Shape-Based Mutual Segmentation — **Source link** ⬈

Tammy Riklin-Raviv, Nir Sochen, Nahum Kiryati

Institutions: Tel Aviv University

Related papers:

- Active contours without edges

- "GrabCut": interactive foreground extraction using iterated graph cuts

- Snakes : Active Contour Models

- Optimal approximations by piecewise smooth functions and associated variational problems

- Cosegmentation of Image Pairs by Histogram Matching - Incorporating a Global Constraint into MRFs

Shape-based Mutual Segmentation

Tammy Riklin-Raviv   Nir Sochen[†]   Nahum Kiryati

School of Electrical Engineering
[†]Department of Applied Mathematics
Tel Aviv University
Tel Aviv 69978, Israel

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.

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 considered ill-posed. 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

1

and smoothness of edges and homogeneity of semantic regions should thus be complemented with a-priori *semantical* 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. [20]. 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 [11] that extracts thin structures such as blood vessels, or to [27] that 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. [7, 8, 23, 24, 25, 29] assume the existence of a well-defined shape prior and use it to extract obscured object boundaries. The statistical approaches, e.g. [3, 5, 12, 18, 29, 33], 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*

2

segmentation.

Recent works, similar in their spirit include, [28], [35] and [37]. In the *cosegmentation* method of [28], 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 [35], uses the best SIFT matches of Harris-Affine features to extract the common objects in image pairs. Similarly to [37], 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 [37] 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 [25].

We use the level set framework [21] 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. Fig. 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 rule [9]. 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[1]. 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 section 2 we present the foundation of the cost functional used in the proposed mutual segmentation algorithm. We review some of the state-of-the-art level-set segmentation methods. The region based approaches of Chan et al [2] and Paragios et al [22]; The Geodesic Active contour method introduced by Kichenassamy et al [14] and Caselles et al [1] and the edge alignment constraint suggested by Vasilevskiy et al [34] and independently by Kimmel et al [16]. We also outline the shape similarity measure suggested in [25] 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 section 4. Experimental results are provided in section 5. We conclude in section 6.

---

[1]A preliminary version of this work was published in [26].

<center>(a)   (b)   (c)</center>

Figure 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).

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 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) $P(C \mid I)$, using Bayes law:

$$P(C \mid I) \propto P(I \mid C)P(C) \,. \tag{1}$$

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 [13] use parametric representation $C = C(p)$. In this formulation the term $-\log P(I \mid C) = \int g(|\nabla I(C)|)ds$ is the "external force", where ds is the

<center>5</center>

arc-length. The term $-\log P(C) = \int \left(\alpha |C_p|^2 + \beta |C_{pp}|^2 \right) dp$ is the "internal force". The maximization over all possible separating curves is done by minimizing $-\log P(C \mid I)$. 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 [21], an evolving curve $C(t)$ is defined as the zero level of a level set function $\phi \colon \Omega \to \mathbb{R}$ at time t:

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

Following [2], we use the Heaviside function of ϕ

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

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

$$P(\phi \mid I) \propto P(I \mid \phi) P(\phi) . \tag{4}$$

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

2.2 Region-based data term

Let $I \colon \Omega \to \mathbb{R}^+$ 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 Ω. In the spirit of the Gestalt laws of similarity and proximity, we define a boundary $C = \partial \omega$ that delimits homogeneous regions in Ω. 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:

$$u^+ = A^+ \int_\omega G^+(I(\mathbf{x}))d\mathbf{x}, \qquad u^- = A^- \int_{\Omega\backslash\omega} G^-(I(\mathbf{x}))d\mathbf{x}. \tag{5}$$

The superscripts $+$ and $-$ correspond to the feature values in $\omega \subset \Omega$ and in Ω/ω, respectively. Hereafter we denote $\mathbf{x} \equiv (x, y)$, $A^+ = 1/\int_\omega d\mathbf{x}$ and $A^- = 1/\int_{\Omega\backslash\omega} d\mathbf{x}$. 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 [2] 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 [31] and in [30]. The definition of $G(I)$ could be further extended as suggested in [36], where $u_+, u_- \in L_2(\mathbb{R}^2)$ are smooth approximations of the regions in I. In this study we use the average gray levels and the variance [36, 19]:

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

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

$$E_{\mathrm{RB}}(\phi) = \sum_{i=1}^{2} \int_\Omega \left[(G_i^+(I(\mathbf{x})) - u_i^+)^2 H(\phi) + (G_i^-(I(\mathbf{x})) - u_i^-)^2(1 - H(\phi))\right]d\mathbf{x} \tag{7}$$

An elegant statistical formulation representing the region-based term was introduced in [38], followed by [22] and reviewed in [6].

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 ϕ evolves via the gradient descent:

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

7

where δ 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 $C\colon [0,1] \to \mathbb{R}^2$ where s is an arc-length parameter.

Let $\nabla I(x,y) = (I_x, I_y)^T = \left(\frac{\partial I(x,y)}{\partial x}, \frac{\partial I(x,y)}{\partial y} \right)^T$ denote the vector field of the image gradients. The Geodesic Active Contour (GAC) term [1] is given by

$$E_{\text{GAC}}(C) = \int_0^1 g_{\text{GAC}}(|\nabla I(C(s))|) ds. \tag{9}$$

The function g_{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 will unorthodoxly split the GAC term, Eq. (9), into two terms. The data term (DGAC) is given by

$$E_{\text{DGAC}}(C) = \int_0^1 g_{\text{DGAC}}(|\nabla I(C(s))|) ds \tag{10}$$

where

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

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

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

with the associated gradient descent equation:

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

The GAC functional includes another geometrical term that will be described in subsection 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 [16, 34]. The edge alignment term (EA) defined in [15] takes the form: $E_{\text{EA}}(C) = -\int_0^L |\langle \nabla I(\mathbf{x}(s)), \vec{n}(s) \rangle| \, ds$, where $< \cdot, \cdot >$ denotes inner product and $\vec{n}(s) = \{-y_s(s), x_s(s)\}$ is the exterior normal to the curve C. The expression for $E_{\text{EA}}(C)$ is an integration of the projection of ∇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:

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

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

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

where ΔI is the Laplacian. This equation is similar to the one derived in [15].

2.4 Syntactic prior: Geometry

The prior probability $P(C)$ in Eq. (1) is determined by the minimum description length criterion (based on the Gestalt principle of simplicity). That is $P(C) \propto \exp(-\nu|C|), \quad \nu > 0$ [6], where $|C| = \int ds = \int |C_p| dp$ is the curve length. An equivalent representation, using the level set formulation, takes the form:

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

This functional measures the length of the curve and usually serves as an indicator for the curve smoothness [2]. Minimizing (16) with respect to ϕ, we obtain the associated Euler Lagrange equation for ϕ:

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

9

Combining E_{LEN} and E_{DGAC}, (defined in Eq.(12)), we get the usual form of the GAC functional [1, 14],

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

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

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

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 $\tilde{\phi} \colon \Omega \to \mathbb{R}$, the statistical formulation is then

$$P(\phi, T \mid I, \tilde{\phi}) \propto P(I \mid \phi)P(\phi \mid \tilde{\phi}, T)P(T)$$

where T is the transformation that aligns ϕ and $\tilde{\phi}$. The inverse of the logarithm of the conditional probability term $-\log P(\phi \mid \tilde{\phi}, T)$ can be substituted by a dissimilarity measure $D(\phi, T(\tilde{\phi}))$ between shape representations $\tilde{\phi}$ and ϕ. Some previous approaches, e.g. [17, 18], used:

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

where ϕ and $T(\tilde{\phi})$ are signed distance functions. This measure, however, is affected by the size of the background area $\Omega \setminus \omega$, see [7] 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 [8]. To avoid these drawbacks several modifications to control the integration domain have been suggested [7, 8, 29]. 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, [25] suggested to use the square difference between the Heaviside functions of ϕ and $\tilde{\phi}$ as a dissimilarity measure between the shape representations, where ϕ and $\tilde{\phi}$ are **not** necessarily distance functions[2].

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

This formulation enables the introduction of the projective alignment term T between the shapes. Furthermore, D does not depend on the size of the background area $\Omega \setminus \omega$.

Denote $L = H(\phi(\mathbf{x}))$ and $\tilde{L}_T = H(T(\tilde{\phi}))$. The functions $L : \Omega \to \{0, 1\}$ and $\tilde{L}_T : \Omega \to \{0, 1\}$ are binary labeling functions. We assume that the labelings L and $\tilde{L}_T$ are independent and identically distributed. Thus, the conditional probability based on the proposed dissimilarity measure can be formulated as:

$$P(\phi \mid \tilde{\phi}, T) = \frac{1}{\mathcal{N}} \exp(-\sum_{\mathbf{x} \in \Omega} (L(\mathbf{x}) - \tilde{L}_T(\mathbf{x}))^2),$$

where $\mathcal{N} > 0$ is a normalizing scalar.

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 apriori 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 ϕ_1 and ϕ_2 are alternately evolved. At even iterations the segmenting level-set is $\phi = \phi_1$ and the prior is given by

---

[2] A significant gain from not enforcing ϕ to be a distance function is the elimination of the process of re-distancing [10, 32].

$\tilde{\phi} = \phi_2$. At odd iterations $\phi = \phi_2$ and $\tilde{\phi} = \phi_1$. The statistical inference problem takes the form:

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

We assume that $P(T)$ is flat such that no projective transformation is a-priori preferred. The assumption that the two views are not unrelated is incorporated by initializing T by the identity transformation. The crucial observation is that $P(C \mid \tilde{C}, T)$ (or equivalently $P(\phi \mid \tilde{\phi}, T)$) which is defined via a dissimilarity measure between curves is *not* symmetric with respect to the two curves. This is so since $P(\phi_1 \mid \phi_2, T)$ is not necessarily equal to $P(\phi_2 \mid \phi_1, T)$. The symmetric shape dissimilarity measure defined in Eq. (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 Figure 2a-b. Both 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 Fig. 2e-f. Note that the left leg is incorrectly segmented in Fig. 2e, while the hoop in Fig. 2f is not segmented at all.

The images in Fig. 2c-d contain superfluous hoops located in different places. When each segmentation is the prior for the other, using the unbiased dissimilarity measure in Eq. (20), the contours of the superfluous hoops in Fig. 2c-d undesirably appear in the segmentation shown in Fig. 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 Fig. 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 Fig. 2c-d, the superfluous hoops are labeled as background, as shown in Fig. 2k-l.

3. **The prior shape is perfect.** Examples for the application of this rule are shown in Fig. 2e-h.

Refer again to the dissimilarity measure in Eq. (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 point-wise exclusive-or (xor) operation integrated over the image domain. We may thus rewrite the functional as follows:

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

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 (ϕ) 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.

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

The associated gradient descent equation for ϕ is then:

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

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_{SHAPE} in (22) depends on the transformation T between the currently evolved level set function ϕ 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

$$\mathcal{H} = \left[\begin{array}{ccc} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{array} \right] \in \mathbb{R}^{3\times3} \tag{24}$$

is the homography matrix. Specifically,

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}, \quad y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \tag{25}$$

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 ϕ. Thus, given ϕ and $\tilde{\phi}$, we would like to infer the homography $\mathcal{H}$ that minimizes the "distance" (22) between ϕ and $\tilde{\phi}_T$. The eight unknown ratios of the homography matrix entries, $\hat{h}_k = h_{ij}/h_{33}$ where $k = 1 \ldots 8$, are re-calculated at each iteration for the currently updated ϕ and $\tilde{\phi}$. The PDEs for $\hat{h}_k$ are obtained by differentiating (22) with respect to each.

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

Derivation of $\frac{\partial T(\tilde{\phi})}{\partial \hat{h}_k}$ can be done as in [25].

At each iteration, the level set functions of the image pair are evolved alternately with the recovery of the transformation that aligns them. In subsection 4.2 we discuss further aspects of this process.

3.3 Unified cost functional

A unified cost functional 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 following form:

$$E(\phi) = W^{\mathrm{RB}} E_{\mathrm{RB}}(\phi) + W^{\mathrm{LEN}} E_{\mathrm{LEN}}(\phi) + W^{\mathrm{GAC}} E_{\mathrm{GAC}}(\phi) + W^{\mathrm{EA}} E_{\mathrm{EA}}(\phi) + W^{\mathrm{SHAPE}} E_{\mathrm{SHAPE}}(\phi), \qquad (27)$$

with the equations (7, 16, 18, 14, 22).

Note that the smoothness term E_{LEN} (eq. 16) is contained in the GAC term $(eq.18)$ as discussed in subsection 2.4. It however appears in the complete energy functional to enable flexibility in the functional construction, in particular when the GAC term should be ignored while the smoothness term is most desirable. Refer for example for Fig. 3. Practically either of the weights W^{GAC} or W^{LEN} is set to zero. The evolution of the level-set functions ϕ 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 ϕ_t are derived using the first variation of the functional (27)

$$\phi_t = W^{\mathrm{RB}} \phi_t^{\mathrm{RB}} + W^{\mathrm{LEN}} \phi_t^{\mathrm{LEN}} + W^{\mathrm{GAC}} \phi_t^{\mathrm{GAC}} + W^{\mathrm{EA}} \phi_t^{\mathrm{EA}} + W^{\mathrm{SHAPE}} \phi_t^{\mathrm{SHAPE}}, \qquad (28)$$

The determination of W^{TERM} is discussed in subsection 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 function is updated at each iteration. The algorithm is quite robust to the selection of initial level-set function $\phi_0(\mathbf{x})$. The only limitation is that image regions labeled as foreground in the first iteration, i.e $\omega_0 = \{\mathbf{x} \,|\, \phi_0(\mathbf{x}) \geq 0\}$, 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-est level set function, 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 parameters setting 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 applied in Eq. (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 ϕ which evolves in time. The dynamics induced by ϕ should be taken into consideration.

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

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

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 Ω. The functional $B(.)$ operates on ϕ_t^{TERM} to bound its dynamic range. Then, the range of $|B(\phi_t^{\text{TERM}})|$ is normalized to $[0, 1]$:

$$W(\phi_t^{\text{TERM}}) = B(\phi_t^{\text{TERM}}) / \max_{\mathbf{x}} |B(\phi_t^{\text{TERM}}(\mathbf{x}))|, \tag{30}$$

where $|B|$ is the absolute value of B. Note that the thresholding operation affects only few values of ϕ_t^{TERM}. Formally, W can be considered as a parameter rather than a function. Since W is recalculated at each iteration it is also 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 the derivatives (26)

may lead to an undesired local minimum as discussed in [25]. 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 [4]. The former search, done only *once*, significantly reduces the search space, constraining it to the region of attraction of the global minimum. The gradient based algorithm, applied in every iteration, tunes of 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 [25].

Since the registration of the evolved level-set functions toward each 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 globally optimal of the transformation parameters. Here we assume that the probability of achieving a minimum, which is not a global one, 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 Section 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 [2], a smooth approximation of the Heaviside function H_ϵ is used:

$$H_\epsilon(\phi) = \frac{1}{2}(1 + \frac{2}{\pi}\arctan(\frac{\phi}{\epsilon}))\tag{31}$$

The evolution of ϕ 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. The homography matrices $\mathcal{H}_{1 \to 2}$ and $\mathcal{H}_{2 \to 1}$ are initialized to the identity matrices.

3. For each image I_i, compute the values u_+ and u_- using Eq. (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 Eq. (25) with the parameters estimated in the preceding time step.

6. Update ϕ using the gradient descent equation (23).

7. Update the transformation parameters h_k using the derivatives (26). The relation $\mathcal{H}_{1 \to 2} = \mathcal{H}_{2 \to 1}^{-1}$ 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 Section 4.

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

Hom. ratios	Fig. No.	$\hat{h_1}$	$\hat{h_2}$	$\hat{h_3}$	$\hat{h_4}$	$\hat{h_5}$	$\hat{h_6}$	$\hat{h_7}$	$\hat{h_8}$
$\mathcal{H}_{1\to2}$	Fig. 2	0.858	0.406	-12.12	-0.298	0.964	-27.93	-0.0002	-0.0009
$\mathcal{H}_{2\to1}^{-1}$		0.868	0.408	-12.11	-0.302	0.969	-27.90	-0.0001	-0.0009
$\mathrm{Err}(\hat{h_i})$		0.0058	0.0025	0.0004	0.0067	0.0026	0.0005	0.3333	0
$\mathcal{H}_{1\to2}$	Fig. 3	1.0502	0.1496	0.67	-0.1271	0.9857	11.41	-0.0001	0.0008
$\mathcal{H}_{2\to1}^{-1}$		1.0573	0.1504	1.2125	-0.1283	1.0027	11.79	-0.0001	0.0007
$\mathrm{Err}(\hat{h_i})$		0.0034	0.0027	0.2882	0.0047	0.0085	0.0164	0	0.0667
$\mathcal{H}_{1\to2}$	Fig. 4	1.0898	-.1315	11.4	0.1276	1.0277	-11.55	-0.0018	0.0003
$\mathcal{H}_{2\to1}^{-1}$		0.9921	-0.1289	10.88	0.2303	1.0272	-11.25	-0.0010	0.0018
$\mathrm{Err}(\hat{h_i})$		0.0469	0.0100	0.0233	0.2870	0.0002	0.0132	0.2857	0.7143
$\mathcal{H}_{1\to2}$	Fig. 5	1.003	-0.1158	8.58	0.0359	0.9188	-0.16	0.0005	-0.0001
$\mathcal{H}_{2\to1}^{-1}$		1.002	-0.1158	8.61	0.0364	0.9189	-0.12	0.0005	-0.0001
$\mathrm{Err}(\hat{h_i})$		0.0005	0	0.0017	0.0069	0.0001	0.1429	0	0
$\mathcal{H}_{1\to2}$	Fig. 6	0.9469	0.1461	17.92	-0.2128	1.031	-29.9	-0.0019	-0.0009
$\mathcal{H}_{2\to1}^{-1}$		0.9503	0.1564	17.97	-0.2240	1.025	-29.6	-0.0018	-0.0007
$\mathrm{Err}(\hat{h_i})$		0.0018	0.0340	0.0014	0.0256	0.0029	0.0050	0.0270	0.1250
$\mathcal{H}_{1\to2}$	Fig. 7	0.9162	0.3374	-12.75	-0.2865	0.9016	8.79	-0.0005	-0.0003
$\mathcal{H}_{2\to1}^{-1}$		0.9183	0.3396	-12.53	-0.2882	0.9034	8.81	-0.0005	-0.0003
$\mathrm{Err}(\hat{h_i})$		0.0011	0.0032	0.0087	0.0030	0.0010	0.0011	0	0

Table 1: Comparison of the entries of the matrices $\mathcal{H}_{1\to2}$ and $\mathcal{H}_{2\to1}^{-1}$ obtained through the registration phase in the mutual segmentation algorithm applied to image pairs presented in Figs. 2-7. The error estimate is calculated according to $\mathrm{Err}(\hat{h_i}) = |\hat{h_i} - \hat{p_i}|/|\hat{h_i} + \hat{p_i}|$, where $\hat{h_i}$ and $\hat{p_i}$ are the i-est entries of the normalized matrices $\mathcal{H}_{1\to2}$ and $\mathcal{H}_{2\to1}^{-1}$, respectively.

tested by a comparison between the $\mathcal{H}_{1\to2}$ and $\mathcal{H}_{2\to1}^{-1}$. Table 1 exemplifies such comparisons done on the homographies recovered for the image pairs shown in Figs. 2-7. The term $\mathrm{Err}(\hat{h_i}) = |\hat{h_i} - \hat{p_i}|/|\hat{h_i} + \hat{p_i}|$ is our suggested error measure for the recovered transformation parameter $\hat{h_i}$, where $\hat{h_i}$ and $\hat{p_i}$ are the i-est entries of the normalized matrices $\mathcal{H}_{1\to2}$ and $\mathcal{H}_{2\to1}^{-1}$, 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 subsection 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 Fig. 3c-d. Fig. 3f-h demonstrate unsuccessful segmentation of each image by itself. Fig. 3g-h display two possible segmentations of the noisy instance of the hand. In the segmentation shown in Fig. 3h the

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

Table 2: The mean of the errors over the entries of the homography matrices $\mathcal{H}_{1\rightarrow2}$ that were estimated in each of the experiments. The estimated errors for each entry of the respective matrices are presented in Table 1.

smoothness term, Eq. (16), has been stressed by multiplying its weight W^{LEN} by 2. The segmenting contour is thus smoother but does not extract precisely the narrow regions between the fingers.

The boot images in Fig. 4a-b were mutually segmented using the proposed algorithm, with $\mu < 1$. The delineating contour (shown in Figure 4d-e) traces precisely the boot 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 Fig. 4f-g. In these figures we used the unbiased dissimilarity measure (Eq. (20) and the evolving segmentation of each image spoiled the segmentation of the other. Fig. 5 demonstrates mutual segmentation of two images of a license plate with corrupted digits. Fig. 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 Fig. 5d-e. For a comparison, fig. 5f-g display undesired segmentation results obtained when each image is segmented by itself. Fig. 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 Fig. 6c-d precisely extract the common object based on its shape. Fig. 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 for 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 topic is under current research.

References

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

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

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

[4] T.F. Coleman. Optimization toolbox: Quasi newton methods. The Mathworks, Inc., 1994–2005. http://www.mathworks.com/access/helpdesk/help/toolbox/optim/ug.

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

[6] D. Cremers, M. Rousson, and R. Deriche. Review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *International Journal of Computer Vision*, published online Aug. 2006. doi:10.1007/s11263-006-8711-1.

[7] D. Cremers and S. Soatto. A pseudo-distance for shape priors in level set segmentation. In *Workshop on Variational, Geometric and Level Set Methods in Computer Vision*, pages 169–176, 2003.

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

[9] A. Duci, A. Yezzi, S. Mitter, and S. Soatto. Region matching with missing parts. In *Proceedings of the European Conference on Computer Vision*, volume 3, pages 48–64, May 2002.

[10] J. Gomes and O.D. Faugeras. Level sets and distance functions. In *Proceedings of the European Conference on Computer Vision*, pages 588–602, 2000.

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

[12] X. Huang, Z. Li, and D. Metaxas. Learning coupled prior-shape and appearance models for segmentation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI04*, volume I, pages 60–69, 2004.

[13] M. Kass, A.P. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, Jan. 1988.

[14] S. Kichenassamy, A. Kumar, P.J. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *Proceedings of the International Conference on Computer Vision*, pages 810–815, 1995.

[15] R. Kimmel. Fast edge integration. In S. Osher and N. Paragios, editors, *Geometric Level Set Methods in Imaging Vision and Graphics*. Springer-Verlag, 2003.

[16] R. Kimmel and A.M. Bruckstein. Regularized laplacian zero crossings as optimal edge integrators. *International Journal of Computer Vision*, 53(3):225–243, 2003.

[17] M. Leventon, O. Faugeraus, W. Grimson, and W. Wells III. Level set based segmentation with intensity and curvature priors. In *Workshop on Mathematical Methods in Biomedical Image Analysis Proceedings*, pages 4–11, June 2000.

[18] M.E. Leventon, W.E.L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 316–323, 2000.

[19] L.M. Lorigo, O. Faugeras, Grimson W.E.L., R. Keriven, R. Kikinis, A. Nabavi, and C.F. Westin. Codimension two-geodesic active contours for the segmentation of tabular structures. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 444–451, 2000.

[20] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–684, 1989.

[21] S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

[22] N. Paragios and R. Deriche. 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, 2002.

[23] T. Riklin-Raviv, N. Kiryati, and N. Sochen. Unlevel-sets: Geometry and prior-based segmentation. In *Proceedings of the European Conference on Computer Vision*, volume 4, pages 50–61, 2004.

[24] T. Riklin-Raviv, N. Kiryati, and N. Sochen. Prior-based segmentation by projective registration and level-sets. In *Proceedings of the International Conference on Computer Vision*, pages 204–211, 2005.

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

[26] T. Riklin-Raviv, N. Sochen, and N. Kiryati. Mutual segmentation with level-sets. In *IEEE Computer Society Workshop on Perceptual Organisation in Computer Vision*, 2006.

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

[28] C. Rother, T. Minka, A. Blake, and V. Kolmogorov. 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*, pages 993–1000, 2006.

[29] M. Rousson and N. Paragios. Shape priors for level set representation. In *Proceedings of the European Conference on Computer Vision*, pages 78–92, 2002.

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

[31] B. Sandberg, T. Chan, and L. Vese. A level-set and gabor based active contour for segmenting textured images. Technical report, UCLA CAM report, July 2002.

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

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

[34] A. Vasilevskiy and K. Siddiqi. Flux maximizing geometric flows. In *Proceedings of the International Conference on Computer Vision*, pages 149–154, 2001.

[35] A. Vedaldi and S. Soatto. Local features, all grown up. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1753–1760, 2006.

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

[37] A. Yezzi, L. Zöllei, and T. Kapur. A variational framework to integrate segmentation and registration through active contours. *Medical Image Analysis*, 7:171–185, 2003.

[38] S.C. Zhu and A.L. Yuille. 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, 1996.
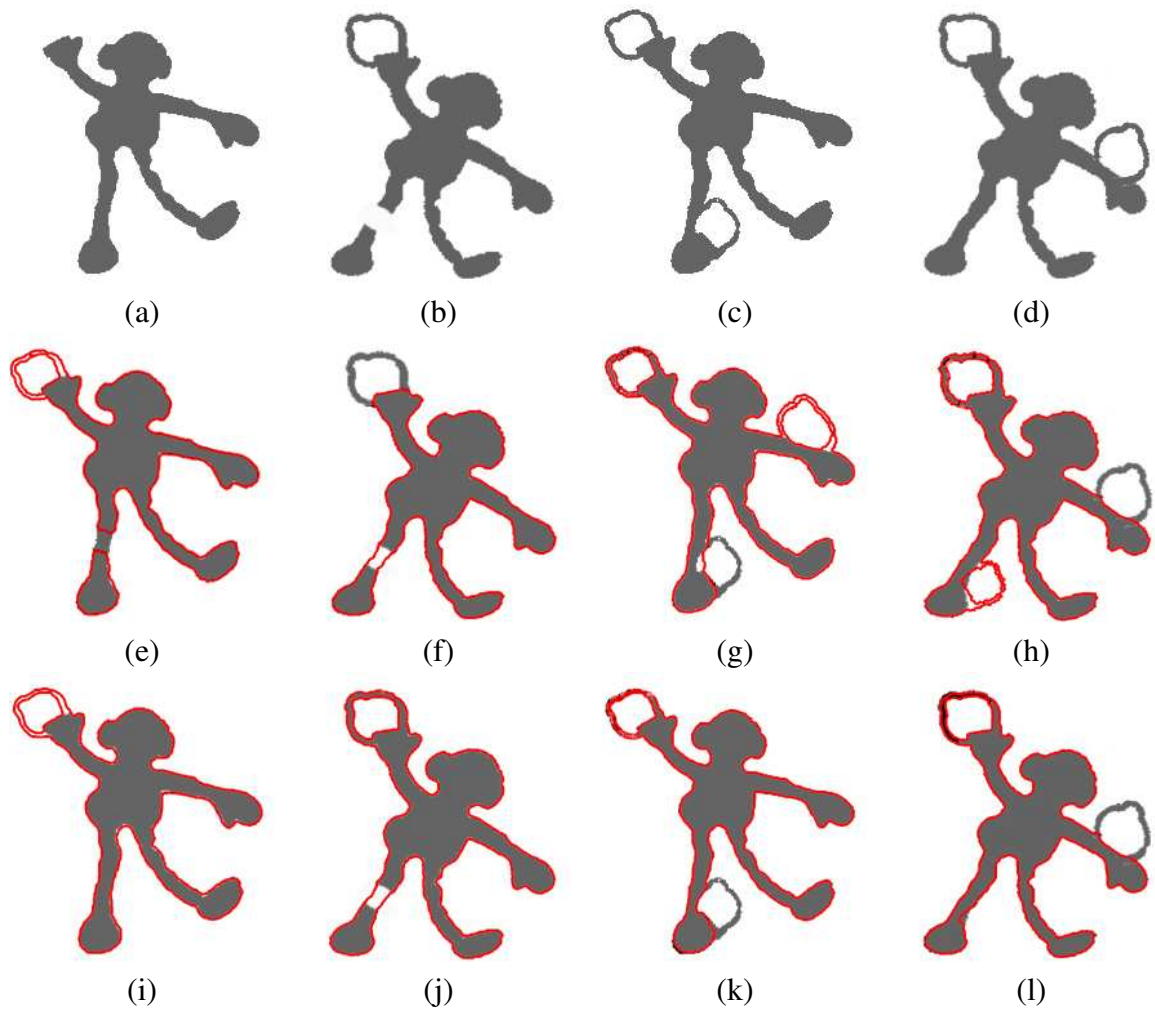
Figure 2: (a,b) Input images with deficiencies (leg, hoop). (c,d) Input images with excess parts (extra hoops). (e) Segmentation (red) of the image in (a) using (b) as a prior. (f) Segmentation of the image in (b) using (a) as a prior. (g) Segmentation of the image in (c) using (d) as a prior. (h) Segmentation 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.
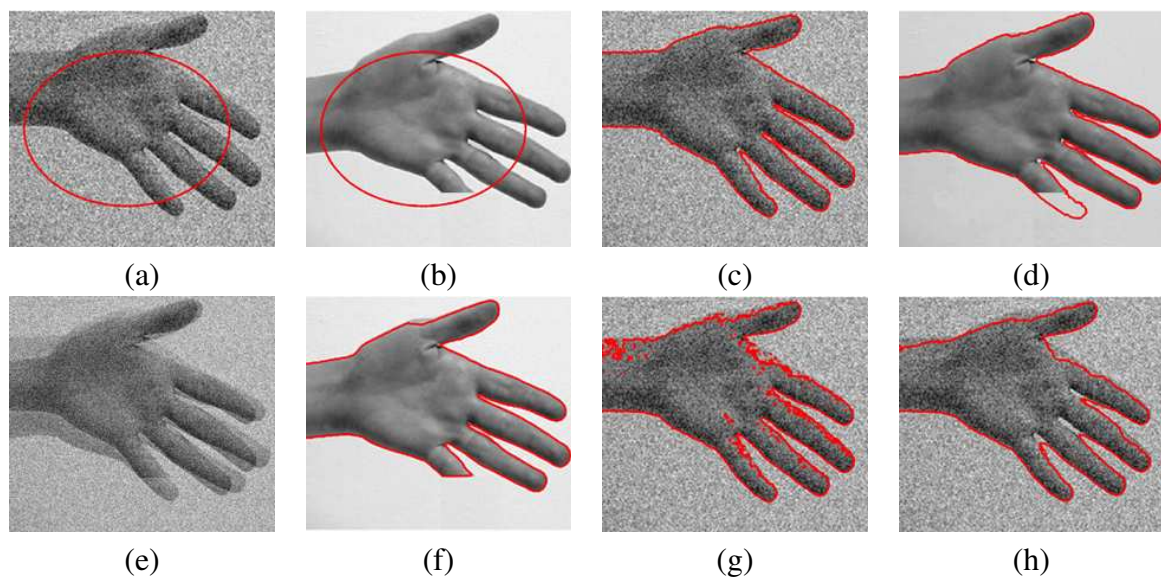
Figure 3: (a,b) Noisy (a) and corrupted (b) images of the same object taken from different view points. The initial contours are drawn in red. (c,d) Successful mutual segmentation results (red). (e) Superposition of the two images to demonstrate the misalignment. (f- h) Segmentation of each image by itself. The noisy image (g,h) was segmented twice with different weights of smoothness term: (g) The contribution of the smoothness term $W_{\text{LEN}}(t)(\phi_t^{\text{LEN}})$ was restricted to $[-1, 1]$ (refer to subsection 4.1 for details). The contour "mistakenly" follows image gradients that are due to noise. (h) The smoothness term $W_{\text{LEN}}(t)(\phi_t^{\text{LEN}})$ 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.
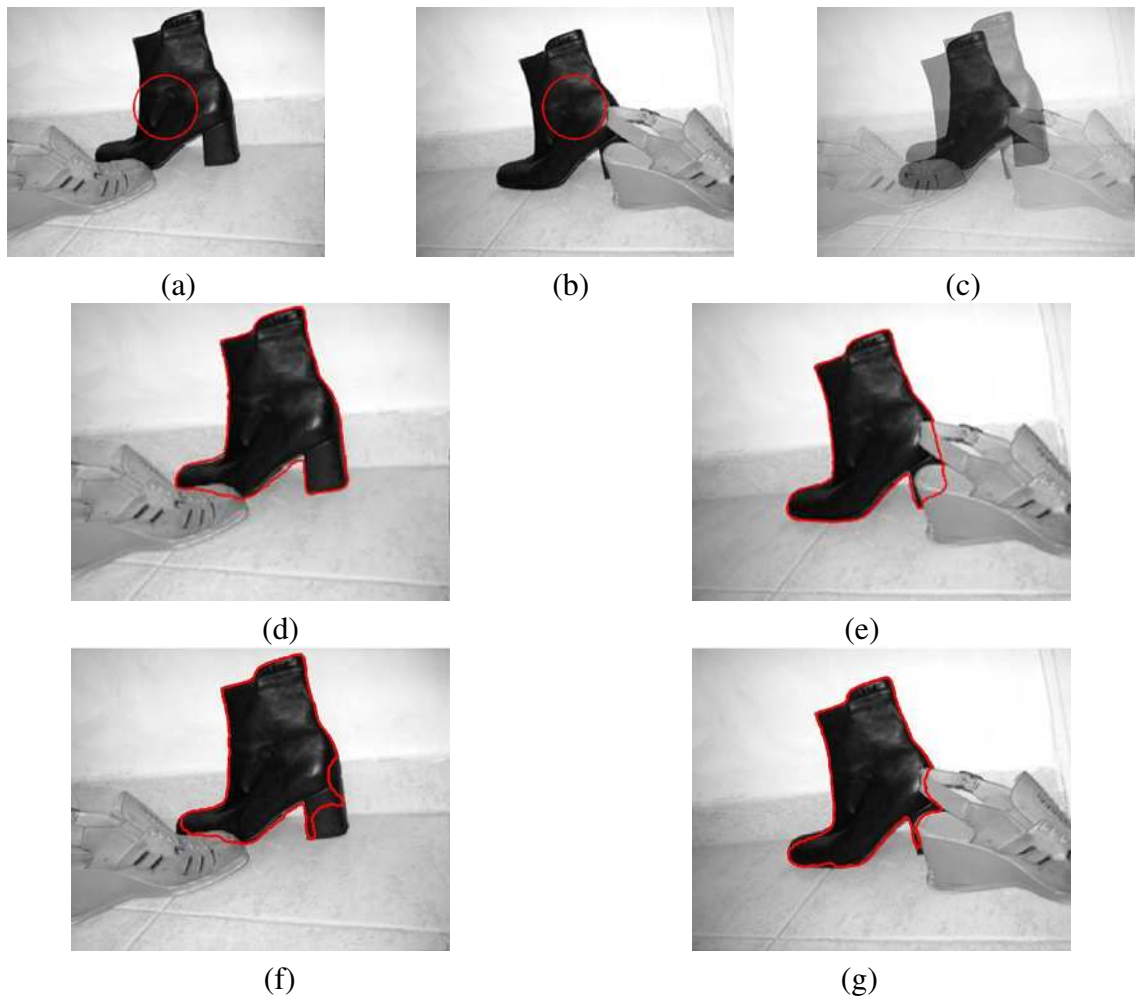
Figure 4: (a,b) Input images with their initial contours (red). The images are of the same object (boot) taken from different viewpoints. (c) Superposition of the two images to demonstrate the misalignment. (f,g) Successful mutual segmentation results (red). (d,e) Segmentation using an unbiased shape dissimilarity measure, Eq.(20). The evolving segmentation of each image spoiled the other.
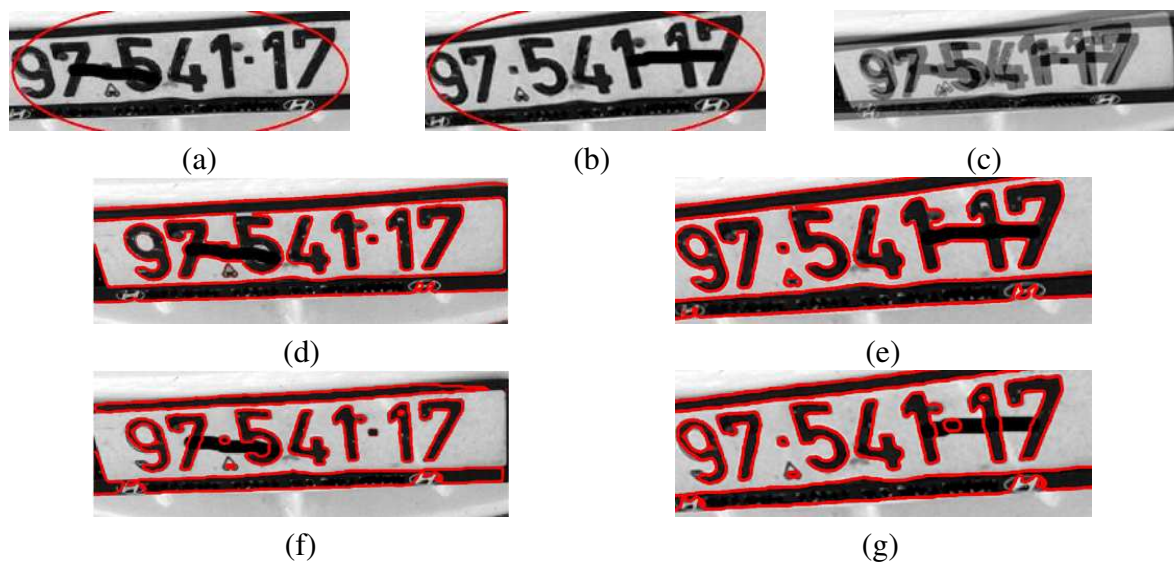
Figure 5: (a)-(b) Input images with their initial contours (red). (c) Superposition of the two images to demonstrate the misalignment. (d)-(e) Segmentation of each license plate image by itself using the Chan-Vese level-set method for segmentation. (f)-(g) Successful *mutual* segmentation of license plate images with corrupted digits taken from two different viewpoints.

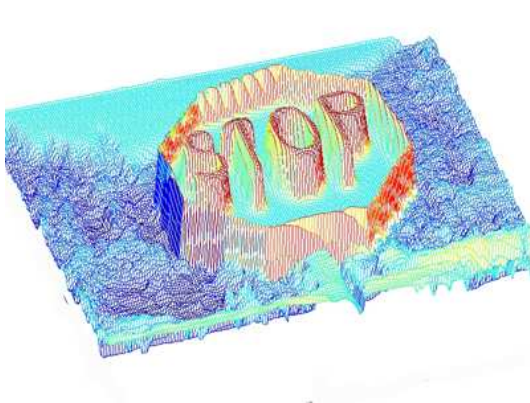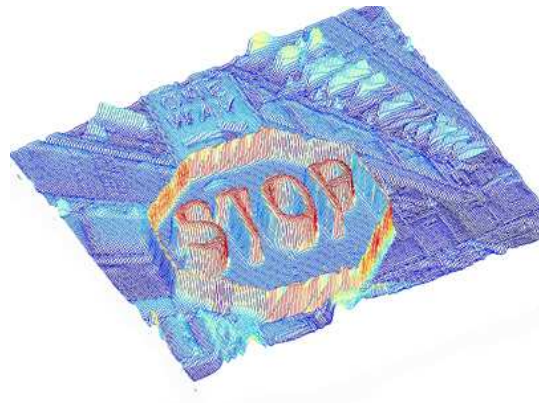Figure 6: (a, b) Different images of stop signs taken from different viewing positions with their initial contours (red).(a) Original image courtesy of Erik Henne, URL: www.erikhenne.indexxxx.com. (b) Original image courtesy of the Friedman archives, URL: www.friedmanarchives.com/Chicago. (c,d) Successful 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.
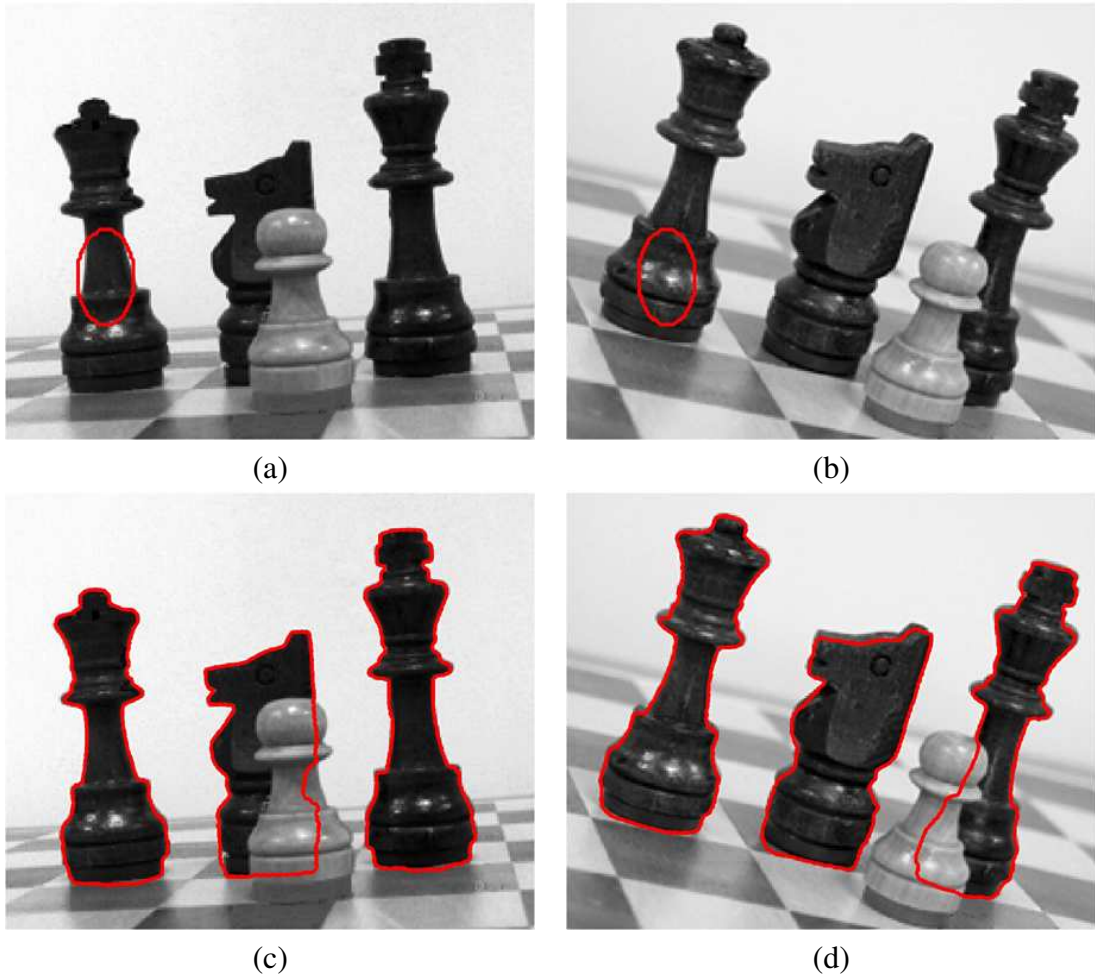
Figure 7: (a,b) Input images with their initial contours (red). (c, d) Successful segmentation of chess pieces using the proposed *mutual* segmentation algorithm. The final delineating contour overcomes clutter and occlusions.