Prior-based Segmentation and Shape Registration in the Presence of Perspective Distortion

Tammy Riklin-Raviv   Nahum Kiryati   Nir Sochen[†]

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

Abstract

Challenging object detection and segmentation tasks can be facilitated by the availability of a reference object. However, accounting for possible transformations between the different object views, as part of the segmentation process, remains difficult. Recent statistical methods address this problem by using comprehensive training data. Other techniques can only accommodate similarity transformations. We suggest a novel variational approach to prior-based segmentation, using a single reference object, that accounts for planar projective transformation. Generalizing the Chan-Vese level set framework, we introduce a novel shape-similarity measure and embed the projective homography between the prior shape and the image to segment within a region-based segmentation functional. The proposed algorithm detects the object of interest, extracts its boundaries, and concurrently carries out the registration to the prior shape. We demonstrate prior-based segmentation on a variety of images and verify the accuracy of the recovered transformation parameters.

1 Introduction

The coupled tasks of segmentation and object detection are essential for the extraction of semantic content from images. Prior knowledge on the shape of interest can significantly facilitate these processes, particularly when the object boundaries are not well defined. However the integration of shape information into the segmentation process is non-trivial. The main difficulty is the need to account for possible transformations between the prior shape and the shape being segmented.

This problem was studied extensively via template matching techniques, see [17, 21, 22, 26, 29, 41, 42] and references therein. When dealing with contours, registration via template matching is equivalent to the Hough Transform [38]. Shape is then commonly represented in a parametric way and the dimension of the representation grows with the complexity of the shape and its degrees of freedom. Template matching and Hough transform are known to yield difficult global optimization problems when the dimension is high, thus incorporation of projective transformations is difficult to accomplish. Moreover, these methods require a substantial set of corresponding points (or features), such that the matched images must have identical texture, see [23] and reference therein.

A different approach focuses on landmark based analysis, in which a training shape set, sampled and aligned, is assumed. In particular, the Active Shape Model [8] and Active Appearance Model [7], that use Principal Component Analysis (PCA) of landmarks to model shape variability, gained popularity for their simplicity and efficiency. However, these methods are semi-automatic, require prior shape analysis and are parameterization dependent.

Variational methods solve segmentation problems by means of energy minimization, integrating top-down and bottom-up information, see [1] and references therein. Specifically, the level-set framework [32] for curve evolution has been successfully applied to extract complex object boundaries, allowing an automatic change in the topology. Prior-based segmentation methods incorporate a representation of a reference shape within the energy functional. Thus, the recovered object boundary should resemble the expected contour, in addition to being constrained by length, smoothness and compatibility with the image gray levels and gradients.

In order to introduce prior shape knowledge and a given group of transformations in the level-set formulation, a shape dissimilarity measure should be provided. Several works, for example [12, 13, 36, 40], use the distance function as the level-set and the square difference between level-sets as the shape dissimilarity measure. A symmetric and unbiased modification of this shape distance (called pseudo distance) has been recently suggested by Cremers and Soatto [10]. However, these similarity measures only account for isometric transformations and scaling, since more general transformations (such as non-isotropic scaling or perspectivity) do not preserve the characteristics of distance functions.

The statistical methodology [4, 9, 19, 24, 25, 36, 40] accounts for transformations beyond similarity and for small non-rigid deformations by using a comprehensive training set. It characterizes the probability distribution of the shapes and then measures the similarity between the evolving object boundary (or level-set function) and representatives of the training data. It is important to note that there is no distinction in this method between transformation-based and deformation-based shape variation. The modes of variation have to account for both. Moreover, the performance depends on the size and coverage of the training set.

None of the existing methods accounts for projective transformations between the prior shape and the shape of interest. The inability to deal with projective transformations is significant. In the presence of projectivity, neither similarity nor (even) the affine model provide reasonable approximation for the transformation between the prior shape and the shape to segment. Figs. 1b-c show the *best* registration of a prior contour of the object shown in Fig. 1a to a given image, assuming similarity and affine transformations respectively. The apparent mismatch inhibits the segmentation process and prohibits accurate reconstruction of the missing parts. In contrast, the registration example in Fig. 1d demonstrates that planar projective transformation is a good approximation even though the elephant shape contour is roughly planar.

We suggest a novel variational approach to prior-based segmentation, that explicitly accounts for planar projective transformation, using a *single* reference object. The segmentation process is carried out concurrently with the registration of the prior shape to the shape of interest. The outcomes of the algorithm include the detection of the object of interest and correct extraction of its boundaries. The planar projective transformation between the two object views is accurately recovered as well. Neither point correspondence nor direct methods [20] are used, thus color or texture compatibility between the prior and the segmented image is needless. This is accomplished by introducing a novel shape-similarity measure, that admits a wide range of transformations, beyond similarity, and using it to generalize the Chan-Vese level-set framework [3]. The proposed region-based segmentation functional includes an explicit expression of the projective homography between the prior shape and the shape to segment.

Employing the parameterization-free shape description, enabled by the level-set formulation, we gain a significant advantage over landmark-based and template matching techniques that represent shapes by collections of points or features. The suggested distance function between the level-set representations of the matched shapes is well defined and is not depend on shapes sampling. Moreover, transformations applied on the domains of the level-set functions, transform the represented shapes correspondingly. This results in an elegant and powerful mathematical formulation to align the prior and the evolving shape, minimizing their dissimilarity measure with respect to the transformation parameters. The graceful merge of the image data with that of the projectively registered prior is the essence of the proposed contribution.

The suggested algorithm is demonstrated on real and synthetic examples, in the presence of perspective distortion. The successful segmentation results and the reliable estimation of the transformation parameters suggest this method as a promising tool for various segmentation and registration applications.

The paper is organized as follows. In section 2 we review the level-set framework of Chan and Vese [3], and outline our prior shape model, in particular the novel shape similarity measure. In section 3 we briefly summarize the prior-based segmentation scheme that was presented in [35]. It has an intuitive geomet-

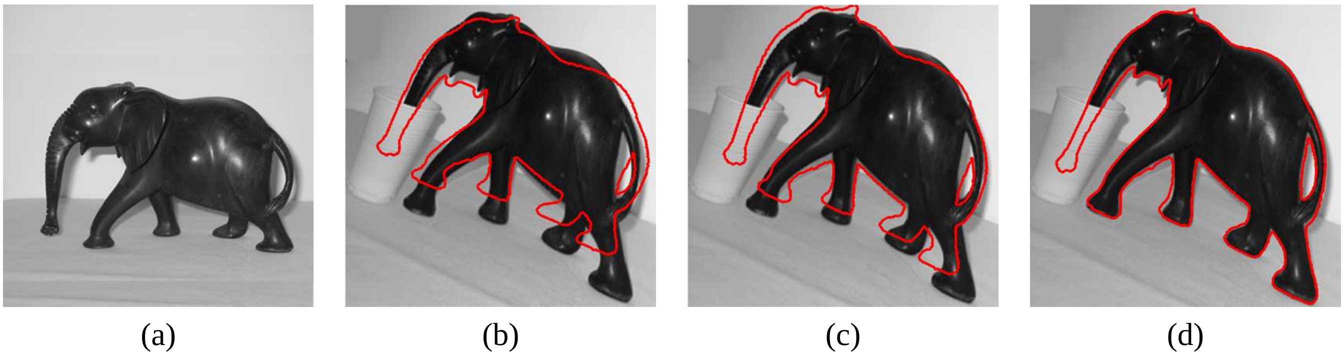


Figure 1: Registration, by the proposed method, of the contour of the object shown in (a) image using : (b) Similarity model (4 transformation parameters). (c) Affine model (6 transformation parameters). (d) Planar projective homography (8 transformation parameters). Note the mismatch in (b) and (c) despite the relatively small change in the camera view. Registration results of (b) and (c) are the best possible with similarity and affine transformations.They have been obtained after an extensive search of the respective parameter spaces. The corresponding transformation parameters are displayed in the last two rows of Table 3.

ric interpretation and a relatively simple mathematical formulation. This method, however, only handles perspectivity (a six-parameter transformation), and is applicable to the limited class of star-shaped objects. Nevertheless, it provides the conceptual basis for the method developed in this paper. Next, we generalize the algorithm to eight-parameter projectivity model and remove the restriction to star-shaped objects. Fundamental concepts from two-view geometry that are necessary to accomplish this generalization are presented in section 4. The embedding of the transformation model within the variational framework, and the minimization of the resulting functional, are considered in section 5. Experimental results are provided in section 6.

2 Variational Framework

2.1 Chan-Vese two-phase model

Mumford and Shah [30] proposed to segment an input image $f\colon \Omega \rightarrow \mathbb{R}$, with respect to a segmenting boundary C and a piecewise smooth approximation of the input image u, by minimizing the functional:

$$E(u,C) = \frac{1}{2}\int_{\Omega}(f-u)^2 d\mathbf{x} + \lambda\frac{1}{2}\int_{\Omega - C}|\nabla u|^2 d\mathbf{x} + \nu|C|\ , \tag{1}$$

where λ and ν are positive scalars. Hereafter $\mathbf{x} = (x,y)$. A reduced form of this functional is simply a restriction of u to a piecewise constant approximation of f, such that each connected component Ω_i, where $\cup_i \Omega_i = \Omega$ and $\Omega_i \cap \Omega_j = \emptyset$, has a constant gray level value u_i. This leads to the minimum partition problem, where the functional takes the following reduced form:

$$E(u,C) = \frac{1}{2}\sum_i \int_{\Omega_i}(f-u_i)^2 d\mathbf{x} + \nu|C| \tag{2}$$

In the level-set formulation for curve evolution [32], a segmenting boundary C in an image plane $\Omega \subset \mathbb{R}^2$, is the zero level-set of a 3D function ϕ, $C = \{\mathbf{x} \in \Omega |\ \phi(\mathbf{x}) = 0\}$. Chan and Vese [3], inspired by the segmentation model of Mumford and Shah (1), suggested to segment an input image f using the Heaviside function of the evolving level-set function $H(\phi)$ as an indicator for object and background regions in the image. Contour length has been also expressed in a region-based manner. For the two-phase case the resultant functional takes the following form:

$$\begin{aligned} E_{CV} &= \int_{\Omega} \left[(f-u_+)^2 H(\phi) + (f-u_-)^2\left(1 - H(\phi)\right)\right. \\ &+ \left.\nu|\nabla H(\phi)|\right] d\mathbf{x}. \end{aligned} \tag{3}$$

The scalars u_+ and u_- are alternately updated with the evolution of the level-set function. They take the average gray level values of the input image in the regions indicated by $\phi \geq 0$ and $\phi < 0$.

$$u_+ = \frac{\int f(\mathbf{x})H(\phi)d\mathbf{x}}{\int H(\phi)d\mathbf{x}} \qquad u_- = \frac{\int f(\mathbf{x})\left(1 - H(\phi)\right)d\mathbf{x}}{\int (1 - H(\phi))d\mathbf{x}} \tag{4}$$

Note that the integration limits of the reformulated functional (3) are now well-defined and known. The gradient descent equation for the evolution of ϕ is derived using the Euler-Lagrange equations for the functional (3):

$$\frac{\partial \phi}{\partial t} = \delta(\phi)\left[\nu \text{ div } \left(\frac{\nabla \phi}{|\nabla \phi|}\right) - (f - u_+)^2 + (f - u_-)^2\right] . \tag{5}$$

As in [3], a smooth approximation of the Heaviside function H_ϵ, rather than a step function, is used. In particular,

$$H_\epsilon(\phi) = \frac{1}{2}(1 + \frac{2}{\pi}\arctan(\frac{\phi}{\epsilon})) \tag{6}$$

and its derivative $\delta_\epsilon(\phi) = dH(\phi)/d\phi$ is

$$\delta_\epsilon(\phi) = \frac{1}{\pi}\frac{\epsilon}{\epsilon^2 + \phi^2}. \tag{7}$$

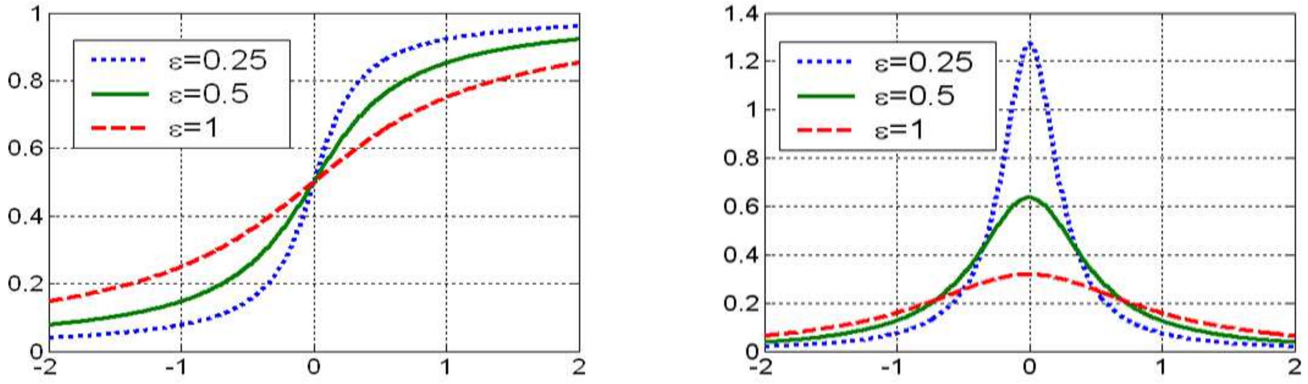
Figs. 2a-b illustrate H_ϵ and δ_ϵ near zero. Thus, the evolutionary change of ϕ, according to (5) is not restricted to its zero level-set.

2.2 Prior shape model

The energy functional (3) can be extended by adding a prior shape term [11]:

$$E(\phi, u_+, u_-) = E_{CV}(\phi, u_+, u_-) + \mu E_{shape}(\phi), \;\; \mu \geq 0. \tag{8}$$

The inclusion of this shape term within the energy functional leads to three related questions:



(a) (b)

Figure 2: Smooth approximation of the Heaviside function H_ϵ (a) and its derivative δ_ϵ (b). Adapted from [3].

1. What should be the representation of the prior shape?
2. What should be the similarity measure between the prior shape and the evolving segmented shape?
3. What is the group of transformations that can be accommodated by the above similarity measure?

The current work provides a comprehensive and innovative solution to these questions. Let $\tilde{\Omega} \subset \mathbb{R}^2$ be the reference image frame. The representation of the prior shape within the energy functional (8) is a $3D$ function $\tilde{\phi}\colon\ \tilde{\Omega} \rightarrow \mathbb{R}$ that embeds the contour $\tilde{C}$ of the known shape:

$$\tilde{C} = \{\mathbf{x} \in \tilde{\Omega} \mid \tilde{\phi}(\mathbf{x}) = 0\}, \tag{9}$$

Positive and the negative values of $\tilde{\phi}$ correspond to object and background regions in $\tilde{\Omega}$ respectively. We suggest two alternative representations of $\tilde{\phi}$, that are discussed and compared in the subsequent sections. The formulation of $\tilde{\phi}$ leads to a definition of a dissimilarity measure with respect to the evolving segmentation. It is a weighted sum of the non-overlapping positive and negative regions of ϕ and $\tilde{\phi}$:

$$E_{shape}(\phi) = \int_\Omega \Big(H_\epsilon(\phi(\mathbf{x})) - H_\epsilon(\tilde{\phi}(\mathbf{x})) \Big)^2 d\mathbf{x} \tag{10}$$

This extension to the functional adds the following term to the evolution equation (5):

$$\delta_\epsilon(\phi)\left(H_\epsilon(\phi) - H_\epsilon(\tilde{\phi})\right) \tag{11}$$

Thus, at each time step, ϕ will be modified in image regions where there is inconsistency between the object-background areas indicated by $H_\epsilon(\phi)$ and $H_\epsilon(\tilde{\phi})$. The change in ϕ is weighted by δ_ϵ, which is illustrated in Fig. 2b for several values of ϵ. The shape-term is further extended to incorporate possible transformations between the prior shape and the shape of interest. This is approached by applying a 3D transformation $\mathcal{T}_p\colon \mathbb{R}^3 \to \mathbb{R}^3$ to $\tilde{\phi}$:

$$E_{shape}(\phi, \mathcal{T}_p) = \int_\Omega \left(H_\epsilon(\phi(\mathbf{x})) - H_\epsilon(\mathcal{T}_p(\tilde{\phi}(\mathbf{x})))\right)^2 d\mathbf{x} \tag{12}$$

The reformulated energy functional becomes:

$$\begin{aligned} E(\phi) &= \int_\Omega \{(f-u_+)^2 H_\epsilon(\phi) + (f-u_-)^2\,(1-H_\epsilon(\phi)) \\ &+ \nu|\nabla H_\epsilon(\phi)| + \mu[H_\epsilon(\phi) - H_\epsilon(\mathcal{T}_p(\tilde{\phi}))]^2\} d\mathbf{x} \end{aligned} \tag{13}$$

The extended gradient descent equation for the evaluation of ϕ is:

$$\begin{aligned} \frac{\partial\phi}{\partial t} &= \delta_\epsilon(\phi)\left[(f-u_-)^2 - (f-u_+)^2 + \nu\,\mathrm{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right)\right. \\ &+ \left.\mu\left(H_\epsilon(\phi) - H_\epsilon(\mathcal{T}_p(\tilde{\phi})\right)\right]. \end{aligned} \tag{14}$$

Note that we do not enforce the evolving level-set function ϕ to resemble $\mathcal{T}_p(\tilde{\phi})$. Instead, we demand similarity of the regions within the respective contours. Thus, ϕ is not necessarily a distance function. Therefore $\mathcal{T}_p$ can be used to accommodate for *planar projective transformations* between the prior contour $\tilde{C}$ and the evolving segmenting contour C.

In section 3 we examine the six-parameter model that was introduced in [35]. It is based on a unique representation of the prior shape $\tilde{\phi}$, which simplifies its perspective mapping to the shape of interest. The eight-parameter model, presented in section 4, generalizes the framework to projectivity.

3 Perspectivity

3.1 Cone of rays

Consider a set of rays, defined by an object in $3D$ space and a camera center. An image is obtained by the intersection of these rays with a plane. This set is usually referred to as a *cone of rays*, although it is not a cone in the classical sense [18]. Now, suppose that this cone of rays is intersected by two planes, as shown in Fig. 3. Then, there exists a perspective transformation H mapping one image onto the other. This means that the images obtained by the same camera center may be mapped to one another by a plane projective transformation [14, 15, 18].

Let f and f' be the image planes of two cameras, having the same camera center, with projection matrices M and M' respectively. For simplicity, the coordinate system of the first camera is chosen to coincide with the world coordinate system, so that $M = K[\ I \mid 0\ ]$, where K is the 3×3 camera calibration matrix and I is the identity matrix. The projection matrix of the second camera is $M' = K'[\ R \mid 0\ ]$, where K' is the calibration matrix and R is the relative rotation between the coordinate systems of the cameras.

Consider two image points $p \in f$ and $p' \in f'$ of a $3D$ world point P. Let $\mathbf{x}$, $\mathbf{x}'$ and $\mathbf{X}$ be their corresponding homogeneous coordinates. Thus, $\mathbf{x} = K[\ I \mid 0\ ]\mathbf{X}$, $\mathbf{x}' = K'[\ R \mid 0\ ]\mathbf{X}$ and the mapping between $\mathbf{x}$ and $\mathbf{x}'$ is $\mathbf{x}' = K'RK^{-1}\mathbf{x} = H\mathbf{x}$, with $H = K'RK^{-1}$. This relation can be simplified by working with a normalized coordinate system[1]. When K and K' only differ by their focal length, $K'K^{-1} =$

---

[1] Internal camera calibration is beyond the scope of this work. Therefore, we will present the internal camera parameters in terms of pixels, assuming square pixels. Moreover, the origin of coordinates in the image plane is set at the camera's principal point. The term *normalized coordinate system* is adopted from [15].

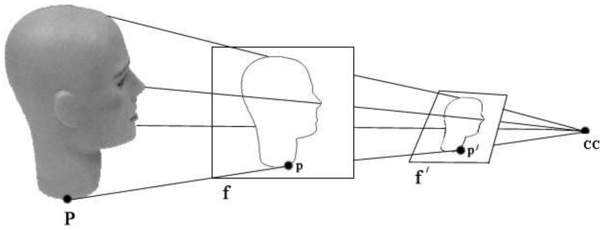


Figure 3: The cone of rays. An image is defined by the intersection of this set of rays with a plane. A ray between the camera center CC and a $3D$ world point P intersects the plane in the image points $p \in f$ and $p' \in f'$. The relation between all such image points is a perspective mapping: $p' = Hp$. Adapted from [18].

$diag(k, k, 1)$ where k is the ratio of the focal lengths.

The transformations of zoom and camera rotation can be expressed as a movement of the image plane while maintaining the camera center fixed. Note that these are particular cases of *perspectivity*. Handling displacements between the centers of the cameras (without parallax) requires extension of the geometric model. These steps are considered in following subsections. We proceed to introduce a prior shape representation derived from the cone of rays.

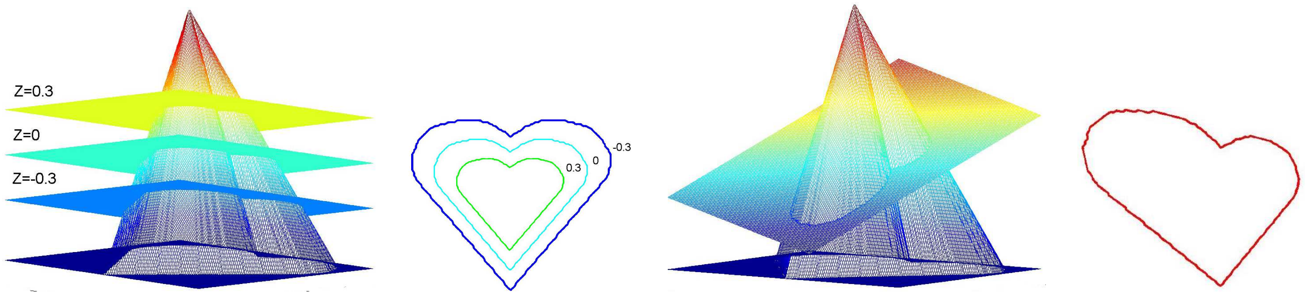
3.2 Generalized cone

A generalized cone[2] or a conical surface, is a ruled surface generated by a moving line (the generator) that passes through a fixed point (the vertex) and continually intersects a fixed planar curve (the directrix) [6]. Let $P_v = (X_v, Y_v, Z_{vertex})$ denote the cone vertex, and let $p_v = (x_v, y_v)$ be the projection of the vertex on the directrix plane. We set, without loss of generality, $X_v = x_v$ and $Y_v = y_v$. Now, consider a directrix, $C = p(s) = (x(s), y(s))$ which is a closed contour, parameterized by arc-length s, of an object shape in the plane $Z = Z_{plane} = 0$. The generalized cone surface is defined by:

$$\tilde{\phi}(r, s) = \tilde{\phi}((1-r)p(s) + rp_v) = (1-r)Z_{plane} + rZ_{vertex} \tag{15}$$

---

[2] The concept of generalized cone (or cylinder) in computer vision has been introduced to model $3D$ objects [2, 28]. Its geometrical properties have been intensively investigated, see [16, 33] and references therein.



(a) (b) (c) (d)

Figure 4: (a) A generalized cone is sliced by three planes, at $Z = 0.3$, $Z = 0$ and $Z = -0.3$. (b) The resulting intersections. (c) A generalized cone is intersected by an inclined plane: $ax + by + cz + d = 0$. (d) The resulting contour.

where r varies smoothly from 1, that corresponds to the vertex, via 0, the directrix, to some convenient negative value.

When the vertex of the generalized cone is located at the camera center, the definition of the generalized cone coincides with that of the cone of rays, presented in subsection 4. It follows that by planar slicing of the generalized cone, one can generate new image views as though they had been taken with a camera under the perspective model[3]. Note, however that the construction of the generalized cone (15), implies that the first image plane is perpendicular to the principal axis. Extension of the allowable transformation between corresponding image points $\mathbf{x}'$ and $\mathbf{x}$ to translation is possible if the object contour is planar and parallel to the first image plane.

3.3 Formulation of the transformation

We will now consider the representation of the transformation $\mathcal{T}_p(\tilde{\phi})$ of the prior shape in the energy functional. The recovery of the transformation parameters, given the prior contour and the curve generated by the zero-crossing of the estimated level-set function, is described subsequently.

Following subsection 2.2, $\tilde{\phi}$ embeds the prior contour according to (15). For simplicity, the vertex height

---

[3] There is, however, one exception to this analogy. The intersection of a cone and a plane is either a closed curve, an open curve or a point. In projective geometry terminology, the latter two correspond to projection of finite points in the first image plane to infinity. We do not consider ideal points and planes at infinity. Phrasing it explicitly, our only concern is the mapping of a given closed curve to another closed curve.

Z_{vertex}, which corresponds to the focal length, will be set to 1. The prior contour $\tilde{C}$ is the cone's directrix.

In order to minimize the energy functional (13), one has to apply a gradient descent process that calls for the evaluation of ϕ simultaneously with the recovery of the transformation $\mathcal{T}_p$ of the function $\tilde{\phi}$. We demonstrate this for transformations $\mathcal{T}_p$ that consist of translation and rotation of the generalized cone, and correspond to scaling, translation, rotation and perspective distortion in the image.

3.3.1 Scaling and general translation

Assume first that the focal length is changed. This approximately corresponds to translation of the image plane along the optical axis. In the image, the effect is scaling. Figs. 4a-b illustrate this notion: as the planar section of the generalized cone is closer to its vertex, the cross-section shape is smaller. Equivalently, the generalized cone can be displaced in the direction of the Z axis, while the intersecting plane remains stationary at $Z = 0$. Formally, in this case, $\mathcal{T}_p(\tilde{\phi}) = \tilde{\phi} + t_z$.

To account also for translation in the image, we displace the generalized cone by $\mathbf{t} = (t_x, t_y, t_z)^T$. The corresponding shape term in the energy functional is then

$$E_{shape}(\phi) = \int_\Omega (H(\phi)(x,y) - H(\tilde{\phi}(x + t_x, y + t_y) + t_z))^2 d\mathbf{x}.$$

3.3.2 Rotation and compound motion

Consider a tilted planar cut of the generalized cone, as shown in Figs. 4c-d. The resulting contour is perspectively deformed, as a function of the inclination of the intersecting plane and its proximity to the vertex of the cone. Equivalently, one may rotate the generalized cone around its vertex, and zero-cross to get the same perspective transformation.

We denote by γ, β and α the Euler rotation angles around the Z,Y and X axes respectively, in that order of rotation. They determine the relative orientation between the first and the second camera coordinate systems. The rotation matrix $R \in \mathbb{R}^{3\times 3}$ operating on a vector $(x, y, z)^T$ takes the form [44]

$$R = \begin{bmatrix} c_\beta c_\gamma & c_\beta s_\gamma & -s_\beta \\ s_\alpha s_\beta c_\gamma - c_\alpha s_\gamma & s_\alpha s_\beta s_\gamma + c_\alpha c_\gamma & s_\alpha c_\beta \\ c_\alpha s_\beta c_\gamma + s_\alpha s_\gamma & c_\alpha s_\beta s_\gamma - s_\alpha s_\gamma & c_\alpha c_\beta \end{bmatrix} \tag{16}$$

where s_α is shorthand for $\sin(\alpha)$ and c_α for $\cos(\alpha)$. General rotation and translation of the generalized cone by R and $\mathbf{t}$ is expressed as $(x', y', \mathcal{T}_p(\tilde{\phi})) = R(x, y, \tilde{\phi}) + \mathbf{t}$. Explicitly,

$$\begin{aligned} x' &= c_\beta c_\gamma x + c_\beta s_\gamma y - s_\beta + t_x \qquad (17) \\ y' &= (s_\alpha s_\beta c_\gamma - c_\alpha s_\gamma)\, x + (s_\alpha s_\beta s_\gamma + c_\alpha c_\gamma)\, y + s_\alpha c_\beta + t_y \\ z' &= (c_\alpha s_\beta c_\gamma + s_\alpha s_\gamma)\, x + (c_\alpha s_\beta s_\gamma - s_\alpha c_\gamma)\, y + c_\alpha c_\beta + t_z \end{aligned}$$

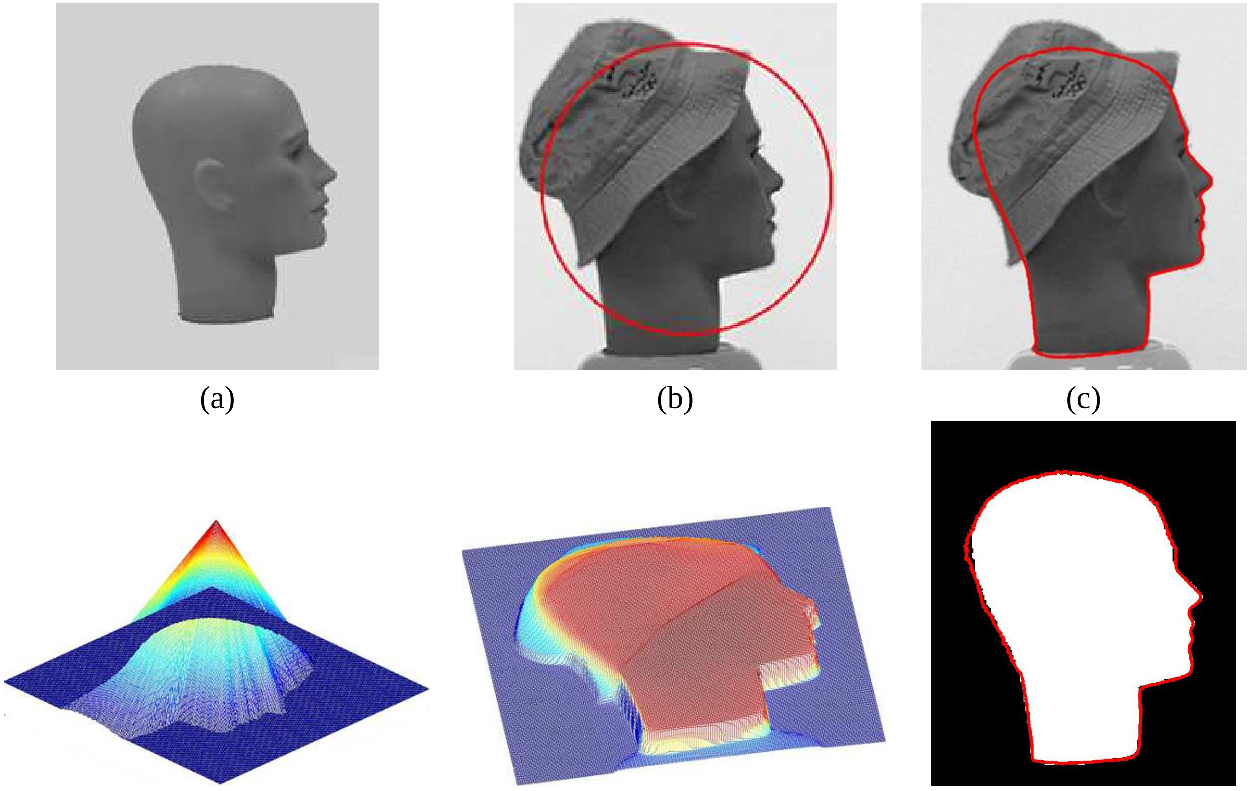
where $z = \tilde{\phi}$ and $z' = \mathcal{T}_p(\tilde{\phi})$. The gradient descent equations of the transformation parameters are presented in [35].

3.4 The six-parameter algorithm

We summarize the six-parameter algorithm for image segmentation and recovery of the transformation between the current and prior object instances, assuming the following setup.

The input consists of two images f and f' of the same object, taken with the same camera, but under different viewing conditions. The boundary $\tilde{C}$ of the object in f is known. The image f' has to be segmented. The world plane is assumed to be parallel to the first image plane f, and f is assumed to be at distance 1 from the camera center. The second image plane f' is tilted and shifted relative to the first one.

1. Given the contour $\tilde{C}$, construct a generalized cone $\tilde{\phi}$, using (15) with $Z_{vertex} = 1$.



(a) (b) (c)

(d) (e) (f)

Figure 5: (a) Prior image (mannequin head). (b) The image to be segmented. The mannequin head is rotated, translated and partly occluded. The initial contour is shown. (c) Successful segmentation using the six-parameter algorithm. (d) The generalized cone based on the prior contour of (a). Only the positive part of the cone is shown. (e) The final state of the level-set function ϕ. (f) Validation: The dark shape is the prior silhouette, transformed according to the recovered transformation parameters. Shown red is the final contour, as in (c).

2. Choose some initial level-set function ϕ, for example a right circular cone. Its zero-level set is the initial segmenting contour.

3. Set initial values (e.g. zero) for α, β , γ , t_x, t_y and t_z.

4. Compute the average gray levels of the current object and background regions, u_+ and u_-, using (4).

5. Translate (by $\mathbf{t}$) and then rotate (by R) the prior shape representation $\tilde{\phi}$, using the coordinate transformation (17) with the current estimate of the transformation parameters.

6. Update ϕ using the gradient descent equation (14).

7. Update the transformation parameters α, β, γ, t_x, t_y and t_z using the derivatives of the cost functional (13) with respect to each parameter.

8. Repeat steps 4-7 until convergence.

The algorithm is exemplified in Fig. 5. The prior image is shown in Fig. 5a; its segmentation is known. Fig. 5b shows a new instance of the mannequin head, rotated and translated with respect to the reference pose. The hat creates significant occlusion. Note that the head is placed on a plastic base of a similar grey-level. The initial segmenting contour, obtained by zero-crossing the initial level-set function (right circular cone) is shown. Successful segmentation using the six-parameter algorithm is shown in Fig. 5c. Note the precise tracing of the profile and the recovery of the occluded crown, despite the perspective distortion. Fig. 5d is the generalized cone that represents the prior shape (a). The final state of the level-set function ϕ is presented in Fig. 5e. Note that ϕ and $\mathcal{T}_p(\tilde{\phi})$ resemble in terms of their Heaviside functions - that is by their zero-crossings (the final contour), but not in their entire shapes. Since the actual transformation was not measured, the recovered transformation parameters are confirmed in Fig. 5f by comparing the final segmenting contour with the prior shape, transformed according to the recovered parameters.

4 From six to eight parameters

4.1 Planar projective homography

To generalize the admissible geometric relation between two corresponding shape contours we review the concept of *planar projective homography*. The equivalence of geometric projectivity and algebraic homography is supported by the theorems of Desargues [37]. Planar projective homography (projectivity) is a mapping $H : \mathbb{P}^2 \to \mathbb{P}^2$ such that the points p_i are collinear if and only if $H(p_i)$ are collinear (projectivity preserves lines) [18, 37].

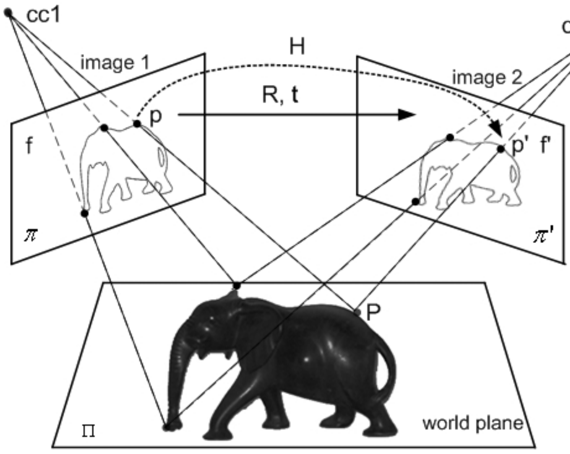


Figure 6: The homography induced by a plane. The ray that intersects the first image plane at a point p meets the world plane Π at a point P. The point P is projected to a point p' in the second image plane. The mapping from p to p' is the homography, denoted by H, induced by the plane Π. Illustrated after [18]. In this research, the corresponding points p and p' are on corresponding planar contours.

The relation between corresponding views of points on a plane Π (world plane) in a 3D space, as is illustrated in Fig. 6, can be modeled by a planar homography induced by the plane. An explicit expression for the induced homography can be derived as follows: Let p and p' be two views of a world point $P \in \Pi$, in two camera frames f and f' respectively. Let $\mathbf{x}$, $\mathbf{x}'$ and $\mathbf{X}$ be their corresponding homogeneous coordinates. As in subsection 3.1, M and M' are the projection matrices of the respective cameras, where $M = K[\ I \mid 0\ ]$. However, since the cameras generally do not have a common center, $M' = K'[\ R \mid \mathbf{t}\ ]$, where $\mathbf{t}$ is the relative translation between the cameras. Thus, $\mathbf{x} = K[\ I \mid 0\ ]\mathbf{X}$ and $\mathbf{x}' = K'[\ R \mid t\ ]\mathbf{X}$.

Let $\mathbf{n}$ be the unit normal vector to the plane Π, and let $d > 0$ denote the distance of Π from the optical center of the first camera. The linear transformation from $\mathbf{x}$ to $\mathbf{x}'$ can be expressed as

$$\mathbf{x}' = K' \left(R + \frac{1}{d}\mathbf{t}\mathbf{n}^T \right) K^{-1}\mathbf{x} = H\mathbf{x}\ . \tag{18}$$

Using the same assumptions as in subsection 3.1, we set K and K' to the identity matrix I. Thus, the planar homography matrix takes the form

$$H = R + \frac{1}{d}\mathbf{t}\mathbf{n}^T. \tag{19}$$

A detailed derivation can be found in [15, 18, 27]. The matrix H is determined by the translation and rotation between the two cameras $\{R, \mathbf{t}\}$, and by the structure parameters $\{\mathbf{n}, d\}$ of the world plane Π. Note that only the ratio $\mathbf{t}/d$ can be recovered from H. We proceed to show how the structure of the homography can be used to recover the relation between the camera frames and thus between the respective shape contours.

4.2 Implicit recovery of the homography

A nonsingular homography between two planes π and π' is determined (up to a scale factor) by four pairs of corresponding points $p \in \pi$ and $p' \in \pi'$, with no triplet of points in either plane being collinear. Consider the homography matrix H,

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \in \mathbb{R}^{3\times 3} \tag{20}$$

and let $\mathbf{x} = (x, y, 1)$ and $\mathbf{x}' = (x', y', 1)$ be the homogeneous representations of the points p and p', such that $\mathbf{x}' = H\mathbf{x}$. The eight unknowns of H (the ratios of its nine entries) can be recovered by solving at least four pairs of equations of the form:

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

Classic approaches recover H by solving an over-determined set of equations like (21). The translation and rotation $(R, \mathbf{t})$ between the image planes, and the scene structure $(\mathbf{n}, d)$, are recovered by decomposition of the known homography matrix (see [15, 18, 27] and references therein). In contrast, our novel approach calculates the homography directly in its *explicit* form (19). Rather than relying on point correspondence, we match two corresponding contours of the shape of interest using calculus of variations. Note that since the recovery of the homography and the segmentation process are simultaneous, only the reference shape is known in advance. The prior shape is registered to the shape being segmented as part of its detection and extraction.

4.3 Explicit recovery of the homography

We now use the explicit formulation of the homography (19) to reformulate equation (21). Consider first the special case in which the world plane and the first image plane coincide, and d is set to 1. In this case, the normal to the world plane, $\mathbf{n}$ is $(0,0,1)$. Substituting d and $\mathbf{n}$ in (19), we obtain the entries of the homography matrix H: $h_{13} = R_{13} + t_x$, $h_{23} = R_{23} + t_y$, $h_{33} = R_{33} + t_z$ and $h_{ij} = R_{ij}$ otherwise.

Generally, the world plane is not perpendicular to the optical axis of the first camera, thus $\mathbf{n} \neq (0,0,1)$. As before, we represent $\mathbf{n}$ using the coordinate system of the first camera. The unit vector $\mathbf{n}$ is obtained by first rotating the vector $(0,0,1)$ by an angle ξ around the y-axis and then by an angle ψ around the x-axis. Hence, $\mathbf{n} = (-\sin\xi, \sin\psi\cos\xi, \cos\psi\cos\xi)$. Substituting R, $\mathbf{t}$ and $\mathbf{n}$ in (19), we obtain the components of the homography matrix H:

$$
\begin{aligned}
h_{11} &= \cos\beta\cos\gamma - \tfrac{t_x}{d}\sin\xi \\
h_{12} &= \cos\beta\sin\gamma + \tfrac{t_x}{d}\sin\psi\cos\xi \\
h_{13} &= -\sin\beta + \tfrac{t_x}{d}\cos\psi\cos\xi \\
h_{21} &= \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma - \tfrac{t_y}{d}\sin\xi \\
h_{22} &= \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma + \tfrac{t_y}{d}\sin\psi\cos\xi \\
h_{23} &= \sin\alpha\cos\beta + \tfrac{t_y}{d}\cos\psi\cos\xi \\
h_{31} &= \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma - \tfrac{t_z}{d}\sin\xi \\
h_{32} &= \cos\alpha\sin\beta\sin\gamma - \sin\alpha\sin\gamma + \tfrac{t_z}{d}\sin\psi\cos\xi \\
h_{33} &= \cos\alpha\cos\beta + \tfrac{t_z}{d}\cos\psi\cos\xi
\end{aligned}
\qquad (22)
$$

Substituting the entries of H in (21), one can explicitly relate shape points between the two views. We will use the calculus of variations to relate the shape contours and recover the translation, rotation and structure parameters.

5 Objective Functional

5.1 Representation of the prior shape

The generalized cone representation of the prior shape, discussed in section 3, is inadequate when the shape contour is seen from two different view points and the world plane is in general position. Furthermore, zero-crossing the transformed generalized cone $H(\mathcal{T}_p(\tilde{\phi}))$, as described in 3.3.2, requires $\mathcal{T}_p(\tilde{\phi})$ to be single-valued. This restricts the group of admissible shape contours and the range of allowable transformations. We therefore present an alternative representation of the prior shape that can be applied to general planar shapes.

Consider a prior image $f : \Omega \rightarrow \mathbb{R}^+$ with labeled object and background regions Ω_+ and Ω_- respectively. The prior shape is represented by a function $\tilde{\phi} \in \mathbb{R}^3$:

$$\tilde{\phi}(x,y) = \begin{cases} 1 & f(x,y) \in \Omega_+ \\ 0 & f(x,y) \in \Omega_- \end{cases} \tag{23}$$

Let $\mathcal{T}_p : (\tilde{\phi}(x,y), x, y) \rightarrow (\tilde{\phi}'(x',y'), x', y')$ denote the projective transformation, that maps a point (x,y) and its $\tilde{\phi}$ value to a (projectively) equivalent point (x',y'), with $\tilde{\phi}'(x',y') = \tilde{\phi}(x,y)$. Thus, the binary function $\tilde{\phi}'(x',y')$ is obtained from $\tilde{\phi}(x,y)$ by coordinate transformation.

Expressions for x' and y' are obtained by substituting the explicit entries for the homography matrix (22) in (21). For example, when $\xi = \psi = 0$ and $d = 1$ the expressions are:

$$x' = \frac{c_\beta c_\gamma x + c_\beta s_\gamma y - s_\beta + t_x}{(c_\alpha s_\beta c_\gamma + s_\alpha s_\gamma)\, x + (c_\alpha s_\beta s_\gamma - s_\alpha c_\gamma)\, y + c_\alpha c_\beta + t_z} \tag{24}$$

$$y' = \frac{(s_\alpha s_\beta c_\gamma - c_\alpha s_\gamma)\, x + (s_\alpha s_\beta s_\gamma + c_\alpha c_\gamma)\, y + s_\alpha c_\beta + t_y}{(c_\alpha s_\beta c_\gamma + s_\alpha s_\gamma)\, x + (c_\alpha s_\beta s_\gamma - s_\alpha c_\gamma)\, y + c_\alpha c_\beta + t_z} \tag{25}$$

The representation (23) of the shape prior is simpler than that using the generalized cone (section 3), but leads to complex expressions for the transformed coordinates x' and y'. For example, the coordinates given in Eqs. (24 - 25) are the quotients of the corresponding expressions in (17). In the proposed algorithm, at each time step one re-evaluates the homography matrix entries (22), based on the estimated transformation parameters. The coordinate transformation $\mathcal{T}_p$ is applied to the representation $\tilde{\phi}$ of the prior shape. The transformed representation $\mathcal{T}_p(\tilde{\phi}(x,y),x,y)$ is substituted in the gradient descent equation for ϕ (14).

5.2 Recovery of the transformation

In order to evolve the level set function ϕ according to (14), one has to simultaneously recover the transformation $\mathcal{T}_p(\tilde{\phi},x,y)$ of the prior level-set function. The transformation parameters (α, β, γ, t_x/d, t_y/d, t_z/d, ψ and ξ) are evaluated via the gradient descent equations obtained by minimizing the energy functional (13) with respect to each of them. The general gradient descent equation for each of the transformation parameters (denoted here by η) is of the form:

$$\frac{\partial \eta}{\partial t} = 2\mu \int_\Omega \delta_\epsilon(\mathcal{T}_p(\tilde{\phi})) \left(H_\epsilon(\phi) - H_\epsilon(\mathcal{T}_P(\tilde{\phi})) \right) \frac{\partial \mathcal{T}_P(\tilde{\phi},\eta)}{\partial \eta} d\mathbf{x} \tag{26}$$

where

$$\begin{aligned} \frac{\partial \mathcal{T}_P(\tilde{\phi},\eta)}{\partial \eta} &= \frac{\partial \mathcal{T}_P(\tilde{\phi})}{\partial x} \left(\frac{\partial x}{\partial x'}\frac{\partial x'}{\partial \eta} + \frac{\partial x}{\partial y'}\frac{\partial y'}{\partial \eta} \right) \\ &+ \frac{\partial \mathcal{T}_P(\tilde{\phi})}{\partial y} \left(\frac{\partial y}{\partial x'}\frac{\partial x'}{\partial \eta} + \frac{\partial y}{\partial y'}\frac{\partial y'}{\partial \eta} \right) \end{aligned} \tag{27}$$

The partial derivatives $\partial \mathcal{T}_p(\tilde{\phi})/\partial x$ and $\partial \mathcal{T}_p(\tilde{\phi})/\partial y$ are computed numerically using the finite difference method. The derivatives $\partial x'/\partial \eta$ and $\partial y'/\partial \eta$ with respect to each transformation parameter η have been derived analytically. From the implicit function theorem we obtain

$$\begin{aligned}
\frac{\partial x}{\partial x'} &= \frac{\frac{\partial y'}{\partial y}}{\frac{\partial y'}{\partial y}\frac{\partial x'}{\partial x} - \frac{\partial y'}{\partial x}\frac{\partial x'}{\partial y}} \\
\frac{\partial x}{\partial y'} &= \frac{\frac{\partial x'}{\partial y}}{\frac{\partial y'}{\partial x}\frac{\partial x'}{\partial y} - \frac{\partial y'}{\partial y}\frac{\partial x'}{\partial x}} \\
\frac{\partial y}{\partial x'} &= \frac{\frac{\partial y'}{\partial x}}{\frac{\partial y'}{\partial x}\frac{\partial x'}{\partial y} - \frac{\partial x'}{\partial x}\frac{\partial y'}{\partial y}} \\
\frac{\partial y}{\partial y'} &= \frac{\frac{\partial x'}{\partial x}}{\frac{\partial y'}{\partial y}\frac{\partial x'}{\partial x} - \frac{\partial y'}{\partial x}\frac{\partial x'}{\partial y}}
\end{aligned} \tag{28}$$

Finally, the equations for $\frac{\partial x'}{\partial x}$, $\frac{\partial x'}{\partial y}$,$\frac{\partial y'}{\partial x}$,$\frac{\partial y'}{\partial y}$ are evaluated by differentiating the expressions (24 - 25) for x' and y' with respect to x and y.

5.3 The eight-parameter algorithm

5.3.1 Algorithm

We summarize the proposed algorithm assuming the following setup. The inputs are two images f and f' of the same object, taken with identical cameras, but from different viewing positions. The segmentation of the reference object in f is known and its contour is assumed to be approximately planar. The segmented reference object is used to construct the prior shape representation $\tilde{\phi}$ according to (23).

1. Choose some initial level-set function ϕ, for example a right circular cone. Its zero level-set is the initial segmenting contour.

2. Set initial values (e.g. zero) for the transformation parameters α, β, γ, t_x, t_y,t_z,ξ and ψ. If d is unknown, set $d = 1$.

3. Compute the average gray level values of the current object and background pixels, u_+ and u_-, using (4).

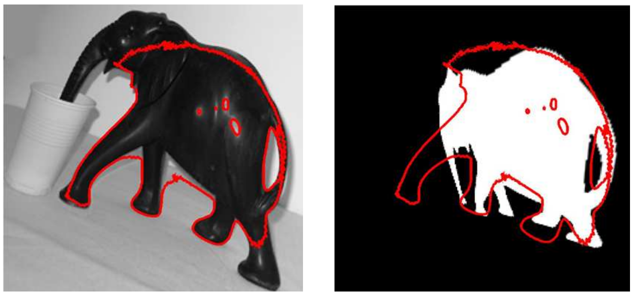
4. Apply a coordinate transformation to the prior shape representation $\tilde{\phi}$ (substitute (22) in (21)) with the current estimate of the transformation parameters.
5. Update ϕ using the gradient descent equation (14).
6. Update the transformation parameters α, β, γ, t_x, t_y, t_z, ψ and ξ as explained in 5.3.2.
7. Repeat steps 3-6 until convergence.

The segmentation and registration steps are interleaved, thus realizing the power of the prior-based segmentation approach. Note that the algorithm directly recovers the transformation parameters, thus the cumbersome task of decomposing the homography matrix is avoided [15, 18, 27].

5.3.2 Optimization

Global optimization problems are common in computer vision. Here, the cost functional (13) may have several local minima with respect to the parameters and the evolving level set function ϕ. Direct update of the parameters via their the derivatives (26) may lead to an undesired local minimum, as exemplified in Fig. 7.

It is known [39] that the global minimum of a general function (or functional) above a continuous domain cannot be found in finite time. Nevertheless, global optimization strategies can be successfully applied to well-behaved functions. In step 6 of the algorithm, the transformation parameters are determined by multidimensional minimization, using the Nelder–Mead downhill simplex algorithm [31] followed by the gradient based Quasi-Newton method [5]. The direct search method of Nelder-Mead is based on evaluating a function at the vertices of a simplex, then iteratively shrinking the simplex as better points are found until some desired bound is obtained. The nine eight-dimensional vertices are randomly initialized within the limits of the parameters. The outcome of the Nelder-Mead algorithm is used as the starting point for Quasi-Newton minimization employing the partial derivatives (26) for fine tuning of the search results. The simplex



(a) (b)

Figure 7: A local minimum. (a) A wrong segmenting contour (red), drawn on the input image. (b) The same contour, drawn on the incorrectly transformed prior shape. To avoid convergence to an incorrect local minimum, we use the Nelder–Mead downhill simplex algorithm followed by the gradient based Quasi-Newton minimization.

algorithm works, in our examples, moderately well. In most cases, the globally minimal cost (corresponding to shape overlap) was identical to the result of quasi-Newton search from the zero starting point. The Quasi-Newton algorithm has a quadratic rate of convergence, but when applied by itself finds the local minimum in the vicinity of its initial point.

The stability of the resulting shape with respect to small errors in the minimizing transformation parameters was examined numerically on several images and transformation examples. The value of the shape term in the functional is taken as a measure for the shape variation. It was found that the minimizing shape varies smoothly with the change in the transformation parameters. The results differ slightly from image to image. In Fig. 8 we depict the results of varying one parameter while keeping the seven others fixed in the ground truth. The degree of stability is defined by the Hessian of the shape term at that point. Few parameters are more sensitive to such a variation. For a better understanding of the stability the correlations (or higher order derivatives) should be calculated. This is under current study.

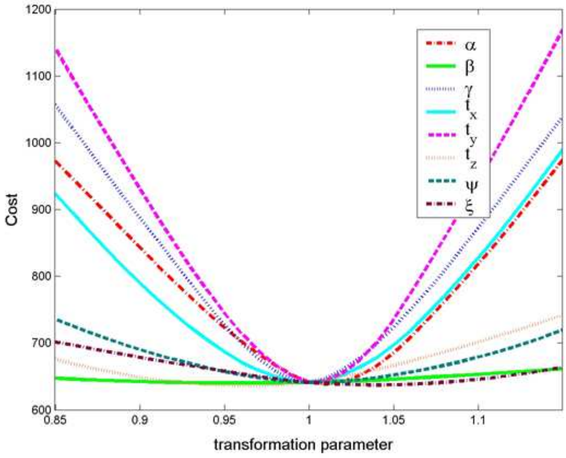


Figure 8: The shape dissimilarity measure (Eq. 12) as a function of each of the transformation parameters while keeping the seven others fixed in the ground truth.

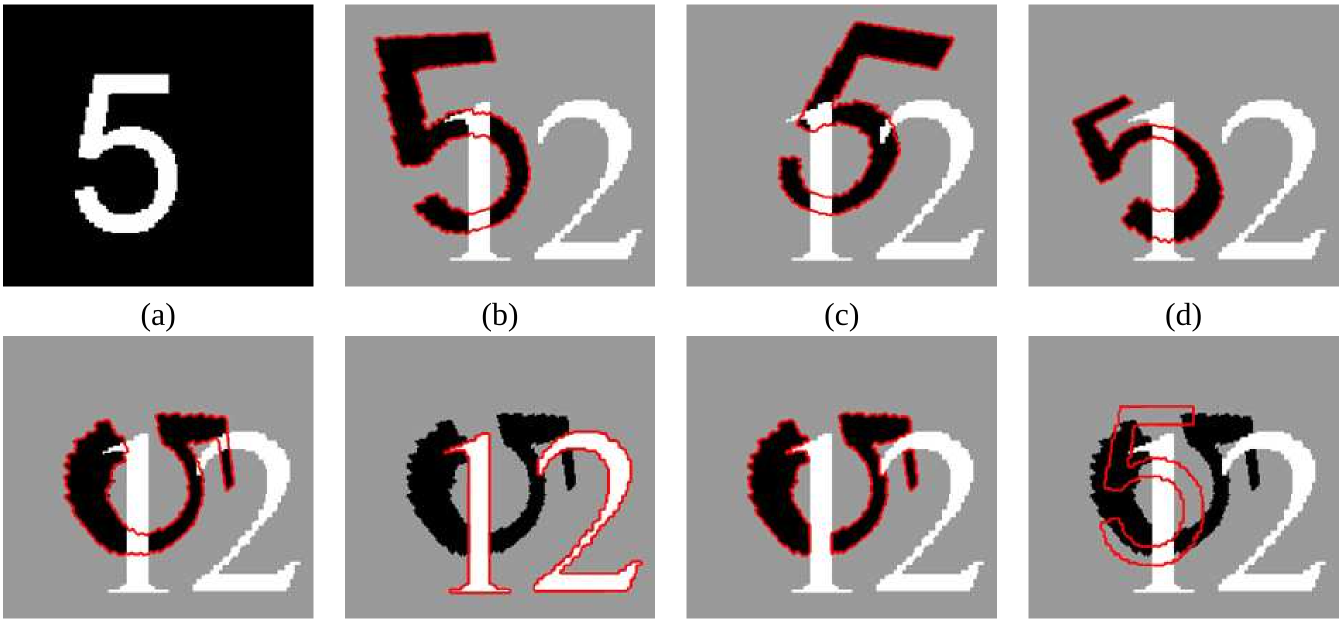
6 Experimental Results

The algorithm developed in this paper is capable of prior-based segmentation and shape registration in the presence of perspective distortion. We demonstrate its operation on a variety of examples. In each example, we show the input image (the image to segment), and the reference image, from which the prior-shape was obtained. In addition to the prior-based segmentation results, we illustrate the mismatch between the input image and the reference image, and verify the transformation parameters estimated by the algorithm. In all the experiments we set $dt = 0.1$, $\epsilon = 1$ and $d = 1$. The contributions of each term in the gradient descent equation of ϕ (14) are normalized to $[-1, 1]$, thus ν and μ are set to 1. Modifying the weight of the prior term is done by normalizing the term to $[-\mu, \mu]$ with $\mu > 0$. Examples for such modifications are shown in Fig. 11f-i. In all the examined examples the number of iterations[4] needed for convergence was between 50 and 100. The reader is also referred to the web page [34] which contains video clips that visually demonstrate the convergence of the segmentation processes for the examples shown in Fig. 10, Fig. 11 and Fig. 14.

We start by presenting two synthetic examples (Fig. 9 and Fig. 10) that allow comparison with the

---

[4] Repetitions of steps 3-6 in the eight parameter algorithm.



(a) (b) (c) (d)

(e) (f) (g) (h)

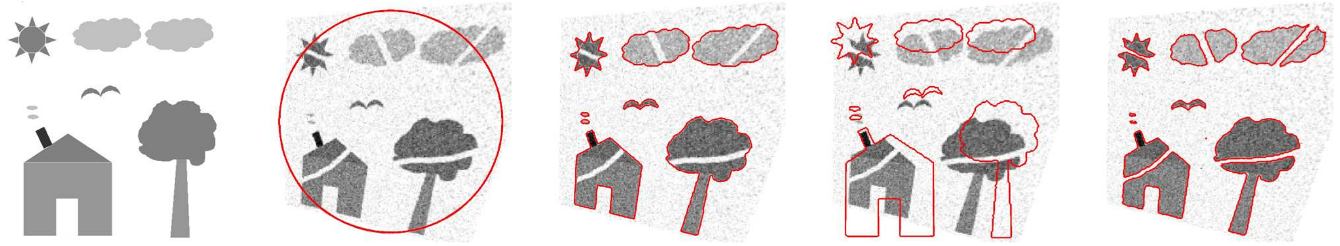
Figure 9: (a) Prior shape. (b)-(e) Successful segmentation results using the prior shape. Note the significant perspective distortions of the shape that simulate quite wide camera view point changes with respect to (a). A comparison of the recovered parameters with the true ones is shown in Table 1. (f)-(g) Segmentation without prior shape term. The final contour may either incorrectly segment the white objects (f) or the black disconnected object (g), depending on the initial contour. (h) The prior shape term is over-stressed. The final segmenting contour ignores the actual image data. This image also demonstrates the misalignment between the prior shape (red) and the input image.

ground truth, and proceed to show segmentation of various real images. The first example (Fig. 9 and Table 1) demonstrates successful segmentation and registration for a wide range of projective transformations between a prior shape, shown in Fig. 9a and its different appearances, presented in Figs. 9b-e. The recovered transformation parameters are compared with the true ones in Table 1. Without the prior shape term, the final contour (red) may either incorrectly segment the white objects (as in Fig 9f) or the black unconnected object (as in Fig 9g), depending on the initial contour. When the prior shape term is over-stressed, the final segmenting contour ignores the actual image data (Fig 9h). This figure also displays the significant mismatch between the prior shape and the shape to segment.

Consider next the synthetic reference image shown in Fig. 10a, that contains several components of different sizes and gray levels. The prior shape was obtained by thresholding. The image to segment is

Transformation parameter	α	β	γ	t_x	t_y	t_z	ψ	ξ
True Fig 9b	0.200^0	-0.200^0	-15.00^0	-5.00	5.00	0.100	0.100^0	-0.200^0
Recovered Fig 9b	0.212^0	-0.214^0	-14.45^0	-5.09	4.66	0.098	0.123^0	-0.269^0
True Fig 9c	0.200^0	0.100^0	20.00^0	-10.00	10.00	-0.050	-0.500^0	-0.500^0
Recovered Fig 9c	0.197^0	0.104^0	20.68^0	-9.82	9.91	-0.040	-0.472^0	-0.495^0
True Fig 9d	-0.300^0	-0.300^0	-50.00^0	-10.00	10.00	0.100	0.700^0	0.700^0
Recovered Fig 9d	-0.302^0	-0.300^0	-49.74^0	-10.19	10.00	0.111	0.718^0	0.678^0
True Fig 9e	-0.300^0	-0.300^0	60.00^0	-10.00	15.00	0.300	-0.300^0	0.700^0
Recovered Fig 9e	-0.299^0	-0.309^0	60.21^0	-10.16	14.96	0.309	-0.281^0	0.701^0

Table 1: Comparison of the recovered and true transformation parameters, for the examples shown in Fig. 9



(a) (b) (c) (d) (e)

Figure 10: (a) A reference image, containing several components of different sizes and gray levels. The prior shape was obtained by thresholding. (b) The input image: a noisy, transformed and corrupted version of (a), with the initial contour (red). (c) Correct segmentation (red). (d) The unregistered prior shape (red) drawn on the input image. The reference (prior) image and the image to segment are clearly misaligned. (e) Segmentation (red) without using the prior shape.

shown in Fig. 10b, together with the initial contour. The input image is a noisy[5], transformed and corrupted version of the image in Fig. 10a. Successful segmentation is demonstrated in Fig. 10c. The recovered transformation parameters are validated by comparison to the true transformation parameters in Table 2. The misalignment between the image to segment and the reference image is shown in Fig. 10d. Without using the prior shape, the input image is difficult to segment, see Fig. 10e.

Real images are used in all other examples, Figs. 11-14. Consider the two views of the toy elephant shown in Figs. 11a-b. As seen in Fig. 11c, the segmenting contour precisely tracks the outline of the elephant, and recovers the missing part of its trunk. This segmenting contour is the zero level-set of the final

---

[5] Zero-mean Gaussian noise with STD equal to 14% of the grey-level dynamic range.

evolving function ϕ shown in Fig. 11d. The accuracy of the recovered transformation is visually verified in Fig. 11e. To demonstrate the influence of the prior shape weight on the final segmentation, we have conducted several experiments with various values of μ, shown in Figs. 11f-i. In Fig. 11f, the prior term is over-stressed ($\mu \gg 1$) thus the data is ignored and the transformation between the prior and the image to segment is not recovered. Note the significant misalignment between the prior shape (red) and the input image. Fig. 11g shows segmentation without prior ($\mu = 0$). When the weight of the shape term is low, as demonstrated in Fig. 11h ($\mu = 0.5$) and in Fig. 11i ($\mu = 0.8$), the final segmenting contour occupies the gap between the object boundaries and the transformed prior. Further stressing the contour length term ($\nu > 1$) a result similar to Fig. 11g is obtained. We compared the homography matrix generated from the estimated transformation parameters with that obtained using 13 manually selected corresponding point pairs, see Table 3. The comparison has been done for the implicit homography parameters that were recovered directly (row 2) and those that were computed from the explicit homography matrix entries according to Eq. 22 (row 3). The values of the parameters recovered assuming similarity transformation (Fig. 1a) and affine transformation (Fig. 1b) are reported in rows 4-5.

Transformation parameter	α	β	γ	t_x	t_y	t_z	ψ	ξ
True values	0.075^0	0.075^0	7.5^0	-10.0	10.0	-0.1	0.075^0	0.15^0
Recovered Values	0.069^0	0.075^0	7.4^0	-10.0	9.6	-0.1	0.086^0	0.14^0

Table 2: Comparison of the recovered and true transformation parameters, for the example shown in Fig. 10.

Fig. 12 demonstrates successful prior-based segmentation and registration of the scissors despite the holes and the specular reflection. This is a challenging example, since the overlap between the reference image and the image to segment is small, and localized in the over-exposed part of the image. In Figs. 13c,g, the bottle-opener is accurately segmented regardless of the significant projective distortion and the over-exposure at the bottom-right part of the opener. Note the difference in the grey-level distribution between the two views. Figs. 13h-i show registration of the prior contour to the given image, assuming similarity and

Homography matrix entry	h_{11}	h_{12}	h_{13}	h_{21}	h_{22}	h_{23}	h_{31}	h_{32}
Point correspondence	0.91	0.28	3.2	-0.31	0.71	34.05	-0.0003	0.0019
Proposed algorithm, implicit homography	0.92	0.27	2.88	-0.29	0.70	32.7	-0.0003	0.0017
Proposed algorithm, explicit homography	0.92	0.27	2.80	-0.30	0.70	33.35	-0.0003	0.0018
Similarity approximation	0.88	0.37	1.81	-0.37	0.88	35.5	0	0
Affine approximation	0.92	0.30	0.09	-0.37	0.78	31.01	0	0

Table 3: Comparison of the homography matrix entries obtained using the proposed algorithm and via manually selected corresponding point pairs, for Fig. 11. The last two rows of the table display the transformation parameters obtained assuming similarity and affine transformations respectively. These values correspond to the registration results shown in Fig. 1b-c. Note, that not all the similarity parameters are independent.

affine transformations respectively. The similarity and affine approximations recover the rotation in the image plane, but fail to capture the perspective deformation. The apparent mismatch inhibits the segmentation process and prohibits accurate reconstruction of the missing parts.

Finally, Figs. 14a-b are a day and night satellite views of Europe respectively. The images were adapted from NASA:*earthobservatory.nasa.gov*. Successful segmentation is demonstrated in Fig. 14c. The coast line of the continent and the islands are precisely extracted, despite the spatial brightness variations. The accuracy of the recovered transformation is visually verified in Fig. 14d. Note the significant mismatch, Fig. 14e, between the two views. Without using the prior shape, segmentation of Fig. 14b is difficult. As seen in Fig. 14f, parts of Scandinavia and the eastern Adriatic shore are incorrectly segmented[6].

7 Discussion

Given an image and a prior shape, segmentation and registration can be viewed as two aspects of a single underlying problem. The essence of the method proposed in this paper is concurrent segmentation and registration, as mutually supporting processes within a unified variational framework. The keys to the success is the introduction of a shape similarity measure that enables the incorporation of the homography formulation

---

[6] A similar segmentation result (no prior), showing similar difficulties, appeared in [3], Fig. 11.

within the objective functional. Explicit encoding of a geometric transformation with such a high descriptive power allowed to break away from the common statistical representation of transformations in prior-based segmentation.

The ability to precisely extract the actual object contour based on the image data and its expected shape is critical in many applications. A robot manipulator guided by the segmentation of Fig. 11g, for example, might break the elephant trunk or attempt to pierce through its specular reflections, mistakenly interpreted as holes. In this sense, segmentation that is based on Fig. 11c is preferable.

The Chan-Vese two-phase segmentation model, currently employed in our scheme, requires the average grey-levels of the foreground and background to be distinct. Nevertheless, as seen in the bottle opener, scissors and Europe examples (Figs. 12-14), the use of the shape prior allows to substantially relax the homogeneity requirements. The suggested method is modular in the sense that the underlying level-set segmentation model can be upgraded. For example, it will be possible to accommodate cluttered background by using the Vese-Chan multi-phase segmentation model [43].

Expressing the planar-projective transformation within the energy functional (13) in terms of relative camera-object motion (R, $\mathbf{t}$) and plane structure ($\mathbf{n}$, d) rather than via the homography matrix is a considerable shortcut toward the recovery of these parameters. The cumbersome task of decomposing the homography matrix [15, 18, 27] is avoided. Also, note that reference images are often taken in controlled conditions. If the prior image plane is perpendicular to the optical axis, the search for the eight homography parameters can be reduced to six by setting the structure parameters (ψ, ξ) to zero, thus simplifying the registration procedure. For non-calibrated cameras the homography can be fully recovered in its implicit form (recovery of $h_{11} \ldots h_{32}$ instead of α, β etc.). Then only eight parameters are recovered and the need for decoupling between the internal and external camera parameters does not arise. The mathematical formulation for the recovery of the (implicit) homography parameters is obtained by substituting the expressions (21) directly in (27) for $\eta = h_{11} \ldots h_{32}$. The homography parameters recovered using this formulation are shown in Table 3, second row.

In principle, the planar projective transformation model requires the observed contour points of the 3D object to be coplanar and to take part in the prior shape outline as well. In practice, this assumption can be relaxed for moderate transformations. As seen in the satellite image example (Fig. 14), an excellent result is obtained even though the coastline of Europe is clearly non-planar. The elephant example (Fig. 11) demonstrates successful application of the suggested method despite the evident violation of the assumption. A comprehensive solution for general 3D objects would require to extend the transformation model beyond planar projective homography, a subject for further study.

In active contour methods that do not employ a prior shape, the selection of the initial contour might be crucial for correct segmentation and designation of segments as object or background. Figs. 9f-g demonstrate possible segmentation results obtained without a shape prior, that are initialization dependent. Prior shape knowledge facilitates the detection of the desired object, reducing the influence of the initialization on the final result, as demonstrated in the successful segmentation results in Figs. 9b-e.

The suggested approach can be extended to deal with non-rigid objects in two ways. If the non-rigidity can be modeled in parametric form, as in the case of articulated objects, the additional parameters could be incorporated within the transformation model. Otherwise, one may integrate the proposed method with a statistical scheme based on a training set, e.g., Rousson and Paragios [36]. This would allow to capture the non-rigidity via the statistical component, and the projective transformation using the approach that we suggest.

The successful segmentation results and the reliable estimation of the transformation parameters encourage future research. Potential applications include perspective-invariant search in image databases, registration and structure recovery in stereo imaging, and object tracking in video sequences.

Acknowledgments

This research was supported by MUSCLE: Multimedia Understanding through Semantics, Computation and Learning, a European Network of Excellence funded by the EC 6th Framework IST Programme.

References

[1] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. Springer, 2002.

[2] T.O. Binford. Visual perception by computer. In *Proc. IEEE Conf. Systems and Control*, December 1971.

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

[4] 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.

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

[6] Cone, in mathematics. The Columbia Encyclopedia, 6th Edition. Columbia University Press, 2001.

[7] G.J. Cootes, T.F.and Edwards and C.J. Taylor. Active appearance models. In *Proceedings of the European Conference on Computer Vision*, volume 2, pages 484–498, 1998.

[8] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995.

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

[10] 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.

[11] D. Cremers, N. Sochen, and C. Schnorr. Towards recognition-based variational segmentation using shape priors and dynamic labeling. In *Intl. Conf. on Scale-Space Theories in Computer Vision*, pages 388–400, June 2003.

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

[13] 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.

[14] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.

[15] O. Faugeras, Q.T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images*. MIT Press, 2001.

[16] D.A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2003.

[17] R.M. Haralick and L.G. Shapiro. *Computer and Robot Vision*, volume II. Addison-Wesley, 1993. pp. 593-594.

[18] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2003.

[19] 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.

[20] M. Irani and P. Anandan. All about direct methods. In W. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*. Springer-Verlag, 1999.

[21] A.K. Jain, Y. Zhong, and M.P. Dubuisson-Jolly. Deformable template models: A review. *Signal Processing*, 71(2):109–129, December 1998.

[22] R. Jain, R. Kasturi, and B.G. Schunck. *Machine Vision*. McGraw-Hill, 1995. pp. 482-483.

[23] F. Jurie and M. Dhome. Real time 3d template matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 791–797, 2001.

[24] 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.

[25] 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.

[26] M.D. Levine. *Vision in Man and Machine*. McGraw-Hill, 1985. pp. 46-53.

[27] Y. Ma, S. Soatto, J. Košecká, and S. S. Sastry. *An Invitation to 3-D Vision*. Springer-Verlag, 2003.

[28] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman, 1982.

[29] M.I. Miller and L. Younes. Group actions, homeomorphisms and matching: a general framework. *International Journal of Computer Vision*, 41:61–84, July 2001. Special issue on statistical and computational theories of vision: Part 2.

[30] 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.

[31] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[32] 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.

[33] K.G. Rao and G. Medioni. Generalized cones: Useful geometric properties. In *Computer Vision, Graphics, and Image Processing*, pages 185–208, 1992.

[34] T. Riklin-Raviv. Segmentation examples. http://www.eng.tau.ac.il/∼tammy/PriorBasedGif.htm, 2005.

[35] 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.

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

[37] C. E. Springer. *Geometry and Analysis of Projective Spaces*. Freeman, 1964.

[38] G.C. Stockman and A.K. Agrawala. Equivalence of hough curve detection to template matching. *Communications of the ACM*, 20(11):820–822, 1977.

[39] A. Törn and A. Žilinskas. *Global Optimization*, volume 350 of *Lecture Notes in Computer Science*. Springer, 1989.

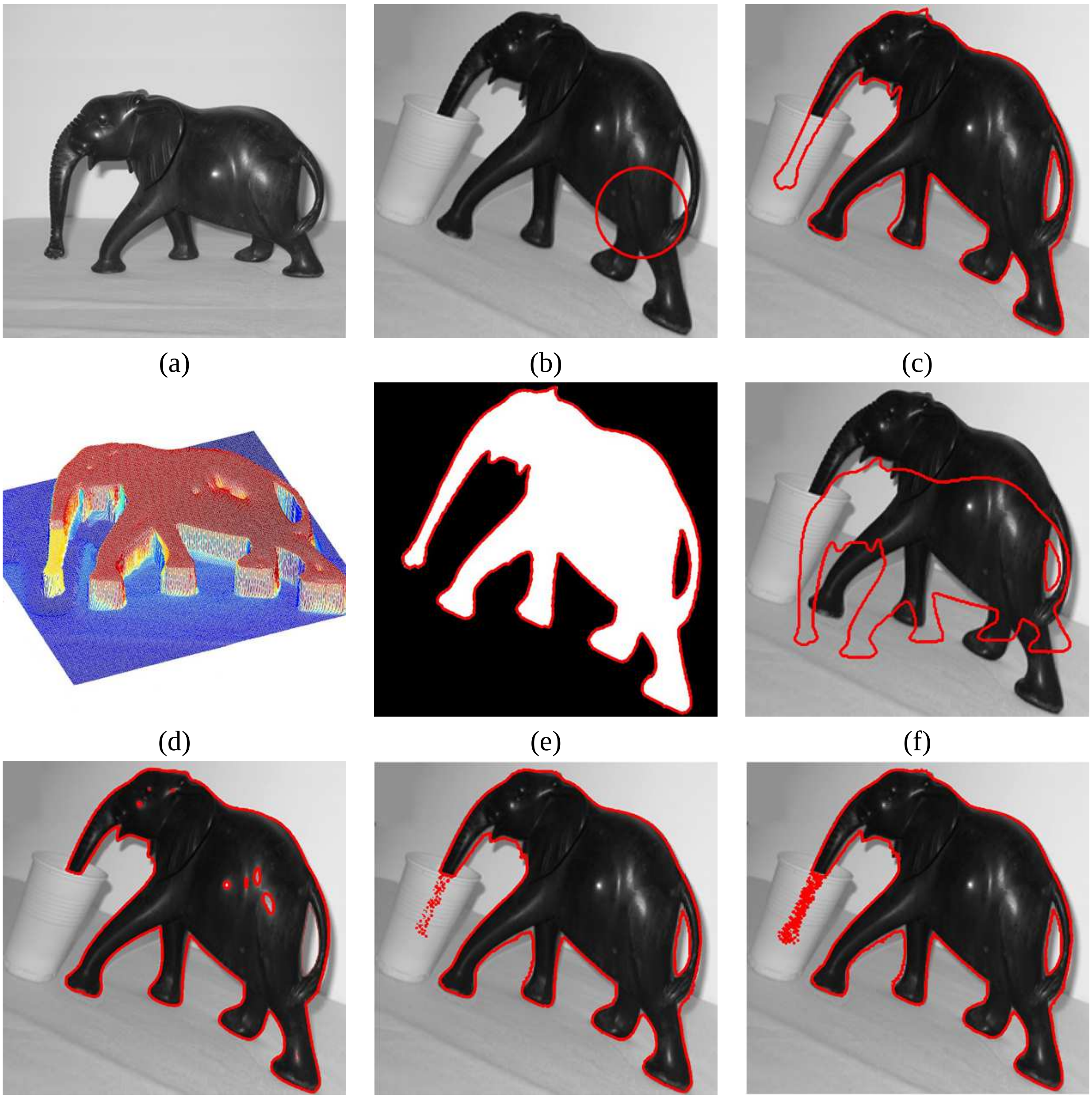
[40] A. Tsai, A. Yezzi, Jr., W.M. Wells, III, C. Tempany, D. Tucker, A. Fan, W.E.L. Grimson, and A.S. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging*, 22(2):137–154, Feb. 2003.

[41] R.C. Veltkamp and M. Hagedoorn. Shape similarity measure, properties and constructions. In *Principles of Visual Information Retrieval*, pages 467–476, London, UK, 2000. Springer-Verlag.

[42] R.C. Veltkamp and M. Hagedoorn. State of the art in shape matching. In *Principles of Visual Information Retrieval*, pages 87–119. Springer-Verlag, London, UK, 2001.

[43] 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.

[44] E.D. Weisstein. Euler angles. From MathWorld– A Wolfram Web Resource., 1999–2005. http://mathworld.wolfram.com/EulerAngles.html.

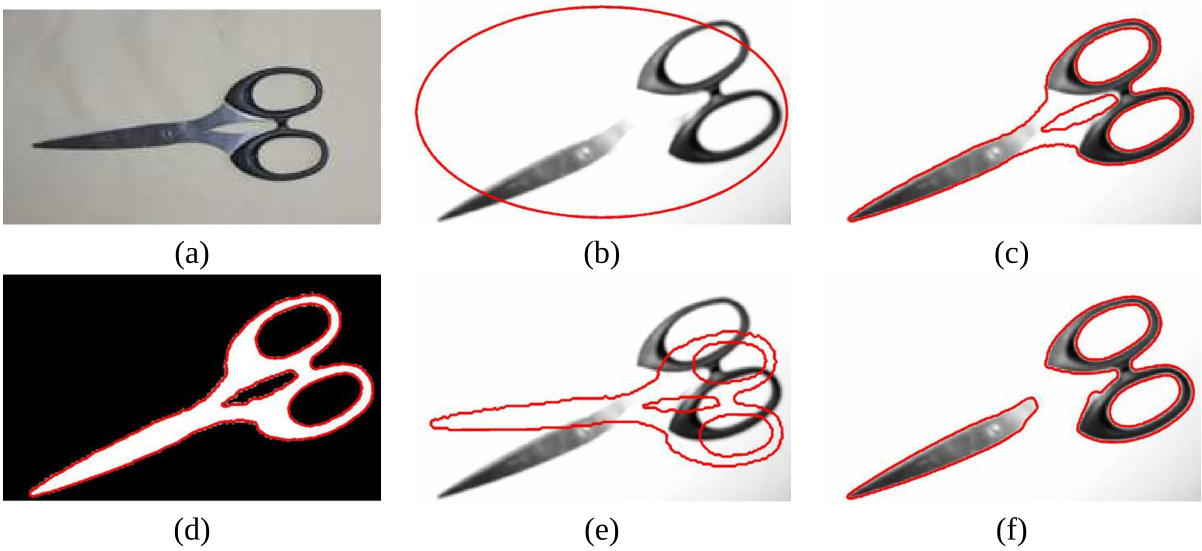


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

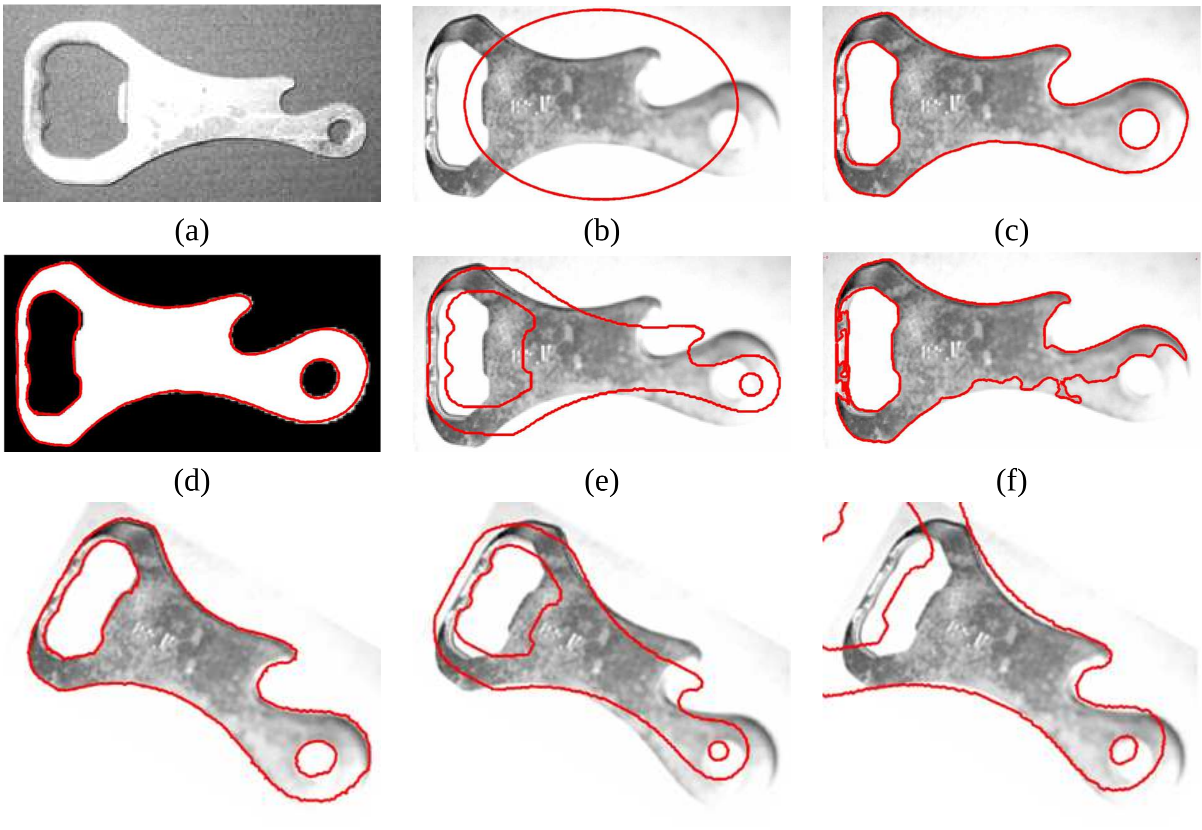
Figure 11: (a) A reference image (elephant), from which the prior shape was obtained. (b) The input image. The initial contour used for the segmentation process is shown (red). (c) Successful segmentation: the final contour is shown (red). (d) The final form of the evolving level-set function ϕ. (e) Verification of the estimated transformation parameters: The final contour as in (c), drawn on the projectively transformed prior shape according to the recovered homography. (f)-(i) The suggested method is demonstrated for several values of prior shape weight (μ). (f) The prior shape is over-stressed ($\mu \gg 1$) thus the data is ignored and the transformation between the prior and the image to segment is not recovered. Note the significant misalignment between the prior contour (red) and the input image. (g) $\mu = 0$. The prior shape does not effect the segmentation result . (h)-(i) Low shape term weight ($\mu < 1$). The final segmenting contour occupies the gap between the object boundaries and the transformed prior contour. Further stressing the contour length will yield result similar to (g). (h) $\mu = 0.5$ (i) $\mu = 0.8$.



(a) (b) (c)

(d) (e) (f)

Figure 12: (a) A reference image (scissors), from which the prior shape was obtained. (b) The input image, taken from a different viewpoint and with missing parts due to over-exposure. (c) Successful segmentation. The missing contour parts are accurately completed. (d) Verification of the estimated transformation parameters: The final contour as in (c), drawn on the projectively transformed prior shape according to the recovered homography. (e) The misalignment between the prior shape (red) and the input image. (f) Segmentation (red) without using the prior shape.

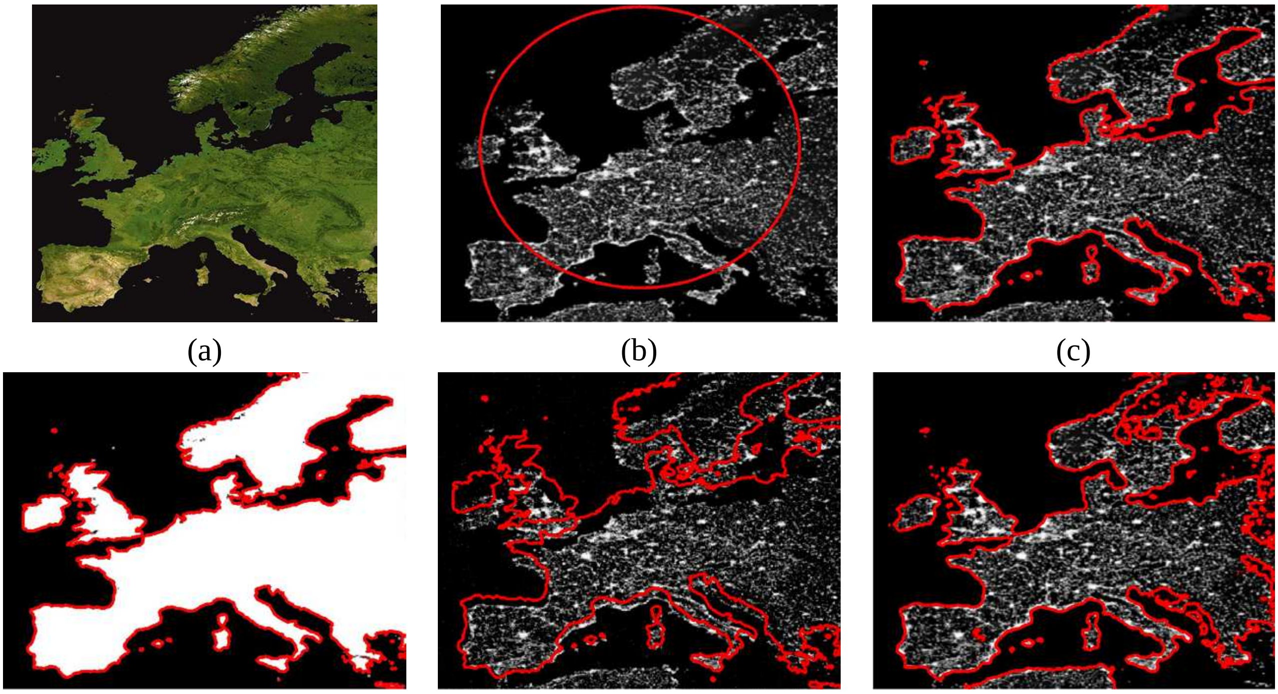


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 13: (a) A reference image (bottle opener), from which the prior shape was obtained. (b) The input image, taken from a different viewpoint and with over-exposure. (c) Successful segmentation. The missing contour parts are accurately completed. (d) Verification of the estimated transformation parameters: The final contour as in (c), drawn on the projectively transformed prior shape according to the recovered homography. (e) The misalignment between the prior shape (red) and the input image. (f) Segmentation (red) without using the prior shape. (g) Successful segmentation with wider camera view, using the reference image (a). (h)-(i) Registration of the prior contour to the given image using (h) similarity approximation and (i) affine approximation. Note the apparent mismatch. The similarity and the affine approximations cannot capture the perspective distortion.



(a) (b) (c)

(d) (e) (f)

Figure 14: (a) A reference image of Europe, from which the prior shape was obtained. (b) Input image: A satellite image of Europe at night. The reference and prior image were adapted from *earthobservatory.nasa.gov*. (c) Successful segmentation: the final contour (red) accurately traces the coast lines. (d) Verification of the estimated transformation parameters: The final contour as in (c), drawn on the projectively transformed prior shape according to the recovered homography. (e) The misalignment between the prior shape (red) and the input image. (f) Segmentation (red) without using the prior shape. The dark areas in Scandinavia and in the eastern Adriatic shore are incorrectly segmented.