

# Shape-Based Mutual Segmentation

Tammy Riklin-Raviv · Nir Sochen · Nahum Kiryati

Received: 26 December 2006 / Accepted: 21 November 2007 / Published online: 29 December 2007  
© Springer Science+Business Media, LLC 2007

**Abstract** We present a novel variational approach for simultaneous segmentation of *two* images of the same object taken from different viewpoints. Due to noise, clutter and occlusions, neither of the images contains sufficient information for correct object-background partitioning. The evolving object contour in each image provides a dynamic prior for the segmentation of the other object view. We call this process *mutual segmentation*. The foundation of the proposed method is a unified level-set framework for region and edge based segmentation, associated with a shape similarity term. The suggested shape term incorporates the semantic knowledge gained in the segmentation process of the image pair, accounting for excess or deficient parts in the estimated object shape. Transformations, including planar projectivities, between the object views are accommodated by a registration process held concurrently with the segmentation. The proposed segmentation algorithm is demonstrated on a variety of image pairs. The homography between each of the image pairs is estimated and its accuracy is evaluated.

**Keywords** Mutual segmentation · Level sets · Shape · Perspective transformation · Planar projective homography

## 1 Introduction

Segmentation is defined as the task of partitioning the image into foreground (object) and background regions. Since the

concept of *object* is subjective and application dependent, segmentation is inherently difficult. Moreover, even when the object of interest is well-defined, segmentation is challenging. This is so since in the imaging process the inherent properties of the inspected object may be corrupted due to noise, occlusion, illumination conditions and more. General *syntactical* assumptions such as continuity and smoothness of edges and homogeneity of semantic regions should thus be complemented with *semantical* prior information on the objects to be segmented.

This paper addresses the segmentation of an image pair of the same object in different appearances and poses. We can then redefine segmentation as the task of extracting the common object in both images. Uncertainty regarding the object boundaries can be resolved having two instances instead of one. Nevertheless, this setting provides only an implicit prior. The “double” segmentation problem is thus far from being trivial.

Mathematical modeling of prior information and its incorporation in object extraction frameworks is an active field of research. Segmentation is commonly inferred by minimizing an objective functional. Thus, the compatibility of the object-background partition with the image data is constrained subject to a general model on natural images, e.g. (Mumford and Shah 1989). Specific information on the segmented object, based on its expected characteristics, is incorporated via additional constraints. These may include shape, color, texture, etc. Model-based approaches embody common properties of the object class. Refer for example to (Holtzman-Gazit et al. 2006) which extracts thin structures such as blood vessels, or to (Rochery et al. 2006) which incorporates geometric information to segment road networks. When the object shape is specified, resemblance of the segmented object to the reference shape can be also imposed. Prior-based segmentation methods, e.g. (Cremers

T. Riklin-Raviv (✉) · N. Kiryati  
School of Electrical Engineering, Tel Aviv University, Tel Aviv  
69978, Israel  
e-mail: tammy@eng.tau.ac.il

N. Sochen  
Department of Applied Mathematics, Tel Aviv University, Tel  
Aviv 69978, Israel

and Soatto 2003; Cremers et al. 2006; Riklin-Raviv et al. 2004, 2005, 2007; Rousson and Paragios 2002) assume the existence of a well-defined shape prior and use it to extract occluded object boundaries. The statistical approaches, e.g. (Chen et al. 2002; Cremers et al. 2003; Huang et al. 2004; Leventon et al. 2000b; Rousson and Paragios 2002; Tsai and Yezzi 2003), capture possible shape variability by employing a set of similar but not identical shape priors. These methods, however, depend on the availability of a comprehensive set of priors or a segmented instance of the object of interest.

The proposed framework is useful when explicit prior knowledge is limited or not available. Instead, two images of the same object are given. Had good segmentation been possible in either image, it could have supported the segmentation of the other. However, in the absence of additional information on the object of interest, segmentation of each image by itself is prone to errors. Specifically, background regions can be mistakenly labeled as foreground (excess) or vice versa (deficiency). The resulting (possibly erroneous) segmentation of one of the images cannot provide a reliable prior for the other.

We therefore suggest a coupled segmentation process for both images, in which the information gained in the evolving segmentation of one image is a *dynamic* prior for the other. The segmentation and thus the prior information are refined along the process. We call this mutually supporting evolution process *mutual segmentation*.

Recent works of similar spirit include (Rother et al. 2006), (Vedaldi and Soatto 2006) and (Yezzi et al. 2003). In the *cosegmentation* method of (Rother et al. 2006), both images are simultaneously partitioned assuming that the common property of the foreground regions is their color probability density function (modeled by histograms) which should be also significantly different from the color distribution of the background. The *unsupervised* segmentation algorithm presented in (Vedaldi and Soatto 2006) uses the best SIFT matches of Harris-Affine features to extract the common objects in image pairs. Similarly to (Yezzi et al.

2003), we presume that the object instances resemble in their *shapes*, thus having the benefit of being specific to the particular object of interest and insensitive to color and (in most cases) illumination variation. Yet, while (Yezzi et al. 2003) handles images corrupted only by noise, where at least one of them nearly contains sufficient information to be segmented by itself, we deal with cluttered images corrupted by noise and occlusions. Possible transformations, including projectivities between the object views, are accommodated by employing a concurrent segmentation and registration process as in (Riklin-Raviv et al. 2007).

We use the level set framework (Osher and Sethian 1988) for segmentation, where the segmenting curve is the zero level of a level set function that evolves subject to some predefined constraints. This parameterization-free representation of the curve allows automatic changes in its topology. The shape of the object being segmented is dynamically encoded by assigning the positive and the negative levels of the level-set function to the foreground and the background image regions respectively. Transformations applied on the level-set coordinate system transform the represented shape accordingly. This enables a meaningful definition of a dissimilarity measure between shapes that accommodates parametric transformations.

The mutual segmentation approach goes beyond the concepts of shape-based segmentation, because a well-defined shape prior is not available and the matching is between two possibly corrupted and noisy images. The main difficulty resides in labeling regions where the aligned images do not overlap. Obviously, erroneous foreground-background classifications spoil the segmentation of both images. Figure 1 exemplifies labeling ambiguity. The regions pointed by the red arrows could be attributed either to the boot (according to one of the images) or to the background (according to the other).

The conflict between two possible interpretations of jointly segmented images has never been addressed before. Note that when more than two object instances are available, this ambiguity can be resolved by applying a majority



**Fig. 1** **a, b** Two instances of a boot. **c** Superposition of the aligned boot images. There is an inherent labeling ambiguity in aligned image regions that correspond to different objects. The regions pointed

by the red arrows could be attributed either to the boot (according to one of the images) or to the background (according to the other)

rule (Duci et al. 2002). Having only two images, we favor the image partitioning that minimizes a biased shape dissimilarity measure between the images. The definition of this biased shape term is one of the important contributions of the proposed study.

We suggest a novel framework for *mutual segmentation* of two images of the same object related by projective transformation.<sup>1</sup> Segmentation is carried out concurrently with registration of the evolving contours. The foundation of the proposed method is the construction of two level set functions, one for each image. Their evolution is controlled by the data contained in the associated image together with the shape information gained in the segmentation process of the other image. The shape information is embedded in a biased shape dissimilarity measure that accommodates either deficient or excess parts in the extracted object shape. This measure is also invariant to planar projective transformations. The outcomes of the proposed algorithm include segmentation of the object appearances in both images and the recovery of the homography that aligns them.

The paper is organized as follows. In Sect. 2 we present the Bayesian rationale of the segmentation problem together with the energy formulation to which it leads. We review some of the state-of-the-art level-set segmentation methods: the region based approaches of Chan and Vese (2001) and Paragios and Deriche (2002), the Geodesic Active contour method introduced by Kichenassamy et al. (1995) and Caselles et al. (1997), and the edge alignment constraint suggested by Vasilevskiy and Siddiqi (2001) and independently by Kimmel and Bruckstein (2003). We also outline the shape similarity measure suggested in (Riklin-Raviv et al. 2007) for prior based segmentation. Section 3 introduces the biased shape dissimilarity measure which plays a key role in the proposed mutual segmentation algorithm. The embedding of the transformation model within the variational framework and the minimization of the resulting functional are also considered. The mutual segmentation algorithm together with the implementation details are presented in Sect. 4. Experimental results are provided in Sect. 5. We conclude in Sect. 6.

## 2 Statistical Set Up and Prior Art

### 2.1 General Principles

Most segmentation approaches, despite their diversity, are motivated by the same Gestalt principles of perception, i.e. simplicity (minimum description length), similarity (homogeneity of semantic regions), continuity, proximity and closure. The image is then partitioned according to a subset

<sup>1</sup>A preliminary version of this work was published in (Riklin-Raviv et al. 2006).

of these rules using various mathematical tools. Commonly, segmentation is obtained by minimizing an objective functional either by discrete (graph-based) approaches such as Markov Random Fields or by the continuous formulation of calculus of variations. While in this paper we use PDEs to solve segmentation problems, Bayesian statistical inference formulation is used to set a common language between the discrete and continuous approaches.

Given an image  $I(\mathbf{x})$  we would like to infer the delineating curve  $C$  between an object and its background. This is done via the maximization of the probability distribution function (PDF)  $P(C|I)$ , using Bayes law:

$$P(C|I) \propto P(I|C)P(C). \quad (1)$$

The prior probability  $P(C)$  will be reformulated to have “syntactic” and “semantic” components. Classical active contours methods, such as the snake algorithm of Kass et al. (1988), use parametric representation  $C = C(p)$ . In this formulation the term  $-\log P(I|C) = \int g(|\nabla I(C)|)ds$  is the “external force”, where  $ds$  is the arc-length. The term  $-\log P(C) = \int (\alpha|C_p|^2 + \beta|C_{pp}|^2)dp$  is the “internal force”. The maximization over all possible separating curves is done by minimizing  $-\log P(C|I)$ . Note that here the prior is syntactic since it dictates the degree of smoothness of the curve and is not directly related to the class of images or objects to be segmented. Our main interest in this work is in a special kind of “semantic” prior that depends on another image of the object.

In the level set framework for curve evolution (Osher and Sethian 1988) an evolving curve  $C(t)$  is defined as the zero level of a level set function  $\phi: \Omega \rightarrow \mathbb{R}$  at time  $t$ :

$$C(t) = \{\mathbf{x} \in \Omega \mid \phi(\mathbf{x}, t) = 0\}. \quad (2)$$

Following (Chan and Vese 2001), we use the Heaviside function of  $\phi$

$$H(\phi(\mathbf{x})) = \begin{cases} 1 & \phi(\mathbf{x}) \geq 0, \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

as an object indicator function, assigning the positive and the negative levels of  $\phi$  to the foreground and the background image regions respectively. We can now rephrase our PDF as

$$P(\phi|I) \propto P(I|\phi)P(\phi). \quad (4)$$

Next we elaborate on the conditional probability term  $P(I|\phi)$ .

### 2.2 Region-Based Data Term

Let  $I: \Omega \rightarrow \mathbb{R}^+$  denote a gray level image, where  $\Omega \subset \mathbb{R}^2$  is the image domain. Let  $\omega \subset \Omega$  be an open subset, not necessarily connected, of the image domain  $\Omega$ . In the spirit of the

Gestalt laws of similarity and proximity, we define a boundary  $C = \partial\omega$  that delimits homogeneous regions in  $\Omega$ . In particular, we use the two-phase formalism, in which the image is partitioned into foreground and background regions. Thus, given a feature  $G$ , e.g. the average gray level, we look for a curve  $C$  that maximizes the difference between two scalars  $u^+$  and  $u^-$  defined as follows:

$$\begin{aligned} u^+ &= A^+ \int_{\omega} G^+(I(\mathbf{x})) d\mathbf{x}, \\ u^- &= A^- \int_{\Omega \setminus \omega} G^-(I(\mathbf{x})) d\mathbf{x}. \end{aligned} \quad (5)$$

The superscripts + and – correspond to the feature values in  $\omega \subset \Omega$  and in  $\Omega / \omega$  respectively. Hereafter we denote  $\mathbf{x} \equiv (x, y)$ ,  $A^+ = 1 / \int_{\omega} d\mathbf{x}$  and  $A^- = 1 / \int_{\Omega \setminus \omega} d\mathbf{x}$ . The possibly different functions  $G^+$  and  $G^-$  are defined on the object and the background domains, respectively.

In general, one can use a set of features  $\{G_i\}$  corresponding to two sets of scalars  $\{u_i^+\}, \{u_i^-\}$ . The features chosen should be related to the expected image homogeneity. In the work of Chan and Vese (2001) the image is approximated by a piecewise constant function whose values are given by  $G_1^+(I(\mathbf{x})) = G_1^-(I(\mathbf{x})) = I(\mathbf{x})$ . Hence  $u_1^+ = \overline{I_{in}}$  and  $u_1^- = \overline{I_{out}}$  are the average gray levels in the object regions and in the background regions respectively. In that formulation it is assumed that the PDF of the gray levels in each region is a Gaussian with  $\sigma = 1$ . More elaborate characterization of the partition can be obtained using color histograms or a Gaussian mixture model. For texture-based segmentation the Gabor filters may be used as in (Sandberg et al. 2002) and in (Sagiv et al. 2006). The definition of  $G(I)$  could be further extended as suggested in (Vese and Chan 2002), where  $u_+, u_- \in C^1(\mathbb{R}^2)$  are smooth approximations of the regions in  $I$ . In this study we use the average gray levels and the variance (Vese and Chan 2002; Lorigo and Faugeras 2000):

$$G_2^+(I) = (I(\mathbf{x}) - \overline{I_{in}})^2, \quad G_2^-(I) = (I(\mathbf{x}) - \overline{I_{out}})^2. \quad (6)$$

We may now express the term  $-\log P(I|\phi)$  via a region based cost functional with a well defined integration domain:

$$\begin{aligned} E_{RB}(\phi) &= \sum_{i=1}^2 \int_{\Omega} [(G_i^+(I(\mathbf{x})) - u_i^+)^2 H(\phi) \\ &\quad + (G_i^-(I(\mathbf{x})) - u_i^-)^2 (1 - H(\phi))] d\mathbf{x}. \end{aligned} \quad (7)$$

An elegant statistical formulation representing the region-based term was introduced in (Zhu and Yuille 1996), followed by (Paragios and Deriche 2002) and reviewed in (Cremers et al. 2007).

The evolving boundary  $C(t)$  is derived from  $\phi(t)$  using (2). For a given  $\phi(t)$  and  $\{G_i\}$ , the feature values  $u_i^+$

and  $u_i^-$  are updated at each iteration according to (5). The level set function  $\phi$  evolves via gradient descent:

$$\phi_t^{\text{RB}} = \delta(\phi) \sum_{i=1}^2 [G_i^-(I(\mathbf{x})) - u_i^-]^2 - (G_i^+(I(\mathbf{x})) - u_i^+)^2, \quad (8)$$

where  $\delta$  is the derivative of the Heaviside function  $H$ .

## 2.3 Edge-Based Data Terms

### 2.3.1 Geodesic Active Contour: Data Part

Edge based segmentation approaches usually define the object boundaries by the local maxima of the image gradients. Let  $C(s) = (x(s), y(s))$  be the parametric description of a planar contour  $C: [0, L] \rightarrow \mathbb{R}^2$  where  $s$  is an arc-length parameter and  $L$  is the length of  $C$ .

Let  $\nabla I(x, y) = (I_x, I_y)^T = (\frac{\partial I(x, y)}{\partial x}, \frac{\partial I(x, y)}{\partial y})^T$  denote the vector field of the image gradients. The Geodesic Active Contour (GAC) term (Caselles et al. 1997) is given by

$$E_{\text{GAC}}(C) = \int_0^L g_{\text{GAC}}(|\nabla I(C(s))|) ds. \quad (9)$$

The function  $g_{\text{GAC}}$  is inversely proportional to  $|\nabla I|$ , thus,  $E_{\text{GAC}}(C)$  is minimized when the curve  $C$  is located on the maxima of the absolute values of the image gradients. We unorthodoxly split the GAC term, (9), into two terms. This splitting will be justified in Sect. 3.3 where both terms will appear in the final energy formulation. The data term (DGAC) is given by

$$E_{\text{DGAC}}(C) = \int_0^L g_{\text{DGAC}}(|\nabla I(C(s))|) ds \quad (10)$$

where

$$g_{\text{DGAC}}(|\nabla I|) = -\frac{|\nabla I|^2}{1 + |\nabla I|^2}. \quad (11)$$

This term vanishes as the gradient magnitudes decrease to zero and attains  $-1$  asymptotically for large gradients. Expressing this term in a level-set framework we obtain

$$E_{\text{DGAC}} = \int_{\Omega} g_{\text{DGAC}}(|\nabla I|) |\nabla H(\phi(\mathbf{x}))| d\mathbf{x}, \quad (12)$$

with the associated gradient descent equation:

$$\phi_t^{\text{DGAC}} = \delta(\phi) \text{div} \left( g_{\text{DGAC}}(|\nabla I|) \frac{\nabla \phi}{|\nabla \phi|} \right). \quad (13)$$

The GAC functional includes another geometrical term that will be described in Sect. 2.4.

### 2.3.2 Edge Alignment Term

Segmentation can be refined by constraining the level set normal directions to align with the directions of the image gradients as suggested in (Kimmel and Bruckstein 2003; Vasilevskiy and Siddiqi 2001). The edge alignment term (EA) defined in (Kimmel 2003) takes the form  $E_{EA}(C) = -\int_0^L |\langle \nabla I(\mathbf{x}(s)), \vec{n}(s) \rangle| ds$ , where  $\langle \cdot, \cdot \rangle$  denotes inner product and  $\vec{n}(s) = \{-y_s(s), x_s(s)\}$  is the exterior normal to the curve  $C$ . The expression for  $E_{EA}(C)$  is an integration of the projection of  $\nabla I$  on the normal  $\vec{n}(s)$  along the curve. A minor contribution of this paper is the level-set formulation of the alignment term:

$$E_{EA} = - \int_{\Omega} \left| \left\langle \nabla I, \frac{\nabla \phi}{|\nabla \phi|} \right\rangle \right| |\nabla H(\phi)| d\mathbf{x} \quad (14)$$

where  $\nabla \phi(\mathbf{x})/|\nabla \phi(\mathbf{x})|$  is normal to the level-set  $\phi$  in  $\mathbf{x}$ . The associated gradient descent equation is

$$\phi_t^{EA} = -\delta(\phi) \text{sign}(\langle \nabla \phi, \nabla I \rangle) \Delta I, \quad (15)$$

where  $\Delta I$  is the Laplacian. This equation is similar to the one derived in (Kimmel 2003).

### 2.4 Syntactic Prior: Geometry

The prior probability  $P(C)$  in (1) is determined by the minimum description length criterion (based on the Gestalt principle of simplicity). That is  $P(C) \propto \exp(-\nu|C|)$ ,  $\nu > 0$  (Cremers et al. 2007), where  $|C| = \int ds = \int |C_p| dp$  is the curve length. An equivalent representation, using the level set formulation, takes the form:

$$|C| = E_{LEN} = \int_{\Omega} |\nabla H(\phi(\mathbf{x}))| d\mathbf{x}. \quad (16)$$

This functional measures the length of the curve and usually serves as an indicator for the curve smoothness (Chan and Vese 2001). Minimizing (16) with respect to  $\phi$ , we obtain the associated Euler Lagrange equation for  $\phi$ :

$$\phi_t^{LEN} = \delta(\phi) \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right). \quad (17)$$

Combining  $E_{LEN}$  and  $E_{DGAC}$ , (defined in (12)), we get the usual form of the GAC functional (Caselles et al. 1997; Kichenassamy et al. 1995),

$$E_{GAC} = \int_{\Omega} g_{GAC}(|\nabla I|) |\nabla H(\phi(\mathbf{x}))| d\mathbf{x}, \quad (18)$$

where  $g_{GAC} = 1 + g_{DGAC} = 1/(1 + |\nabla I|^2)$ . The gradient descent equation is

$$\phi_t^{GAC} = \delta(\phi) \text{div} \left( g_{GAC}(|\nabla I|) \frac{\nabla \phi}{|\nabla \phi|} \right). \quad (19)$$

### 2.5 Semantic Prior: Shape Term

In the previous subsections we described the edge-based, region-based and smoothness constraints that control the segmentation based on the image data and on general assumptions on natural images. Here and in the next section we present several forms of the shape term, denoted by  $E_{SHAPE}$ , that are derived from the semantic knowledge on the object of interest.

Denoting a prior shape representation by  $\tilde{\phi}: \Omega \rightarrow \mathbb{R}$ , the statistical formulation is then

$$\begin{aligned} P(\phi, T|I, \tilde{\phi}) &\propto P(I|\phi, \tilde{\phi}, T) P(\phi|\tilde{\phi}, T) P(T, \tilde{\phi}) \\ &= P(I|\phi) P(\phi|\tilde{\phi}, T) P(T) \end{aligned}$$

where  $T$  is the transformation that aligns  $\phi$  and  $\tilde{\phi}$  and pairwise independence of  $I, \tilde{\phi}$  and  $T$  is assumed. The logarithm of the inverse conditional probability term  $-\log P(\phi|\tilde{\phi}, T)$  can be substituted by a dissimilarity measure  $D(\phi, T(\tilde{\phi}))$  between shape representations  $\tilde{\phi}$  and  $\phi$ . Some previous approaches, e.g. (Leventon et al. 2000a, 2000b), used:

$$D(\phi, T(\tilde{\phi})) = \int_{\Omega} (\phi(\mathbf{x}) - T(\tilde{\phi}(\mathbf{x})))^2 d\mathbf{x}$$

where  $\phi$  and  $T(\tilde{\phi})$  are signed distance functions. This measure is affected, however, by the size of the background area  $\Omega \setminus \omega$ , see (Cremers and Soatto 2003) and references therein. Moreover, in a cluttered image, when the weight of this measure in the segmentation functional is high, objects that do not correspond to the prior shape are ignored (Cremers et al. 2006). To avoid these drawbacks several modifications to control the integration domain have been suggested (Cremers and Soatto 2003; Cremers et al. 2006; Rousson and Paragios 202). In all dissimilarity measures between level-set functions (represented as signed distance functions) the alignment term  $T$  is restricted to be an isometry. This restriction guarantees that the transformed level-set function  $\tilde{\phi}_T = T(\tilde{\phi})$  will preserve the form of a signed distance function.

Recently, we suggested (Riklin-Raviv et al. 2007) to use the square difference between the Heaviside functions of  $\phi$  and  $\tilde{\phi}$  as a dissimilarity measure between the shape representations, where  $\phi$  and  $\tilde{\phi}$  are *not* necessarily distance functions<sup>2</sup>

$$D(\phi, \tilde{\phi}|T) = \int_{\Omega} [H(\phi(\mathbf{x})) - H(T(\tilde{\phi}(\mathbf{x})))]^2 d\mathbf{x}. \quad (20)$$

This formulation enables the introduction of the projective alignment term  $T$  between the shapes. Furthermore,  $D$  does

<sup>2</sup>A significant gain from not enforcing  $\phi$  to be a distance function is the elimination of the process of re-distancing (Gomes and Faugeras 2000; Strain 1999).

not depend on the size of the background area  $\Omega \setminus \omega$ . The shape term (20) is suitable when the prior  $\tilde{\phi}$  is perfect and constant in time. In the following section, a different dissimilarity measure is suggested, which is more appropriate when  $\tilde{\phi}$  is the evolving (and imperfect) segmentation of the other image.

### 3 Mutual Segmentation with Projectivity

In this paper we consider the segmentation of two images,  $I_1$  and  $I_2$  that depict two imperfect (occluded, noisy etc.) instances of an object. The segmentation is done by performing a joint maximum posteriori probability (MAP) approximation via an alternating maximization. Indeed, since a reliable prior  $\tilde{\phi}$  is not available, each evolving contour in its turn is employed as a prior for the other. The two level set functions  $\phi_1$  and  $\phi_2$  are alternately evolved. At even iterations the segmenting level-set is  $\phi = \phi_1$  and the prior is given by  $\tilde{\phi} = \phi_2$ . At odd iterations  $\phi = \phi_2$  and  $\tilde{\phi} = \phi_1$ . The statistical inference problem takes the form:

$$P(\phi, \tilde{\phi}, T | I_1, I_2) \propto P(I_1, I_2 | \phi, \tilde{\phi}, T) P(\phi | \tilde{\phi}, T) P(\tilde{\phi}) P(T).$$

We assume that  $P(T)$  is flat such that no projective transformation is a-priori preferred. The assumption that the two views are related is incorporated by initializing  $T$  by the identity transformation. The crucial observation is that  $P(C | \tilde{C}, T)$  (or equivalently  $P(\phi | \tilde{\phi}, T)$ ) which is defined via a dissimilarity measure between curves is *not* symmetric with respect to the two curves. This is so since  $P(\phi_1 | \phi_2, T)$  is not necessarily equal to  $P(\phi_2 | \phi_1, T)$ . The symmetric shape dissimilarity measure defined in (20) is thus not suitable for mutually evolving level-set functions. In the next subsection we present a biased shape dissimilarity term which is more appropriate.

#### 3.1 Biased Shape Dissimilarity Measure

Consider the image pair in Figs. 2a–b. Both images have deficiencies. In Fig. 2a the hoop is absent. In Fig. 2b a portion of the creature's leg was erased. When the final segmentation of Fig. 2a is the prior for the segmentation of Fig. 2b and vice versa, the imperfections of each segmentation spoil the other as shown in Figs. 2e–f. Note that the left leg is incorrectly segmented in Figs. 2e, while the hoop in Fig. 2f is not segmented at all.

The images in Figs. 2c–d contain superfluous hoops located in different places. When each segmentation is the prior for the other, using the unbiased dissimilarity measure in (20), the contours of the superfluous hoops in Figs. 2c–d undesirably appear in the segmentation shown in Figs. 2h, g respectively.

The discrimination between integral object parts (leg, hoop) and other surrounding objects (superfluous hoops) raises a fundamental question which extends beyond the scope of the current work and relates to perceptual organization of images in general. Given more than two images, this difficulty can be tackled by employing a majority decision rule to determine the object-background partition. However, for mutual segmentation of two images, another decision tool or source of information is needed. For simplicity, we assume either of the following “world states”:

1. **The images to segment have (mostly) deficiencies.** Application of this rule is demonstrated in Figs. 2i–j to resolve labeling ambiguities between Fig. 2a and Fig. 2b. Thus, for example, the missing part in the leg of the creature shown in Fig. 2b has been labeled as part of the object.
2. **The images to segment have (mostly) excess parts.** When this rule is applied to Figs. 2c–d, the superfluous hoops are labeled as background, as shown in Figs. 2k–l.
3. **The prior shape is perfect.** Examples for the application of this rule are shown in Figs. 2e–h.

Refer again to the dissimilarity measure in (20). The cost functional integrates the non-overlapping object-background regions in both images indicated by  $H(\phi)$  and  $H(T(\tilde{\phi}))$ . This is equivalent to a pointwise exclusive-or (xor) operation integrated over the image domain. We may thus rewrite the functional as follows:

$$D(\phi, \tilde{\phi} | T) = \int_{\Omega} [H(\phi)(1 - H(\tilde{\phi}_T)) + (1 - H(\phi))H(\tilde{\phi}_T)]d\mathbf{x}. \quad (21)$$

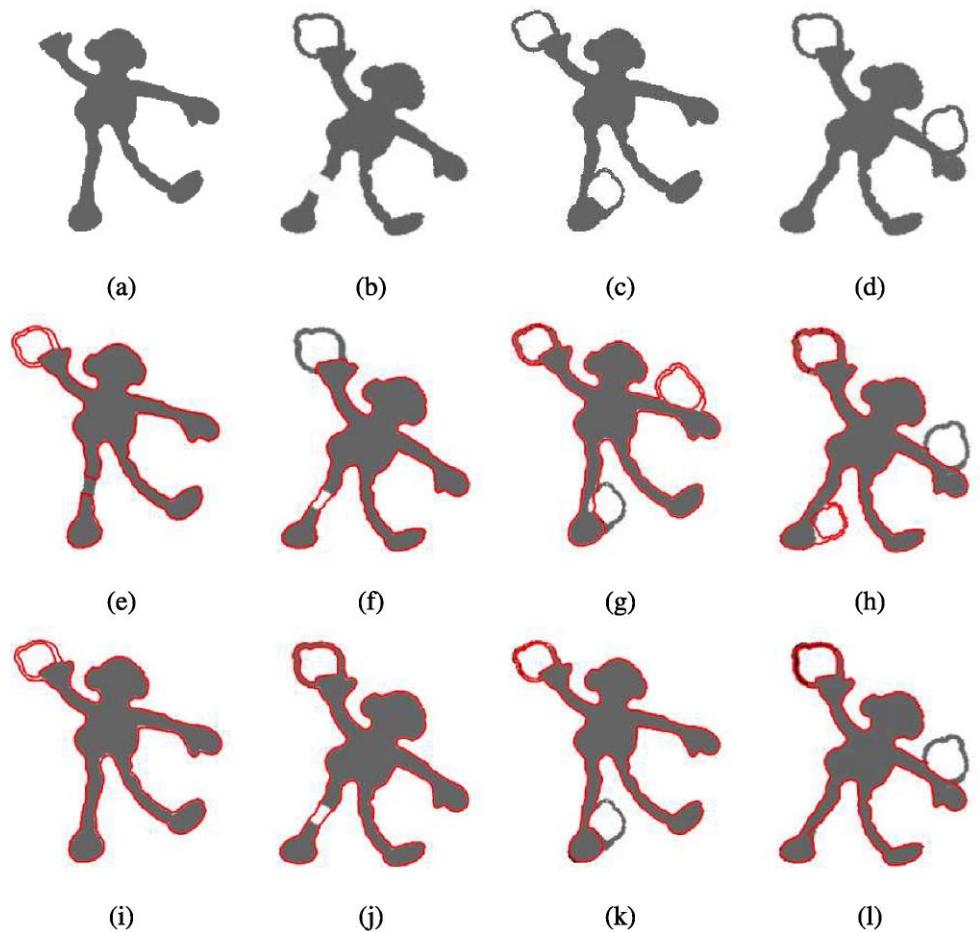
To simplify the expression we denote  $T(\tilde{\phi}) \equiv \tilde{\phi}_T$ . Note that the expressions (20) and (21) are identical, since  $H(\phi)$  is equal to  $(H(\phi))^2$  when  $H(\phi)$  is a strictly binary function. There are two types of disagreement between the labeling of  $H(\phi)$  and  $H(\tilde{\phi}_T)$ . The left term in (21) does not vanish if there exist image regions labeled as *object* by the image data ( $\phi$ ) and labeled as *background* by the shape prior ( $\tilde{\phi}_T$ ). The right term in (21) does not vanish if there exist image regions labeled as *background* by the image data and labeled as *object* by  $\tilde{\phi}_T$ . Inserting a weight parameter  $\mu \geq 0$ , the relative contributions of the terms are changed.

$$E_{\text{SHAPE}}(\phi, \tilde{\phi}, T) = \int_{\Omega} [\mu H(\phi)(1 - H(\tilde{\phi}_T)) + (1 - H(\phi))H(\tilde{\phi}_T)]d\mathbf{x}. \quad (22)$$

The associated gradient descent equation for  $\phi$  is then

$$\phi_t^{\text{SHAPE}} = \delta(\phi)[H(\tilde{\phi}_T) - \mu(1 - H(\tilde{\phi}_T))]. \quad (23)$$

**Fig. 2** (Color online) **a, b** Input images with deficiencies (leg, hoop). **c, d** Input images with excess parts (extra hoops). **e** Segmentation (red) of the image in (a) using (b) as a prior. **f** Segmentation of the image in (b) using (a) as a prior. **g** Segmentation of the image in (c) using (d) as a prior. **h** Segmentation of the image in (d) using (c) as a prior. **i, j** Mutual segmentation results for images (a) and (b) respectively. **k, l** Mutual segmentation results for images (c) and (d) respectively. The images are related by projective transformation. The recovered parameters are shown and compared in Table 1



Now, if excess parts are assumed, the left penalty term should be dominant, thus  $\mu > 1$ . Otherwise, if deficiencies are assumed, the right penalty term should be dominant and  $\mu < 1$ .

### 3.2 Projective Invariance

The mutual shape term  $E_{\text{SHAPE}}$  in (22) depends on the transformation  $T$  between the evolving level set function  $\phi$  and the level-set function associated with the other image  $\tilde{\phi}$ . We assume that  $T$  can be modeled by planar projective transformation. Let  $C(t)$  and  $\tilde{C}(t)$  be the two *planar* active contours associated with the images  $I$  and  $\tilde{I}$ , respectively. Recall that  $C(t)$  and  $\tilde{C}(t)$  are the zero levels of the level set functions  $\phi(t)$  and  $\tilde{\phi}(t)$  (respectively) at time  $t$ . Let  $p \in C$  and  $p' \in \tilde{C}$  denote corresponding points on  $C$  and  $\tilde{C}$ . Their homogeneous coordinates  $\mathbf{x}$  and  $\mathbf{x}'$  are related by *planar projective homography*, i.e.  $\mathbf{x}' = \mathcal{H}\mathbf{x}$  where

$$\mathcal{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \in \mathbb{R}^{3 \times 3} \quad (24)$$

is the homography matrix. Specifically,

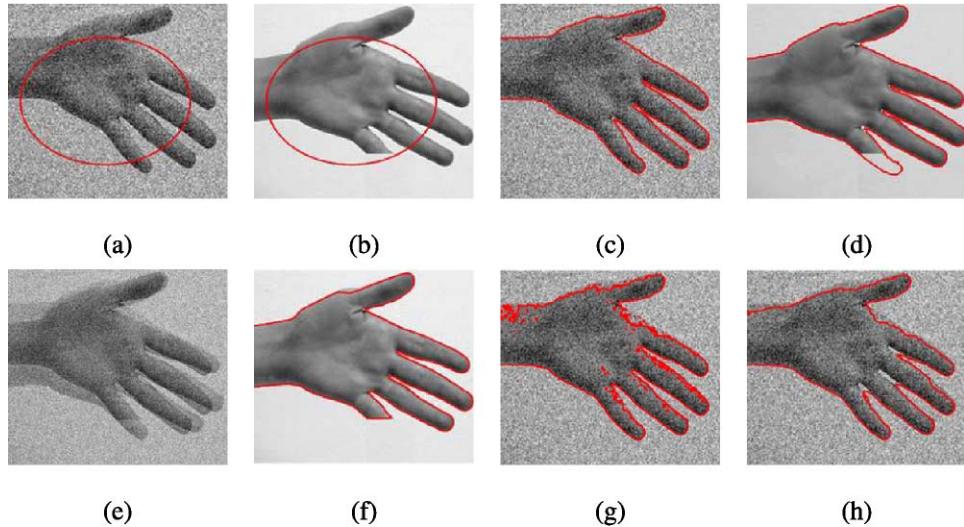
$$\begin{aligned} x' &= \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}, \\ y' &= \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}. \end{aligned} \quad (25)$$

Equivalently we can define  $T(\phi(\mathbf{x})) \equiv \phi(\mathbf{x}') = \phi(\mathcal{H}\mathbf{x})$ , where  $\mathcal{H}$  is a linear transformation applied on the homogeneous coordinate of  $\phi$ . Thus, given  $\phi$  and  $\tilde{\phi}$ , we would like to infer the homography  $\mathcal{H}$  that minimizes the “distance” (22) between  $\phi$  and  $\tilde{\phi}_T$ . The eight unknown ratios of the homography matrix entries,  $\hat{h}_k = h_{ij}/h_{33}$  where  $k = 1, \dots, 8$ , are re-calculated at each iteration for the currently updated  $\phi$  and  $\tilde{\phi}$ . The PDEs for  $\hat{h}_k$  are obtained by differentiating (22) with respect to each

$$\frac{\partial \hat{h}_k}{\partial t} = \int_{\Omega} \delta(T(\tilde{\phi}))[(1 - H(\phi)) - \mu H(\phi)] \frac{\partial T(\tilde{\phi})}{\partial \hat{h}_k} d\mathbf{x}. \quad (26)$$

The derivation of  $\partial T(\tilde{\phi})/\partial \hat{h}_k$  can be done as in (Riklin-Raviv et al. 2007).

At each iteration, the level set functions of the image pair are evolved alternately with the recovery of the transforma-



**Fig. 3** (Color online) **a, b** Noisy (**a**) and corrupted (**b**) images of the same object taken from different view points. The initial contours are drawn in red. **c, d** Successful mutual segmentation results (red). **e** Superposition of the two images to demonstrate the misalignment. **f–h** Segmentation of each image by itself. The noisy image (**g**), (**h**) was segmented twice with different weights of smoothness term: **g** The

contribution of the smoothness term  $W_{LEN}(t)(\phi_t^{LEN})$  was restricted to  $[-1, 1]$  (refer to Sect. 4.1 for details). The contour “mistakenly” follows image gradients that are due to noise. **h** The smoothness term  $W_{LEN}(t)(\phi_t^{LEN})$  was further stressed, i.e. its contributions were multiplied by two. The segmenting contour (red) is smoother but the gaps between the fingers are not well extracted

tion that aligns them. In Sect. 4.2 we discuss further aspects of this process.

### 3.3 Unified Cost Functional

A unified cost functional that is based on the data of the images to segment, on general assumptions with respect to properties of natural images and on the segmentation curves both images, takes the form

$$\begin{aligned} E(\phi) = & W^{RB} E_{RB}(\phi) + W^{LEN} E_{LEN}(\phi) \\ & + W^{DGAC} E_{DGAC}(\phi) + W^{EA} E_{EA}(\phi) \\ & + W^{SHAPE} E_{SHAPE}(\phi), \end{aligned} \quad (27)$$

with (7), (16), (12), (14), (22).

Note that the GAC term (see (18)) is split into the smoothness term  $E_{LEN}$  (see (16)) and the DGAC term  $E_{DGAC}$  (see (12)) where each has its own weight. This allows more flexibility in the functional construction, in particular when the edge based term (DGAC) should be ignored while the smoothness term is most desirable. Refer for example to Fig. 3.

The evolution of the level-set functions  $\phi$  in either of the images is determined at each iteration by  $\phi(t + \Delta t) = \phi(t) + \phi_t \Delta t$ . The associated gradient descent equations  $\phi_t$  are derived using the first variation of the functional (27)

$$\begin{aligned} \phi_t = & W^{RB} \dot{\phi}_t^{RB} + W^{LEN} \dot{\phi}_t^{LEN} + W^{DGAC} \dot{\phi}_t^{DGAC} \\ & + W^{EA} \dot{\phi}_t^{EA} + W^{SHAPE} \dot{\phi}_t^{SHAPE}. \end{aligned} \quad (28)$$

The terms  $\dot{\phi}_t^{TERM}$  are obtained by slight modification of the gradient descent terms  $\phi_t^{TERM}$  determined by (8), (17), (13), (15), (23). This issue and the determination of the weights  $W^{TERM}$  for the different terms in (28) are discussed in Sect. 4.1.

## 4 Implementation

The essence of the proposed method is the simultaneous evolution of two level-set functions. Each evolves on its corresponding image and is controlled by the data of the associated image and by the level-set function associated with the other image. The planar projective transformation between these two level-set functions is updated at each iteration. The algorithm is quite robust to the selection of the initial level-set function  $\phi_0(\mathbf{x})$ . The only limitation is that image regions labeled as foreground in the first iteration, i.e.  $\omega_0 = \{\mathbf{x} \mid \phi_0(\mathbf{x}) \geq 0\}$ , will contain a significant portion of the object to be segmented, such that the calculated image features will approximately characterize the object region. Formally, we assume that  $G^+(I(\omega_0)) \approx G^+(I(\hat{\omega}))$ , where  $\hat{\omega}$  is the actual object region in the image. When there exists an estimate of the average gray levels of either the foreground or the background image regions, this restriction can be eliminated.

We run the algorithm until the following stopping condition is met:

$$\max(d(\phi_1), d(\phi_2)) < s$$

where  $s$  is a predefined threshold and

$$d(\phi_i) = \sum_{\mathbf{x} \in \Omega} |H(\phi_i^{t+\Delta t}(\mathbf{x})) - H(\phi_i^t(\mathbf{x}))|, \quad i = 1, 2.$$

Here,  $\phi_i^{t+\Delta t}(\mathbf{x})$  is the  $i$  level set function corresponding to image  $i$  at time  $t + \Delta t$ .

#### 4.1 Setting the Weights of the Energy Terms

When the solution to an image analysis problem is obtained by minimizing a cost functional, the issue of setting the relative weights of the energy terms is unavoidable. However, in the absence of a satisfying method for the determination of these parameters, this subject is usually marginalized. A guaranteed but time consuming approach is to set the weight parameters by an exhaustive search on the parameters grid, followed by qualitative examination of the solutions obtained. In the context of image processing, algorithms for setting parameters are considered robust if the search space can be significantly reduced or if the parameters should be only slightly tuned for a class of similar images. Being tedious and subjective this heuristic is not suitable for most real world applications.

In contrast, we propose a heuristic that adaptively determines the relative weight of the contributions of the terms of the functional to the evolution of the level set function, as expressed in (28). The proposed heuristic is based on the assumption that the contributions of the energy terms in each iteration should be balanced. The following observations are considered:

1. The relative weight between the length term and the area terms (region based terms) is squared as the image size is increased.
2. The weight of region based term is affected by the units of the feature chosen to characterize the regions. The simplest example is the gray level range which is usually chosen to be either  $[0, 1]$  or  $[0, 255]$ .
3. The terms that are based on the image gradients may have high dynamic range due to noise or sharp discontinuities.
4. All terms are affected by the instantaneous form of the level set function  $\phi$  which evolves in time. The dynamics induced by  $\phi$  should be taken into consideration.

The suggested scheme for automatic and dynamic weight setting is as follows. Let

$$\check{\phi}_t^{\text{TERM}}(\mathbf{x}) = B(\phi_t^{\text{TERM}}(\mathbf{x}))$$

$$= \begin{cases} U_B & \text{if } \phi_t^{\text{TERM}}(\mathbf{x}) > U_B, \\ L_B & \text{if } \phi_t^{\text{TERM}}(\mathbf{x}) < L_B, \\ \phi_t^{\text{TERM}}(\mathbf{x}) & \text{otherwise} \end{cases} \quad (29)$$

where

$$U_B = \text{std}(\phi_t^{\text{TERM}}(\mathbf{x})), \quad L_B = -U_B.$$

Here,  $\text{std}(\phi_t(\mathbf{x}))$  stands for the standard deviation of  $\phi_t(\mathbf{x})$  over  $\Omega$ . The functional  $B(\cdot)$  operates on  $\check{\phi}_t^{\text{TERM}}$  to bound its dynamic range. Next, the range of  $|\check{\phi}_t^{\text{TERM}}|$  is normalized

$$W^{\text{TERM}} = 1 / \max_{\mathbf{x}} |\check{\phi}_t^{\text{TERM}}(\mathbf{x})|. \quad (30)$$

Note that the clipping (see (29)) affects only extreme values of  $\phi_t^{\text{TERM}}$ , that is  $\check{\phi}_t^{\text{TERM}}(\mathbf{x}) = \phi_t^{\text{TERM}}(\mathbf{x})$  for most  $\mathbf{x} \in \Omega$ . Since  $W$  is recalculated at each iteration it is time dependent. This formulation enables an automatic and adaptive determination of the weights of the energy terms.

#### 4.2 Recovery of the Transformation Parameters

Minimizing the cost functional (22) with respect to the eight unknown ratios of the homography matrix entries is a complicated computational task. Direct update of the parameters via their derivatives (26) may lead to an undesirable local minimum as discussed in (Riklin-Raviv et al. 2007). We thus suggest to perform a rough search in the 8 dimensional parameter space, working on a coarse to fine set of grids, before applying the gradient based Quasi-Newton method (Coleman 1994–2005). The former search, done only *once*, significantly reduces the search space and is intended to lead to the region of attraction of the global minimum. The gradient based algorithm, applied in every iteration, tunes the search result based on the updated level-set functions. Further discussion on the application of the Quasi-Newton optimization and its significance for the recovery of the transformation parameters can be found in (Riklin-Raviv et al. 2007).

Since the registration of each evolving level-set function toward the other is done independently, the mean error (or max error) between the entries of the matrices  $\mathcal{H}_{1 \rightarrow 2}$  and  $\mathcal{H}_{2 \rightarrow 1}^{-1}$  (or  $\mathcal{H}_{2 \rightarrow 1}$  and  $\mathcal{H}_{1 \rightarrow 2}^{-1}$ ) could be used as a criterion for obtaining optimal recovery of the transformation parameters. Here we assume that the probability of reaching a local minimum that satisfies  $\mathcal{H}_{1 \rightarrow 2} \approx \mathcal{H}_{2 \rightarrow 1}^{-1}$  is small. Table 1 and Table 2 present the estimated errors in the recovered homographies based on this comparison. The exact formalism of the error estimate is discussed in Sect. 5.

#### 4.3 Numerical Considerations

We use the finite difference method, in which a grid of nodes spaced by a parameter  $h$  is set up over the image domain. The differential operator is approximated by finite difference equations operating on neighboring nodes. Following (Chan and Vese 2001), a smooth approximation of the Heaviside function  $H_\epsilon$  is used:

$$H_\epsilon(\phi) = \frac{1}{2} \left( 1 + \frac{2}{\pi} \arctan \left( \frac{\phi}{\epsilon} \right) \right). \quad (31)$$

**Table 1** Comparison of the entries of the matrices  $\mathcal{H}_{1 \rightarrow 2}$  and  $\mathcal{H}_{2 \rightarrow 1}^{-1}$  obtained through the registration phase in the mutual segmentation algorithm applied to image pairs presented in Figs. 2–7. The error esti-

mate is calculated according to  $\text{Err}(\hat{h}_i) = |\hat{h}_i - \hat{p}_i|/|\hat{h}_i + \hat{p}_i|$  where  $\hat{h}_i$  and  $\hat{p}_i$  is the  $i$ th entry of the normalized matrices  $\mathcal{H}_{1 \rightarrow 2}$  and  $\mathcal{H}_{2 \rightarrow 1}^{-1}$  respectively

| Hom. ratios                          | Fig. No. | $\hat{h}_1$ | $\hat{h}_2$ | $\hat{h}_3$ | $\hat{h}_4$ | $\hat{h}_5$ | $\hat{h}_6$ | $\hat{h}_7$ | $\hat{h}_8$ |
|--------------------------------------|----------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| $\mathcal{H}_{1 \rightarrow 2}$      | Fig. 2   | 0.858       | 0.406       | -12.12      | -0.298      | 0.964       | -27.93      | -0.0002     | -0.0009     |
| $\mathcal{H}_{2 \rightarrow 1}^{-1}$ |          | 0.868       | 0.408       | -12.11      | -0.302      | 0.969       | -27.90      | -0.0001     | -0.0009     |
| $\text{Err}(\hat{h}_i)$              |          | 0.0058      | 0.0025      | 0.0004      | 0.0067      | 0.0026      | 0.0005      | 0.3333      | 0           |
| $\mathcal{H}_{1 \rightarrow 2}$      | Fig. 3   | 1.0502      | 0.1496      | 0.67        | -0.1271     | 0.9857      | 11.41       | -0.0001     | 0.0008      |
| $\mathcal{H}_{2 \rightarrow 1}^{-1}$ |          | 1.0573      | 0.1504      | 1.2125      | -0.1283     | 1.0027      | 11.79       | -0.0001     | 0.0007      |
| $\text{Err}(\hat{h}_i)$              |          | 0.0034      | 0.0027      | 0.2882      | 0.0047      | 0.0085      | 0.0164      | 0           | 0.0667      |
| $\mathcal{H}_{1 \rightarrow 2}$      | Fig. 4   | 1.0898      | -0.1315     | 11.4        | 0.1276      | 1.0277      | -11.55      | -0.0018     | 0.0003      |
| $\mathcal{H}_{2 \rightarrow 1}^{-1}$ |          | 0.9921      | -0.1289     | 10.88       | 0.2303      | 1.0272      | -11.25      | -0.0010     | 0.0018      |
| $\text{Err}(\hat{h}_i)$              |          | 0.0469      | 0.0100      | 0.0233      | 0.2870      | 0.0002      | 0.0132      | 0.2857      | 0.7143      |
| $\mathcal{H}_{1 \rightarrow 2}$      | Fig. 5   | 1.003       | -0.1158     | 8.58        | 0.0359      | 0.9188      | -0.16       | 0.0005      | -0.0001     |
| $\mathcal{H}_{2 \rightarrow 1}^{-1}$ |          | 1.002       | -0.1158     | 8.61        | 0.0364      | 0.9189      | -0.12       | 0.0005      | -0.0001     |
| $\text{Err}(\hat{h}_i)$              |          | 0.0005      | 0           | 0.0017      | 0.0069      | 0.0001      | 0.1429      | 0           | 0           |
| $\mathcal{H}_{1 \rightarrow 2}$      | Fig. 6   | 0.9469      | 0.1461      | 17.92       | -0.2128     | 1.031       | -29.9       | -0.0019     | -0.0009     |
| $\mathcal{H}_{2 \rightarrow 1}^{-1}$ |          | 0.9503      | 0.1564      | 17.97       | -0.2240     | 1.025       | -29.6       | -0.0018     | -0.0007     |
| $\text{Err}(\hat{h}_i)$              |          | 0.0018      | 0.0340      | 0.0014      | 0.0256      | 0.0029      | 0.0050      | 0.0270      | 0.1250      |
| $\mathcal{H}_{1 \rightarrow 2}$      | Fig. 7   | 0.9162      | 0.3374      | -12.75      | -0.2865     | 0.9016      | 8.79        | -0.0005     | -0.0003     |
| $\mathcal{H}_{2 \rightarrow 1}^{-1}$ |          | 0.9183      | 0.3396      | -12.53      | -0.2882     | 0.9034      | 8.81        | -0.0005     | -0.0003     |
| $\text{Err}(\hat{h}_i)$              |          | 0.0011      | 0.0032      | 0.0087      | 0.0030      | 0.0010      | 0.0011      | 0           | 0           |

**Table 2** The mean of the errors over the entries of the homography matrices  $\mathcal{H}_{1 \rightarrow 2}$  that were estimated in each of the experiments. The estimated errors for each entry of the respective matrices are presented in Table 1

| Fig. No. | Mean error |
|----------|------------|
| Fig. 2   | 0.0440     |
| Fig. 3   | 0.0488     |
| Fig. 4   | 0.1726     |
| Fig. 5   | 0.019      |
| Fig. 6   | 0.0278     |
| Fig. 7   | 0.0023     |

The evolution of  $\phi$  at each time step is weighted by the derivative of the regularized form of the Heaviside function:

$$\delta_\epsilon(\phi) = \frac{dH_\epsilon(\phi)}{d\phi} = \frac{1}{\pi} \frac{\epsilon}{\epsilon^2 + \phi^2}.$$

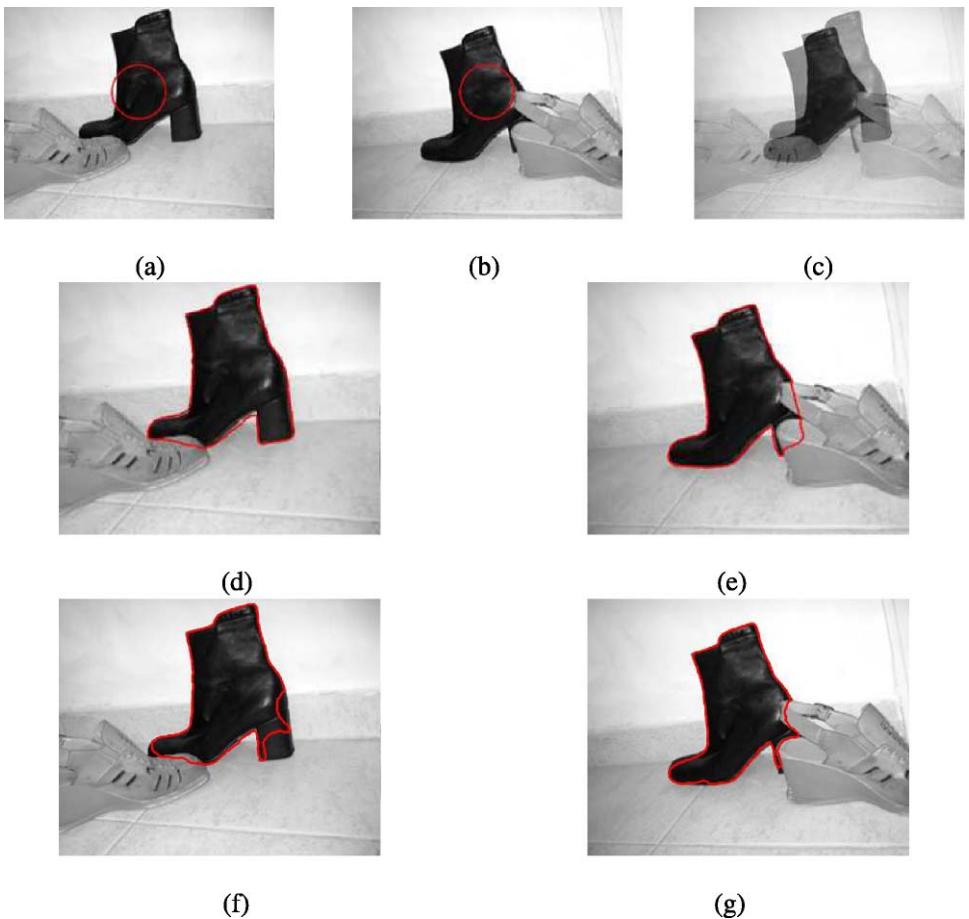
#### 4.4 Algorithm

We summarize the proposed algorithm assuming the following setup. The input is two images  $I_1$  and  $I_2$  of the same object, taken from different viewpoints. The object contours are approximately coplanar. Two level-set functions  $\phi_i$ ,  $i = 1, 2$  that correspond to images  $I_i$  are alternately

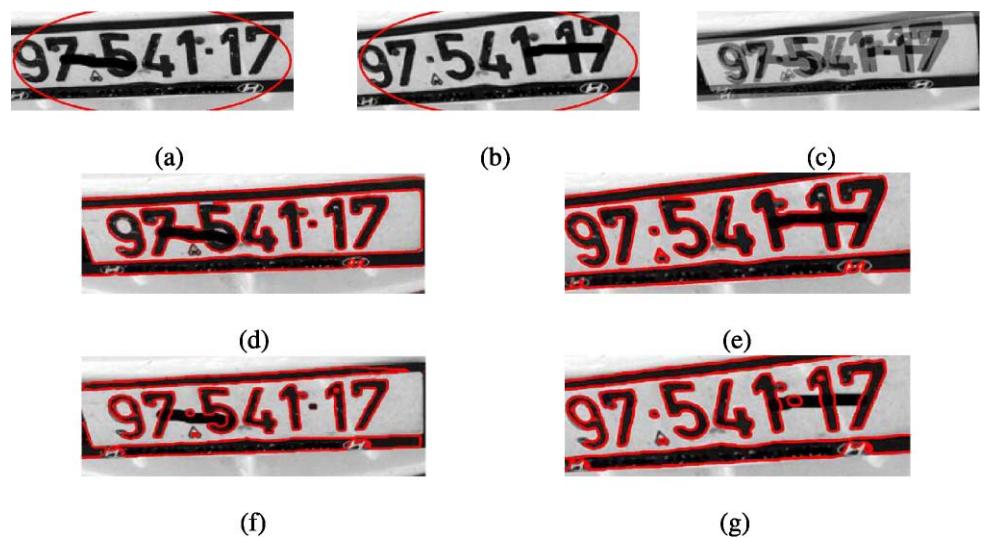
evolved, based on the data of the corresponding image and the other level-set function.

1. Choose initial level-set functions  $\phi_i$ ,  $i = 1, 2$ , for example standard circular (or elliptic) cones. The intersections of the initial level-sets with the corresponding image domains form the initial contours.
2. Initialize the homography matrices  $\mathcal{H}_{1 \rightarrow 2}$  and  $\mathcal{H}_{2 \rightarrow 1}$  to the identity matrix.
3. For each image  $I_i$ , compute the values  $u_+$  and  $u_-$  using (5), based on the current object-background partition defined by the corresponding level-set function.
4. At even iterations evolve the level-set function  $\phi = \phi_1$  using the other level-set function  $\tilde{\phi} = \phi_2$  as a prior. At odd iterations evolve the level-set function  $\phi = \phi_2$  using the other level-set function  $\tilde{\phi} = \phi_1$  as a prior.
5. Apply the corresponding projective transformation on the instantaneous prior  $\tilde{\phi}$  using (25) with the parameters estimated in the preceding time step.
6. Update  $\phi$  using the gradient descent equation (23).
7. Update the transformation parameters  $h_k$  using the derivatives (26). The relation  $\mathcal{H}_{1 \rightarrow 2} = \mathcal{H}_{2 \rightarrow 1}^{-1}$  can be used to speed up (or enhance) the recovery of the transformation parameters.
8. Repeat steps 3–7 until convergence. A convergence criterion is stated in Sect. 4.

**Fig. 4** (Color online) **a, b** Input images with their initial contours (red). The images are of the same object (boot) taken from different viewpoints. **c** Superposition of the two images to demonstrate the misalignment. **f, g** Successful mutual segmentation results (red). **d, e** Segmentation using an unbiased shape dissimilarity measure, (20). The evolving segmentation of each image spoiled the other



**Fig. 5** (Color online) **a, b** Input images with their initial contours (red). **c** Superposition of the two images to demonstrate the misalignment. **d, e** Segmentation of each license plate image by itself using the Chan-Vese level-set method for segmentation. **f, g** Successful mutual segmentation of license plate images with corrupted digits taken from two different viewpoints



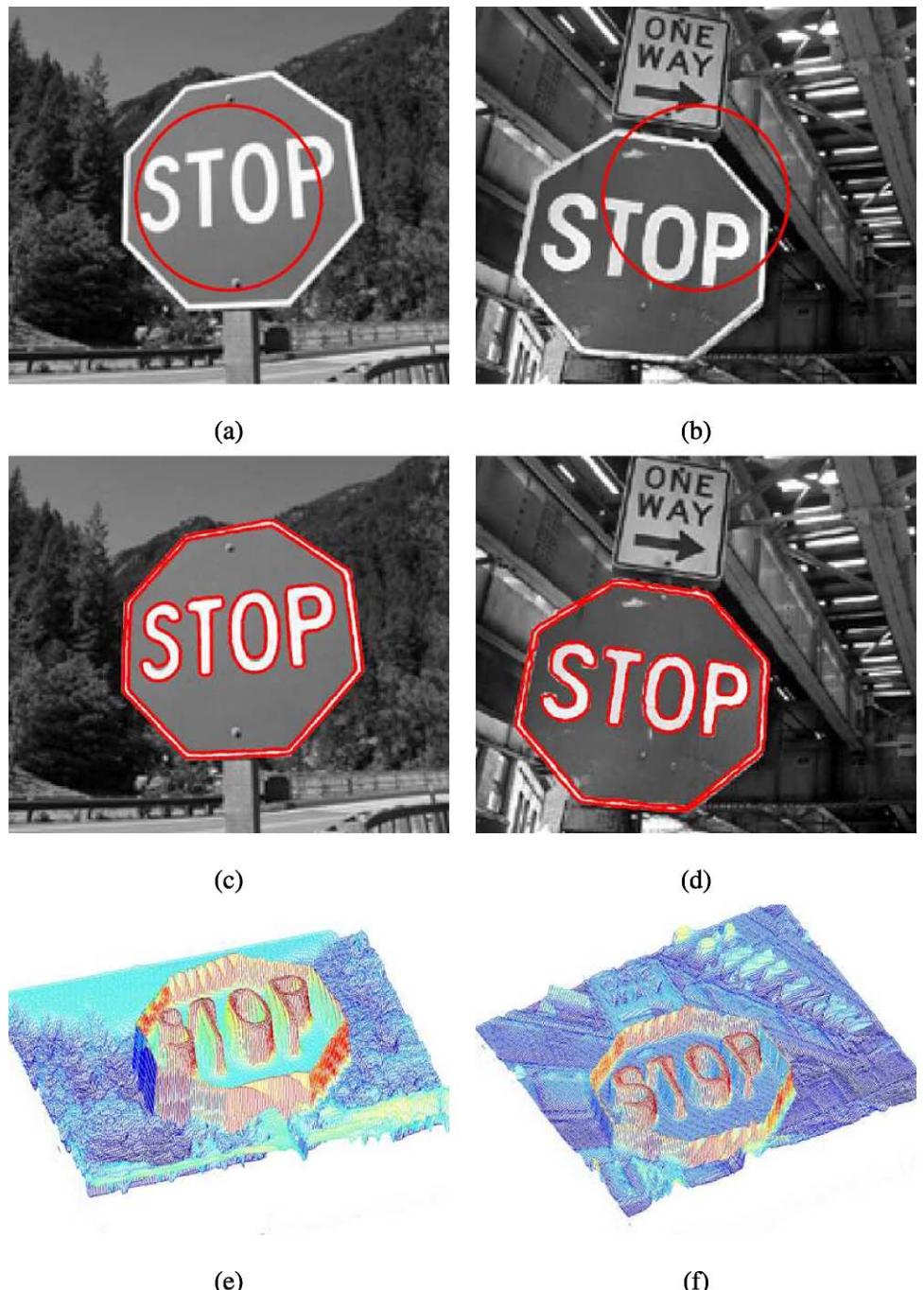
## 5 Experiments

We exemplify the mutual segmentation algorithm on image pairs related by projective transformations. The input images are shown with the initial and final segmenting contours. The mismatch between the respective object views is demonstrated by superposition of the images. The

accuracy of recovered homographies is tested by a comparison between the  $\mathcal{H}_{1 \rightarrow 2}$  and  $\mathcal{H}_{2 \rightarrow 1}^{-1}$ . Table 1 exemplifies such comparisons done on the homographies recovered for the image pairs shown in Figs. 2–7. The term  $\text{Err}(\hat{h}_i) = |\hat{h}_i - \hat{p}_i|/|\hat{h}_i + \hat{p}_i|$  is our suggested error measure for the recovered transformation parameter  $\hat{h}_i$ , where

**Fig. 6** (Color online)

**a, b** Different images of stop signs taken from different viewing positions with their initial contours (red). **a** Original image courtesy of Erik Henne, URL: [www.erikhenne.indexxxx.com](http://www.erikhenne.indexxxx.com). **b** Original image courtesy of the Friedman archives, URL: [www.friedmanarchives.com/Chicago](http://www.friedmanarchives.com/Chicago). **c, d** Successful detection of stop signs using the proposed *mutual* segmentation algorithm. **e, f** Final level set functions of the stop-sign images obtained by using the *mutual* segmentation algorithm

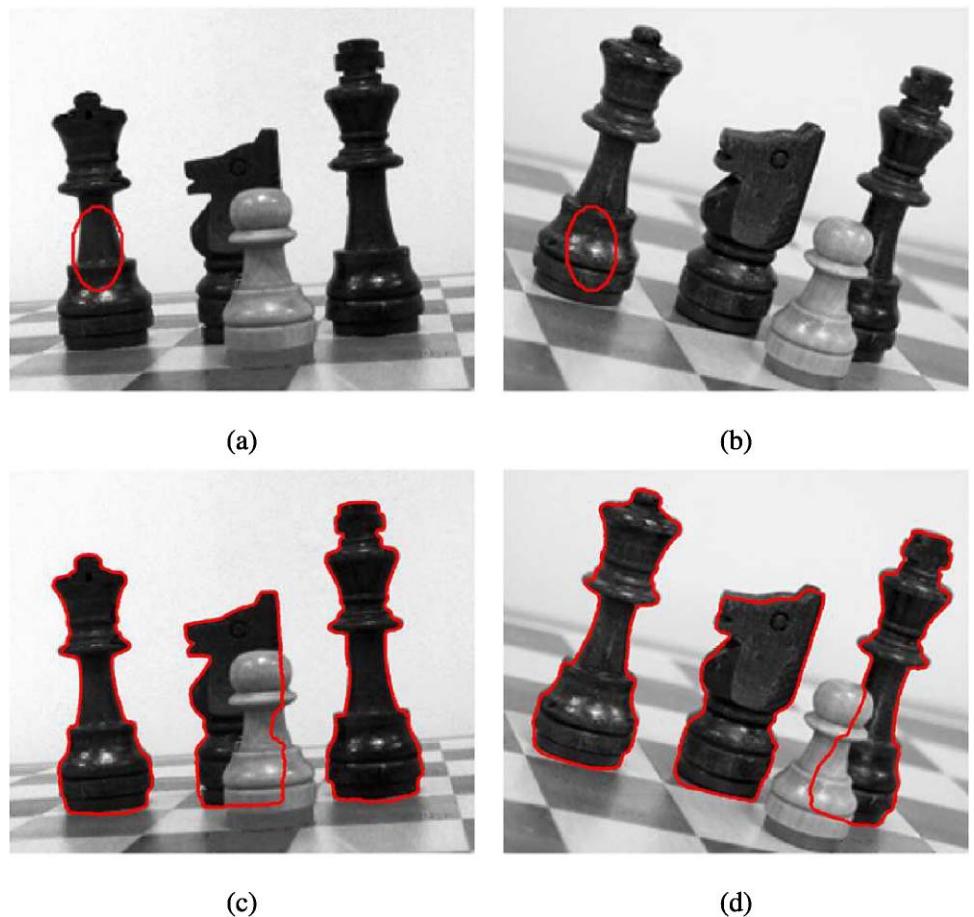


$\hat{h}_i$  and  $\hat{p}_i$  are the  $i$ -est entries of the normalized matrices  $\mathcal{H}_{1 \rightarrow 2}$  and  $\mathcal{H}_{2 \rightarrow 1}^{-1}$  respectively. Table 2 presents the mean error for each of the estimated homography matrices. In all the experiments we set  $dt = 0.1$  and  $\epsilon = 1$ . The weights of the gradient descent terms (8) are adaptively determined as described in Sect. 4.1. Figure 3 shows two images of a hand taken from two different view points. The misalignment between the hand instances is shown in Fig. 3e. Successful segmentation of both images, using the mutual segmentation algorithm with  $\mu < 1$ , is demonstrated in Figs. 3c–d.

Figures 3f–h demonstrate unsuccessful segmentation of each image by itself. Figures 3g–h display two possible segmentations of the noisy instance of the hand. In the segmentation shown in Fig. 3h the smoothness term, (16), has been stressed by multiplying its weight  $W^{\text{LEN}}$  by 2. The segmenting contour is thus smoother but does not extract precisely the narrow regions between the fingers.

The boot images in Figs. 4a–b were mutually segmented using the proposed algorithm with  $\mu < 1$ . The delineating contour (shown in Figs. 4d–e) precisely traces the boot

**Fig. 7** (Color online) **a, b** Input images with their initial contours (red). **c, d** Successful segmentation of chess pieces using the proposed *mutual* segmentation algorithm. The final delineating contour overcomes clutter and occlusions



boundaries while correctly completing the occluded parts. The misalignment between the boot instances is shown in Fig. 4c. The necessity of the biased shape dissimilarity measure is demonstrated in Figs. 4f–g. In these figures we used the unbiased dissimilarity measure (see (20)) and the evolving segmentation of each image spoiled the segmentation of the other. Figure 5 demonstrates mutual segmentation of two images of a license plate with corrupted digits. Figures 5a–b present the license plate images with the initial contour (red). The misalignment between the images is shown in Fig. 5c. Assuming excess parts we set  $\mu > 1$ . Successful mutual segmentation results are shown in Figs. 5d–e. For a comparison, Figs. 5f–g display undesired segmentation results obtained when each image is segmented by itself. Figure 6 shows the ability of the proposed method to detect the common object (stop sign) in an image pair despite the clutter. The images have been downloaded from different web sites (see caption). Note that the poses and surroundings of the stop signs are different. As in all other examples, the algorithm was tested on gray level images, thus the color cue is not used. The contours (red) in Figs. 6c–d precisely extract the common object based on its shape. Figures 6e–f present the final level-set functions. Mutual segmentation of the chess pieces shown in Fig. 7 demonstrates

the ability of the algorithm to deal with both clutter and partial occlusions. Note that though the objects segmented are definitely not planar, the comparisons of the homographies between the object contours in Table 1 show that homography is a reasonable approximation of the transformation.

## 6 Discussion

We presented a method for concurrent, mutually-supporting segmentation of two images of the same object, taken from different view points. Having *two* images instead of one provides redundancy that is employed by using each instance to guide the segmentation of the other. Unlike previous methods, the concept of a perfect shape prior is replaced by information gathered from incomplete instances.

Segmentation is metaphorically similar to cliff climbing. Prior-based segmentation is analogous to situations where someone climbs first and secures a rope to the cliff. If this is not possible, the combined effort of at least a duo is needed. The two climb in turns: at each stage one person holds the cliff and helps the other climb. The main contribution of this paper is the formulation of this duo shape term, that enables solution of the mutual segmentation problem.

Having two object instances is helpful in regions where the aligned images agree, but there is an inherent ambiguity where they don't. In this paper, we address this ambiguity via the biased shape dissimilarity measure. Note that if more than two images are available, the ambiguity can be resolved by majority rule. This is a possible topic for future research.

**Acknowledgements** We thank the anonymous referees for their constructive comments. This research was supported by the A.M.N. Foundation and by MUSCLE: Multimedia Understanding through Semantics, Computation and Learning, a European Network of Excellence funded by the EC 6th Framework IST Programme. T. Riklin-Raviv was also supported by the Yizhak and Chaya Weinstein Research Institute for Signal Processing.

## References

Caselles, V., Kimmel, R., & Sapiro, G. (1997). Geodesic active contours. *International Journal of Computer Vision*, 22(1), 61–79.

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

Chen, Y., Tagare, H. D., Thiruvenkadam, S., Huang, F., Wilson, D., Gopinath, K. S., Briggs, R. W., & Geiser, E. A. (2002). Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision*, 50(3), 315–328.

Coleman, T. F. (1994–2005). Optimization toolbox: quasi-Newton methods. The Mathworks, Inc. <http://www.mathworks.com/access/helpdesk/help/toolbox/optim/ug>.

Cremers, D., & Soatto, S. (2003). A pseudo-distance for shape priors in level set segmentation. In *Workshop on variational, geometric and level set methods in computer vision* (pp. 169–176).

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

Cremers, D., Sochen, N., & Schnörr, C. (2006). Multiphase dynamic labeling for variational recognition-driven image segmentation. *International Journal of Computer Vision*, 66(1), 67–81.

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

Duci, A., Yezzi, A., Mitter, S., & Soatto, S. (2002). Region matching with missing parts. In *Proceedings of the European conference on computer vision* (Vol. 3, pp. 48–64).

Gomes, J., & Faugeras, O. D. (2000). Level sets and distance functions. In *Proceedings of the European conference on computer vision* (pp. 588–602).

Holtzman-Gazit, M., Kimmel, R., Peled, N., & Goldsher, D. (2006). Segmentation of thin structures on volumetric medical images. *IEEE Transactions on Image Processing*, 15(2), 354–363.

Huang, X., Li, Z., & Metaxas, D. (2004). Learning coupled prior-shape and appearance models for segmentation. In *Medical image computing and computer-assisted intervention, MICCAI04* (Vol. I, pp. 60–69).

Kass, M., Witkin, A. P., & Terzopoulos, D. (1988). Snakes: active contour models. *International Journal of Computer Vision*, 1(4), 321–331.

Kichenassamy, S., Kumar, A., Olver, P. J., Tannenbaum, A., & Yezzi, A. (1995). Gradient flows and geometric active contour models. In *Proceedings of the international conference on computer vision* (pp. 810–815).

Kimmel, R. (2003). Fast edge integration. In S. Osher & N. Paragios (Eds.), *Geometric level set methods in imaging vision and graphics*. Berlin: Springer.

Kimmel, R., & Bruckstein, A. M. (2003). Regularized Laplacian zero crossings as optimal edge integrators. *International Journal of Computer Vision*, 53(3), 225–243.

Leventon, M., Faugeras, O., Grimson, W., & Wells III, W. (2000a). Level set based segmentation with intensity and curvature priors. In *Workshop on mathematical methods in biomedical image analysis proceedings* (pp. 4–11), June 2000.

Leventon, M. E., Grimson, W. E. L., & Faugeras, O. (2000b). Statistical shape influence in geodesic active contours. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (Vol. I, pp. 316–323).

Lorigo, L. M., Faugeras, O., Grimson, W. E. L., Keriven, R., Kikinis, R., Nabavi, A., & Westin, C. F. (2000). Codimension two-geodesic active contours for the segmentation of tabular structures. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 444–451).

Mumford, D., & Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42, 577–684.

Osher, S., & Sethian, J. A. (1988). Fronts propagating with curvature-dependent speed: algorithms based on Hamilton–Jacobi formulations. *Journal of Computational Physics*, 79, 12–49.

Paragios, N., & Deriche, R. (2002). Geodesic active regions: a new paradigm to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation*, 13, 249–268.

Riklin-Raviv, T., Kiryati, N., & Sochen, N. (2004). Unlevel-sets: geometry and prior-based segmentation. In *Proceedings of the European conference on computer vision* (Vol. 4, pp. 50–61).

Riklin-Raviv, T., Kiryati, N., & Sochen, N. (2005). Prior-based segmentation by projective registration and level-sets. In *Proceedings of the international conference on computer vision* (pp. 204–211).

Riklin-Raviv, T., Sochen, N., & Kiryati, N. (2006). Mutual segmentation with level-sets. In *IEEE computer society workshop on perceptual organisation in computer vision*.

Riklin-Raviv, T., Kiryati, N., & Sochen, N. (2007). Prior-based segmentation and shape registration in the presence of projective distortion. *International Journal of Computer Vision*, 72(3), 309–328.

Rochery, M., Jermyn, I. H., & Zerubia, J. (2006). Higher-order active contours. *International Journal of Computer Vision*, 69(1), 27–42.

Rother, C., Minka, T., Blake, A., & Kolmogorov, V. (2006). Cosegmentation of image pairs by histogram matching—incorporating a global constraint into mrfs. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 993–1000).

Rousson, M., & Paragios, N. (2002). Shape priors for level set representation. In *Proceedings of the European conference on computer vision* (pp. 78–92).

Sagiv, C., Sochen, N., & Zeevi, Y. Y. (2006). Integrated active contours for texture segmentation. *IEEE Transactions on Image Processing*, 15(6), 1633–1646.

Sandberg, B., Chan, T., & Vese, L. (2002). *A level-set and Gabor based active contour for segmenting textured images* (Technical report). UCLA CAM report, July 2002.

Strain, J. (1999). Fast tree-based redistancing for level set computations. *Journal of Computational Physics*, 152(2), 664–686.

Tsai, A., Yezzi, A. Jr., Wells, W.M. III, Tempany, C., Tucker, D., Fan, A., Grimson, W.E.L., Willsky, A.S., (2003). A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging*, 22(2), 137–154.

Vasilevskiy, A., & Siddiqi, K. (2001). Flux maximizing geometric flows. In *Proceedings of the international conference on computer vision* (pp. 149–154).

Vedaldi, A., & Soatto, S. (2006). Local features, all grown up. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1753–1760).

Vese, L. A., & Chan, T. F. (2002). A multiphase level set framework for image segmentation using Mumford and shah model. *Journal of Computer Vision*, 50(3), 271–293.

Yezzi, A., Zöllei, L., & Kapur, T. (2003). A variational framework to integrate segmentation and registration through active contours. *Medical Image Analysis*, 7, 171–185.

Zhu, S. C., & Yuille, A. L. (1996). Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9), 884–900.