A VARIATIONAL FRAMEWORK FOR PARTIALLY OCCLUDED IMAGE SEGMENTATION USING COARSE TO FINE SHAPE ALIGNMENT AND SEMI-PARAMETRIC DENSITY APPROXIMATION

Lin Yang

Electrical and Computer Engineering
Rutgers University
Piscataway, NJ 08854

*David J. Foran**

Center of Biomedical Imaging and Informatics
The Cancer Institute of New Jersey
UMDNJ-Robert Wood Johnson Medical School
Piscataway, NJ 08854

ABSTRACT

In this paper, we propose a variational framework which combines top-down and bottom-up information to address the challenge of partially occluded image segmentation. The algorithm applies shape priors and divides shape learning into shape mode clustering and non-rigid transformation estimation to handle intraclass and interclass coarse to fine variations. A semi-parametric density approximation using adaptive meanshift and L_2E robust estimation is used to model the likelihood. A set of real images is used to show the good performance of the algorithm.

Index Terms— Image Segmentation, Shape Modeling, Density Approximation

1. INTRODUCTION

Active contour models [1] have been widely used since their introduction. Currently many efforts are focused on object-based segmentation using shape priors. Leventon et. al [2] applied principle component analysis (PCA) to obtain training shape modes and presented them as signed functions. Dambreville et. al [3] applied kernel PCA for shape learning in geodesic active contour models. Cremers et. al [4] introduced shape priors into the level-set formulation, where the transformation and rotation parameters depend upon the level-set function ϕ. Shape symmetry information has been used in the level-set framework in [5] on symmetric object segmentation.

In this paper, we exploit the idea of introducing shape priors into active contour models but using a coarse to fine shape learning approach. We apply the algorithm to address the challenge of partially occluded image segmentation. The deformable model we present is based on the Bayesian rule and numerically implemented using level-set. Adaptive meanshift clustering [6] and L_2E robust estimation [7] are used to perform likelihood density approximation. The Elliptical Fourier Descriptor (EFD) [8] is applied on the training shapes followed by an agglomerate clustering to determine the shape modes. Non-rigid transformation estimation is used to align the shape mode to the edge of the testing image's likelihood map.

What distinguishes our work from previous ones are: we propose 1) *A novel coarse to fine shape model*, which divides the shape learning into EFD shape clustering and non-rigid transformation estimation. The *coarse* intraclass variations are captured by the learned shape modes while the *fine* interclass variations are addressed by the non-rigid transformation estimation; 2) *A novel semi-parametric density approximation method*. Considering the trade-off between accuracy and storage, we apply adaptive mean-shift and L_2E robust estimation for density approximation. The likelihoods and the shape priors are tightly coupled into the level-set framework which provides smooth boundaries and can handle topological changes by its nature.

2. FEATURES AND SHAPE CLUSTERING

2.1. Features

The Luv colors and textures are used to describe the bottom-up information which is extracted using a set of linear filters - filter banks. We modified the MR filter bank [9] to compute the filter response. The feature vector is composed using two LoG filter responses on the L channel with $\sigma = 1, 2$, six Gaussian filtering responses on the L, u and v channels with $\sigma = 1, 2$ and the MR maximal bar and edge response within six different directions, $\theta = 0, \pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6$, with $\sigma = 1, 2$ and 4. In total, each image pixel is represented by a 10 dimensional feature vector. Figure 1 shows the filtering results after applying the proposed filter bank to one testing image.

*This research was funded, in part, by grants from the NIH through contracts 5R01LM007455-03 from the National Library of Medicine and 1R01EB003587-01A2 from the National Institute on Biomedical Imaging and Bioengineering.

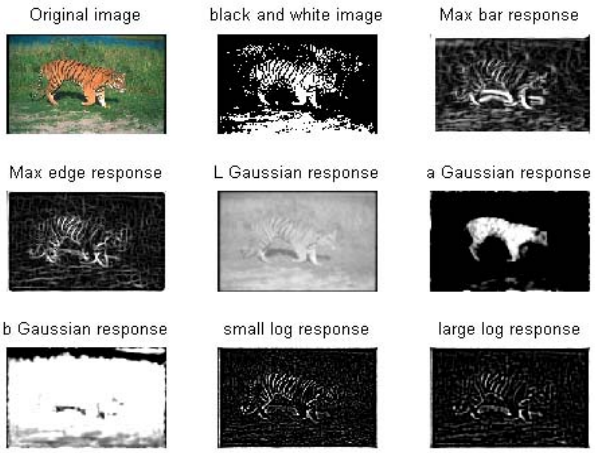

Fig. 1. The filtering response using our modified MR filter bank.

2.2. Elliptic Fourier Descriptor (EFD) Based Shape Clustering

For each human-delineated contour, we apply Elliptic Fourier Descriptor (EFD) [8] to calculate the Fourier coefficients and use the first eight normalized harmonics. Each harmonic contains four coefficients. Agglomerative clustering was applied on the set of the coefficients to determine the cluster centers, which best represent the shape modes. The advantages of using the EFD are: 1) The EFD is a good measure for global properties of shapes, which is preferred in our algorithm because small interclass variations are handled by non-rigid transformation estimation. This is explained in detail in *section 3.3*; 2) The normalized EFD descriptions are invariant to rotation, translation and scaling. Fig 2 shows the six shape modes of tigers reconstructed using $8 * 4$ EFD coefficients.

3. SEGMENTATION FRAMEWORK

3.1. Bayesian Segmentation Model

Define $\mathbf{f} \in \mathbf{F}$ in R^d as the random variable describing the feature vector, where $\mathbf{F}$ is the set of all features and d is the dimension of the feature space. In order to build the connection between the image partition and contours, we define a partitioning operator ϑ with $\vartheta(I)$ describing the partition of image I. According to Bayesian rule, the segmentation can be modeled as the maximum a posterior (MAP) estimation

$$p(\vartheta(I)|f) \propto p(f|\vartheta(I))p(\vartheta(I)) \tag{1}$$

where $p(f|\vartheta(I))$ is the likelihood of bottom-up features f given $\vartheta(I)$. The $p(\vartheta(I))$ is the prior probability of the image partition which is used to model the shape priors. The final $\vartheta(I)$ can be solved by maximizing the posterior probability $p(\vartheta(I)|f)$.

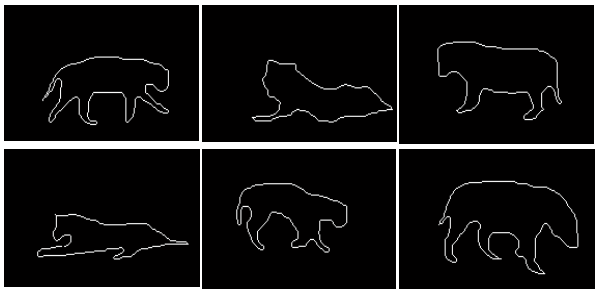

Fig. 2. Some "ground-truth" contour examples of tiger images.

3.2. Kernel Based Density Approximation Using Adaptive Mean Shift and L_2E Robust Estimation

In order to calculate the posterior probability $p(\vartheta(I)|f)$, we need to compute the likelihood $p(f|\vartheta(I))$. Given the likelihood $p(f|\vartheta(I))$ to be an arbitrary density function, kernel density estimation (KDE) is often used. The flexibility to represent any complicated probability density is KDE's major advantage. On the other hand, the high memory usage and computational complexity limit its practical use.

In this paper, we propose to model the likelihood $p(f|\vartheta(I))$ using a semi-nonparametric approximation which is the linear combination of M Gaussian distributions. We show that density approximation using adaptive mean-shift and L_2E robust estimation performs almost as well as KDE but requires far fewer parameters. The semi-nonparametric Gaussian mixture model (GMM) is distinguished from the normal GMM because the mean, variance, weights and number of Gaussian distributions are not known in advance. The likelihood is

$$p(\mathbf{f}|\vartheta(I)) = \sum_{i=1}^{N} \frac{\alpha_i \exp\left(-\frac{1}{2}(\mathbf{f}-\mathbf{f}_i)\mathbf{\Sigma}_i^{-1}(\mathbf{f}-\mathbf{f}_i)'\right)}{2^d \pi^{d/2} \left|\widehat{\mathbf{\Sigma}}_i\right|^{1/2}} \tag{2}$$

where $i = 0, ..., N$ denotes the total number of points in the image. The α_i is the weight of the ith Gaussian, $Normal(\mathbf{f}_i, \mathbf{\Sigma}_i)$. The sum of α_i is equal to one. Assume that after applying adaptive mean-shift based mode detection, there is totally M unique modes with weights $\widehat{\alpha}_i$ as the proportion of the number of points in the ith cluster, where $\widehat{\mathbf{f}}_i$ is the stationary mode point. The L_2E robust estimation can be applied to estimate the variance $\widehat{\mathbf{\Sigma}}_i$

$$L_2E(\widehat{\theta}_i) = \arg \min_{\widehat{\theta}_i = [\widehat{\mathbf{f}}_i, \widehat{\mathbf{\Sigma}}_i]} \left[\varphi(\widehat{\theta}_i)\right] \tag{3}$$

where $\varphi(\widehat{\theta}_i)$ is

$$\frac{m - 2^{d+1}\pi^{d/2} \left|\widehat{\mathbf{\Sigma}}_i\right|^{1/2} \sum_{j=1}^{m} \phi(\mathbf{f}_j|\widehat{\mathbf{f}}_i, \widehat{\mathbf{\Sigma}}_i)}{2^d m \pi^{d/2} \left|\widehat{\mathbf{\Sigma}}_i\right|^{1/2}} \tag{4}$$

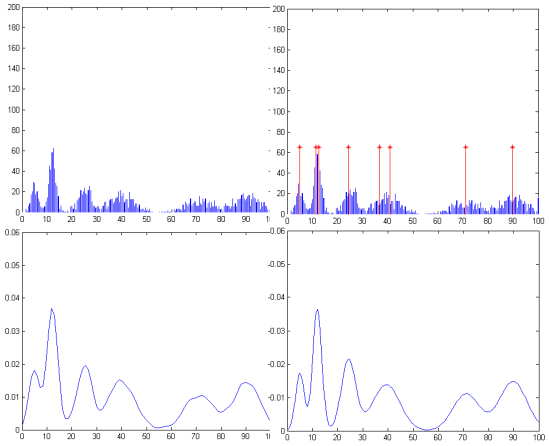

Fig. 3. The standard KDE and our proposed semi-parametric density approximation results upon a six Gaussian mixture density.

where $j = 1...m$ is the points in the ith cluster and $\phi(\mathbf{f}_j|\widehat{\mathbf{f}}_i, \widehat{\mathbf{\Sigma}}_i)$ is defined as

$$\phi(\mathbf{f}_j|\widehat{\mathbf{f}}_i, \widehat{\mathbf{\Sigma}}_i) = \frac{\exp\left(-\frac{\left(-\frac{1}{2}(\mathbf{f}-\widehat{\mathbf{f}}_i)\widehat{\mathbf{\Sigma}}_i^{-1}(\mathbf{f}-\widehat{\mathbf{f}}_i)'\right)}{\sqrt{(2\pi)^{d/2}\left|\widehat{\mathbf{\Sigma}}_i\right|}}\right)}{2^d\pi^{d/2}\left|\widehat{\mathbf{\Sigma}}_i\right|^{1/2}} \quad (5)$$

The final density approximation can be written as

$$p(\mathbf{f}|\vartheta(I)) = \sum_{i=1}^{M} \frac{\widehat{\alpha}_i \exp\left(-\frac{1}{2}(\mathbf{f}-\widehat{\mathbf{f}}_i)\widehat{\mathbf{\Sigma}}_i^{-1}(\mathbf{f}-\widehat{\mathbf{f}}_i)'\right)}{2^d\pi^{d/2}\left|\widehat{\mathbf{\Sigma}}_i\right|^{1/2}} \quad (6)$$

In stead of keeping all the N sets of parameters in (2), the likelihood $p(\mathbf{f}|\vartheta(I))$ is modeled using M sets of parameters and $M \ll N$. We test our density approximation method on a range of densities and it provides similar *pdf* as the KDE result using much less storage space. Fig 3 shows an example. The original six mixture Gaussian distributions and the modes detected by the adaptive mean-shift algorithm are shown in upper-left and upper-right of Fig 3. The bottom-left is the standard KDE using Epanetchnikov kernel. The bottom-right illustrates the our semi-parametric density approximation. The testing image's likelihood maps are generated using a Bayesian classifier over the learned *pdf* calculated using (6).

3.3. Shape Alignment

Based on EFD clustering, we can calculate the shape modes for each object class. However, the shape modes shown in Fig 2 can not be used directly as shape priors in the level set framework without alignment. Therefore, a mapping T which aligns the shape mode to the edge of the testing image's likelihood map has to be computed.

In order to determine the transformation T, gradient decent is used during each iteration of the level set function in [10] and on the nested Euler-Lagrange equations in [4]. One common problem of these methods is their assumptions of affine or even rigid mapping T. Instead, we generalize T to be the thin plate spline (TPS) transformation [11]. Affine transformation has proven to be a special case of TPS. Given two point sets, the TPS is estimated by minimizing

$$E_{TPS} = \sum_i \|T(w_i) - v_i\|^2 + \lambda I_f \quad (7)$$

where

$$I_f = \iint \left[\left(\frac{\partial^2 f}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 f}{\partial y^2}\right)^2\right] dxdy \quad (8)$$

with w_i denote the points on the training shape modes and v_i denote the points on the edges of the calculated testing image's likelihood map.

Because T is a nonrigid mapping which has infinite number of solutions for the first term in 7, the smoothness term λI_f is used to regularize the ambiguity. Each shape mode is matched to the edges of the testing image's likelihood map by minimizing (7). The final mapping T is computed using the shape mode with the minimal matching error.

3.4. The Variational Framework

Define C as the contour to separate the object from the background, $p(C)$ describes the shape prior of the object. We calculate the non-rigid transforms T with the minimal matching error and use $T(C)$ to denote the aligned shape prior. Define $\vartheta(I_{x,y})$ as the labeling of the xth column and yth row of the original image and Ω_1/Ω_2 as the region in which $I_{x,y}$ locates inside/outside $T(C)$. Denote label $\cdot^o$ as the object and $\cdot^b$as background and $pos_{x,y}$ as the coordinates. Because the pixels inside the aligned shape $T(C)$ have more chances to be labeled as object, the $p(\vartheta(I))$ in equation (1) can be linked with $p(C)$ using the softmax function and signed Chamfer distance

$$\begin{aligned} p(\vartheta^o(I_{x,y})) &= \frac{1}{1+\exp(-dist(x,y,T(C)))}, \\ p(\vartheta^b(I_{x,y})) &= 1 - p(\vartheta^o(I_{x,y})) \end{aligned} \quad (9)$$

where

$$\begin{aligned} dist(x,y,T(C)) &= \min \|pos_{x,y} - T(C)\| \text{ on } \Omega_1, \\ dist(x,y,T(C)) &= -\min \|pos_{x,y} - T(C)\| \text{ on } \Omega_2 \end{aligned} \quad (10)$$

The maximization of equation (1) is equal to minimize its negative logarithm. Rewriting (1) in the energy integral equation and extending the approach in [1]

$$
\begin{aligned}
E(\phi) = \gamma \int_{\Omega} |\nabla u(\phi(x,y)|\, dxdy \\
- \lambda_1 \int_{\Omega} u(\phi(x,y) \log p(f|\vartheta^o(I))dxdy \\
- \lambda_2 \int_{\Omega} (1 - u(\phi(x,y)) \log p(f|\vartheta^b(I))dxdy \\
- \lambda_3 \int_{\Omega} u(\phi(x,y) \log p(\vartheta^o(I))dxdy \\
- \lambda_4 \int_{\Omega} (1 - u(\phi(x,y)) \log p(\vartheta^b(I))dxdy \qquad (11)
\end{aligned}
$$

where the $u(\phi(x,y)$ is the heaviside step function defined as $u(t) = \frac{1}{2}\left(1 + \frac{2}{\pi} \arctan\left(\frac{t}{\varepsilon}\right)\right)$. The $\gamma \geqslant 0$ is used to adjust the effect of the length of the contour C. The $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \geqslant 0$ are used to adjust the weights of the bottom-up information and the top-down shape constraints. The minimization is performed using the Euler-Largrange equation with the gradient of the embedding function ϕ defined as

$$
\begin{aligned}
\frac{\partial \phi}{\partial t} = \delta(\phi)\left(\gamma div\left(\frac{\nabla \phi}{|\nabla \phi|}\right)\right) \\
+ \delta(\phi)(\lambda_2 \log p(f|\vartheta^b(I)) - \lambda_1 \log p(f|\vartheta^o(I))) \\
+ \delta(\phi)(\lambda_4 \log p(\vartheta^o(I)) - \lambda_3 \log p(\vartheta^b(I))) \qquad (12)
\end{aligned}
$$

where $\delta(\phi) = u'(\phi)$. For all the experiments we use $\gamma = 0.01$ and all the λ are set to be 1.

4. EXPERIMENTS

In the limited space available here, we describe the performance of our algorithm on a set of tiger images taken from COREL database [12] and GOOGLE search. Two types of occlusions are shown here. In Fig 4a, the tree which blocked part of the tiger is mislabeled as object; the man-made white strip which divides the whole tiger into two parts is shown in Fig4d. The segmentation results without shape priors are shown on Fig4b and Fig4e. Applying the shape priors calculated using EFD shape clustering and non-rigid transformation estimation, more accurate results are obtained in Fig4c and Fig4f for both cases. In total, we use 15 tiger images for likelihood density approximation/shape clustering in the training stage and 20 partially occluded images for testing. The overall pixel-wise segmentation accuracy we obtained is 91.27%. Note that without applying shape prior the accuracy dropped to 85.01%.

5. CONCLUSION

In this paper, we propose an effective algorithm and apply it on the partially occluded image segmentation. The MATLAB implementation is completed in 20 seconds on a 192×128 image using a PC with P4 1.5G and 2 G memory. The speedup of the level-set evolution is due to the rough initial position provided by the aligned shape prior and bottom-up segmentation. The algorithm can be further optimized using C++.

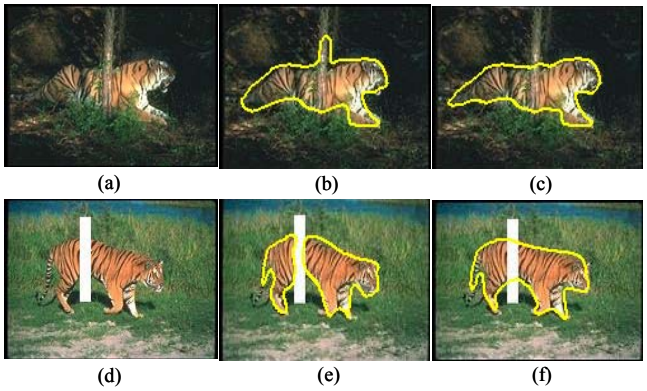

Fig. 4. The image segmentation results of two occluded tiger images.

6. REFERENCES

[1] T. F. Chan and L. A. Vese, "Active contours withour edges," *ITIP*, vol. 10, no. 2, pp. 266–277, 2001.

[2] M. E. Leventon, W. E. L. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," *CVPR*, vol. 1, pp. 316–323, 2000.

[3] S. Dambreville, Y. Rathi, and A. Tannenbaum, "Shape-based approach to robust image segmentation using kernel PCA," *CVPR*, vol. 1, pp. 977–984, 2006.

[4] D. Cremers, S. J. Osher, and S. Soatto, "Kernel density estimation and intrinsic alignment for shape priors in level set segmentation," *IJCV*, vol. 69, pp. 335–351, 2006.

[5] T. R. Raviv, N. Kiryati, and N. Sochen, "Segmentation by level sets and symmetry," *CVPR*, vol. 1, pp. 1015–1022, 2006.

[6] D. Comanicu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *PAMI*, vol. 24, no. 5, pp. 603–619, 2002.

[7] D. W. Scott, "Parametric statistical modeling by minimum integrated square error," *Technometrics*, vol. 43, pp. 274–285, 2001.

[8] F. P. Kuhl and C. R. Giardina, "Elliptic Fourier features of a closed contour," *CGIP*, vol. 18, pp. 236–258, 1982.

[9] M. Varma and A. Zisserman, "Classifying images of materials: Achieving viewpoint and illumination independence," *ECCV*, vol. 3, pp. 255–271, 2002.

[10] Y. Chen, H. D. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, A. Geiser, K. Gopinath, and R. Briggs, "Using prior shapes in geometric active contours in a variational framework," *IJCV*, vol. 50, no. 3, pp. 315–328, 2002.

[11] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *CVIU*, vol. 89, no. 2, pp. 114–141, 2003.

[12] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *PAMI*, vol. 25, no. 8, pp. 1027–1037, 2002.