

Pairwise Similarities across Images for Multiple View Rigid/Non-Rigid Segmentation and Registration

Luca Bertelli Marco Zuliani B.S. Manjunath
Electrical and Computer Engineering Department
University of California, Santa Barbara
93106 Santa Barbara, CA
{lbertelli, zuliani, manj}@ece.ucsb.edu

Abstract

A variational approach for the background/foreground segmentation of multiple views of the same scene is presented. The main novelty is the introduction of cost functions based on pairwise similarity between pixels across different images. These cost functions are minimized within a level set framework. In addition, a warping model (rigid or non-rigid) between the emerging foregrounds in the different views is imposed, thus avoiding the introduction of a specific shape term in the cost function to handle occlusions. The thin plate spline (TPS) warping is for the first time employed within the level set framework to model non-rigid deformations. The minimization of these cost functions leads to simultaneous segmentation and registration of the different views. Examples of segmentations of a variety of objects are shown and possible applications are proposed.

1. Introduction

Image segmentation is one of the primary tasks in computer vision. It can be, in fact, considered the first step towards any further analysis and processing.

A fundamental variational approach to region-based image segmentation was presented by Mumford and Shah in [14, 15], where they minimized a functional to approximate the image in a piecewise smooth way, penalizing at the same time the excessive length of the contours between regions. Later on, Chan and Vese minimized this functional within the level set framework [17], for both piecewise constant [4] and smooth approximations of the image. Other relevant work in variational segmentation include region based methods [30, 29, 25] and edge driven approaches [3, 11].

In the recent past many attempts have been made to include shape information in the variational segmentation framework [5, 23, 19, 20, 12, 6, 7, 9]. These methods are

motivated by the need to segment partly occluded objects. In most of the cases a shape term is added to the variational cost function and the minimization is performed with respect to both the specific segmentation term and the prior term. The prior shape term is used to penalize the presence of non-overlapping areas between the known desirable shape and the evolving shape. Some models assume the existence of a well defined shape prior [5, 23, 19, 8, 18, 21], which is then utilized to recover the boundaries of the occluded or partially missing objects along with the transformation that aligns the shape prior on top of the target image. All these methods involve only rigid registrations and are not suitable for coping with non-rigid deformations of the prior. To address this issue, in [7, 6, 12, 24] statistical inferences are made by exploiting a set of reference shapes, that form a basis for the admissible shapes. Recently, a prior based level set framework to model non-rigid deformations using a deformation vector field has been proposed in [26].

In practical applications the information about the shape may not be directly available. Multiple views of the same scene can be instead exploited to correctly segment and identify objects, even if in some of the views the objects of interest are occluded. Relevant work in this field includes [20], where two images are concurrently segmented and registered and the emerging shape in one image is used as a prior for the second one. In [28, 26] multiple images are segmented evolving a single contour as well as the mapping of that contour into each image. No direct connection between the different views is defined, except for the common contour. In [10] a set of similar shapes is used to complete missing parts applying majority rules.

In this paper we introduce a direct connection between the multiple views of the same scene, by defining novel variational cost functions based on pixel pairwise dissimilarities across images. After imposing a warping constraint between the views, and therefore between the emerging segmentations, we minimize these cost functions by evol-

ing only one contour and registering it across the different views. In this way, we avoid the need of introducing a specific shape term in the cost function and, at the same time, we exploit the prior in a more complete way, taking into account also intensity or color information of the emerging regions and not just using it as a binary mask. Finally, we introduce thin plate spline interpolation within the level set framework to cope with non-rigid deformation of the object shape, which, to the best of our knowledge, is the first attempt in this direction.

1.1. Segmentation based on Pixel Pairwise Dissimilarities

Pairwise pixel dissimilarities can be integrated within a variational approach, see for example [22]. A corresponding cost function can be written as:

$$E(C) = - \int_{\mathbf{x}_1 \in R_o(C)} \int_{\mathbf{x}_2 \in R_i(C)} w(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \quad (1)$$

where $\mathbf{x}_1, \mathbf{x}_2$ are points in the image domain, $R_i(C)$ and $R_o(C)$ are respectively the regions inside and outside of the curve C and $w(\mathbf{x}_1, \mathbf{x}_2)$ is the pairwise dissimilarity between point \mathbf{x}_1 and point \mathbf{x}_2 . Minimizing $E(C)$ with respect to C , we obtain a partitioning of the image, which maximizes the dissimilarity between regions $R_i(C)$ and $R_o(C)$. The curve evolution corresponding to the steepest descent minimization of (1) is given by:

$$\frac{\partial C}{\partial t} = \left(\int_{\mathbf{x} \in R_o(C(t))} w(\mathbf{c}, \mathbf{x}) d\mathbf{x} - \int_{\mathbf{x} \in R_i(C(t))} w(\mathbf{c}, \mathbf{x}) d\mathbf{x} \right) \vec{N} \quad (2)$$

where \mathbf{c} is a point on the curve C . An intuitive interpretation of this evolution force is the following: the similarity of every point \mathbf{c} belonging to the curve C is compared to $R_i(C)$ and $R_o(C)$. If \mathbf{c} is more similar (less dissimilar) to the interior of C the curve expands, otherwise it shrinks.

The rest of this paper is organized as follows: in Section 2 we extend the above formulation to the multi-view case. In Section 3, imposing a warping constraint between the emerging backgrounds and foreground of the multiple views, we provide an effective way to handle occlusions. The different warping models for rigid as well as for non-rigid registration are described in Section 4. After presenting experimental results and proposing potential applications in Section 5, we conclude in Section 6.

2. Multiple View Background/Foreground Separation

Consider two images I_1 and I_2 , and the dissimilarity measure $w(\mathbf{x}_1, \mathbf{x}_2)$, where $\mathbf{x}_i = (x_i, y_i)$ is a 2D point in the image I_i . Define:

$$diss(A_1, A_2) \triangleq \int_{\mathbf{x}_1 \in A_1} \int_{\mathbf{x}_2 \in A_2} w(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \quad (3)$$

and

$$diss(B_1, B_2) \triangleq \int_{\mathbf{x}_1 \in B_1} \int_{\mathbf{x}_2 \in B_2} w(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \quad (4)$$

where A_i, B_i form a partitioning such that $A_i \cup B_i = \Omega_i$ and Ω_i represents the whole image domain (of image I_i). We can formulate a multiple-view cost function based on pairwise dissimilarity as:

$$E_D = diss(A_1, A_2) + diss(B_1, B_2) \quad (5)$$

The regions in (5) can be represented within a level set framework. Define a 3D surface ϕ_i such that its zero level set is the curve C_i , which is the boundary between foreground and background. We can write:

$$C_i = \{\mathbf{x} \in \Omega_i | \phi_i(\mathbf{x}) = 0\} \quad (6)$$

$$A_i = \{\mathbf{x} \in \Omega_i | \phi_i(\mathbf{x}) > 0\} \quad (7)$$

$$B_i = \{\mathbf{x} \in \Omega_i | \phi_i(\mathbf{x}) < 0\} \quad (8)$$

Now using the Heaviside function $H(z)$, equal to 1 if $z > 0$ and 0 if $z < 0$, we can rewrite (5) as follows:

$$E_D = \int_{\Omega_1} \int_{\Omega_2} w(\mathbf{x}_1, \mathbf{x}_2) H(\phi_1(\mathbf{x}_1)) H(\phi_2(\mathbf{x}_2)) d\mathbf{x}_2 d\mathbf{x}_1 + \int_{\Omega_1} \int_{\Omega_2} w(\mathbf{x}_1, \mathbf{x}_2) (1 - H(\phi_1(\mathbf{x}_1))) (1 - H(\phi_2(\mathbf{x}_2))) d\mathbf{x}_2 d\mathbf{x}_1 \quad (9)$$

The cost function in (9) has to be minimized with respect to both ϕ_1 and ϕ_2 . Let us first minimize with respect to ϕ_1 . The gradient projection method minimizing $\int f(\phi_i(\mathbf{x})) d\mathbf{x}$, using t as the descent variable leads to:

$$\frac{\partial \phi_i}{\partial t} = - \frac{\partial f}{\partial \phi_i} \quad (10)$$

In our case we can manipulate (9) so that:

$$f(\phi_1(\mathbf{x}_1)) = \int_{\Omega_2} w(\mathbf{x}_1, \mathbf{x}_2) H(\phi_1(\mathbf{x}_1)) H(\phi_2(\mathbf{x}_2)) d\mathbf{x}_2 + \int_{\Omega_2} w(\mathbf{x}_1, \mathbf{x}_2) (1 - H(\phi_1(\mathbf{x}_1))) (1 - H(\phi_2(\mathbf{x}_2))) d\mathbf{x}_2 \quad (11)$$

Applying (10) to (11) yields:

$$\begin{aligned} \frac{\partial \phi_1(\mathbf{x}_1)}{\partial t} = & - \int_{\Omega_2} w(\mathbf{x}_1, \mathbf{x}_2) \delta(\phi_1(\mathbf{x}_1)) H(\phi_2(\mathbf{x}_2)) d\mathbf{x}_2 \\ & + \int_{\Omega_2} w(\mathbf{x}_1, \mathbf{x}_2) \delta(\phi_1(\mathbf{x}_1)) \left(1 - H(\phi_2(\mathbf{x}_2)) \right) d\mathbf{x}_2 \end{aligned} \quad (12)$$

Rearranging the terms, we can rewrite (12) as:

$$\begin{aligned} \frac{\partial \phi_1(\mathbf{x}_1)}{\partial t} = & \delta(\phi_1(\mathbf{x}_1)) \left[\int_{\Omega_2} w(\mathbf{x}_1, \mathbf{x}_2) \left(1 - H(\phi_2(\mathbf{x}_2)) \right) d\mathbf{x}_2 \right. \\ & \left. - \int_{\Omega_2} w(\mathbf{x}_1, \mathbf{x}_2) H(\phi_2(\mathbf{x}_2)) d\mathbf{x}_2 \right] \end{aligned} \quad (14)$$

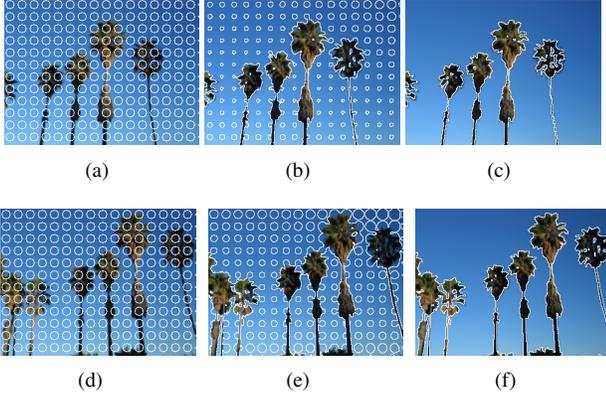


Figure 1. Mutual Segmentation of two views according to (14). (a-b) Initialization of ϕ_1 and ϕ_2 in multiple sub-parts. (c-d) Intermediate phase of the evolution. (e-f) Final segmentation result.

Every point on the curve C_1 in image I_1 is therefore compared in similarity with the points inside and outside C_2 in image I_2 . The curve is then expanded or shrunk accordingly. Similar derivation can be conducted for ϕ_2 . The algorithm evolves alternatively ϕ_1 and ϕ_2 until convergence is reached. Fig. 1 shows some snapshots of this concurrent evolution process. The evolution equation in (14) can be normalized with respect to the areas of the evolving regions, obtaining therefore comparisons in terms of average similarity rather than total similarity.

The cost function presented in (5) for the two-image case can be easily extended to more than two images considering the cross-image dissimilarities between pixels of a reference image, I_1 for instance, and all the other views of the same scene, i.e. I_i for $i = 2 \dots n$. We can therefore rewrite (5) as:

$$E_D = \sum_{i=2}^n \text{diss}(A_1, A_i) + \text{diss}(B_1, B_i) \quad (15)$$

where n is the number of views. The curve evolution equations for each ϕ_i can be derived following the same approach described above.

Similar to all the active contours techniques, a regularization term is needed in the cost function to prevent oversegmentation. This is done by adding a term proportional to the length of the zero level set of ϕ_i (see [17, 4, 16]):

$$E_L = \mu \sum_{i=1}^n \int_{\Omega_i} |\nabla H(\phi_i(\mathbf{x}))| d\mathbf{x} \quad (16)$$

Minimizing (16) we obtain the associated steepest descent equation for each ϕ_i :

$$\frac{\partial \phi_i(\mathbf{x})}{\partial t} = \mu \delta(\phi_i(\mathbf{x})) \text{div} \left(\frac{\nabla \phi_i(\mathbf{x})}{|\nabla \phi_i(\mathbf{x})|} \right) \quad (17)$$

3. Shape Prior Without Shape

The model described so far allows the two level set function ϕ_i to freely evolve, one independently of the other one. This is not suitable for handling occlusion. If in one of the two views the object of interest is occluded, a desirable segmentation result should be able to preserve and recover its shape. A common approach, extensively utilized in the literature (see for example [23, 19, 20, 5]), consists of introducing a shape term in the cost function, that penalizes the presence of non-overlapping areas between the shape prior and the evolving shape. Therefore, such a shape term can be written as:

$$E_S = \nu \int_{\Omega} \left(H(\phi(\mathbf{x})) - H(T_{\mathcal{P}} \circ \phi_p(\mathbf{x})) \right)^2 d\mathbf{x} \quad (18)$$

where ϕ_p is a labeling function for the prior shape (i.e. ϕ_p is positive inside of the shape and negative elsewhere) and $T_{\mathcal{P}}$ is a suitable geometric transformation which registers ϕ_p with ϕ . The main drawback of this approach is that the choice of the parameter ν , the weight of the shape term, is image dependent and can be critical to obtain good segmentation results. In addition the shape is used as a binary mask, without exploiting the information content of the prior image in terms of intensity or color (see Fig. 4).

In this section we reformulate the cost function in (9) imposing a projective warping on the shape of the evolving regions. By doing this, the prior shape information is embedded into the cost function in an implicit way (hence, “Shape Prior Without Shape”), eliminating the need of additional tuning parameters.

First, let us assume that two views are related through a warping model $\mathbf{W} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, parameterized in terms of the parameters vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$. Therefore, a point \mathbf{x} in one view is mapped onto the transformed view according to $\mathbf{x}' = \mathbf{W}(\mathbf{x}, \boldsymbol{\theta})$.

Now consider a function $\phi(\mathbf{x}_1)$, whose zero level set function defines a curve in image I_1 . The function $\phi(\mathbf{W}(\mathbf{x}_2, \boldsymbol{\theta}))$ defines a zero level set curve in the transformed view I_2 . We can therefore rewrite (9), substituting ϕ_1 and ϕ_2 with $\phi(\cdot)$ and its warped version $\phi(\mathbf{W}(\cdot, \boldsymbol{\theta}))$:

$$E = \int_{\Omega_1} \int_{\Omega_2} w(\mathbf{x}_1, \mathbf{x}_2) H(\phi(\mathbf{x}_1)) H(\phi(\mathbf{W}(\mathbf{x}_2, \boldsymbol{\theta}))) d\mathbf{x}_1 d\mathbf{x}_2 + \int_{\Omega_1} \int_{\Omega_2} w(\mathbf{x}_1, \mathbf{x}_2) \left(1 - H(\phi(\mathbf{x}_1)) \right) \left(1 - H(\phi(\mathbf{W}(\mathbf{x}_2, \boldsymbol{\theta}))) \right) d\mathbf{x}_1 d\mathbf{x}_2 \quad (19)$$

Now the minimization has to be performed with respect to both ϕ and the parameters vector $\boldsymbol{\theta}$, which characterizes the warping, yielding:

$$\frac{\partial \phi(\mathbf{x}_1)}{\partial t} = \delta(\phi(\mathbf{x}_1)) \left[\int_{\Omega_2} w(\mathbf{x}_1, \mathbf{x}_2) \left(1 - H(\phi(\mathbf{W}(\mathbf{x}_2, \boldsymbol{\theta}))) \right) d\mathbf{x}_2 - \int_{\Omega_2} w(\mathbf{x}_1, \mathbf{x}_2) H(\phi(\mathbf{W}(\mathbf{x}_2, \boldsymbol{\theta}))) d\mathbf{x}_2 \right] \quad (20)$$

and

$$\frac{\partial \theta}{\partial t} = \int_{\Omega_1} \int_{\Omega_2} \left[w(\mathbf{x}_1, \mathbf{x}_2) \left(1 - 2H(\phi(\mathbf{x}_1)) \right) \delta(\phi(\mathbf{W}(\mathbf{x}_2, \theta))) \cdot \left[\nabla \phi(\mathbf{W}(\mathbf{x}_2, \theta)) \frac{\partial \mathbf{W}(\mathbf{x}_2, \theta)}{\partial \theta} \right]^T \right] d\mathbf{x}_1 d\mathbf{x}_2 \quad (21)$$

4. Warping Model

Almost all techniques present in the literature, concerning shape based level set methods, involve only rigid registration models as the transformation between the prior shape and the target image (see [5, 23, 19, 28, 18]). In a recent work [26], the authors modeled non rigid deformation between boundaries, using a deformation vector field. Within the framework presented in this paper, we have experimented two classes of transformations: *rigid* (affine and homographic) and *non-rigid* (thin plate splines). To the best of our knowledge this work represents the first attempt to introduce thin plate spline warp within a level set framework.

4.1. Rigid Warping

Within the rigid transformation class, we experimented using affine and homographic warpings (parameterized respectively by $\theta \in \mathbb{R}^6$ and $\theta \in \mathbb{R}^8$):

$$\mathbf{x}' = A\mathbf{x} + T = \begin{bmatrix} s_x & \alpha \\ 0 & s_y \end{bmatrix} \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \mathbf{x} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (22)$$

and

$$\begin{bmatrix} \mathbf{x}' \\ 1 \end{bmatrix} \sim \mathcal{H} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} \quad (23)$$

where $\mathcal{H} \in \mathbb{R}^{3 \times 3}$ describes a planar projective transformation. The rigid warping models do not need regularization.

4.2. Non-Rigid Warping

The thin plate spline warp [2] is a composition of an affine transformation and a non-rigid warping:

$$\mathbf{x}' = [A \ T \ W] \begin{bmatrix} \mathbf{x} \\ 1 \\ U(\|\mathbf{x} - \hat{\mathbf{x}}^{(\cdot)}\|) \end{bmatrix} \quad (24)$$

where $A \in \mathbb{R}^{2 \times 2}$ and $T \in \mathbb{R}^{2 \times 1}$ describe the affine/rigid transformation, while the weight matrix $W \in \mathbb{R}^{2 \times K}$ and set of K control points $\hat{\mathbf{x}} = (\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_K)$ characterize the non-rigid deformation. $U(r) = r^2 \log r^2$ is the radial basis function of the spline. Constraints on the square-integrability of the second derivatives of the spline-based interpolation functions give the following additional relationship:

$$W [\hat{\mathbf{x}}^T \ \mathbf{1}] = W P = \mathbf{0} \quad (25)$$

where $\mathbf{1} \in \mathbb{R}^{K \times 1}$ and $\mathbf{0} \in \mathbb{R}^{2 \times 3}$ (matrices with respectively only one and zero entries). The unknown transformation

parameters in A, T and W can be uniquely determined if the transformed control points $\hat{\mathbf{x}}'$ are available. In this case, combining (24) and (25), one could write:

$$[\hat{\mathbf{x}}' \ \mathbf{0}] = [A \ T \ W] \begin{bmatrix} P^T \ O \\ Z \ P \end{bmatrix} \quad (26)$$

where $Z_{ij} = U(\|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j\|)$ and O is a zero matrix $\in \mathbb{R}^{3 \times 3}$. The warping model in (24) can be therefore parameterized only in terms of the transformed control points $\hat{\mathbf{x}}'$, via a linear mapping [13]:

$$\mathbf{x}' = [\hat{\mathbf{x}}' \ \mathbf{0}] \begin{bmatrix} P^T \ O \\ Z \ P \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x} \\ 1 \\ U(\|\mathbf{x} - \hat{\mathbf{x}}^{(\cdot)}\|) \end{bmatrix} \quad (27)$$

This linear relation makes extremely efficient the computation of the Jacobian of the warp with respect to the parameter vector (in this case the transformed control points) $\frac{\partial \mathbf{W}(\mathbf{x}, \theta)}{\partial \theta}$ in (21). Regularization can be obtained adding to the cost function a term proportional to the bending energy as in [27]. In this work we rather used the approach suggested in [13], consisting in replacing the matrix Z with $Z + \lambda \mathbf{I}$, where \mathbf{I} is the identity matrix. This is equivalent to adding a further constraint on the TPS smoothing. The bigger λ , the more rigid is the warping.

5. Experimental Results

In this section we present segmentation results on a variety of images to demonstrate the capabilities of the proposed model. In all the experiments dissimilarities are evaluated as L2 distances in the CIE-Lab color space. The curve evolution has been implemented using a semi-implicit finite difference scheme¹. The curvature parameter μ is chosen between 2000 and 2500.

Rigid Registration The first experiment (Fig. 2) is meant to show robustness to increasing levels of occlusion. The reference image not occluded is shown in Fig. 2(a), while in Fig. 2(c-f) we show the segmentations of the occluded views. The reliability of the recovered homography parameters is demonstrated in Table 1, where they are compared with ground truth values obtained by point correspondence with sub-pixel accuracy (\mathcal{H} is normalized so that $\theta_9 = 1$).

The second experiment (Fig. 3) consists of two different outdoor views of a tower (related by affinity), where an occlusion has been generated in the second view by putting the camera string in front of the objective. The proposed algorithm is able to recover the boundaries in both views.

The following experiment, presented in Fig. 4, highlights one of the novel assets of the proposed method. Images of a hand with increasing occlusions are displayed.

¹ H and δ are regularized as follows: $H(z) = \frac{1}{2} [1 + \frac{2}{\pi} \arctan(\frac{z}{\epsilon})]$ and $\delta(z) = \frac{1}{\pi} \frac{\epsilon}{\epsilon^2 + z^2}$

Table 1. Homography parameters for the experiments in Fig. 2. The Ground Truth is calculated using point correspondences with sub-pixel accuracy. The mean μ and standard deviation σ are evaluated using the results of the segmentation for the four different occlusion pairs ((a-c), (a-d), (a-e), (a-f)).

\mathcal{H} parameters	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
Ground Truth	0.945568	0.166809	0.000002	-0.203228	0.904125	-0.000298	52.586173	-0.061305
Proposed Method μ	0.943369	0.167061	-0.000007	-0.199886	0.905191	-0.000270	52.761128	0.160094
Proposed Method σ	0.020410	0.014676	0.000093	0.007183	0.004465	0.000050	0.517181	1.763819

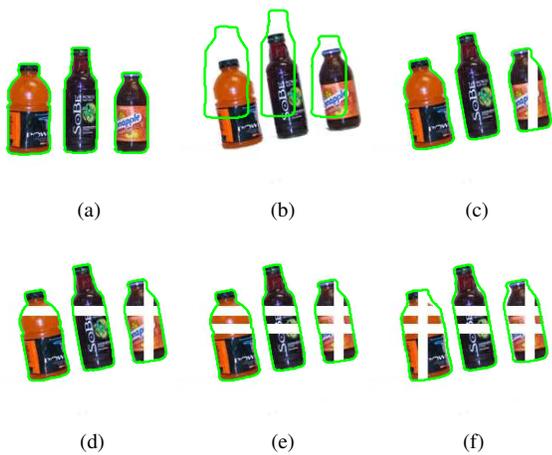


Figure 2. An image of bottles and its projective transformation with increasing level of synthetically generated occlusions. (a) Segmented reference (b) Superposition of boundaries of one view onto the second one to demonstrate misalignment. (c-f) Segmentation of the occluded views. The homography parameters are compared with the ground truth in Table 1.

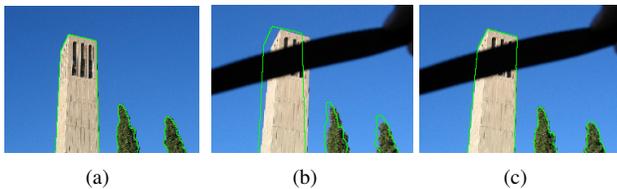


Figure 3. Images of a tower. (a) Segmented reference (b) Superposition of boundaries of one view onto the second one to demonstrate misalignment. (c) Segmentation of the occluded view.

The peculiar aspect of these images is that the occluding object has shape similar to the object of interest. The distinctive feature which can help to discriminate between the two is the color (or intensity) information. Fig. 4(b-c) demonstrate how the proposed model is capable to successfully recover the boundaries of the occluded hand, even with a significant level of occlusion. In comparison, Fig. 4(f) shows that using just a binary shape (as suggested in [23], [5], see Fig. 4(e)), which does not take into account the

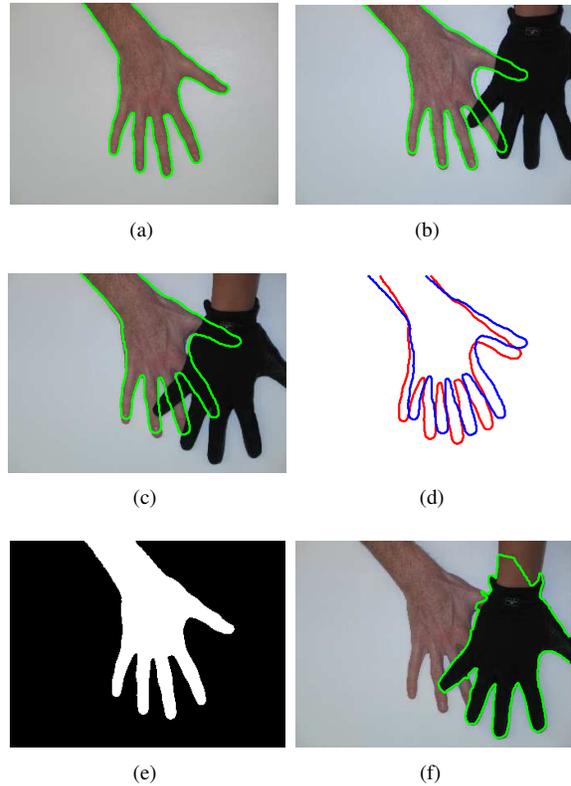


Figure 4. Images of hands. (a) Segmented reference image. (b-c) Segmented occluded views. (d) Superposition of the views (a) and (b or c) to demonstrate misalignment. (e) Binary prior used to segment the image in (f). (f) Segmentation using the Chan-Vese algorithm (vector valued version) using the binary shape prior from (e) (as in [23]). Since intensity or color information about the prior is not utilized, the algorithm selects the hand with the glove since it is darker and not occluded.

color information of the prior, leads to mis-segmentation. In this particular case, the algorithm selects the hand with the glove since it is darker and not occluded.

As a potential application of the framework introduced in this paper, we present several results on automatic segmentation and registration of aerial images. Fig. 5(a-b) show two aerial views of a pier, one of which occluded by clouds, segmented and registered using the proposed method. The

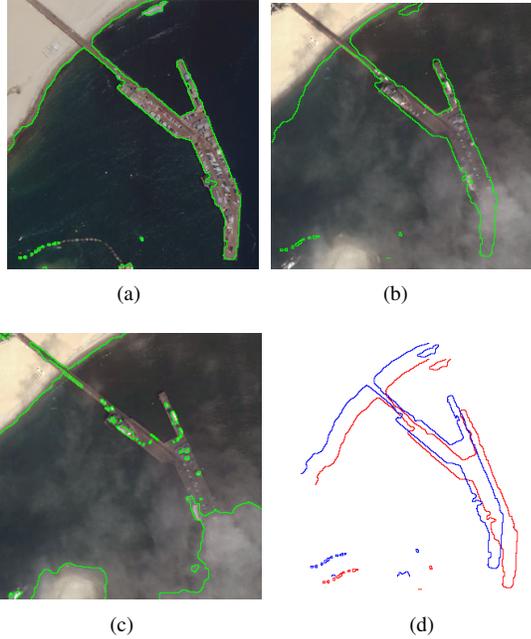


Figure 5. (a-b) Automatic segmentations and registration of two views of a pier. The second one is occluded by clouds. (c) Segmentation of the occluded view by itself. (d) Superposition of segmentation results to demonstrate misalignment.

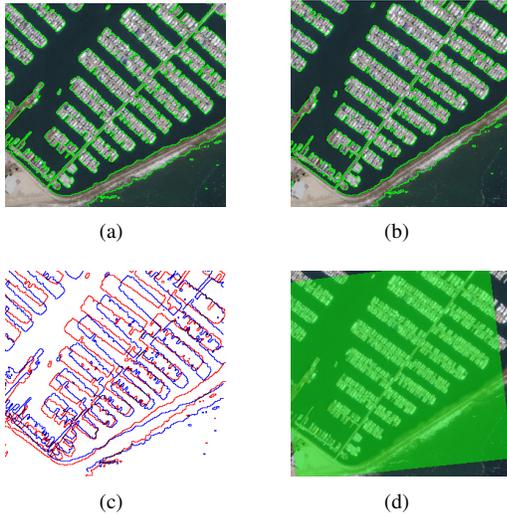


Figure 6. (a-b) Two aerial views of a harbor automatically segmented and registered (c) Superposition of the segmentation results to demonstrate misalignment. (d) Overlapping area between the two views.

segmentation of the occluded image by itself is shown in Fig. 5(c). An important consideration regarding this image pair is that feature point based registration methods are likely to produce inaccurate results. In fact, the valid point matches are arranged in an ill-conditioned configuration (along the pier line) and the homography estimate

Table 2. Homography parameters for the experiments in Fig. 6. Ground Truth is evaluated using point correspondences.

\mathcal{H} parameters	Ground Truth	Proposed Method
θ_1	0.98236679	0.98325788
θ_2	0.12069941	0.12113208
θ_3	0.00000302	0.00000447
θ_4	-0.12144333	-0.12156598
θ_5	0.97370151	0.97438497
θ_6	-0.00006800	-0.00006625
θ_7	18.53983280	18.48239430
θ_8	-25.02214483	-25.12943160

computed via the DLT (direct linear transform) algorithm is biased. This suggests possible advantages of registration using region based methods.

In the last example (Fig. 6) we present successful segmentation and registration of two aerial views of a harbor. The accuracy of the recovered homography parameters is demonstrated in Table 2, where they are compared with the ground truth, evaluated using point correspondences. To verify the accuracy we introduce the following distance measure between two transformations. Let $\mathbf{W} : (\mathbf{x}; \boldsymbol{\theta}) \mapsto \mathbf{x}' = \mathbf{W}(\mathbf{x}; \boldsymbol{\theta})$ be the warping (parametrized by $\boldsymbol{\theta}$) which maps the point \mathbf{x} to the point \mathbf{x}' . We define the distance between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ (with respect to the point configuration \mathcal{X}) as:

$$D_{\mathcal{X}}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{1}{|\mathcal{X}|} \sum_{\mathbf{x} \in \mathcal{X}} \|\mathbf{W}(\mathbf{x}; \boldsymbol{\theta}) - \mathbf{W}(\mathbf{x}; \boldsymbol{\theta}')\| \quad (28)$$

where $|\cdot|$ denotes cardinality. This tells on average how close are the mappings via \mathbf{W} using the parameters $\boldsymbol{\theta}$ or $\boldsymbol{\theta}'$. Using as \mathcal{X} the four corners of the image, for the two homographies in Table 2 we obtain $D_{\mathcal{X}} = 0.126$ pixels, which demonstrates the accuracy of the proposed method.

Non-Rigid Registration The first experiment with non-rigid deformation of the object of interest is shown in Fig. 7. The deformed view of Fig. 7(c) is synthetically generated applying a non-rigid warping to Fig. 7(a), such that the ten control points (denoted by a blue cross) are mapped onto the ten correspondent points in Fig. 7(c) (again denoted by a blue cross). The magenta crosses represent the transformed points according to the recovered warping. The distance measure defined in (28), using the ten points as the configuration \mathcal{X} , turns out to be 2.9246. The stiffness parameter λ defined in Section 4.2, as coefficient of the regularization term, plays an important role in order to accommodate for local deformation and, at the same time, preserving the information about the original shape, gaining therefore robustness to occlusion.

The second set of experiments involves real images,

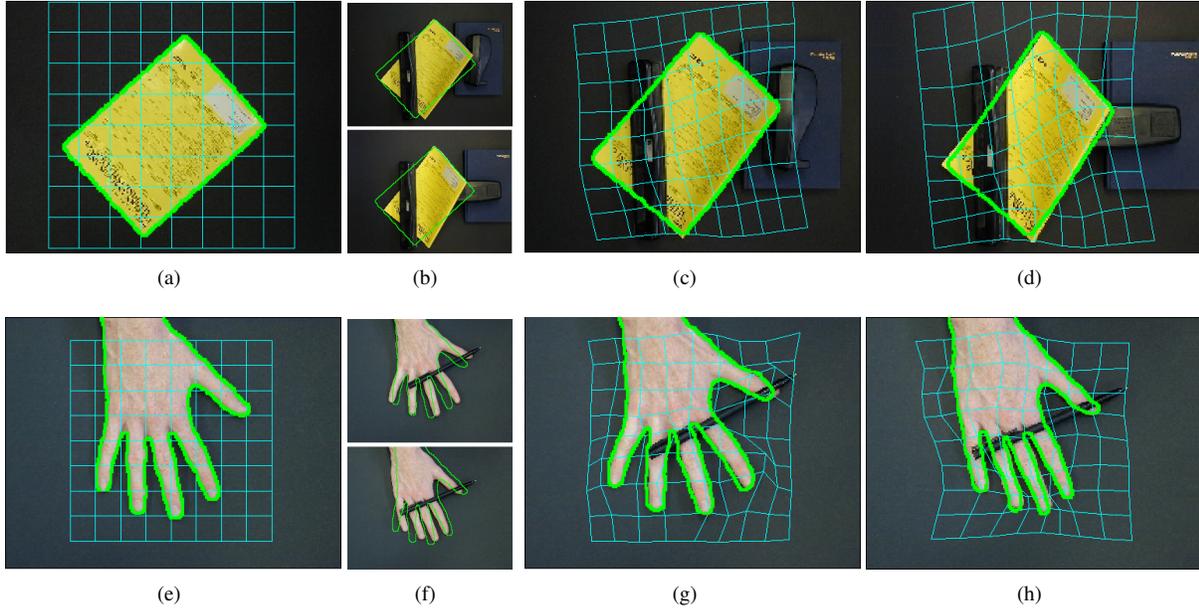


Figure 8. (a,e) Segmented reference (in cyan is the 9×9 control points grid). (b,f) Superposition of the views to demonstrate misalignment. (c-d,g-h) Successful segmentation of the occluded views (in cyan is the grid representing the estimated warp).

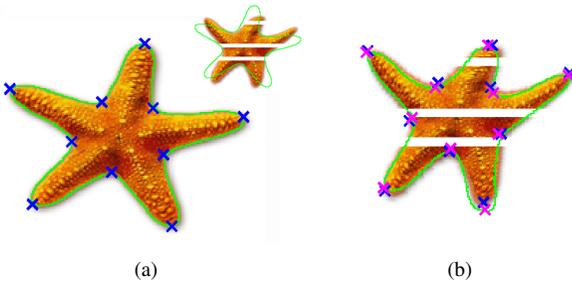


Figure 7. (a) Segmented reference. The blue crosses are the control points used to synthetically warp the image. Top right: overlapping of the two views to demonstrate misalignment. (b) Segmentation of the deformed and occluded view. The blue crosses represent the ground truth and magenta ones are estimated using the proposed algorithm.

where the object of interest is first deformed and then occluded using other objects. A grid of 9×9 control points (shown in cyan in Fig. 8(a) and Fig. 8(e)) is used. The recovered warping is depicted in Fig. 8(c-d) and Fig. 8(g-h) along with the final segmentation result.

Numerical Considerations The stability of numerical techniques utilized in our algorithm is an important issue that deserves further investigation. A consideration related to the proposed model can be made in terms of the dynamic range of the warping parameters. For the rigid warping models, the parameters have very different dynamic ranges. The steepest descent search in the parameter space can be

therefore extremely slow. A solution to this consists in pre-multiplying the transpose Jacobian of the warp $\frac{\partial \mathbf{W}(\mathbf{x}, \theta)^T}{\partial \theta}$ in (21), by a diagonal matrix. The entries of this matrix can be estimated using the reciprocal diagonal elements of the Hessian of the cost [1]. The problem does not arise in the case of TPS warping, where control points are used as transformation parameters and therefore they have the same dynamic range (pixels). This suggests possible advantages of indirect estimation of transformations, using a set of control points.

6. Conclusion

In this paper we presented a novel variational approach to the background/foreground segmentation of multiple views of the same scene. The main contributions of this work can be summarized as follows:

- We introduced variational cost functions, based on pairwise pixel dissimilarity across the different images, and we minimized these cost functions within the level set framework.
- We imposed a warping constraint between the emerging foregrounds in the different images. In this way we handled presence of occlusions, avoiding the introduction of a specific shape term in the cost function.
- The prior image is utilized in a more complete way, exploiting also intensity and color information, and not only as a binary mask.

- We introduced the use of the thin plate spline warp within the level set framework to model non-rigid deformations of the object of interest across the different views.

Acknowledgments This research was supported by NSF ITR grant #0331697.

References

- [1] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999. 7
- [2] F. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989. 4
- [3] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–80, February 1997. 1
- [4] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–77, February 2001. 1, 3
- [5] T. F. Chan and W. Zhu. Level set based shape prior segmentation. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1164–1170, 2005. 1, 3, 4, 5
- [6] Y. Chen, H. Tagare, S. Thiruvankadam, F. Huang, D. Wilson, K. Gopinath, R. Briggs, and E. Geiser. Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision*, 50(3):315–328, 2002. 1
- [7] D. Cremers. Dynamical statistical shape priors for level set based tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1262–1273, August 2006. 1
- [8] D. Cremers, N. Sochen, and C. Schnorr. Multiphase dynamic labeling for variational recognition-driven image segmentation. In Springer, editor, *European Conference on Computer Vision (ECCV)*, pages 74–86, 2004. 1
- [9] D. Cremers, F. Tischhauser, J. Weickert, and C. Schnorr. Diffusion snakes: introducing statistical shape knowledge into the mumford-shah functional. *International Journal of Computer Vision*, 50(3):295–313, 2002. 1
- [10] A. Duci, A. Yezzi, S. Mitter, and S. Soatto. Region matching with missing parts. In *European Conference on Computer Vision (ECCV)*, volume 3, pages 48–64, 2002. 1
- [11] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, pages 321–31, 1988. 1
- [12] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 316–323, 2000. 1
- [13] J. Lim and M.-H. Yang. A direct method for modeling non-rigid motion with thin plate spline. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 1196–1202, 2005. 4
- [14] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22–6, 1985. 1
- [15] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Mathematics*, 42:577–685, 1989. 1
- [16] S. Osher and R. Fedkiw. Level set methods: an overview and some recent results. *Journal of Computational Physics*, 169(2):463–502, 2001. 3
- [17] 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. 1, 3
- [18] T. R. Raviv, N. Kiryati, and N. Sochen. Unlvelsets: Geometry and prior-based segmentation. In *European Conference on Computer Vision (ECCV)*, pages 50–61, 2004. 1, 4
- [19] T. R. Raviv, N. Kiryati, and N. Sochen. Prior-based segmentation and shape registration in the presence of projective distortion. *International Journal of Computer Vision (IJCV)*, *Accepted for publication*, 2006. 1, 3, 4
- [20] T. R. Raviv, N. Sochen, and N. Kiryati. Mutual segmentation with level sets. In *Proc. IEEE CVPR Workshop on Perceptual Organization in Computer Vision (POCV)*, 2006. 1, 3
- [21] M. Rousson and N. Paragios. Shape priors for level set representation. In *European Conference on Computer Vision (ECCV)*, pages 78–92, 2002. 1
- [22] B. Sumengen and B. Manjunath. Graph partitioning active contours (GPAC) for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):509–521, April 2006. 2
- [23] N. K. T. Riklin Raviv and N. Sochen. Prior-based segmentation by projective registration and level sets. In *IEEE International Conf. on Computer Vision (ICCV)*, pages pp 204–211, October 2005. 1, 3, 4, 5
- [24] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. Grimson, and A. Willsky. A shape based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging*, 22(2):137–154, 2003. 1
- [25] A. Tsai, A. J. Yezzi, and A. S. Willsky. Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Transactions on Image Processing*, pages 1169–86, August 2001. 1
- [26] G. Unal and G. Slabaugh. Coupled pdes for non-rigid registration and segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages pp. 168–175, 2005. 1, 4
- [27] A. Vedaldi and S. Soatto. Local features, all grown up. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1753–1760, 2006. 4
- [28] A. Yezzi, L. Zollei, and T. Kapur. A variational framework to integrate segmentation and registration through active contours. *Medical Image Analysis*, 7:171–185, 2003. 1, 4
- [29] A. J. Yezzi, A. Tsai, and A. Willsky. A statistical approach to snakes for bimodal and trimodal imagery. In *International Conference on Computer Vision (ICCV)*, pages 898–903, 1999. 1
- [30] S. C. Zhu and A. Yuille. Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 884–900, September 1996. 1