A New Approach to Image Segmentation

Antonio Succi and Vincent Torre

Dipartimento di Fisica dell'Università di Genova

Via Dodecaneso 33, 16146 Genova, Italy

e-mail: succi@genova.infn.it

Introduction

The segmentation of an image ([1], [3], [2]) is a major problem in Machine Vision, usually representing the initial step for high-level processing stages such as recognition. Its purpose is to partition the image into distinct regions, each corresponding to an area which is homogeneous in some sense; these regions are usually assumed to represent distinct relevant “objects” in the scene. This problem has been addressed in several ways, including very specialized and *ad hoc* procedures.

In this paper a general scheme based on a “mechanical” (deterministic) approach is outlined; segmentation is defined as a solution to the problem of finding a configuration of static equilibrium for an ensemble of “bubbles” interacting with external fields in various ways. In this aspect, the method has points in common with a version of the Mumford & Shah approach ([4]) and with *snake* models ([7], [6]).

Minimization of the energy functional for the bubble ensemble is achieved by means of a deterministic Iterated Conditional Mode (ICM) algorithm in a multi-scale (pyramidal) scheme ([5]); this choice was motivated by its good properties in terms of convergence speed and relative insensitivity to the presence of local minima of the functional.

The segmentation is achieved in two steps. In the first one a coarse segmentation is produced and the maximum number of different bubbles is determined; in the second stage the segmentation is refined and a precise localization of the bubble boundaries is obtained.

The proposed method has been shown to be able to segment a large variety of images, both static and dynamic. Experiments have been carried out with good results on synthetic and natural textures, aerial images, biomedical images, road images and road sequences.

1 The Segmentation Model

1.1 Energy Functional

Let us consider an elastic 2D bubble whose generic boundary element is assumed to experience three kinds of forces: a hydrostatic force directed along the normal, an external force with an arbitrary direction (referred to as an “electrostatic” force) and finally a pair of tangential forces caused by surface tension and acting on the element endpoints.

The newtonian equation for the equilibrium of the bubble boundary can be obtained as the Euler-Lagrange equation associated to the following energy functional:

$$U[\phi] = -p \int_{\Omega} d^2 \mathbf{x} + \int_{\mathcal{P}} \mathcal{F}(\|\phi'(\xi)\|) d\xi + \int_{\mathcal{P}} V(\phi(\xi)) d\xi \quad (1)$$

where $\mathcal{P}$ is the parameter domain, Ω is the region whose boundary is represented by the curve $\phi(\xi)$, V is a potential for the external force field $\mathbf{f}$, p is the hydrostatic pressure (positive if inflating), and $\mathcal{F}$ is the potential for the tension forces.

The second term (surface tension) is a regularizing term, which promotes smooth boundaries. The third (electrostatic) term will be used to take account of intensity edges, by making V an attractive potential generated by intensity edges. The first (hydrostatic) term represents the internal energy of the bubble: its effect (with $p > 0$) is to promote configurations such that the bubble covers the largest surface available. In the absence of other terms the bubble inflates indefinitely. In order to obtain a meaningful image segmentation, the internal energy term must be appropriately modified so as to promote configurations in which the bubble doesn't cover just the largest area, but rather the largest *homogeneous* area. The internal energy takes the following form:

$$U_{int} = \Phi_{\Omega}[\psi] \quad (2)$$

where Φ_{Ω} is a functional with a support in the region inside the closed curve $\partial\Omega$, and ψ is the image intensity field. The choice of this functional defines the criteria of homogeneity of an area of the image.

The energy functional for a many-bubble system becomes:

$$U = \sum_k \Phi_{\Omega_k}[\psi] + \frac{1}{2} \int_{\mathcal{P}_k} \mathcal{F}(\|\phi'_k(\xi)\|) d\xi + \frac{1}{2} \sum_k \int_{\mathcal{P}_k} V(\phi_k(\xi)) d\xi \quad (3)$$

where k index runs over the set of bubbles.

1.2 External Forces And Surface Tension

The external force term can be defined by assuming the following form for the external potential:

$$V(\mathbf{x}) = -\epsilon(G_{\sigma} * E)(\mathbf{x}) \quad (4)$$

where E is the edge map extracted from the original image, ϵ is a weight coefficient, and G_{σ} is a bidimensional gaussian function with width σ and unitary amplitude:

$$G_{\sigma}(\mathbf{x}) = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}\|^2\right) \quad (5)$$

Expression (4) represents a short-range attractive potential field generated by the intensity edges; it causes the bubble boundaries to be attracted by intensity edges which are sufficiently close.

A suitable choice for the regularizing term (surface tension) is obtained by taking the function $\mathcal{F}$, describing the elastic properties of the boundary, to be linear:

$$U_{s.t.} = \lambda \int_{\mathcal{P}} \|\phi'(\xi)\| d\xi \quad (6)$$

This is the form corresponding to the surface tension energy of a physical bubble of fluid.

Other useful choices for the function $\mathcal{F}$ are quadratic or higher-degree polynomials; the corresponding energy term is the continuum limit of a chain of linear (Hooke) or higher order springs. As with snakes, a curvature-dependent term could be added, thus obtaining a higher degree of regularity for the boundaries.

2 Implementation Details

Segmentation of an image is achieved by minimizing the energy functional for the bubble ensemble with respect to bubble creation/destruction, merging of adjacent bubbles, deformation of bubble boundaries, and variation of the internal parameters (if *a priori* models are employed). Minimization is performed by means of a pyramidal deterministic scheme as in [5]; at a generic resolution, the bubble ensemble domain is partitioned in a set of pixel blocks of side n , each bubble being represented by a subset of these blocks.

In the case in which no *a priori* model is employed, the computation of internal energy requires the subdivision of each bubble into windows where image attributes are evaluated. The window size is held fixed at a value that allows a meaningful computation of attributes.

The multiscale optimization algorithm is organized as two distinct modules operating at different resolutions. The first one terminates when the block size is the same as the window size, thus generating a coarse approximation of the segmented image. The second one takes this coarse segmentation as its input state and carries on the minimization process up to a resolution in which the block size is 2×2 pixels. The coarse segmentation defines an upper limit for the number of regions appearing in the final result, since no new regions are generated by the high-resolution stage. The two stages differ in some points, as explained in the following.

In the coarse segmentation stage, energy minimization is performed with respect to the complete set of possible variations (deformation, creation, merging). The minimization procedure is structured as follows.

First, a raster scan is performed over the set of blocks in which the image is partitioned at a given resolution. For each block, “virtual” updates are performed on its label and the consequent energy variations are computed; the following updates are considered:

- the new label is one of the labels of the adjacent blocks (local deformation of the boundary);
- the new label is one not appearing in the current configuration (creation of a new bubble);
- the new label is the same as the old one (no operation).

The update corresponding to the lower (most negative) energy variation is actually performed. The scan is iterated until the number of blocks whose label is actually updated becomes lower than a given percentage of the total block number; being deterministic, the process is convergent.

After the block-wise minimization, a region merging is performed: first, the energy variations induced by the merging of adjacent pairs of bubbles are computed. Bubble pairs are sorted in order of the associated energy variations, and then merged in this order; at each merge the energy variations are recomputed and the list is updated. The process terminates when no bubble pair gives a negative energy variation when virtually merged.

The energy functional is represented in this stage by the internal energy term alone, the other terms being considered as accessory at low resolution.

In the high-resolution stage the energy functional is written in its complete form. Region creation is not allowed: energy minimization is performed only with respect to local updates of the bubble boundaries and possibly region merging, that can be optionally inhibited. The minimization procedure operates along the same lines as before, the main differences being that the block scan in the block-wise minimization is performed only on the bubble boundaries, and no new labels are considered in the virtual updates.

When making use of *a priori* models, each bubble is characterized not only by its shape but also by the value of its model-related parameters, which become involved in the minimization problem together with the bubble boundaries. This is true in particular for segmentation in the sense of motion.

It is difficult to perform the minimization with respect to the complete set of variables of the problem, starting from an arbitrary state. The task is easier if an approximate state is already available, under the form of a low-resolution segmentation. In this case an approximate value θ^* for the bubble parameters is first computed by minimizing only the internal energy term with respect to these variables, having fixed the bubble configuration to the one identified by the approximate segmentation:

$$\theta_k^* = \arg \min_{\theta_k} \Phi_{\Omega_k}[\psi, \theta_k] \quad (7)$$

After completing the initial parameter computation, the minimization with respect to bubble deformations is performed with the parameters being fixed to the computed approximate value. When the convergence is reached, the parameters are recomputed and the next step in the multiresolution process is started; this defines a sub-optimal algorithm, whose convergence properties are strongly conditioned by the choice of the starting state (presegmentation). The algorithm terminates when convergence is reached at the resolution of 2×2 pixels.

The approximate segmentation representing the starting state can be obtained in several ways; a possibility is to use the low-resolution stage with no *a priori* models, but other alternatives are possible. In the experiments, simple thresholding methods have often been employed.

3 Experimental Results

Only a little number of the results of the experiments performed with our segmentation method are here reported, due to lack of space.

Figure 1-A shows an echocardiographic image; for this image it was useful to have a binary segmentation into cavities and tissues. This was achieved by employing a binary presegmentation as input of the high-resolution stage. The presegmentation was obtained by a thresholding method at low resolution; the threshold was determined as the main valley of the bimodal histogram of the whole image.

Figure 1-B shows the final segmentation at 2×2 resolution. The boundaries of the detected regions are shown superimposed to the original image. The image attributes used were the first three moments of the histogram computed over 8×8 windows. Parameters were set to $\lambda = 2$, $\epsilon = 1$. Figure 2-A shows a detail

Fig. 1. Echocardiographic image and its segmentation.

of an aerial image obtained by a SAR image after a suitable preprocessing to reduce the amount of speckle noise due to coherent illumination.

The simple two-parameters model proved to be useful to discriminate the different regions present in this image. Figures 2-B shows the final segmentation at 2×2 resolution. Parameters were set to $\lambda = 3$, $\epsilon = 1$.

Figure 3 shows an example of motion segmentation. Figure 3-A shows a frame of a road sequence taken by a camera mounted on a vehicle moving toward the horizon; the car on the left is approaching. Segmentation was performed on a limited square zone of the image surrounding the approaching car. The pre-segmented image (not shown) was obtained by a thresholding method applied on a sparse optical flow field: a threshold was put on the modulus of the difference between the computed optical flow and the quadratic theoretical flow corresponding to the planar motion on the tarred road.

Figure 3-B shows the result of the final segmentation; parameters were set to $\lambda = 3$, $\epsilon = 2$.

Fig. 2. SAR aerial image and its segmentation.

Fig. 3. A frame of a road sequence and the result of motion segmentation.

References

1. R. M. Haralick and L. G. Shapiro, "Image segmentation techniques", *Comput. Vis. Graph. Image Process.*, vol. 29, 1985
2. E. Riseman and M. Arbib, "Segmentation of static scenes", *Comput. Graphics Image Process.*, vol. 6, 1977
3. S. Zucker, "Region growing: Childhood and adolescence", *Comput. Graphics Image Process.*, vol. 5, 1976
4. D. Munford and J. Shah, "Optimal approximation by piecewise smooth functions and variational problems", *Communication on Pure and Applied Mathematics*, vol. 42, No 5, 1988
5. A. Ackah-Miezan and A. Galalowicz, "Discrete models for energy-minimizing segmentation", *Proceedings of the 4th International Conference on Computer Vision*, 1993
6. L. D. Cohen and I. Cohen, "A finite element method applied to new active contour models and 3D reconstruction from cross sections", *Technical report 1245*, INRIA, June 1990
7. A. Kass, A. Witkin and D. Terzopoulos, "Snakes: active contour models", *Int. J. Computer Vision.*, vol. 1, No. 3, 1988