INCORPORATING COMPLEX STATISTICAL INFORMATION IN ACTIVE CONTOUR-BASED IMAGE SEGMENTATION

Junmo Kim†, *John W. Fisher III*†, *Mujdat Cetin*†, *Anthony Yezzi, Jr.*‡, *and Alan S. Willsky*†

†Laboratory for Information and Decision Systems, Massachusetts Institute of Technology
77 Massachusetts Ave., Cambridge, MA 02139, USA
‡ School of Electrical and Computer Engineering, Georgia Institute of Technology
Atlanta, GA 30332, USA

ABSTRACT

We propose an information-theoretic method for multi-phase image segmentation, in an active contour-based framework. Our approach is based on nonparametric density estimates, and is able to solve problems involving arbitrary probability densities for the region intensities. This is achieved by maximizing the mutual information between the region labels and the image pixel intensities, in order to segment up to 2^m regions using m curves. The method does not require any prior training regarding the regions of interest, but rather learns the probability densities during the evolution process. We present some illustrative experimental results, demonstrating the power of the proposed segmentation approach.

1. INTRODUCTION

A number of active contour-based variational techniques have recently been developed and used in image segmentation. These methods are based on fairly simple statistical models for the intensities of the regions to be segmented. For example, either simple Gaussian intensity models are assumed, or a particular discriminative feature (such as the intensity mean or variance) is used [5].

This work considers more general problems where the regions to be segmented may not be separable by a simple discriminative feature, or by using simple Gaussian probability densities. We present an information-theoretic segmentation approach, which can deal with a variety of intensity distributions. Some previous techniques which have relations to our approach include the region competition method of [6], and the supervised texture segmentation method of [4]. Our strategy is different from previous methods in three major ways. First, unlike e.g. [6], our approach is based on nonparametric statistics. Secondly, unlike e.g. [4], our technique requires no training. Thirdly, the optimization problem we pose is based on a new information-theoretic cost functional utilizing mutual information. In particular, we cast the segmentation problem as the maximization of the mutual information between the region labels and the image pixel intensities.

We have previously described a two-region version of this approach [3]. In this paper, we provide an extension of this technique for images with more than two regions, by incorporating the multi-phase segmentation formulation of [1] into our information-theoretic, nonparametric segmentation framework. Our method uses m level set functions to segment up to 2^m regions, and the resulting curve evolution equation (motion equation) turns out to be a natural generalization of nonparametric region competition. The nonparametric aspect of our approach makes it especially appealing in applications where there is little or no prior information about the statistical properties of the regions to be segmented. We present the effectiveness of the proposed segmentation strategy on a number of synthetic and real scenes.

This work was supported by the Office of Naval Research under Grant N00014-00-1-0089, and the Air Force Office of Scientific Research under Grant F49620-00-0362.

2. INFORMATION THEORETIC APPROACH TO IMAGE SEGMENTATION: MULTI-PHASE FRAMEWORK

2.1. Problem Statement

We consider an n-ary (i.e. n-region) image segmentation problem, where $R_1, \dots, R_n$ denote the true unknown regions, and the image intensity at pixel x, denoted by $G(x)$, is drawn from the density p_i if $x \in R_i$, where p_i's are unknown. Figure 1(a) illustrates this image model when $n = 4$.

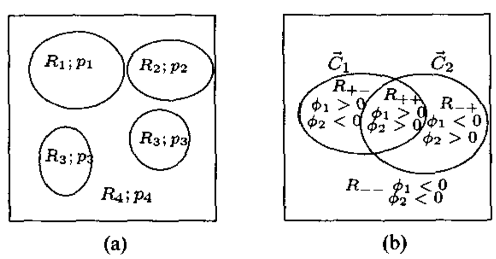

Fig. 1. Multi-phase segmentation image model. (a): Illustration of the case $n = 4$: true regions $R_1, \dots, R_4$, with the associated distributions $p_1, \dots, p_4$. (b): Illustration of the two curves ($\vec{C}_1, \vec{C}_2$), the regions $R_{++}, R_{+-}, R_{-+}, R_{--}$ partitioned by the curves.

The goal of n-ary image segmentation by curve evolution is to move a set of curves $\{\vec{C}_1, \dots, \vec{C}_m\}$ (equivalently, a set of level set functions $\{\phi_1, \dots, \phi_m\}$) such that these curves partition the image domain into the true regions $R_1, \dots, R_n$. Each curve C_i partitions the image domain into the two regions, the region inside the curve and the region outside the curve (ϕ_i does the same thing by its sign). Thus the m level set functions partition the image domain into 2^m regions, each of which we label by the signs of the level set functions in that region. For instance, when $m = 2$, we have 4 regions, $R_{++}, R_{+-}, R_{-+}, R_{--}$ as illustrated in Figure 1(b). More formally, we define R_{++} as a closure of the open set $\{\mathbf{x}|\phi_1(\mathbf{x}) > 0, \phi_2(\mathbf{x}) > 0\}$. R_{+-}, R_{-+}, and R_{--} are similarly defined.

Given the partitioning by the curves $\mathbf{C} \triangleq \{\vec{C}_i\}_{i=1}^m$, we can label each pixel x by its label $L_\mathbf{C}(x)$. For instance, if $x \in R_{++}$, $L_\mathbf{C}(x) = L_{++}$. More formally, this partitioning of the image domain by the curves $\mathbf{C}$ gives us a label

$$L_\mathbf{C} : \Omega \to \{L_{++\cdots+}, \ldots, L_{--\cdots-}\},$$

which is a mapping from the image domain Ω to a set of 2^m labeling symbols $\{L_{++\cdots+}, \ldots, L_{--\cdots-}\}$ defined as follows:

$$L_\mathbf{C}(x) = L_{s(i)} \text{ if } x \in R_{s(i)}, 1 \le i \le 2^m, \tag{1}$$

where $s(i)$ is the ith element in the set $\{++\cdots+, \ldots, --\cdots-\}$. By this correspondence between labels and curves, image segmentation by curve evolution is equivalent to the 2^m-ary labeling problem.

2.2. Mutual Information between the Image Intensity and the Label

As mentioned before, we have a combination of candidate segmenting curves $\mathbf{C}$, and $R_1, \ldots, R_n$ are the true unknown regions. Now suppose that X is a uniformly distributed random location in the image domain Ω. In this case, the label $L_\mathbf{C}(X)$ is a random variable that depends on the curves $\mathbf{C}$. It takes the value $L_{s(i)} \in \{L_{++\cdots+}, \ldots, L_{--\cdots-}\}$ with probability $\frac{|R_{s(i)}|}{|\Omega|}$, where $|R_{s(i)}|$ denotes the area of the region $R_{s(i)}$. On the other hand, the image intensity $G(X)$ is a random variable that depends on the true regions $\{R_i\}_{i=1}^n$ with the density $p_{G(X)} = \frac{|R_i|}{|\Omega|} p_i$.

Now let us consider the mutual information $I(G(X); L_\mathbf{C}(X))$ between the label and the intensity

$$\begin{aligned} &I(G(X); L_\mathbf{C}(X)) \\ &= h(G(X)) - \sum_{i=1}^{2^m} Pr(L_\mathbf{C}(X) = L_{s(i)}) h(G(X)|L_\mathbf{C}(X) = L_{s(i)}) \end{aligned}$$

where the differential entropy $h(Z)$ of a continuous random variable Z with support S is defined by $h(Z) = -\int_S p_Z(z) \log p_Z(z) dz$. The entropies $h(G(X))$ and $h(G(X)|L_\mathbf{C}(X) = L_{s(i)})$ are functionals of $p_{G(X)}$ and $p_{G(X)|L_\mathbf{C}(X)=L_{s(i)}}$, respectively. The conditional distribution is given as follows:

$$\begin{aligned} &p_{G(X)|L_\mathbf{C}(X)=L_{s(i)}} \\ &= \sum_{j=1}^{n} Pr(X \in R_j | L_\mathbf{C}(X) = L_{s(i)}) p_{G(X)|X\in R_j, L_\mathbf{C}(X)=L_{s(i)}} \\ &= \sum_{j=1}^{n} \frac{|R_{s(i)} \cap R_j|}{|R_{s(i)}|} p_j \end{aligned} \tag{2}$$

Each conditional entropy measures the degree of heterogeneity in each region determined by the curves $\mathbf{C}$.

Using the data processing inequality [2], we can show that the mutual information $I(G(X); L_\mathbf{C}(X))$ is maximized if and only if the curves $\mathbf{C}$ give the correct segmentation,[1] i.e. $\{R_i\}_{i=1}^n = \{R_{s(i)} | 1 \le i \le 2^m\}$. We omit the proof here. This result suggests that mutual information is a reasonable criterion for the segmentation problem we have formulated. However, in practice, we really cannot compute the mutual information $I(G(X); L_\mathbf{C}(X))$ since the regions $\{R_i\}_{i=1}^n$ and the probability densities $p_1, \ldots p_n$ are unknown.

We thus need to estimate the mutual information as follows:

$$\begin{aligned} &\hat{I}(G(X); L_\mathbf{C}(X)) \\ &= \hat{h}(G(X)) - \sum_{i=1}^{2^m} Pr(L_\mathbf{C}(X) = L_{s(i)}) \hat{h}(G(X)|L_\mathbf{C}(X) = L_{s(i)}) \end{aligned} \tag{3}$$

This in turn requires us to estimate the densities $p_{G(X)}$ and $p_{G(X)|L_\mathbf{C}(X)=L_{s(i)}}$.

2.3. The Energy Functional

Finally, we combine the mutual information estimate with a regularization term, and the resulting energy functional to minimize is then given by

$$E(\mathbf{C}) = -\hat{I}(G(X); L_\mathbf{C}(X)) - \alpha \sum_{i=1}^{m} \oint_{\vec{C}_i} ds, \tag{4}$$

where the second term acts as a curve length penalty.

3. NONPARAMETRIC DENSITY ESTIMATION AND GRADIENT FLOWS

We now describe our density estimation process needed in (3), and the gradient flow to minimize $E(\mathbf{C})$ of (4). For notational convenience, we consider the case where $m = 2$, but the development could easily be generalized to any m.

3.1. Estimation of the Differential Entropy

Based on (3), we have $2^m = 4$ conditional entropies to estimate, namely, $\hat{h}(G(X)|L_\mathbf{C}(X) = L_{++}), \ldots, \hat{h}(G(X)|L_\mathbf{C}(X) = L_{--})$. In order to compute $\hat{h}(G(X)|L_\mathbf{C}(X) = L_{++})$, we use the following Parzen density estimate of $p_{R_{++}} \triangleq p_{G(X)|L_\mathbf{C}(X)=L_{++}}$:

$$\hat{p}_{R_{++}}(z) = \frac{1}{|R_{++}|} \int_{R_{++}} K(z - G(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \tag{5}$$

Thus $\hat{h}(G(X)|L_\mathbf{C}(X) = L_{++})$ is given by

$$\begin{aligned} &\hat{h}(G(X)|L_\mathbf{C}(X) = L_{++}) \\ &= -\frac{1}{|R_{++}|} \int_{R_{++}} \log \hat{p}_{R_{++}}(G(\mathbf{x}))\, d\mathbf{x} \\ &= -\frac{1}{|R_{++}|} \int_{R_{++}} \log \left(\frac{1}{|R_{++}|} \int_{R_{++}} K(G(\mathbf{x}) - G(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right) d\mathbf{x}, \end{aligned}$$

where we have approximated the entropy $h(G(X)|L_\mathbf{C}(X) = L_{++})$, which is the expected value of the logarithm of $p_{R_{++}}$, by the sample mean of $\log p_{R_{++}}$. Similarly, $\hat{h}(G(X)|L_\mathbf{C}(X) = L_{+-})$,

[1] For the sake of notational simplicity we assume $n = 2^m$ in our development here, but the approach is applicable to the case where $n < 2^m$, as well.

$\hat{h}(G(X)|L_C(X) = L_{-+})$, and $\hat{h}(G(X)|L_C(X) = L_{--})$ are given by nested region integrals over R_{+-}, R_{-+}, and R_{--}, respectively.

3.2. The Gradient Flow for the Information-Theoretic Energy Functional

Based on the first variation of the energy functional $E(\mathbf{C})$ in (4), we obtain the following coupled motion equations[2]:

$$\begin{aligned}\frac{\partial \vec{C}_1(p)}{\partial t} = \vec{N}_1 \Big[&-\alpha\kappa_1 + \chi(\vec{C}_1(p) \in R_{++}) \log \hat{p}_{++}(G(\vec{C}_1(p))) \\ &- \chi(\vec{C}_1(p) \in R_{-+}) \log \hat{p}_{-+}(G(\vec{C}_1(p))) \\ &+ \chi(\vec{C}_1(p) \in R_{+-}) \log \hat{p}_{+-}(G(\vec{C}_1(p))) \\ &- \chi(\vec{C}_1(p) \in R_{--}) \log \hat{p}_{--}(G(\vec{C}_1(p))) \Big] \end{aligned} \quad (6)$$

$$\begin{aligned}\frac{\partial \vec{C}_2(p)}{\partial t} = \vec{N}_2 \Big[&-\alpha\kappa_2 + \chi(\vec{C}_2(p) \in R_{++}) \log \hat{p}_{++}(G(\vec{C}_2(p))) \\ &- \chi(\vec{C}_2(p) \in R_{+-}) \log \hat{p}_{+-}(G(\vec{C}_2(p))) \\ &+ \chi(\vec{C}_2(p) \in R_{-+}) \log \hat{p}_{-+}(G(\vec{C}_2(p))) \\ &- \chi(\vec{C}_2(p) \in R_{--}) \log \hat{p}_{--}(G(\vec{C}_2(p))) \Big], \end{aligned} \quad (7)$$

where $\vec{C}_1$ and $\vec{C}_2$ are parameterized by $p \in [0, 1]$, and $\chi()$ is an indicator function such that $\chi(\vec{C}_1(p) \in R_{++})$ is 1 if the pixel $\vec{C}_1(p)$ is in R_{++} and 0 otherwise.

For the purpose of illustration, let us observe how these general equations (6), (7) specialize when the topology of the curves looks like Figure 1(b). For this case, we have the following non-parametric region competitions:

$$\begin{aligned}\frac{\partial \vec{C}_1}{\partial t} &= \vec{N}_1 \left[-\alpha\kappa_1 + H(\phi_2(\vec{C}_1)) \log \frac{\hat{p}_{++}(G(\vec{C}_1))}{\hat{p}_{-+}(G(\vec{C}_1))} \right. \\ &\qquad \left. + (1 - H(\phi_2(\vec{C}_1)) \log \frac{\hat{p}_{+-}(G(\vec{C}_1))}{\hat{p}_{--}(G(\vec{C}_1))} \right] \end{aligned} \quad (8)$$

$$\begin{aligned}\frac{\partial \vec{C}_2}{\partial t} &= \vec{N}_2 \left[-\alpha\kappa_2 + H(\phi_1(\vec{C}_2)) \log \frac{\hat{p}_{++}(G(\vec{C}_2))}{\hat{p}_{+-}(G(\vec{C}_2))} \right. \\ &\qquad \left. + (1 - H(\phi_1(\vec{C}_2)) \log \frac{\hat{p}_{-+}(G(\vec{C}_1))}{\hat{p}_{--}(G(\vec{C}_2))} \right] \end{aligned} \quad (9)$$

where $H(\cdot)$ is the Heaviside function ($H(\phi)$=1 if $\phi \geq 0$ and $H(\phi) = 0$ if $\phi < 0$).

Equations (8), (9) involve log likelihood ratio tests comparing the hypotheses that the observed image intensity $G(\vec{C}_i)$ at a given point on the active contour $\vec{C}_i$ belongs to one region or the other.

As illustrated in Figure 1(b), $\vec{C}_1$ delineates either the boundary between R_{++} and R_{-+}, or the boundary between R_{+-} and R_{--}, when $\vec{C}_1$ lies inside or outside the curve $\vec{C}_2$, respectively. Equation (8) exactly reflects this situation and reveals the region competition between regions adjacent to the curve $\vec{C}_1$. Similarly, Equation (9) shows the region competition between regions adjacent to the curve $\vec{C}_2$.

[2]These expressions are approximate, since they contain only the dominant terms contributing to the curve evolution. For the complete motion equations (for the $n = 2$ case), see [3]

4. EXPERIMENTAL RESULTS

First, we demonstrate our information-theoretic, multi-phase segmentation method on a synthetic image of geometric objects. The image shown in Figure 2(a) contains three regions (circle/rectangle, ellipse/hexagon, and background) with Gaussian distributions with different means. Hence, in this case we have $m = 2$, $n = 3$. The initial, intermediate, and final stages of our curve evolution algorithm are shown in Figure 2, with the inside of the first (solid) and second (dashed) curves capturing the ellipse/hexagon region, and the background region respectively. Figure 3(a) contains an example with four regions (circle, ellipse, hexagon, and background), hence $m = 2, n = 4$, with again Gaussian distributions with different means. The first (solid) curve has the circle and ellipse in it and the second (dashed) curve has the circle and the hexagon in it. Equivalently, R_{++}, R_{+-}, R_{-+}, and R_{--} capture the circle, the ellipse, the hexagon, and the background, respectively. Note that, methods such as [5] would also work in these simple examples, but would require the selection of an appropriate statistic (in this case the mean) a priori, whereas our method does not. The Mumford Shah-based multi-phase technique of [1], would also work in this case. Figure 4(a) contains an example with four regions having Gaussian distributions with different variances. Again, R_{++}, R_{+-}, R_{-+}, and R_{--} capture the circle, the ellipse, the hexagon, and background, respectively.

In addition, our approach is directly applicable in problems involving more challenging intensity distributions, as we demonstrate on a real-image example next. Figure 5(a) shows the image of a zebra on a background ($m = 1$, $n = 2$). Note that the foreground intensities appear to have a bimodal density, whereas the background appears unimodal. Techniques, such as [5, 1] are not suited to this kind of problem. The final result of our approach is shown in Figure 5(d). This identical image has also been used in [4]. Note that unlike the training-based approach of [4], our method achieves an accurate segmentation without any supervision.

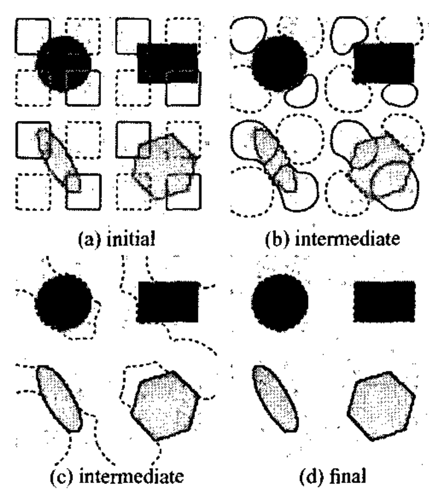

Fig. 2. Evolution of the curve on a synthetic image; three regions with different mean intensities.

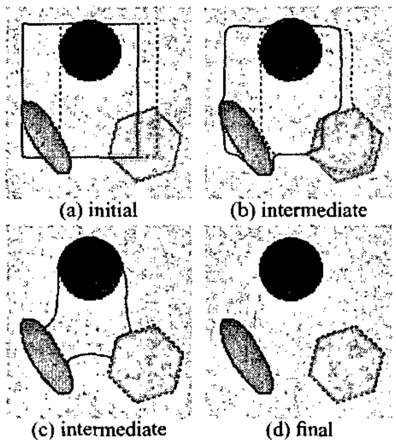

Fig. 3. Evolution of the curve on a synthetic image; four regions with different mean intensities.

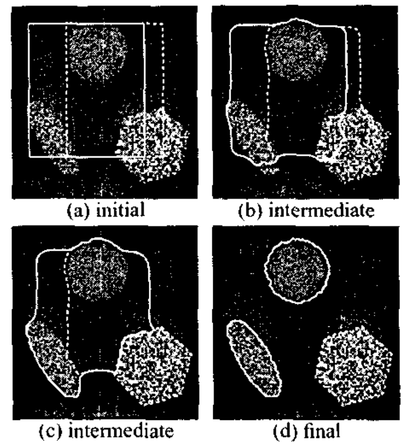

Fig. 4. Evolution of the curve on a synthetic image; four regions with different intensity variances.

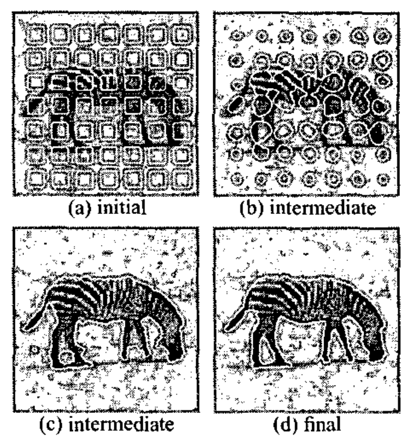

Fig. 5. Evolution of the curve on a zebra image. (Input image: courtesy of Nikos Paragios)

5. CONCLUSION

We have presented a multi-phase, information theoretic n-ary segmentation technique based on nonparametric density estimates, which is able to solve challenging segmentation problems in an unsupervised fashion. The technique is general in the sense that it could in principle be applied to any segmentation problem where pixel intensity distributions could be used to discriminate between different regions. On the other hand, the method can also solve simple special cases, e.g. when a particular discriminative feature is sufficient for the segmentation task. We have presented some preliminary experimental results illustrating the flavor of this technique, and our current work involves the validation of the proposed approach on a variety of images.

6. REFERENCES

[1] T. F. Chan and L. A. Vese. An efficient variational multiphase motion for the mumford-shah segmentation model. In *Proc. Asilomar Conf. on Signals, Systems, and Computers*, pages 490–494, 2000.

[2] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991.

[3] J. Kim, J. W. F. III, A. Yezzi, Jr., M. Cetin, and A. S. Willsky. Nonparametric methods for image segmentation using information theory and curve evolution. In *Proc. IEEE Conf. on Image Processing*, 2002.

[4] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *Int. J. Computer Vision*, 2002.

[5] A. Yezzi, Jr., A. Tsai, and A. Willsky. A statistical approach to snakes for bimodal and trimodal imagery. In *Int. Conf. on Computer Vision*, pages 898–903, 1999.

[6] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(9):884–900, September 1996.