This paper appeared in Proceedings of the 2002 IEEE International Conference on Image Processing, Rochester, NY, September 2002.

NONPARAMETRIC METHODS FOR IMAGE SEGMENTATION USING INFORMATION THEORY AND CURVE EVOLUTION

Junmo Kim[†]*, John W. Fisher III*[†]*, Anthony Yezzi, Jr.*[‡]*, Mujdat Cetin*[†]*, 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

In this paper, we present a novel information theoretic approach to image segmentation. We cast the segmentation problem as the maximization of the mutual information between the region labels and the image pixel intensities, subject to a constraint on the total length of the region boundaries. We assume that the probability densities associated with the image pixel intensities within each region are completely unknown a priori, and we formulate the problem based on nonparametric density estimates. Due to the nonparametric structure, our method does not require the image regions to have a particular type of probability distribution, and does not require the extraction and use of a particular statistic. We solve the information-theoretic optimization problem by deriving the associated gradient flows and applying curve evolution techniques. We use fast level set methods to implement the resulting evolution. The evolution equations are based on nonparametric statistics, and have an intuitive appeal. The experimental results based on both synthetic and real images demonstrate that the proposed technique can solve a variety of challenging image segmentation problems.

1. INTRODUCTION

Image segmentation has been an important problem in image analysis with applications to pattern recognition, object detection, and medical image analysis. Thus, there has been a considerable amount of work on image segmentation including those using curve evolution techniques [2, 3, 4, 7, 8, 10, 11, 14, 15]. For example, Paragios et al. [8] developed a parametric model for analysis and segmentation of textured images. Yezzi et al. [15] developed a segmentation technique using a particular discriminative statistical feature such as the mean or the variance of image regions. These and many other recent works (such as [11]) have been inspired by the region competition model of Zhu and Yuille [16].

In all the work mentioned above, the typical statistical model for the underlying image was in a parametric form. However, this parametric approach is not robust in the sense that its performance is severely affected when the parametric model is not correct.

In response to the need for robustness in statistical analysis, nonparametric methods [9] have been widely used in machine learning problems. Nonparametric methods estimate the underlying distributions from the data without any assumptions about the structures of the distributions. On the other hand, mutual information has been used as a tool to solve a variety of problems such as MR-CT image registration [13], 3-D pose alignment [12], and measuring global and local spatial correspondence [1].

In this paper, we propose a novel approach to image segmentation. Here we focus on images with two regions, but the method can be generalized to multi-region images. We segment a given image into the foreground and the background by evolving a closed curve with curve length penalty so that we maximize the nonparametric estimate of the mutual information between the binary (foreground region inside the curve/ background region outside the curve) label determined by the curve and the image pixel intensity. The resulting curve evolution formula involves a nonparametric likelihood ratio and other terms explaining the change of density estimates due to the evolution of the curve. To compute the density estimates, we use the fast calculation methods proposed in [6].

The remainder of this paper is organized as follows. Section 2 presents the novel information theoretic objective functional for image segmentation. Section 3 then derives our curve evolution-based approach to minimizing this objective functional. We then present experimental results in Section 4, using both synthetic and real images. Finally, we conclude in Section 5 with a summary.

2. INFORMATION THEORETIC APPROACH TO IMAGE SEGMENTATION

In this section, we state the problem, the assumptions, and present our information theoretic segmentation criterion.

2.1. Image Model

The image model we are dealing with has two unknown regions R_1 and R_2 with the associated unknown distributions p_1 and p_2. The image intensity at pixel x denoted by $I(x)$ is drawn from p_1 if $x \in R_1$ and from p_2 if $x \in R_2$. The left-hand side of Figure 1 illustrates the image model.

The goal of curve evolution is to move the curve $\vec{C}$ such that it matches the boundary between R_1 and R_2, i.e. the region inside the curve R and the region outside the curve R^c converge to R_1 and R_2 respectively.

2.2. Mutual Information between Image Intensity and the Label

We present an information theoretic energy functional based on mutual information between a binary region label and the intensity val-

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.

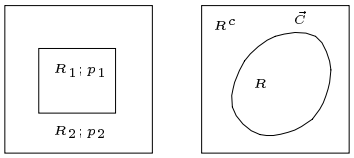


Fig. 1. Left: Illustration of the foreground region (R_1), the background region (R_2), and the associated distributions (p_1 and p_2). Right: Illustration of the curve ($\vec{C}$), the region inside the curve (R), and the region outside the curve (R^c).

ues of an image. We define the binary label determined by the curve $\vec{C}$ as a mapping from the image domain Ω to $\{F, B\}$ denoted by $L : \Omega \to \{F, B\}$ as follows:

$$L(x) = \begin{cases} F & \text{if } x \in R \\ B & \text{if } x \in R^c \end{cases} \quad (1)$$

Let X be a random variable which is uniformly distributed over the image domain Ω, then $L(X)$ becomes a binary random variable taking value F or B with probability $\frac{|R|}{|\Omega|}$ and $\frac{|R^c|}{|\Omega|}$ respectively, where $|\cdot|$, the cardinality of a set, is given by the area of the set. Note that $L(X)$ conveys information about the image intensity $I(X)$ at a random location via X. The mutual information between the image intensity at X and the label at X is given formally as follows:

$$\begin{aligned} & I(I(X); L(X)) \\ &= h(I(X)) - h(I(X)|L(X)) \\ &= h(I(X)) - Pr(L(X) = F)h(I(X)|L(X) = F) \\ &\quad - Pr(L(X) = B)h(I(X)|L(X) = B), \end{aligned} \quad (2)$$

where the differential entropy [5] of a continuous random variable Z with a support S is defined by

$$h(Z) = -\int_S p_Z(z) \log p_Z(z) dz \quad (3)$$

2.3. Utility of Mutual Information as a Segmentation Statistic

Since I(X), X, L(X) form a Markov chain, by the data processing inequality,

$$I(I(X); L(X)) \quad \leq \quad I(I(X); X), \quad (4)$$

where equality holds if and only if I(X), L(X), X form a Markov chain, i.e., I(X) and X are conditionally independent given L(X). If $L(\cdot)$ is not the correct segmentation, then knowing $L(X)$ is not enough to determine whether the distribution of $I(X)$ is p_1 or p_2, and thus $I(X)$ is not independent of X. Therefore, $I(I(X); L(X))$ is maximized if and only if $L(\cdot)$ gives the correct segmentation.

2.4. The Energy Functional

Since $I(I(X); L(X))$ is a functional of the unknown densities p_1 and p_2, we need to estimate the mutual information:

$$\begin{aligned} & \hat{I}(I(X); L(X)) \\ &= \hat{h}(I(X)) - Pr(L(X) = F)\hat{h}(I(X)|L(X) = F) \\ &\quad - Pr(L(X) = B)\hat{h}(I(X)|L(X) = B) \end{aligned} \quad (5)$$

We combine the mutual information estimate with the typical regularization penalizing the length of curve. This regularization prevents the formation of fractal segmenting curves. The resulting energy functional to minimize is then given by

$$E(\vec{C}) = -\hat{I}(I(X); L(X)) + \alpha \oint_{\vec{C}} ds, \quad (6)$$

where $\oint_{\vec{C}} ds$ is the length of the curve and α is a scalar parameter.

3. NONPARAMETRIC DENSITY ESTIMATION AND GRADIENT FLOWS

This section derives the curve evolution formula for minimizing the energy functional (6) using nonparametric Parzen density estimates.

3.1. Estimation of the Differential Entropy

The expression (5) involves differential entropy estimates and we use nonparametric Parzen density estimates in order to estimate the differential entropies.

Since $\hat{h}(I(X))$ is independent of the curve or the label, we just consider $\hat{h}(I(X)|L(X) = F)$ and $\hat{h}(I(X)|L(X) = B)$, which are given as follows:

$$\begin{aligned} & \hat{h}(I(X)|L(X) = F) \\ &= -\frac{1}{|R|}\int_R \log \hat{p}_R(I(\mathbf{x}))\, d\mathbf{x} \quad (7) \\ &= -\frac{1}{|R|}\int_R \log\left(\frac{1}{|R|}\int_R K(I(\mathbf{x}) - I(\hat{\mathbf{x}}))d\hat{\mathbf{x}}\right) d\mathbf{x}, \quad (8) \end{aligned}$$

where (7) is an approximation of the entropy using weak law of large numbers, and (8) uses a continuous version of the Parzen density estimate [9] of $p_R \stackrel{\text{def}}{=} p_{I(X)|L(X)=F}$. In (8), the kernel is $K(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2\sigma^2}}$, where σ is a scalar parameter. Similarly,

$$\begin{aligned} & \hat{h}(I(X)|L(X) = B) \\ &= -\frac{1}{|R^c|}\int_{R^c} \log\left(\frac{1}{|R^c|}\int_{R^c} K(I(\mathbf{x}) - I(\hat{\mathbf{x}}))d\hat{\mathbf{x}}\right) d\mathbf{x} \quad (9) \end{aligned}$$

3.2. Gradient Flows for General Nested Region Integrals

Note that (8) and (9) have nested region integrals. For a general nested region integral of the form

$$\int_R f(\varepsilon(\mathbf{x}, t))\, d\mathbf{x} \quad \text{where} \quad \varepsilon(\mathbf{x}, t) = \int_R g(\mathbf{x}, \hat{\mathbf{x}})\, d\hat{\mathbf{x}}, \quad (10)$$

we have derived the gradient flow (the negative of the gradient), which is given by

$$\frac{\partial \vec{C}}{\partial t} = -\left[f\left(\varepsilon(\vec{C})\right) + \int_R f'\left(\varepsilon(\mathbf{x})\right) g(\mathbf{x}, \vec{C})\, d\mathbf{x}\right] \vec{N}, \quad (11)$$

where $\vec{N}$ is the outward unit normal vector.

3.3. The Gradient Flows for the Information Theoretic Energy Functional

Now based on (11), (8) and (9), the gradient flow for $E(\vec{C})$ of (6) is obtained as follows:

$$\frac{\partial \vec{C}}{\partial t} = [\log \frac{\hat{p}_R(I(\vec{C}))}{\hat{p}_{R^c}(I(\vec{C}))} + \frac{1}{|R|}\int_R \frac{K\left(I(\mathbf{x}) - I(\vec{C})\right)}{\hat{p}_R(I(\mathbf{x}))}\, d\mathbf{x} - \frac{1}{|R^c|}\int_{R^c} \frac{K\left(I(\mathbf{x}) - I(\vec{C})\right)}{\hat{p}_{R^c}(I(\mathbf{x}))}\, d\mathbf{x}]\vec{N} - \alpha\kappa\vec{N},$$

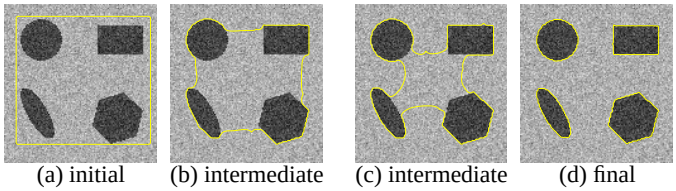
where κ is the curvature of the curve and $-\alpha\kappa\vec{N}$ is the gradient flow for the curve length penalty, whose derivation can be found in [14].

The first term of this gradient flow is a likelihood ratio test which compares the hypotheses that the observed image intensity $I(\vec{C})$ at a given point on the active contour $\vec{C}$ belongs to the foreground region R or the background region R^c based upon the current estimates of the distributions p_R and p_{R^c}. The second and third terms respond to the changes incurred on the distributions p_R and p_{R^c} by moving a given point on the active contour.

These last two terms distinguish this active contour model from those obtained using coordinate descent, in which alternating iterations of estimating the distribution parameters inside and outside the curve are followed by likelihood ratio tests to evolve the curve as in the "Region Competition" algorithm of Zhu and Yuille [16]. In such algorithms, changes in the distributions are not directly coupled with likelihood ratio tests. In contrast, the mathematical structure of our nonparametric estimators are built directly into the curve evolution equation through the last two terms.

Since the evaluation of the density estimate at each pixel takes $O(\#\text{ of pixels})$ time, calculation of the gradient flow takes $O((\#\text{ of pixels})^2)$ time. We reduced the computational complexity to $O(\#\text{ of pixels})$ time using the fast Gauss transform [6] in calculating the density estimates.

4. EXPERIMENTAL RESULTS

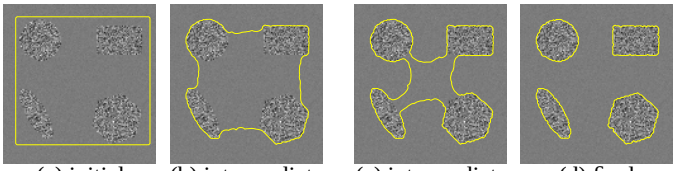


(a) initial (b) intermediate (c) intermediate (d) final

Fig. 2. Evolution of the curve on a synthetic image: the different mean case.

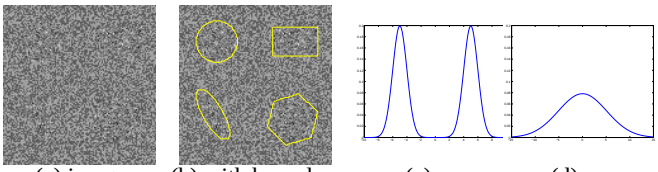
We present experimental results on synthetic images and a real image of a leopard. Three synthetic images are generated by three sets of distributions: two Gaussian distributions with different means, two Gaussian distributions with different variances, and two distributions with the same mean and the same variance.

Figure 2 shows the result for the first case, where the two distributions for the foreground and the background have different means and the same variance.



(a) initial (b) intermediate (c) intermediate (d) final

Fig. 3. Evolution of the curve on a synthetic image: the different variance case.



(a) image (b) with boundary (c) p_1 (d) p_2

Fig. 4. Example image with two regions (boundaries marked in (b)), where the foreground has a bimodal density p_1, and the background has a unimodal density p_2. The two densities p_1 and p_2 have the same mean and the same variance.

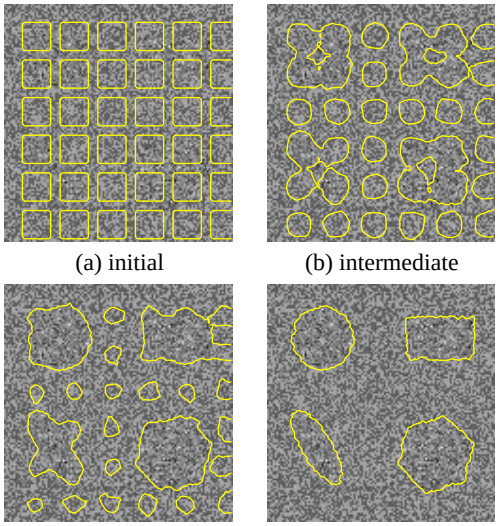
Figure 3 shows the result for the second case, where the two distributions for the foreground and the background have different variances and the same mean.

For these two cases, the method of Yezzi et al. [15] would require the selection of the appropriate statistic a priori, whereas our method does not.

Now we consider a more challenging image shown in Figure 4(a). The two underlying distributions are illustrated in Figure 4(c) and Figure 4(d). Since the two distributions have the same mean and same variance, it is hard even for a human observer to separate the foreground from the background. In order to let readers see the foreground, we show the actual boundaries by a curve in Figure 4(b). For this kind of image, the methods based on means and variances such as that proposed by Yezzi et al. [15] would no longer work.

Figure 5 shows our segmentation results. As shown in Figure 5(a), we have used an automatic initialization with multiple seeds. For this kind of image, using a simple initialization as in the examples of Figure 2 and Figure 3 leads to a large number of iterations, and in some cases the curve may get stuck in a local optimum. The power of the multiple-seed initialization is that it observes entire regions and the evolution of the curve occurs globally. Figure 5(b) and Figure 5(c) show the intermediate stages of the evolution, where the seeds in the background region gradually shrink at each iteration whereas those in the foreground region grow. Figure 5(d) gives the segmentation result.

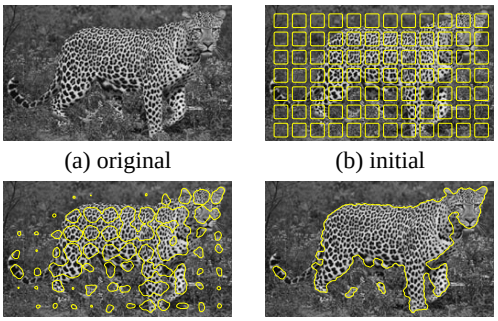
We now report the result for a leopard image, which is similar to the case of bimodal versus unimodal density example of Figure 5. Figure 6(d) shows the segmentation result. The final curve captures the main body of the leopard and some parts of its tail and legs. The parts of the tail and the legs that are missing look similar to the background, which makes a perfect segmentation difficult. Paragios et al. [8] performed a similar experiment on a leopard image. Their supervised texture segmentation algorithm requires an image patch taken from the leopard and an image patch taken from the background in advance as an input to the algorithm. It is no-



(a) initial (b) intermediate

(c) intermediate (d) final

Fig. 5. Evolution of the curve on a synthetic image: bimodal versus unimodal densities.



(a) original (b) initial

(c) intermediate (d) final

Fig. 6. Evolution of the curve on a leopard image.

ticeable that our method, which is unsupervised, can segment this complex image as accurately as their supervised algorithm.

5. CONCLUSION

We have developed a new information theoretic image segmentation method based on nonparametric statistics and curve evolution. We have formulated the segmentation problem as one of maximizing the mutual information between the region labels and the pixel intensities, subject to curve length constraints. We have derived the curve evolution equations for the optimization problem posed in our framework. Due to the nonparametric aspect of our formulation, the proposed technique can automatically deal with a variety of segmentation problems, in which many currently available curve evolution-based techniques would either completely fail or at least require the a priori extraction of representative statistics for each region. We use fast techniques for the implementation of curve evolution and nonparametic estimation, which keep the computational complexity at a reasonable level. Our preliminary experimental results have shown the strength of the proposed technique in accurately segmenting real and synthetic images.

We have recently extended our method to problems involving more than two regions. Our current work involves use of spatially dependent probability density functions for accurate texture modeling and segmentation.

6. REFERENCES

[1] F. Bello and A. Colchester. Measuring global and local spatial correspondence using information theory. In *Proceedings of the First International Conference on Medical Computing and Computer-Assited Intervention*, 1998.

[2] V. Caselles, F. Catte, T. Col, and F. Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66:1–31, 1993.

[3] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic snakes. *Int. J. Computer Vision*, 1998.

[4] T. Chan and L. Vese. Active contours without edges. *IEEE Trans. on Image Processing*, 10(2):266–277, February 2001.

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

[6] L. Greengard and J. Strain. The fast Gauss transform. *SIAM J. Sci. Stat. Comput.*, 12(1):79–94, 1991.

[7] R. Malladi, J. Sethian, and B. Vemuri. Shape modeling with front propagation: a level set approach. *IEEE Trans. Pattern Anal. Machine Intell.*, 17:158–175, 1995.

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

[9] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[10] R. Ronfard. Region-based strategies for active contour models. *Int. J. Computer Vision*, 13:229–251, 1994.

[11] C. Samson, L. Blanc-Feraud, G. Aubert, and J. Zerubia. A level set method for image classification. In *Int. Conf. Scale-Space Theories in Computer Vision*, pages 306–317, 1999.

[12] P. Viola and W. M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[13] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.

[14] A. Yezzi, Jr., S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum. A geometric snake model for segmentation of medical imagery. *IEEE Trans. on Medical Imaging*, 16(2):199–209, April 1997.

[15] 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.

[16] 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.