SHAPE GRADIENT FOR MULTI-MODAL IMAGE SEGMENTATION USING MUTUAL INFORMATION

A.Herbulot †, *S.Jehan-Besson* ‡, *M.Barlaud* †,*G.Aubert* §

†Laboratoire I3S - CNRS UNSA, 2000 route des Lucioles, 06903 Sophia-Antipolis FRANCE
‡Laboratoire GREYC-Image, 6 Bd Marechal Juin, 14050 Caen Cedex FRANCE
§Laboratoire J.A.Dieudonne - CNRS, Parc Valrose, 06108 Nice Cedex 2 FRANCE

ABSTRACT

This paper deals with video and image segmentation using region based active contours. We propose to search for an optimal domain with regards to a criterion based on information measures such as entropy of mutual information. We use a general derivation framework based on the notion of shape gradient. This general derivation is applied to criteria based on information theory, such as mutual information for the segmentation of sequences of images. Finally, we present experimental results on color video sequences showing the efficiency of the proposed method for face segmentation.

1. INTRODUCTION

The notion of entropy has first been introduced by Shannon [1] and further developed in the information theory framework whose principles can be found in [2]. Information measures such as entropy or mutual information can be efficiently managed for image and video segmentation [3] or medical image registration [4, 5]. As far as segmentation is concerned, a region may be characterized using the average quantity of information, namely the entropy, carried out by the intensity or using mutual information for features combination.

We propose here to embed information measures such as the entropy or the mutual information into a variational framework. We search for an optimal domain with regards to a global criterion including region-based and boundary-based terms. A local shape minimizer of this criterion may be reached using deformable domains, namely region-based active contours. The basic idea is to obtain, from the derivation of the criterion, a Partial Differential Equation (PDE) that drives an initial region towards a local shape minimum of the error criterion. Classically, we propose to make it evolve in the direction of a gradient. However, since the set of image regions, i.e. the set of regular open domains in $\mathbf{R}^n$, does not have a structure of vector space, we can not use gradient descent methods in a straightforward fashion. We propose to use shape gradients coming from shape optimization theory [6] to bear on the problem. Such an approach has been detailed in [7, 8] and is here further developped for the minimization of information measures using non parametric probability distribution functions (pdf) of image features following the work in [3].

Let us recall that active contours have been introduced in [9], and further extended in [10]. Region-based active contours have then appeared for the minimization of additional region features [11].Our approach is based on shape gradients which allow us to derive easily complex criteria involving non parametric probability distributions as in [8] and so a non parametric estimation of the entropy. Non parametric probability distributions have been used for region-based active contours in [8] and in [12] for mutual information estimation. We here propose to derive different information measures including non approximated entropy and mutual information with combination of color channels. We show that these criteria are very efficient for the segmentation of an almost homogeneous region : the human face, as shown in experimental results.

In this document, first section recalls the problem of optimization of region and boundary functionals with active contours. We then define a criterion based on mutual information for image segmentation in section 3 and take benefit of derivation tools to compute the evolution equation of the active contour. This framework is then extended to multimodal color segmentation in section 4 and applied to face segmentation in section 5.

2. PROBLEM STATEMENT

Let us consider an image with intensity $I(x)$ at pixel x. We note:

- $p_\Omega(I(x))$ the probability to have the intensity $I(x)$ with x in the region Ω. We can estimate the pdf

PART OF THIS WORK WAS SUPPORTED BY THE EUROPEAN NETWORK OF EXCELLENCE FP6-507609 SIMILAR.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

through the use of the Parzen method:

$$p_\Omega(I(x)) = \frac{1}{|\Omega|}\int_\Omega K(I(x) - I(\hat{x}))d\hat{x} \quad (1)$$

where K is chosen as a gaussian kernel with 0-mean and σ-variance.

- φ a function: $\mathcal{R}^+ \to \mathcal{R}^+$ of this probability (for example $\varphi(r) = -r\ln(r)$ for the computation of the entropy).

Let us define the following criterion:

$$J(\Omega) = \int_\Omega \varphi(p_\Omega(I(x)))dx \quad (2)$$

This criterion is derived using shape derivation tools as detailed in [7] and we obtain the following PDE for curve evolution [3]:

$$\frac{\partial\Gamma}{\partial\tau} = \Big(\varphi(p_\Omega(I(x))) + A(x,\Omega)\Big)\mathbf{N} \quad (3)$$

where $\mathbf{N}$ is the unit inward normal and $A(x,\Omega)$ is a term coming from the dependance of the criterion to the domain Ω. This term will be detailed in the following sections.

Now let us define a weighted criterion:

$$J(\Omega) = |\Omega| \int_\Omega \varphi(p_\Omega(I(x)))dx \quad (4)$$

The corresponding evolution equation is then:

$$\frac{\partial\Gamma}{\partial\tau} = \Big(|\Omega|\,\varphi(p_\Omega(I(x))) + B(x,\Omega)\Big)\mathbf{N} \quad (5)$$

with $B(x,\Omega)$ a term that will be detailed below.

3. IMAGE SEGMENTATION

3.1. A criterion based on Mutual Information

Let X and Y denote two random variables with marginal probability distributions $p_X(x)$ and $p_Y(y)$. $p_{XY}(x,y)$ is the joint probability distribution. $H(X)$ and $H(Y)$ denote the entropy of X and Y respectively, and $H(X,Y)$ their joint entropy.
The *Mutual Information* noted $MI(X,Y)$, or relative entropy, measures the degree of dependence of X and Y by measuring the Kullback-Leibler distance between the joint distribution and the product of the distributions:

$$MI(X,Y) = \int_{\Omega_X}\int_{\Omega_Y} p_{XY}(x,y)\ln\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}\,dx\,dy \quad (6)$$

We can rewrite this definition by using the entropies $H(X)$ and $H(Y)$, the joint entropy $H(X,Y)$ and the conditional entropy $H(X/Y)$:

$$\begin{aligned} MI(X,Y) &= H(X) + H(Y) - H(X,Y) \quad (7)\\ &= H(X) - H(X/Y) \end{aligned}$$

3.2. Segmentation criterion

Following the work of [12], we may define the binary label L determined by the curve Γ as a mapping from the image domain to $\{R_{in}, R_{out}\}$:

$$L(x) = \begin{cases} R_{in} & \text{if } x \in \Omega_{in} \\ R_{out} & \text{if } x \in \Omega_{out} \end{cases}$$

We consider the mutual information between the label and the image intensity:

$$MI(I(X), L(X)) = H(I(X)) - H(I(X)|L(X)) \quad (8)$$

with X a random variable uniformly distributed over the image domain.
The mutual information is maximized if $R_{in} = \Omega_{in}$ and $R_{out} = \Omega_{out}$, ie if the segmentation is correct. The functional to minimize is then given by:

$$J(\Omega_{in},\Omega_{out},\Gamma) = -MI(I(X),L(X)) + \int_\Gamma \lambda ds \quad (9)$$

where λ is a regularization parameter.
Mutual information can be rewritten like follows:

$$\begin{aligned} MI(I(X),L(X)) = H(I(X)) &\\ - Pr(L(X) = R_{in}).H(I(X)|L(X) = R_{in}) &\\ - Pr(L(X) = R_{out}).H(I(X)|L(X) = R_{out}) & \end{aligned}$$

The term $H(I(X))$ is independent of the curve so we just consider the two other terms which leads to:

$$H(I(X)|L(X) = R_i) = \int_{\Omega_i} -p_{\Omega_i}(I(x))\ln p_{\Omega_i}(I(x))dx \quad (10)$$

with $p_{\Omega_i} = p(I(X)|L(X) = R_i)$ where $i = in$ or out

Thus, the criterion becomes:

$$\begin{aligned} J(\Omega_{in},\Omega_{out},\Gamma) &= \frac{|\Omega_{in}|}{|D|}\int_{\Omega_{in}} -p_{\Omega_{in}}(I(x))\ln p_{\Omega_{in}}(I(x))\,dx \\ &+ \frac{|\Omega_{out}|}{|D|}\int_{\Omega_{out}} -p_{\Omega_{out}}(I(x))\ln p_{\Omega_{out}}(I(x))\,dx \\ &+ \int_\Gamma \lambda\,ds \quad (11) \end{aligned}$$

with $|D| = |\Omega_{in}| + |\Omega_{out}|$. The function φ from equation (4) is as follows:

$$\varphi(p_{\Omega_i}(I(x))) = -p_{\Omega_i}(I(x))\ln p_{\Omega_i}(I(x))$$

3.3. Derivation of the criterion

Taking benefit of shape derivation tools [7, 3], we deduce the following evolution equation from the computation of shape gradients:

$$\frac{\partial\Gamma}{\partial\tau}(\hat{x}) = \Big[C(\Omega_{in}) - C(\Omega_{out}) + \lambda.\kappa\Big]\mathbf{N} \quad (12)$$

$$\begin{aligned} \text{with } C(\Omega_i) &= \frac{|\Omega_i|}{|D|}\big[-p_{\Omega_i}(I(\hat{x}))\ln p_{\Omega_i}(I(\hat{x})) \quad (13)\\ &- \frac{1}{|\Omega_i|}\int_{\Omega_i} K(I(x) - I(\hat{x}))(\ln p_{\Omega_i}(I(x)) + 1)\,dx\big] \end{aligned}$$

where κ is the curvature of the curve Γ. The term $B(\hat{x}, \Omega)$ from equation (5) is then given by:

$$B(\hat{x}, \Omega) = -\frac{1}{\mid \Omega \mid} \int_{\Omega} K(I(x) - I(\hat{x}))(\ln p_{\Omega}(I(x)) + 1)\, dx$$

3.4. Criterion using an approximation of the entropy

Instead of the definition of the entropy, we can consider an approximation using weak law of large numbers as in [12]. Thus we have:

$$H(I(X)|L(X) = R_i) = \frac{1}{\mid \Omega_i \mid} \int_{\Omega_i} - \ln p_{\Omega_i}(I(x)) dx \qquad (14)$$

With this approximation, the criterion is given by:

$$\begin{aligned} J(\Omega_{in}, \Omega_{out}, \Gamma) &= \frac{1}{\mid D \mid} \int_{\Omega_{in}} - \ln p_{\Omega_{in}}(I(x))\, dx \qquad (15) \\ &+ \frac{1}{\mid D \mid} \int_{\Omega_{out}} - \ln p_{\Omega_{out}}(I(x))\, dx + \int_{\Gamma} \lambda\, ds \end{aligned}$$

The function φ from equation (2) is then given by:

$$\varphi(p_{\Omega_i}(I(x))) = - \ln p_{\Omega_i}(I(x))$$

So, using an approximation of the entropy, we obtain the same equation as in [12] but using efficient shape derivation tools.

$$\frac{\partial \Gamma}{\partial \tau}(\hat{x}) = \Big[C(\Omega_{in}) - C(\Omega_{out}) + \lambda.\kappa\Big]\mathbf{N} \qquad (16)$$

$$\text{with } C(\Omega_i) = \frac{1}{\mid D \mid}\big[- \ln p_{\Omega_i}(I(\hat{x})) - \frac{1}{\mid \Omega_i \mid} \int_{\Omega_i} \frac{K(I(x) - I(\hat{x}))}{p_{\Omega_i}(I(\hat{x}))}\, dx\big] \qquad (17)$$

The term $A(\hat{x}, \Omega)$ from equation (3) is then given by:

$$A(\hat{x}, \Omega) = -\frac{1}{\mid \Omega \mid} \int_{\Omega} \frac{K(I(x) - I(\hat{x}))}{p_{\Omega}(I(\hat{x}))}\, dx$$

4. MULTIMODAL IMAGE SEGMENTATION

In this section, we extend our work to multimodal images segmentation. We consider two images I_1 and I_2 obtained from a different color channel. We take benefit of joint probabilities between the different channel. We define $p_{R_i}(I_1(x), I_2(x))$ as the probability to get the intensity $I_1(x)$ in the first image and $I_2(x)$ in the second one at pixel x, where x is in the region R_i.

4.1. Criterion

We define a binary label L determined by the curve Γ, as in section 3. This label is identical for each image because the curve is the same on each image.

Let us consider n random variables $X_1, X_2, ..., X_n$. The *chain rule* for mutual information is the following one:

$$MI(X_1,X_2, \dots, X_n) = \sum_{i=1}^{n-1} MI(X_i, X_n | X_{i-1}, X_{i-2}, \dots, X_1) \qquad (18)$$

Applying (18) we have:

$$MI(I_1(X), I_2(X), L(X)) = H(I_1(X), I_2(X)) - H(I_1(X), I_2(X)|L(X)) \qquad (19)$$

As in the previous section, the term $H(I_1(X), I_2(X))$ is independent of the curve. The second term (the conditional entropy) can be rewritten as follows:

$$\begin{aligned} &H(I_1(X), I_2(X)|L(X)) = \\ &\quad Pr(L(X) = R_{in}).H(I_1(X), I_2(X)|L(X) = R_{in}) \qquad (20) \\ &+ Pr(L(X) = R_{out}).H(I_1(X), I_2(X)|L(X) = R_{out}) \end{aligned}$$

The conditional entropies are given respectively by:

$$H(I_1(X), I_2(X)|L(X) = R_i) = \int_{\Omega_i} -p_{\Omega_i}(I_1(x), I_2(x) \ln p_{\Omega_i}(I_1(x), I_2(x)) dx \qquad (21)$$

with $p_{\Omega_i} = p(I_1(X), I_2(X)|L(X) = R_i)$, $i = in$ or out.

And the criterion becomes:

$$J(\Omega_{in}, \Omega_{out}, \Gamma) = H(I_1(X), I_2(X)|L(X)) + \int_{\Gamma} \lambda\, ds \qquad (22)$$

4.2. Derivation of the criterion

By deriving the criterion with the same method as in previous section, we obtain the following equation evolution:

$$\frac{\partial \Gamma}{\partial \tau}(\hat{x}) = \Big[C(\Omega_{in}) - C(\Omega_{out}) + \lambda.\kappa\Big]\mathbf{N} \qquad (23)$$

$$\begin{aligned} \text{with } C(\Omega_i) = \frac{\mid \Omega_i \mid}{\mid D \mid}\big[&- p_{\Omega_i}(I_1(x), I_2(x)) \ln p_{\Omega_i}(I_1(x), I_2(x)) \\ &- \frac{1}{\mid \Omega_i \mid} \int_{\Omega_i} K_{XY}(I_1(x) - I_1(\hat{x}), I_2(x) - I_2(\hat{x})). \\ &(\ln p_{\Omega_i}(I_1(x), I_2(x)) + 1)\, dx\big] \end{aligned}$$

where κ is the curvature of the curve Γ and K_{XY} is a gaussian kernel. As in the previous section, the second term of $C(\Omega_i)$ is the term B of equation (5).

4.3. Criterion with an approximation of entropy

We replace probabilities by joint probabilities in equation (15) and we obtain the following criterion:

$$\begin{aligned} J(\Omega_{in}, \Omega_{out}, \Gamma) &= \frac{1}{\mid D \mid} \int_{\Omega_{in}} - \ln p_{\Omega_{in}}(I_1(x), I_2(x))\, dx \\ &+ \frac{1}{\mid D \mid} \int_{\Omega_{out}} - \ln p_{\Omega_{out}}(I_1(x), I_2(x))\, dx \\ &+ \int_{\Gamma} \lambda\, ds \qquad (24) \end{aligned}$$

Similarly we compute the following evolution equation:

$$\frac{\partial \Gamma}{\partial \tau}(\hat{x}) = \Big[C(\Omega_{in}) - C(\Omega_{out}) + \lambda.\kappa\Big]\mathbf{N} \quad (25)$$

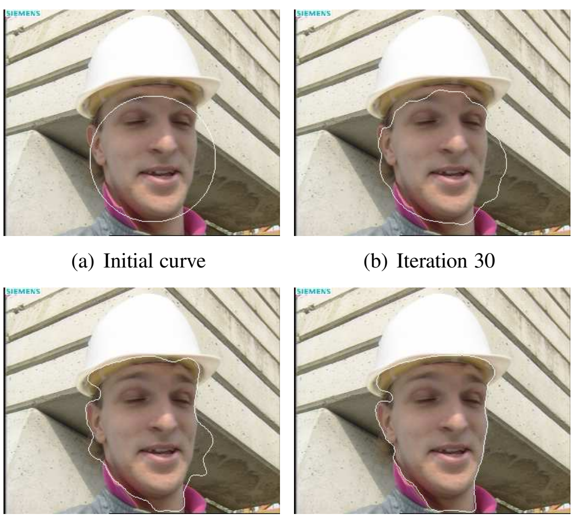
$$\text{with } C(\Omega_i) = -\ln p_{\Omega_i}(I_1(\hat{x}), I_2(\hat{x})) - \frac{1}{|\Omega_i|}\int_{\Omega_i} \frac{K_{XY}(I_1(x) - I_1(\hat{x}), I_2(x) - I_2(\hat{x}))}{p_{\Omega_i}(I_1(\hat{x}), I_2(\hat{x}))}\,dx \quad (26)$$

The second term of $C(\Omega_i)$ is the term A of eq. (3).

5. EXPERIMENTAL RESULTS

In these experiments, we use an efficient parametric method to implement the evolution equation based on smoothing B-splines (see [13]). Such an approach combines a very low computational cost compared to the level set method and a global robustness to noisy data thanks to the internal regularization parameter.

We get interested in the segmentation of a slightly textured region : the face. We test our method on the sequence *foreman*. Color images are in the RGB color space. We choose the two channels R and B and we ignore the G one. The evolution equation (23) is implemented for segmentation from the definition of the criterion (22) applied to the channels R and B. We quantify the histograms using a uniform step quantization, identical for the two components and we estimate them with the Parzen method using a parameter σ between 2 and 5. **Fig. 1.** shows the evolution of the curve.



(a) Initial curve (b) Iteration 30

(c) Iteration 150 (d) Final segmentation

Fig. 1. Evolution of the curve with the criterion (22)

6. CONCLUSION

In this paper we present a general framework based on information measures for image segmentation using region-based active contours. Non parametric measures of the entropy or the mutual information are embedded into a variational framework. The evolution equation of an active contour is then deduced from the computation of shape gradients. This general framework is extended for multimodal color image segmentation and applied efficiently to face segmentation.

7. REFERENCES

[1] C.E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, 1948.

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

[3] A. Herbulot, S. Jehan-Besson, M. Barlaud, and G. Aubert, "Shape gradient for image segmentation using information theory," in *ICASSP*, may 2004, vol. 3, pp. 21–24.

[4] W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Medical Image Analysis*, vol. 1, no. 1, pp. 35–51, 1996.

[5] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, 1997.

[6] M.C. Delfour and J.P. Zolésio, *Shape and geometries*, Advances in Design and Control, SIAM, 2001.

[7] S. Jehan-Besson, M. Barlaud, and G. Aubert, "DREAM^{2}S: Deformable regions driven by an eulerian accurate minimization method for image and video segmentation," *IJCV*, vol. 53, no. 1, pp. 45–70, 2003.

[8] G. Aubert, M. Barlaud, O. Faugeras, and S. Jehan-Besson, "Image segmentation using active contours: Calculus of variations or shape gradients ?," *SIAM Applied Mathematics*, vol. 1, no. 2, pp. 2128–2145, 2003.

[9] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, pp. 321–332, 1988.

[10] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997.

[11] S. Zhu and A. Yuille, "Region competition: unifying snakes, region growing, and bayes/MDL for multiband image segmentation," *PAMI*, vol. 18, pp. 884–900, september 1996.

[12] J. Kim, J.W. Fisher, M. Cetin, A.J. Yezzi, and A.S. Willsky, "Incorporating complex statistical information in active contour-based image segmentation," in *ICIP*, 2003, pp. II: 655–658.

[13] F. Precioso, M. Barlaud, T. Blu, and M. Unser, "Smoothing B-spline active contour for fast and robust image and video segmentation," in *ICIP*, Barcelona, Spain, September 2003, IEEE, vol. 1, pp. 137–140.