance, we have two choices. We can either define a distance  $\text{dist}(S_0, S_1)$  with respect to the whole shapes  $S_i$  or with respect to their boundaries  $\partial S_i$ . In the first case we speak of *region-based* distances and in the latter case we speak of *contour-based* distances. While contour-based distances proved to be very descriptive [17], it is difficult to incorporate them into image segmentation tasks. The goal of this section is to overcome this limitation of contour-based distances by approximating them in a regional sense.

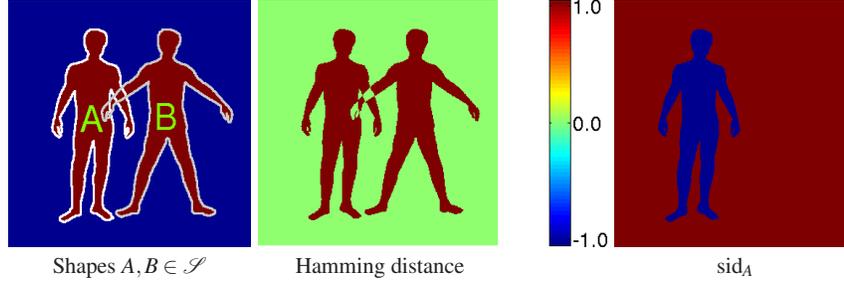
To study relationships between  $S$  and  $\partial S$ , we use the following representations.

**Definition 1.** Given a shape  $S \in \mathcal{S}$ , we denote the *indicator function*, the *signed indicator function*, the *distance function* and the *signed distance function* (cf. Fig. 4) as  $\mathbb{1}_S, \text{sid}_S, \text{df}_S, \text{sdf}_S: \Omega \rightarrow \mathbb{R}$  and define them via

$$\begin{aligned} \mathbb{1}_S(x) &:= \begin{cases} 1 & , x \in S \\ 0 & , x \notin S \end{cases} & \text{sid}_S(x) &:= \begin{cases} -1 & , x \in S \\ +1 & , x \notin S \end{cases} \\ \text{df}_S(x) &:= \min_{s \in \partial S} \|x - s\| & \text{sdf}_S(x) &:= \text{sid}_S(x) \cdot \text{df}_S(x). \end{aligned}$$

In Section 3.1 we will study the Hamming distance  $\text{dist}_H(\cdot, \cdot)$  and show its restrictions for the trust region sub-problem (6). In Section 3.2, we will study a contour-based distance  $\text{dist}_{L^2}(\cdot, \cdot)$  and explore its regional approximation  $\text{dist}_2(\cdot, \cdot)$ .

In particular, we will show that both,  $\text{dist}_H(u_0, u)$  and  $\text{dist}_2(u_0, u)$  are affine in  $u$  and can therefore be easily incorporated into (6).  $\text{dist}_{L^2}(u_0, u)$  on the other hand is not affine in  $u$  and cannot be used in the trust region framework. For that reason, we have to approximate it with the distance  $\text{dist}_2(u_0, u)$  that is affine in  $u$ .



**Fig. 5** The Hamming distance  $\text{dist}_H(A, B)$  of two shapes  $A, B \in \mathcal{S}$  is the area of its symmetric difference. Using the signed indicator function  $\text{sid}_A$ , this distance becomes affine in  $B$  (cf. (9)).

### 3.1 Regional Hamming Distance and its Restrictions

The Hamming distance of two shapes  $A, B \in \mathcal{S}$  is defined as the area of its symmetric difference  $A \triangle B := (A \setminus B) \sqcup (B \setminus A)$ :

$$\text{dist}_H(A, B) := \text{area}(A \triangle B). \quad (8)$$

Using the signed indicator function  $\text{sid}_A$ , we can rewrite the Hamming distance as

$$\text{dist}_H(A, B) = \int_B \text{sid}_A(x) \, dx - \int_A \text{sid}_A(x) \, dx. \quad (9)$$

To see that Equations (8) and (9) describe the same function, note that

$$\begin{aligned} \int_B \text{sid}_A(x) \, dx - \int_A \text{sid}_A(x) \, dx &= \int_{\sqcup(B \setminus A)} \text{sid}_A(x) \, dx - \int_{\sqcup(A \setminus B)} \text{sid}_A(x) \, dx \\ &= \int_{(B \setminus A)} 1 \, dx - \int_{(A \setminus B)} (-1) \, dx = \text{area}(A \triangle B). \end{aligned}$$

The advantage of the formulation (9) is that it can be integrated into the trust region sub-problem (6). Using the notation  $\langle f, S \rangle := \langle f, \mathbb{1}_S \rangle = \int_S f(x) \, dx$ , we obtain

$$\text{dist}_H(A, B) = \langle \text{sid}_A, B \rangle + C, \quad C := -\langle \text{sid}_A, A \rangle,$$

which is affine in  $B$ . We will show in the following that the Hamming distance is a shape distance that is disadvantageous for shape prior-based image segmentation. For this reason we want to study different shape distances.

*Example 1.* Let us consider two different shapes  $A, B \in \mathcal{S}$  and the energy function  $E(S) := \langle \text{sid}_A, S \rangle$ . Its unique minimizer is  $S^* = A$ . Adding a weighted shape prior with respect to  $B$  leads to the energy

$$\begin{aligned} E_\lambda(S) &= (1 - \lambda) \cdot E(S) + \lambda \cdot \text{dist}_H(B, S) \\ &= \langle (1 - \lambda) \text{sid}_A + \lambda \text{sid}_B, S \rangle + \lambda C, \quad C := -\langle \text{sid}_B, B \rangle. \end{aligned}$$

If we denote with  $S_\lambda^*$  a global minimum of  $E_\lambda(S)$ , we obtain a mapping  $m : \lambda \mapsto S_\lambda^*$  that starts at  $S_0^* = A$  and ends at  $S_1^* = B$ . One major disadvantage of the used Hamming distance is that  $m$  is not a continuous morphing (cf. 1<sup>st</sup> plot of Fig. 6).

**Theorem 1.** *If we define a mapping  $m : [0, 1] \rightarrow \mathcal{S}$  as above, the following holds:*

$$\begin{aligned} m(\lambda) &= A, & \text{if } 0 \leq \lambda < \frac{1}{2} \\ m(\lambda) &= B, & \text{if } \frac{1}{2} < \lambda \leq 1 \end{aligned}$$

*Proof.* A minimizer of  $E_\lambda$  is easily found by thresholding  $(1 - \lambda) \text{sid}_A + \lambda \text{sid}_B$  at 0. The following observation

$$(1 - \lambda) \text{sid}_A(x) + \lambda \text{sid}_B(x) = \begin{cases} -1 + 2\lambda & , \text{ if } x \in A \setminus B \\ 1 - 2\lambda & , \text{ if } x \in B \setminus A \\ -1 & , \text{ if } x \in A \cap B \\ 1 & , \text{ if } x \notin A \sqcup B \end{cases}$$

proves the theorem. □

Because of this theorem, we cannot use  $\text{dist}_H$  in (6) in order to *push* the segmentation towards a specific shape. As mentioned in Remark 2, continuous morphings are essential for a successful trust region computation. With the Hamming distance we can only encode a hard constraint. In order to handle soft constraints, we will explore next a contour-based distance and its region-based approximation.

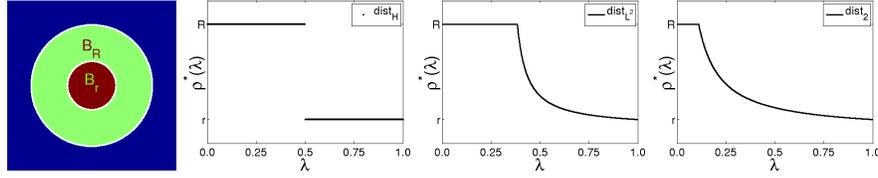
### 3.2 $L^2$ Contour Distance and its Regional Approximation

An  $L^2$  distance between two shapes  $A, B \in \mathcal{S}$  can be formulated as

$$\text{dist}_{L^2}(A, B) := \left( \int_{\partial B} \min_{x \in \partial A} \|x - s\|^2 ds \right)^{\frac{1}{2}}. \quad (10)$$

This distance only considers the shapes' boundaries. The interior of the shapes is completely ignored. In order to simplify the study of this distance, we will only consider concentric balls  $B_\rho$  of radius  $\rho > 0$ . For these examples, the distance can be computed analytically. Given two concentric balls of radius  $r$  and  $R$ , we obtain

$$\text{dist}_{L^2}(B_r, B_R)^2 = \int_{\partial B_R} (R - r)^2 ds = 2\pi R \cdot (R - r)^2.$$



**Fig. 6** *Image:* As shapes we consider concentric balls  $B_\rho$  of radius  $\rho$ . Given the radii  $0 < r \leq R$ , we consider an energy  $E_\lambda(\rho) = (1 - \lambda) \langle \text{sid}_{B_R}, B_\rho \rangle + \lambda \text{dist}(B_r, B_\rho)$ . The first term favors  $\rho = R$  and the second term favors  $\rho = r$ . The minimizer  $\rho^*$  of  $E_\lambda$  depends on  $\lambda$ . *Plots:* Using  $\text{dist}_H$  leads to a non-continuous function  $\rho^*(\lambda)$ . For  $\text{dist}_{L^2}$  and  $\text{dist}_2$ ,  $\rho^*(\lambda)$  becomes continuous.

The distance  $\text{dist}_{L^2}(\cdot, \cdot)$  is not symmetric and thus not a metric. Analogously to Section 3.1, we want to study the influence that a  $\text{dist}_{L^2}^2$ -based shape prior has on image segmentation.

*Example 2.* Let us consider the radii  $0 < r \leq R$ . The unique minimizer of the energy  $E(\rho) = \langle \text{sid}_{B_R}, B_\rho \rangle$  is obtained for  $\rho^* = R$ . Adding  $B_r$  as a shape prior, results in the following energy

$$\begin{aligned}
 E_\lambda(\rho) &= (1 - \lambda) \langle \text{sid}_{B_R}, B_\rho \rangle + \lambda \text{dist}_{L^2}(B_\rho, B_r) \\
 &= (1 - \lambda) \langle \text{sid}_{B_R}, B_\rho \rangle + \lambda \int_{\partial B_r, x \in \partial B_\rho} \min \|x - s\|^2 ds \\
 &= (1 - \lambda) \langle \text{sid}_{B_R}, B_\rho \rangle + \lambda \cdot 2\pi r(\rho - r)^2 \\
 &= \begin{cases} \lambda \cdot 2\pi r(\rho - r)^2 - (1 - \lambda)\pi\rho^2 & , \text{ if } \rho \leq R \\ \lambda \cdot 2\pi r(\rho - r)^2 + (1 - \lambda)\pi(\rho^2 - 2R^2) & , \text{ if } \rho > R \end{cases}
 \end{aligned}$$

The global minimum of  $E_\lambda$  is (cf. 2<sup>nd</sup> plot of Fig. 6)

$$\rho^*(\lambda) = \begin{cases} R & , \lambda \leq \frac{1}{2r+1} \\ \min\left(r + \frac{(1-\lambda)r}{2\lambda r - (1-\lambda)}, R\right) & , \lambda > \frac{1}{2r+1}. \end{cases}$$

First of all, this means that  $\rho^*$  continuously changes from  $\rho^*(0) = R$  to  $\rho^*(1) = r$ . We are therefore able to continuously push the segmentation to a certain shape prior. Secondly, there is a small range for  $\lambda$  where the shape prior is ignored. This means that a strong data term always overrules the shape prior. Both of these properties are important for the trust region sub-problem (6). A major disadvantage of  $\text{dist}_{L^2}^2$  over the Hamming distance is the fact that it cannot be incorporated into (6) in such a way that results in an energy of the form (1). This is because  $\text{dist}_{L^2}^2(A, B)$  is not affine in  $B$ . Therefore, we seek in the following an affine approximation of  $\text{dist}_{L^2}^2$ .

In order to compute  $\text{dist}_{L^2}^2(A, B)$ , an explicit matching  $\xi: \partial B \rightarrow \partial A$  between the shapes' boundaries is computed, where  $\xi(s) := \text{argmin}_{x \in \partial A} \|x - s\|$ . If we denote the straight line from  $\xi(s)$  to  $s$  as

$$\ell_s : [0, 1] \rightarrow \Omega \quad \ell_s(t) := (1-t) \cdot \xi(s) + t \cdot s,$$

we observe  $\text{df}_A(\ell_s(t)) = t \cdot \|\xi(s) - s\|$ . This leads to

$$\begin{aligned} \text{dist}_{L^2}^2(A, B)^2 &= \int_{\partial B} \min_{x \in \partial A} \|x - s\|^2 \, ds = \int_{\partial B} \|\xi(s) - s\|^2 \, ds \\ &= \int_{\partial B} \int_0^1 2t \|\xi(s) - s\|^2 \, dt \, ds \\ &= \int_{\partial B} \int_0^1 2 \text{df}_A(\ell_s(t)) \cdot \|\ell'_s(t)\| \, dt \, ds \\ &= \int_{\partial B} \int_{\ell_s} 2 \text{df}_A(x) \, dx \, ds \end{aligned}$$

In the last equation, we rewrote the equation in means of the line integral evaluated along the line  $\ell_s$ , which still depends on  $\xi(s)$ . Since  $\xi$  is in general difficult to compute, we want to replace the integration domain  $(s, t) \mapsto (1-t) \cdot \xi(s) + t \cdot s$  with a simpler domain. Note that if both  $A$  and  $B$  are concentric circles, the integration domain is exactly  $A \triangle B$ . Therefore, we will approximate  $\text{dist}_{L^2}^2$  via

$$\begin{aligned} \text{dist}_2(A, B) &:= \int_{B \setminus A} 2 \text{df}_A(x) \, dx + \int_{A \setminus B} 2 \text{df}_A(x) \, dx \\ &= \int_{B \setminus (B \cap A)} 2 \text{sdf}_A(x) \, dx - \int_{A \setminus (B \cap A)} 2 \text{sdf}_A(x) \, dx \\ &= \int_B 2 \text{sdf}_A(x) \, dx - \int_A 2 \text{sdf}_A(x) \, dx \end{aligned} \quad (11)$$

This distance can be easily integrated into the sub-problem (6), because it is affine in  $B$ , similar to the Hamming distance formulation (9). The main difference between these two distances is that instead of  $\text{sid}_A$  we use the signed distance function  $\text{sdf}_A$ .

Note that in general,  $\text{dist}_2$  does not approximate  $\text{dist}_{L^2}^2$  very well. First of all, the integration domain  $(s, t) \mapsto (1-t) \cdot \xi(s) + t \cdot s$  does not always coincide with  $A \triangle B$ . Even if it does, the explicit parameterization of the integration domain is partly ignored. Only the variation in the direction of  $\ell_s$  is considered correctly. As a result, the distance between two concentric balls becomes

$$\begin{aligned} \text{dist}_2(B_r, B_R) &= \int_{B_R} 2(|x| - r) \, dx - \int_{B_r} 2(|x| - r) \, dx = 4\pi \left( \frac{R^3}{3} - \frac{R^2 r}{2} + \frac{r^3}{6} \right) \\ &= 2\pi R(R-r)^2 \cdot \left( 1 + \frac{r-R}{3R} \right). \end{aligned}$$

The scaling factor  $1 + \frac{r-R}{3R}$  is the result of the reparametrization and is only negligible if  $|R-r| \ll r$ . Only in that sense can we speak of  $\text{dist}_2$  as an approximation of  $\text{dist}_{L^2}^2$ . Note that even for balls,  $\text{dist}_2$  is only a zeroth order approximation for  $\text{dist}_{L^2}^2$ .

In order to see whether  $\text{dist}_2$  is as useful for shape prior-based image segmentation as  $\text{dist}_{L^2}^2$ , let us take another look at Example 2 of Page 11. If we replace  $\text{dist}_{L^2}^2$  with  $\text{dist}_2$ , the energy  $E_\lambda$  becomes

$$\begin{aligned} E_\lambda(\rho) &= (1 - \lambda) \langle \text{sid}_{B_R}, B_\rho \rangle + \lambda \text{dist}_2(B_r, B_\rho) \\ &= (1 - \lambda) \int_{B_\rho} \text{sid}_{B_R}(x) \, dx + \frac{2\lambda\pi}{3} (2\rho^3 - 3\rho^2r + r^3) \\ &= \begin{cases} \frac{2\lambda\pi}{3} (2\rho^3 - 3\rho^2r + r^3) - (1 - \lambda)\pi\rho^2 & , \text{if } \rho \leq R \\ \frac{2\lambda\pi}{3} (2\rho^3 - 3\rho^2r + r^3) + (1 - \lambda)\pi(\rho^2 - 2R^2) & , \text{if } \rho > R \end{cases} \end{aligned}$$

and its global optimum is realized at (cf. 3<sup>rd</sup> plot of Fig. 6)

$$\rho^*(\lambda) = \begin{cases} R & , \lambda \leq \frac{1}{1+2(R-r)} \\ r + \frac{1-\lambda}{2\lambda} & , \lambda > \frac{1}{1+2(R-r)}. \end{cases}$$

As for  $\text{dist}_{L^2}^2$ ,  $\rho^*$  starts at  $\rho^*(0) = R$  and changes continuously to  $\rho^*(1) = r$ . It also remains at the initial solution  $\rho^* = R$  for a certain range of  $\lambda$ . Therefore, we consider  $\text{dist}_2$  as a good compromise between  $\text{dist}_H$  and  $\text{dist}_{L^2}^2$  to be used in the trust region framework as proposed in [14].

To use a shape distance that depends on the shapes' signed distance function is not a new concept. Rousson and Paragios [22] used the distance

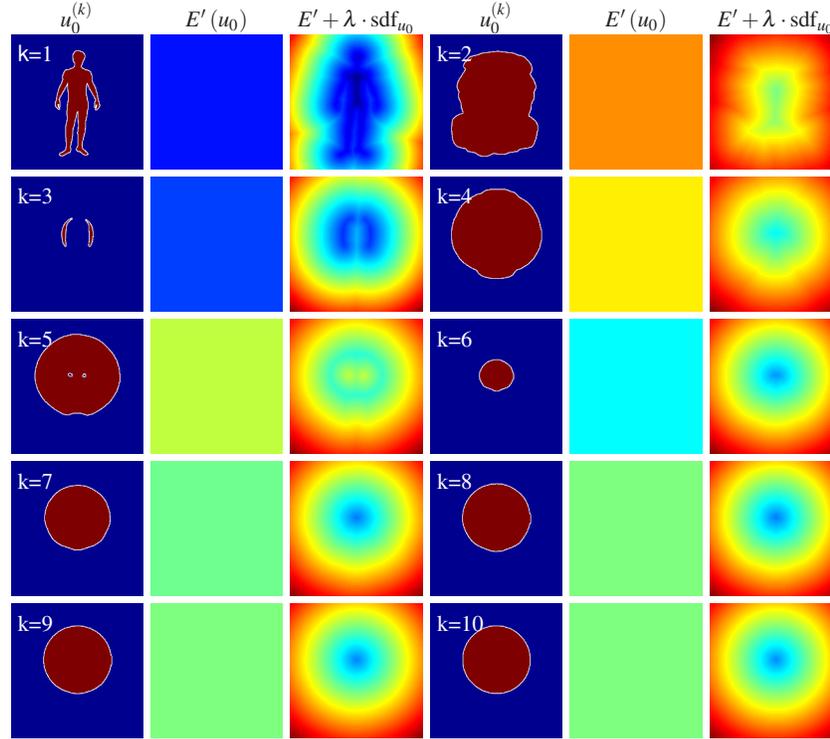
$$\text{dist}_{\text{sdf}}(A, B) = \left( \int_{\Omega} (\text{sdf}_A(x) - \text{sdf}_B(x))^2 \, dx \right)^{\frac{1}{2}}$$

to penalize shape dissimilarity.  $\text{dist}_{\text{sdf}}$  depends in contrast to  $\text{dist}_2$  on the size of the image domain  $\Omega$ , *e.g.*,  $\text{dist}_{\text{sdf}}(B_r, B_R) = (R - r)^2 \text{area}(\Omega)$ . It is therefore not a general, domain-independent shape measure. In addition, we cannot use  $\text{dist}_{\text{sdf}}$  as shape distance for the trust region sub-problem, because it depends on computing the signed distance function of *both shapes*. As a result,  $\text{dist}_{\text{sdf}}(A, S)$  is not affine in  $S$ . Therefore, the sub-problem (6) does not become an energy of the form (1).

The distance  $\text{dist}_2$  is very different in that respect. If we use  $\text{dist}_2(A, \cdot)$  in (6),  $A$  is known and  $\text{sdf}_A$  can be pre-computed. This makes  $\text{dist}_2$  much easier to handle than  $\text{dist}_{\text{sdf}}$ . To our knowledge,  $\text{dist}_2$  was first applied to a computer vision application by Boykov et al. in [2].

## 4 Experiments

In the following we present two applications of the trust region method. To solve the subproblem (6) we use the primal-dual method of [5]. Since only a few iterations are necessary, we will present most of the iterations. By doing so, we substantiate the theoretical results in Section 3 with practical examples.



**Fig. 7** In this toy example we explore the volume constraint  $(\langle 1, u \rangle - V)^2$ , where  $V = \frac{1}{2} |\Omega|$  represents 50% of the image domain's area. Using a man shaped initialization (1<sup>st</sup> image of top row), the method computes in a few iterations a circle (last image of last row). Each row represents two iterations of the trust region method. **1<sup>st</sup>, 4<sup>th</sup> column:** At each iteration  $k$ , we start with a current solution  $u_0^{(k)}$ . **2<sup>nd</sup>, 5<sup>th</sup> column:** The derivative  $E'(u_0)$  could define a gradient descent. The global optimum of this energy is a trivial solution ( $\Omega$  or  $\emptyset$ ). **3<sup>rd</sup>, 6<sup>th</sup> column:**  $E'(u_0) + \lambda \cdot \text{sdf}_{u_0}$  is the data term for the Lagrangian formulation (6) of the trust region approach [14].  $\lambda$  is chosen automatically by the trust region method. **4<sup>th</sup>, 1<sup>st</sup> column:**  $\hat{u}$  is the global optimizer of the Lagrangian trust region sub-problem. It becomes  $u_0$  of the next iteration. Note that in the beginning we can experience big jumps with respect to the segmentation. Nonetheless, the energy decreases in each iteration until we reach a local minimum of the original energy, which for this toy example is a global optimum.

#### 4.1 Volume Constraint

We consider the energy  $E_{\text{Vol}}(u) = (\langle 1, u \rangle - V)^2 + \text{len}(u)$ , which penalizes the deviation of the volume  $\langle 1, u \rangle$  from the target volume  $V > 0$ . The additional length term  $\text{len}(u)$  guarantees that the global minimum  $u^*$  of  $E_{\text{Vol}}$  describes a circle of radius  $r$  that satisfies  $2\pi \cdot r^3 - 2Vr + 1 = 0$ . For this toy example, we set the target volume  $V$  to cover 50% of the image domain  $\Omega$ . In Fig. 7 we show how the trust region method finds the global optimum in just a few iterations. While the shape from one

iteration to the next changes drastically in the beginning, the energy  $E_{\text{Vol}}$  decreases in each iteration and moves the shape to the global optimum of the energy.

Besides the energy, we also show the derivative  $E'$  of the regional energy  $E(u) = (\langle 1, u \rangle - V)^2$ . If the current solution  $u_0^k$  is smaller than the target volume,  $E'(u_0^k)$  is constantly negative (blue or cyan in Fig. 7) in the image domain and encourages larger segmentations. If on the other hand  $u_0^k$  is larger than the target volume,  $E'(u_0^k)$  is constantly positive (orange and yellow in Fig. 7) and encourages smaller segmentations. Without the distance constraint in (6), the approximation  $E'(u) + \text{len}(u)$  would either choose  $\emptyset$  or  $\Omega$  as  $u_0^{(k+1)}$ . Together with the scaled signed distance function that originates from the  $\text{dist}_2$  distance, we are able to smoothly change the shape in a way that the overall energy  $E_{\text{Vol}}$  decreases.

## 4.2 Distribution Constraint

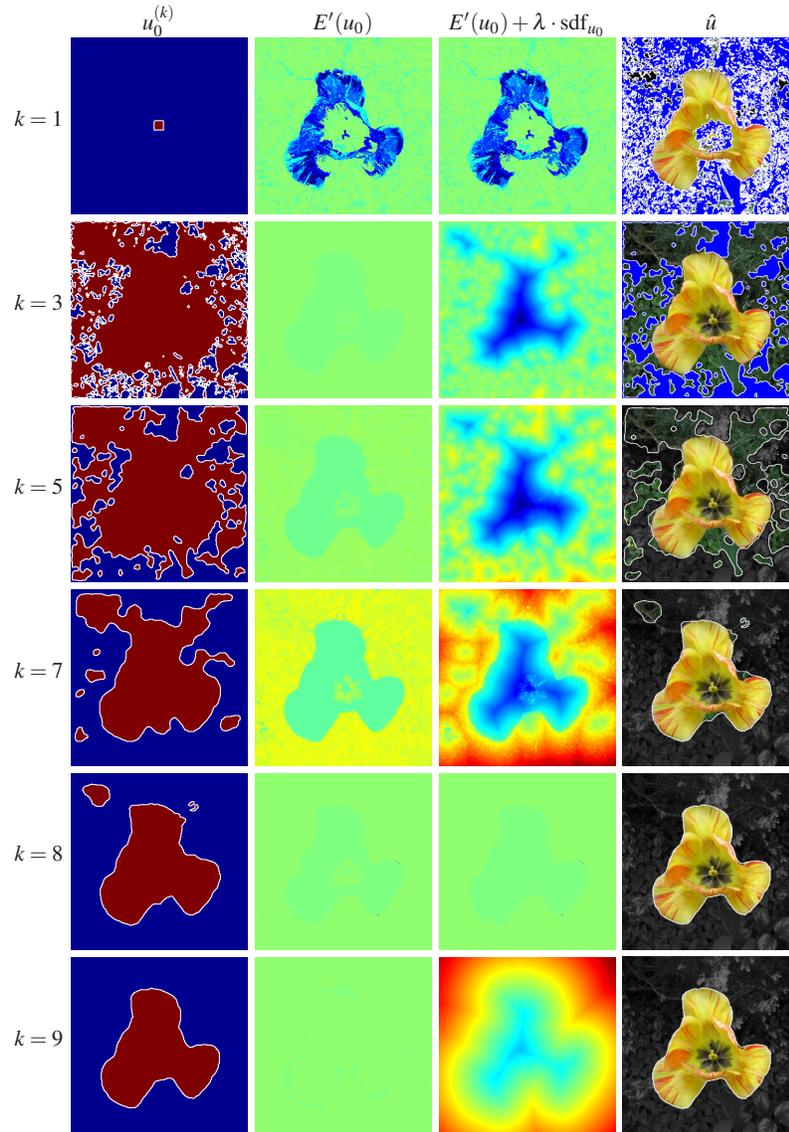
We consider the energy function (4) as introduced in Section 2.2. For this application, we assume knowledge about the object and describe it with 512 color models (8 per color channel). The results are presented in Fig. 8. As in the previous experiment, the data term of the approximation  $\tilde{E}$  is in general not very informative (cf. 2<sup>nd</sup> column of Fig. 7). Only in combination with the distance  $\text{dist}_2$  do we obtain a data term (cf. 3<sup>rd</sup> column of Fig. 8) that helps to improve the segmentation (cf. 4<sup>th</sup> column of Fig. 8).

## 5 Summary

In this chapter we demonstrated that the choice of a shape distance influences the result for image segmentation applications, even at the absence of any shape prior. The importance of the chosen shape distance becomes apparent if we want to deal with local optimization. In particular, we analyzed the behavior with respect to three different distances in the context of the fast trust region image segmentation framework of [14]. In order to obtain good segmentation results, we advocate the use of the distance  $\text{dist}_2$ .

## References

1. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on PAMI* **26**(9), 1124–1137 (2004)
2. Boykov, Y., Kolmogorov, V., Cremers, D., Delong, A.: An integral solution to surface evolution PDEs via Geo-Cuts. In: A. Leonardis, H. Bischof, A. Pinz (eds.) *Europ. Conf. on Computer Vision, LNCS*, vol. 3953, pp. 409–422. Springer, Graz, Austria (2006)



**Fig. 8** We explore the segmentation task (4) that uses a distribution of 512 entries. Using a square as initialization (1<sup>st</sup> image of top row), the method computes in a few iterations the segmentation of the flower (last image of last row). Each row represents one iteration of the trust region method. **1<sup>st</sup> column:** At each iteration  $k$  we start with a current solution  $u_0$ . **2<sup>nd</sup> column:** The derivative  $E'(u_0)$  could define a gradient descent, but it provides only for a weak data term. **3<sup>rd</sup> column:**  $E'(u_0) + \lambda \cdot \text{sdf}_{u_0}$  is the data term for the Lagrangian formulation (6) of the trust region approach [14].  $\lambda$  is chosen automatically by the trust region method. **4<sup>th</sup> column:**  $\hat{u}$  is the global optimizer of the Lagrangian trust region sub-problem. For visualization purposes we set the background in the first two rows to blue and in the remaining rows to the gray-scale of the original image.

3. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Transactions on PAMI* **23**(11), 1222–1239 (2001)
4. Brox, T., Rousson, M., Deriche, R., Weickert, J.: Unsupervised segmentation incorporating colour, texture, and motion. In: N. Petkov, M.A. Westenberg (eds.) *Computer Analysis of Images and Patterns, LNCS*, vol. 2756, pp. 353–360. Springer, Groningen, The Netherlands (2003)
5. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Im. Vis.* **40**(1), 120–145 (2011)
6. Chan, T., Esedoğlu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics* **66**(5), 1632–1648 (2006)
7. Cremers, D., Osher, S.J., Soatto, S.: Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *Int. J. of Computer Vision* **69**(3), 335–351 (2006)
8. Cremers, D., Schmidt, F.R., Barthel, F.: Shape priors in variational image segmentation: Convexity, Lipschitz continuity and globally optimal solutions. In: *IEEE Int. Conf. on Computer Vision and Pattern Recognition*. Anchorage, Alaska (2008)
9. Cremers, D., Soatto, S.: A pseudo-distance for shape priors in level set segmentation. In: N. Paragios (ed.) *IEEE 2nd Int. Workshop on Variational, Geometric and Level Set Methods*, pp. 169–176. Nice (2003)
10. Delong, A., Gorelick, L., Schmidt, F.R., Veksler, O., Boykov, Y.: Interactive segmentation with super-labels. In: *Int. Conf. on Energy Minimization Methods for Computer Vision and Pattern Recognition, LNCS*, vol. 6819, pp. 147–162. Springer, Saint Petersburg, Russia (2011)
11. Delong, A., Osokin, A., Isack, H.N., Boykov, Y.: Fast approximate energy minimization with label costs. *Int. J. of Computer Vision* **96**(1), 1–27 (2012)
12. Gao, T., Koller, B.P.D.: A segmentation-aware object detection model with occlusion handling. In: *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 1361–1368. Colorado Springs, Colorado (2011)
13. Gorelick, L., Ayed, I.B., Schmidt, F.R., Boykov, Y.: An experimental comparison of trust region and level sets. arXiv <http://arxiv.org/abs/1311.2102> (2013)
14. Gorelick, L., Schmidt, F.R., Boykov, Y.: Fast trust region for segmentation. In: *IEEE Int. Conf. on Computer Vision and Pattern Recognition*. Portland, Oregon (2013)
15. Leventon, M., Grimson, W., Faugeras, O.: Statistical shape influence in geodesic active contours. In: *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 316–323. Hilton Head Island, SC (2000)
16. Li, C., Xu, C., Gui, C., Fox, M.D.: Level set evolution without re-initialization: A new variational formulation. In: *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 430–436 (2005)
17. Ling, H., Jacobs, D.W.: Shape classification using the innerdistance. *IEEE Transactions on PAMI* **29**(02), 286–299 (2007)
18. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics* **42**, 577–685 (1989)
19. Nieuwenhuis, C., Strelakovskiy, E., Cremers, D.: Proportion priors for image sequence segmentation. In: *IEEE Int. Conf. on Computer Vision*, pp. 1769–1776. Sydney, Australia (2013)
20. Nocedal, J., Wright, S.J.: *Numerical Optimization*. Springer, Berlin (2006)
21. Osher, S.J., Sethian, J.A.: Fronts propagation with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations. *Journal of Computational Physics* **79**, 12–49 (1988)
22. Rousson, M., Paragios, N.: Shape priors for level set representations. In: A. Heyden, et al. (eds.) *Europ. Conf. on Computer Vision, LNCS*, vol. 2351, pp. 78–92. Springer (2002)
23. Schmidt, F.R., Boykov, Y.: Hausdorff distance constraint for multi-surface segmentation. In: *Europ. Conf. on Computer Vision, LNCS*, vol. 7572, pp. 598–611. Springer, Florence, Italy (2012)
24. Soares, J.V.B., Cesar, J.J.G.L.R.M., Jelinek, H.F., Cree, M.J.: Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Trans Med Imaging* **25**(9), 1214–1222 (2006)
25. Tang, M., Gorelick, L., Veksler, O., Boykov, Y.: Grabcut in one cut. In: *IEEE Int. Conf. on Computer Vision*, pp. 1769–1776. Sydney, Australia (2013)