SEGMENTING MULTIPLE FAMILIAR OBJECTS UNDER MUTUAL OCCLUSION

Qilong Zhang and Robert Pless

Department of Computer Science & Engineering, Washington University
One Brookings Drv. St. Louis, MO, 63130, USA

ABSTRACT

We address the problem of segmenting multiple similar objects by optimizing a Chan-Vese-like [1] functional with respect to a mixture of level set functions. We solve the variational formulation under this model allowing for similarity transforms. This allows shape priors to be enforced even in the presence of mutual occlusion, lifting the limitation in [2]. We show numerical results on example images to demonstrate the promise of our approach.

Index Terms— image segmentation, variational methods, shape priors, level set methods, mutual occlusion

1. INTRODUCTION

A fundamental problem in image processing is segmenting an image into regions and their boundaries. The level set method [3] has become an important tool in this task, because it is able to naturally present shapes with complex boundaries and topologies. In recent years, level set methods have been extended to enforce priors on the extracted shapes, which are particularly important in medical image segmentation. This paper extends this to the common cases where there are multiple known objects and there is mutual occlusion.

An important variational approach underlying level set-based segmentation method is derived from the Mumford-Shah functional [4], in which image segmentation is posed as finding an optimal piecewise smooth approximation of the given image and a set of boundaries with minimal length between contiguous regions. Since the seminal work of Osher and Sethian [3], level set-based active contours [5, 6, 7] have been increasingly popular in image segmentation. A level set implementation of piecewise constant Mumford-Shah functional was proposed in [1]. These models all require that some low level features distinguish the region of interest from the background, and often use edges consistency or homogeneity of intensity, color, texture or motion. They may fail to segment meaningful objects from images in the presence of missing or misleading information due to noise, clutter and occlusion. Searching for objects whose shape is known a priori, what we call familiar objects, can be improved by incorporating the prior on the shape model into the level set frameworks [8, 9, 10, 11] as additional shape-driven term.

A limitation of these approaches is that they introduce the shape priors into level set framework in such a way that only *one* object can be segmented per image. They do not permit the simultaneous segmentation of multiple independent familiar objects, except for [2] which explicitly labels exclusive image regions for each independent object, therefore fails if objects overlap. In this context, our goal is to segment all objects of familiar shape simultaneously even in the presence of mutual occlusion. This work is, to the best of our knowledge, the first to apply a shape prior to extracting multiple, potentially overlapping shapes in the scene.

The idea of this work is to solve for a level set as a mixture of basis functions. Using multiple basis functions allows the extraction of multiple objects; constraining each basis function to be consistent with a shape prior improves the extraction of familiar objects, and an extra coupling term is added between mixtures to keep the basis components from evolving to be identical with each other.

The remainder of this paper is organized as follows: in Section 2, we review the level set formulation of the Mumford-Shah functional proposed by Chan and Vese [1], as well as the segmentation model integrating shape priors. Section 3 presents our main contribution. Section 4 demonstrates our experimental results. Conclusions and future work are presented in Section 5.

2. BACKGROUNDS

2.1. Region-based Segmentation with Level Sets

The basic idea of the level set method is to implicitly represent contours C in the image plane as the zero-level of a Lipschitz function $\phi : \Omega \rightarrow R$.

In [1], Chan and Vese proposed a level set-based formulation of the Mumford-Shah functional. In particular, a two-phase segmentation of an image f can be generated by minimizing the following functional:

$$E^{CV}(c_1, c_2, \phi) = \lambda_1 \int_\Omega |f - c_1|^2 H_\phi \, d\mathbf{x} + \lambda_2 \int_\Omega |f - c_2|^2 (1 - H_\phi) d\mathbf{x} + \mu \int_\Omega |\nabla H_\phi| d\mathbf{x} \quad (1)$$

where H_ϕ denotes the Heaviside step function of ϕ, and scalar variables c_1 and c_2 model the intensity values of the image regions inside and outside C respectively, and may be updated as the estimate of the contour C is updated. The first two terms enforce a minimal intensity variance in the segmented regions, and the last term penalizes a large length of the separating boundary.

Using the calculus of variations, one can recover the following evolution equation which incorporates an artificial time parameter t and converges to minimize $E^{CV}(\phi)$:

$$\frac{\partial \phi}{\partial t} = \delta_\phi \left[\mu \nabla \cdot \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \lambda_1 (f - c_1)^2 + \lambda_2 (f - c_2)^2 \right] \quad (2)$$

where δ_ϕ is the dirac delta function of ϕ. The scalars c_1 and c_2 are updated in alternation with the level set evolution to take on the mean intensity values of the input image f with the regions $\phi > 0$ and $\phi < 0$, respectively.

2.2. Incorporating Shape Priors

In many applications of image segmentation, some prior knowledge about the shape of the expected objects is available. A straightforward incorporation of the shape prior in the Chan-Vese segmentation model can be generally formalized as a modification to the Chan-Vese energy functional:

$$E(c_1, c_2, \phi) = E^{CV}(c_1, c_2, \phi) + \nu E_{shape}(\phi) \quad (3)$$

where $\nu > 0$ is a weighting parameter which determines the influence of the shape prior, and E_{shape} is the shape constraint energy that restricts the space of possible shapes to segment. Specifically, E_{shape} penalizes the dissimilarity between the shape embodied by the evolving level set function ϕ and the shape prior.

The shape prior can be derived from a single or a collection of reference shapes, and is implicitly represented by a signed distance function [12, 9] . For simplicity, we will only consider a single training shape in our following presentation. Nevertheless, our proposed model can be easily extended to more involved statistical shape priors of the form given in [8, 11].

Even when the shape of objects of interest is known, often their scales and poses are unknown. We can encode the similarity transform $A = (s, \theta, T)$ of the shape prior ψ_0 as:

$$\begin{aligned} A(\mathbf{x}) &= s \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \mathbf{x} + T, \\ \psi(\mathbf{x}) &= \frac{1}{s} \psi_0(A(\mathbf{x})), \ \forall \mathbf{x} \in \Omega \end{aligned} \quad (4)$$

where $s > 0$ is scaling factor, θ represents the rotation angle between shapes, and T is the displacement. Accordingly, the E_{shape} can be simply formulated as follows:

$$E_{shape}(\phi, \psi) = \int_\Omega (\phi(\mathbf{x}) - \psi(\mathbf{x}))^2 d\mathbf{x} \quad (5)$$

Since ψ is related to ψ_0 by some similarity transformation A, we may also write E_{shape} in terms of ψ_0 with unknown variables (s, θ, T).

The above model introduces a shape prior in such a way that only objects of interest similar to the shape prior can be recovered, and all unfamiliar image structures are suppressed. Therefore, it enables recovery of the preferred object in the presence of large image artifacts. However, this formulation solves for a single level set consistent with the shape prior. If there are several objects of such shape in the scene, this model finds at most *one*, and may not find any.

To simultaneously segment an image consisting of multiple objects of a familiar shape, Cremers et al. [2] introduced a multiphase dynamic labelling scheme, in which multiple preferred objects can be recovered by jointly generating a segmentation (by a level set function) and a recognition-driven partition of the image domain (by a vector-valued labelling function) which indicates where to enforce certain shape priors. However, this approach restricts each pixel to be associated with only one object. Therefore, in cases of mutual occlusion, this approach will fail to recover all familiar objects. In the next section, we cope with this limitation by presenting a novel method in a way which permits the simultaneous segmentation of multiple independent familiar objects even in the presence of mutual occlusion.

3. OUR METHOD

Now we consider a given image consisting of multiple objects $\{\mathcal{O}_1, \mathcal{O}_2, \cdots, \mathcal{O}_n\}$ of familiar shape. Instead of partitioning image domain into mutual exclusive regions, we allow each pixel to be associated with multiple objects or the background. Specifically, we try to find a set of characteristic functions $\{\chi_i\}$ such that:

$$\chi_i(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \mathcal{O}_i; \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

To define $\{\chi_i\}$, we associate one level set per object in such a way objects are allowed to overlap with each other within the image. These level set components may both be positive on the area of overlap, and enforce the prior on the shapes of objects extracted from the image. We first consider the case of segmenting two objects within an input image, then we generalize to simultaneous segmentation of n independent familiar objects.

Two familiar objects. Suppose we are given an image f with two familiar objects, and for simplicity, assume that these are consistent with the same shape prior ψ_0 and share a similar intensity value. Then simultaneous segmentation of two familiar objects with respect to the given shape prior is solved by minimizing the following energy functional:

$$
\begin{aligned}
E(c1, c2, \Phi, \Psi) &= \lambda_1 \int_\Omega (f - c_1)^2 H_{\chi_1 \vee \chi_2}\, d\mathbf{x} \\
&+ \lambda_2 \int_\Omega (f - c_2)^2 (1 - H_{\chi_1 \vee \chi_2})\, d\mathbf{x} \\
&+ \mu \sum_{i=1}^{2} \int_\Omega |\nabla H_{\phi_i}| d\mathbf{x} + \omega \int_\Omega H_{\chi_1 \wedge \chi_2}\, d\mathbf{x} \\
&+ \nu \sum_{i=1}^{2} \int_\Omega (\phi_i - \psi_i)^2 d\mathbf{x} \qquad (7)
\end{aligned}
$$

with

$$
H_{\chi_1 \vee \chi_2} = H_{\phi_1} + H_{\phi_2} - H_{\phi_1} H_{\phi_2},\ H_{\chi_1 \wedge \chi_2} = H_{\phi_1} H_{\phi_2}
$$

where $\Phi = (\phi_1, \phi_2)$ and $\Psi = (\psi_1, \psi_2)$. The fourth term penalizes the overlapping area between the two segmenting regions, and it prevents the two evolving level set functions ϕ_1 and ϕ_2 from becoming identical.

Minimizing the energy functional (7) alternatingly with respective to the dynamic variables, yields the associated Euler-Lagrange equations, parameterizing the decent direction by an artificial time $t > 0$ as follows:

$$
\begin{aligned}
\frac{\partial \phi_i}{\partial t} &= \delta_{\phi_i} [\left(-\lambda_1 (f - c_1)^2 + \lambda_2 (f - c_2)^2\right) (1 - H_{\phi_j}) \\
&\quad + \mu \nabla \cdot \left(\frac{\nabla \phi_i}{|\nabla \phi_i|}\right) - \omega H_{\phi_j}] - 2\nu (\phi_i - \psi_i), \\
\frac{\partial \theta_i}{\partial t} &= 2\nu \int_\Omega (\phi_i - \psi_i)(\nabla \psi_i \cdot \nabla_\theta A_i)\, d\mathbf{x}, \\
\frac{\partial T_i}{\partial t} &= 2\nu \int_\Omega (\phi_i - \psi_i)(\nabla \psi_i \cdot \nabla_T A_i)\, d\mathbf{x}, \\
\frac{\partial s_i}{\partial t} &= 2\nu \int_\Omega (\phi_i - \psi_i)(-\frac{\psi_i}{s} + \nabla \psi_i \cdot \nabla_s A_i)\, d\mathbf{x}, \\
&\quad i, j \in \{1, 2\},\ \ i \neq j \qquad (8)
\end{aligned}
$$

where $\psi_i = \psi_0(A_i(\mathbf{x}))/s_i$. Similar to the Chan-Vese model, we update c_1, c_2 for each iteration as follows:

$$
c_1 = \frac{\int_\Omega f H_{\chi_1 \vee \chi_2}\, d\mathbf{x}}{\int_\Omega H_{\chi_1 \vee \chi_2}\, d\mathbf{x}},\ \ c_2 = \frac{\int_\Omega f (1 - H_{\chi_1 \vee \chi_2})\, d\mathbf{x}}{\int_\Omega (1 - H_{\chi_1 \vee \chi_2})\, d\mathbf{x}} \qquad (9)
$$

General cases of $n > 2$. Our proposed method can be generalized for simultaneous segmentation of $n > 2$ independent objects familiar to the shape prior ψ_0. The energy in this more general case will be:

$$
\begin{aligned}
E(c1, c2, \Phi, \Psi) &= \lambda_1 \int_\Omega (f - c_1)^2 H_{\vee_{i=1}^{n} \chi_i}\, d\mathbf{x} \\
&+ \lambda_2 \int_\Omega (f - c_2)^2 (1 - H_{\vee_{i=1}^{n} \chi_i})\, d\mathbf{x} \\
&+ \mu \sum_{i=1}^{n} \int_\Omega |\nabla H_{\phi_i}| d\mathbf{x} + \omega \sum_{i \neq j} \int_\Omega H_{\chi_i \wedge \chi_j}\, d\mathbf{x} \\
&+ \nu \sum_{i=1}^{n} \int_\Omega (\phi_i - \psi_i)^2 d\mathbf{x} \qquad (10)
\end{aligned}
$$

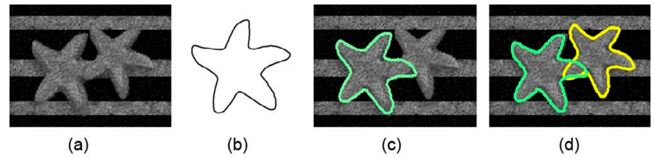

Fig. 1. Starfish example. (a) an image consisting of two starfish with horizontal strips. (b) the shape prior. (c) the result from the method proposed in [2]. (d) our method: the final segmented contour for each starfish is illustrated as yellow and green respectively.

where $\Phi = \{\phi_i\}$ and $\Psi = \{\psi_i\}$. Minimizing the functional (10) with respect to the unknowns can be obtained in the same fashion as in Equation (8) and (9).

A multi-resolution approach. When allowing the shape prior to be invariant to similarity transforms, the computational complexity may be high. A coarse-to-fine approach [13], which begins the process on a low resolution image, allows for an efficient implementation. The algorithm may be summarized as three steps: (1) initialize the input image and shape prior at the lowest resolution; (2) at the current resolution level, do gradient decent optimization of the functional (10) with respect to the unknowns c_1, c_2, Φ and Ψ , for a defined number of iterations or until convergence; (3) increase image resolution as well as the resolution of evolving level sets, and repeat the previous until the finest level of resolutions is done.

4. EXPERIMENTAL RESULTS

In this section we present the experimental results from our proposed segmentation model on various synthetic and real images. We use through our experiments the following parameter settings: $\lambda_1 = 1$, $\lambda_2 = 2$, $\mu = 0.2$, $\omega = 0.5$ and $\nu = 0.5$ unless otherwise stated. The automatic selection of optimal parameters is under our future investigation.

Consider the image shown in Fig. 1.a of two overlapping starfish cluttering with horizontal strips. Only the contour in Fig. 1.b is known in advance and used as shape prior. Fig. 1.c shows the result obtained with the model [2] for two known objects. Due to the mutual occlusion, the starfish favored by the image data is correctly segmented, while the other one is suppressed by the competition process of dynamic labelling. Our result is demonstrated Fig. 1.d, in which the final segmented contours for both starfish are drawn as yellow and green respectively. This demonstrates that our proposed model can correctly recover from background clutter both starfish which occlude one another. Figure (2) shows our experiment on segmenting jets on formation flying. In this case, three jets are present in the scene and perceptually share a same shape which is known in advance. Note almost half part of the jet in the middle is occluded by the front one, which makes the segmentation task very challenging. Our model successfully extracted and reconstructed all three jets from the image.

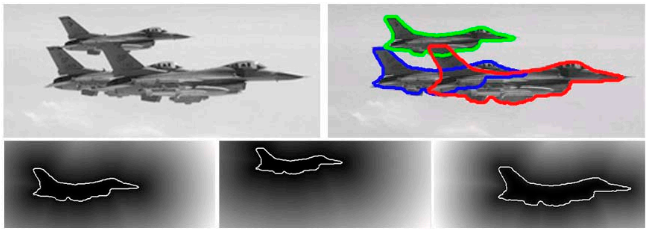

Fig. 2. Jets example. The top row shows an image of three jets on formation flying and the corresponding segmentation results from our model. On the bottom are respectively the plots of the final level set functions $\phi_{1\sim3}$ with zero iso-contour overlaid.

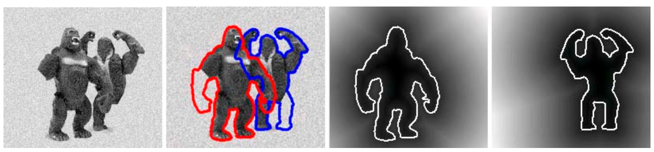

Fig. 3. Missing parts example. The left two frames shows the input image of two overlapping toys with missing parts and the segmentation results from our proposed model. The right two frames illustrate respectively the final level set functions $\phi_{1\sim2}$ with zero iso-contour overlaid.

The last example as in Fig. (3) demonstrates that our method successfully recovered two objects with different shape priors. The image contains two overlapping toys with missing parts. The simultaneous segmentation of both toys is obtained by extending our model to integrate multiple competing shape priors. The toy on the right is correctly reconstructed despite prominent occlusion by the one in front.

In the experiments, we found that our method becomes more sensitive to the initial contour position on more complicated images. This is due to the non-convexity of the Mumford-Shah functional from which our model stems. Another reason is that our similarity transform optimization is also a local optimization scheme. Moreover, the complexity becomes high when the number of objects increases. Improving these is a subject of our further investigations.

5. CONCLUSION

In this paper, we present a novel level set based variational model for simultaneous segmentation of multiple similar objects using shape prior. In contrast to existing shape prior segmentation models, our method permits the reconstruction of familiar objects even they partially *occlude* each other. In addition, invariance of the shape prior with respect to similarity transformations of the level set function is also incorporated. Experimental results confirm that our algorithm converges empirically even for fairly large object overlaps and substantial transformations in the presence of significant images artifacts.

6. REFERENCES

[1] T. Chan and L. Vese, "An active contour model without edges," in *Proc. Int. Conf. Scale-Space Theories in Comput. Vision*. 1999, pp. 141–151, Springer-Verlag.

[2] D. Cremers, N. Sochen, and C. Schnorr, "A multiphase dynamic labeling model for variational recognition-driven image segmentation," *Int. J. Comput. Vision*, vol. 66, no. 1, pp. 67–81, January 2006.

[3] S. Osher and J. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations," *J. Computational Physics*, vol. 79, pp. 12–49, 1988.

[4] D. Mumford and J. Shah, "Optimal approximations of piecewise smooth functions and associated variational problems," *Comm. on Pure and Appl. Mathematics*, pp. 577–685, 1989.

[5] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vision*, vol. 22, no. 1, pp. 61–79, 1997.

[6] N. Paragios and R. Deriche, "Geodesic active regions and level set methods for supervised texture segmentation," *Int. J. Comput. Vision*, vol. 46, no. 3, pp. 223–247, 2002.

[7] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi, "Gradient flows and geometric active contour models," in *Proc. Int. Conf. Comput. Vision*, 1995, pp. 810–815.

[8] M. Leventon, E. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *Proc. Conf. Comput. Vision and Pattern Recognition*, 2000, pp. 1316–1323.

[9] M. Rousson and N. Paragios, "Shape priors for level set representations.," in *Proc. European Conf. Comput. Vision*, 2002, pp. 78–92.

[10] Y. Chen and et al., "Using prior shapes in geometric active contours in a variational framework.," *Int. J. Comput. Vision*, vol. 50, no. 3, pp. 315–328, 2002.

[11] D. Cremers and et al., "Diffusion snakes: Introducing statistical shape knowledge into the mumford-shah functional.," *Int. J. Comput. Vision*, vol. 50, no. 3, pp. 295–313, 2002.

[12] N. Paragios, M. Rousson, and V. Ramesh, "Matching distance functions: A shape-to-area variational approach for global-to-local registration," in *Proc. European Conf. Comput. Vision*. 2002, pp. 775–789, Springer-Verlag.

[13] M. Corvi and G. Nicchiotti, "Multiresolution image registration," in *Proc. Int. Conf. Image Processing (Vol. 3)*, Washington, DC, USA, 1995, p. 3224, IEEE Computer Society.