

Region Competition: Unifying Snakes, Region Growing, Energy/Bayes/MDL for Multi-band Image Segmentation

S.C. Zhu, T.S. Lee and A.L. Yuille

Division of Applied Sciences, Harvard University
Cambridge, MA 02138 USA
zhu@metatron.harvard.edu

ABSTRACT

We present a novel statistical and variational approach to image segmentation based on a new algorithm named *region competition*. This algorithm is derived by minimizing a generalized Bayes/MDL (Minimum Description Length) criterion using the variational principle. We show that existing techniques in early vision such as, snake/balloon models, region growing, and Bayes/MDL are addressing different aspects of the same problem and they can be unified within a common statistical framework which combines their advantages. We analyze how to optimize the precision of the resulting boundary location by studying the statistical properties of the region competition algorithm and discuss what are good initial conditions for the algorithm. Our method is generalized to color and texture segmentation and is demonstrated on grey level images, color images and texture images.

1. INTRODUCTION

Image segmentation is a critical problem of early vision and hence is one of the most intensively studied areas. In the past two decades researchers have developed several alternative approaches. These approaches can be roughly classified into four classes: (i) Local filtering approaches such as the Canny edge detector [2], (ii) Snake [8] and Balloon methods [4], (iii) Region growing and merging techniques [12] [15], and (iv) Global optimization approaches based on energy functions [14] or Bayesian [6] [1] and MDL criteria [10] [7]. A common property of these approaches is that they all make *hypotheses* about the image, *test* features and make *decisions* by applying thresholds explicitly or implicitly. A major difference between the four approaches lies in the domains on which the hypotheses, tests, and decisions are based. All these approaches all have certain drawbacks. The filtering approach only makes use of local information and cannot guarantee continuous closed edge contours. Snake/balloon models make use only of information along

It is a pleasure to acknowledge many suggestions and helpful discussions with David Mumford, and Y.N.Wu. The first author was supported by the NSF grant DMS-91-21266 to David Mumford. This research was supported in part by the Brown/Harvard/MIT Center for Intelligent Control Systems with U.S. Army Research Office grant number DAAL03-86-K-0171. The authors would also like to thank ARPA for an Air Force contract F49620-92-J-0466.

their boundaries and require good initial estimates to yield correct convergence. An advantage of region growing is that it does make use of the image statistics inside regions, however it often generates irregular boundaries and small holes. In addition, all these three methods lack a global criterion for segmenting the entire image. By contrast, energy/Bayes/MDL have global criteria but it is often very difficult to find their minima.

We have developed a common statistical framework for image segmentation. Our algorithm, which we call *region competition*, is derived by minimizing a generalized Bayes/MDL criterion using the variational principle, and it combines the most attractive features of snake/balloon model and region growing. Thus we show that three of the standard approaches – (i) snakes and balloons, (ii) region growing by using statistical tests such as Fisher's test, (iii) Energy, Bayesian and MDL methods – are addressing different aspects of the same problem and can be unified within this framework. Our work is close in spirit to recent work by [3] and [11].

We analyze the precision of the resulting boundaries by studying the statistical behavior of the region competition algorithm. Then we discuss the criteria for defining "good" initial conditions, and demonstrate an interesting domino effect.

Our approach is applied both for obtaining global segmentation of images and for finding individual regions and is illustrated on two grey level images. We then generalized our approach to deal with color, by using a novel color model which helps detect specularities, and to texture. Results are demonstrated for both color and texture segmentation.

This paper is organized as follows. Section (2) sets the scene by briefly describing Snakes and Balloons, Region Growing and Energy/Bayes/MDL. In section (3) we introduce our approach for grey level images and describe the implementation results. In section (4), we analyze the precision of the boundary location and discuss how to choose the windows and initial conditions. Section (5) extends our approach to color images, thereby allowing us to filter out intensity gradients, and gives results for these cases. In section (6), we apply our method to texture images. Finally section (7) discusses limitations and possible extensions.

2. SNAKES, REGION GROWING AND ENERGY/BAYES/MDL

This section will briefly review the properties of snakes, region growing and Energy/Bayes/MDL.

2.1. Snakes and Balloons

A snake [8] is an active contour defined by $\Gamma(s) = (x(s), y(s))$, where s is the arc length of the contour¹. Here we assume that $\Gamma(s)$ is the closed boundary of a region R (i.e. $\Gamma(s) = \partial R$).

A typical energy for a snake is:

$$E[\Gamma(s)] = \oint_{\Gamma(s)} \left\{ \frac{1}{2}(\alpha |\Gamma_s|^2 + \beta |\Gamma_{ss}|^2) - \lambda |\vec{\nabla} I \cdot \vec{\nabla} I| \right\} ds \quad (1)$$

which we minimize by steepest descent:

$$\frac{d\Gamma(s)}{dt} = -\frac{\delta E}{\delta \Gamma} = -\alpha \Gamma_{ss} + \beta \Gamma_{ssss} + \lambda \vec{\nabla} |\vec{\nabla} I \cdot \vec{\nabla} I|, \quad (2)$$

where $\Gamma_{ss} = (\ddot{x}(s), \ddot{y}(s))$, etc.

The balloon models[4] are motivated by the desire to drive the snake automatically to a good initial position. They introduce an additional force term $\nu \vec{n}(s)$ to Equation (2) which pushes the contour out (or in) along its normal $\vec{n}(s) = (\dot{y}, -\dot{x})$.

This additional force term can be derived from an additional energy term $E_{add}[\Gamma(s)] = -\nu \iint_R dA$. So the energy for the balloon is:

$$E[\Gamma(s)] = \oint_{\Gamma} \left\{ \frac{1}{2}(\alpha |\Gamma_s|^2 + \beta |\Gamma_{ss}|^2) - \lambda |\vec{\nabla} I \cdot \vec{\nabla} I| \right\} ds - \nu \iint_R dA \quad (3)$$

Thus the balloon tries to maximize its area while smoothing its bounding contour and maximizing the intensity gradient along the contour.

2.2. Region Growing and Merging

Region merging works by building up complicated regions by combining smaller regions based on a statistical similarity test. A popular choice is Fisher's test. For example, suppose there are two adjacent regions R_1 and R_2 , where $n_1, n_2, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2$ are the sizes, sample means and sample variances of R_1, R_2 respectively. Then in order to decide whether or not to merge them, we can look at the squared *Fisher distance*:

$$\frac{(n_1 + n_2)(\hat{\mu}_1 - \hat{\mu}_2)^2}{n_1 \hat{\sigma}_1^2 + n_2 \hat{\sigma}_2^2} = \frac{n \hat{\sigma}^2}{n_1 \hat{\sigma}_1^2 + n_2 \hat{\sigma}_2^2} - 1, \quad (4)$$

where $n = n_1 + n_2$ and $\hat{\sigma}^2$ is the sample variance of the mixture region. If this statistic is below a certain threshold then the regions are merged.

Region growing can be considered as a special case of region merging, where R_1 is the growing region, while R_2 is a single pixel at the boundary of R_1 , i.e. $n_2 = 1$ and n_1 is very large (say $n_1 > 100$). In this case we can treat

¹The contour can also be represented parametrically, for example by B-splines or Fourier coefficients, but we will not consider such variants in this paper.

$\mu = \hat{\mu}_1, \sigma^2 = \hat{\sigma}_1^2$, and $\mu_2 = I_{(x,y)}$ (the intensity at point (x, y)), and approximate the squared Fisher distance by: $\frac{(I - \mu)^2}{\sigma^2}$.

A variant of region growing [12] is to fit the intensity within each region to a parameterized model, such as a plane or a quadratic form. Then tests like Equation (4) can be applied to the residuals of the boundary pixels after fitting.

Although region growing algorithms are very intuitive they can rarely be proven to converge to the minimum of some global cost function, and the resulting regions may have noisy boundaries. Another drawback results from the Fisher's test, see Equation (4), which cannot distinguish between two distributions with the same means but different variances. This problem will be discussed in a later section.

2.3. Energy, Bayes and MDL

Both Bayes and MDL specify ways for segmenting an entire image using an energy function criteria. These two approaches are motivated by different considerations but it is straightforward to transfer an MDL criterion into a Bayesian one and vice versa[10].

Bayesian estimation is a common approach to image segmentation [6]. The observed image is modeled as being a degraded version of an ideal image which is assumed to be piecewise smooth. For example, the energy function used in Mumford and Shah [14], and in Blake and Zisserman [1] (i.e. the *weak membrane model*) is:

$$E[f, \Gamma] = \mu \iint_R (f - I)^2 dx dy + \lambda \iint_{R-\Gamma} \|\nabla f\|^2 dx dy + \nu |\Gamma|, \quad (5)$$

where I is the input image, f is the output image, and Γ labels the discontinuities. It is easy to see that when $\lambda \mapsto \infty$ this reduces to a cartoon model:

$$E[\{f_i\}, \Gamma] = \sum_i \iint_{R_i} \frac{(f_i - I)^2}{\sigma^2} dx dy + \nu |\Gamma|, \quad (6)$$

where f_i is constant within each region R_i and $\sigma^2 = \frac{1}{\mu}$ is constant over the entire image.

A typical MDL criterion occurs in Equation (7). It differs from Equation (6) by letting the σ^2 's be unknown variables which are assumed to be constant within each region.

It is usually very difficult to minimize the energy functions resulting from Bayes/MDL models. Algorithms such as simulated annealing [6], graduated non-convexity [1] and deterministic annealing [5] are perhaps the most successful.

3. REGION COMPETITION FOR GREY LEVEL IMAGES

3.1. From MDL to A Unified Framework

The goal of image segmentation is to partition the image into subregions with homogeneous properties (intensity, color or texture), such regions will typically correspond to objects or object parts in the scene. Before describing the algorithm, we first need a definition of homogeneity.

In this section a region is considered to be homogeneous if its intensity values are consistent with having been generated by one of a family of pre-specified probability distributions $P(I : \alpha)$, where α are the parameters of the distribution. We assume that the probability models may be different in different regions. In later sections, we will describe how the probability distributions $P(I : \alpha)$ can be generalized to multi-band features extracted from the images, and to the residues after fitting certain parameterized models in which case α may include the parameters specifying those physical models. Now suppose that the entire image domain R has been initially segmented into M piecewise "homogeneous" underlying regions $\{R_i\}, i = 1, 2, \dots, M$, i.e. $R = \cup_{i=1}^M R_i$, $R_i \cap R_j = \emptyset$, if $i \neq j$. Let ∂R_i be the boundary of region R_i , where we define the direction of ∂R_i to be counter-clockwise, i.e. when we travel along the boundary, R_i is on the left hand side. Let $\Gamma = \cup_{i=1}^M \Gamma_i$ be the edges or segmentation boundaries of the entire image with $\Gamma_i = \partial R_i$.

Our MDL criterion corresponds to a global energy functional which is the continuum limit of Leclerc's [10] for an appropriate choice of the family of probability distributions. This gives:

$$E[\Gamma, \{\alpha_i\}] = \sum_{i=1}^M \left\{ \frac{\mu}{2} \oint_{\partial R_i} ds - \iint_{R_i} \log P(I_{(x,y)} : \alpha_i) dx dy + \lambda \right\}, \quad (7)$$

where the first term within the braces is the length of the boundary curve ∂R_i for region R_i . Without prior knowledge about the shape of the region, we simply assume that the code length is proportional to the curve length where μ is the code length for unit arc length. Since all edge segments are shared by two adjacent regions, we divide the first term by a factor of 2. The second term is the sum of the cost for coding the intensity of every pixel (x, y) inside region R_i according to a distribution $P(I_{(x,y)} : \alpha_i)$ in the sense of optimal coding theory. λ is the code length needed to describe the distribution and coding system for region R_i , we simply assume λ is the same for all regions.

Because the energy E in equation (7) depends on two groups of variables – the segmentation Γ and the parameters α_i 's, we propose a greedy algorithm which minimizes the energy functional by iterating two stages.

In the first stage, we fix Γ , in other words, we fix R_i and solve for the α_i , $i = 1, 2, \dots, M$ to minimize the description cost for each region. This corresponds to setting:

$$\alpha_i^* = \arg \min_{\alpha_i} \left\{ - \iint_{R_i} \log P(\alpha_i | I_{(x,y)}) dx dy \right\}, \quad \forall i \quad (8)$$

in the discrete case:

$$\alpha_i^* = \arg \max_{\alpha_i} P(\alpha_i | \{I_{(x,y)}, \forall (x,y) \in R_i\}) \quad \forall i$$

In other words, the α_i 's are estimated by maximizing the conditional probabilities. For many distributions this can be done analytically. For example, the α_i^* 's are simply the sample mean and sample variance over pixels inside R_i if $P(I : \alpha_i)$ is a Gaussian distribution.

In the second stage, we fix the $\{\alpha_i\}$, thus $P(I : \alpha)$ becomes the conditional probability distribution $P(I | \alpha)$

and then do steepest descent with respect to Γ . For any point $\vec{v} = (x, y)$ on the edge Γ :

$$\frac{d\vec{v}}{dt} = - \frac{\delta E[\Gamma, \{\alpha_i\}]}{\delta \vec{v}} \quad (9)$$

where the right hand side is (minus) the functional derivative of the energy E .

Taking the functional derivative by applying the variational principle [16], yields the motion equation for point \vec{v} :

$$\frac{d\vec{v}}{dt} = \sum_{k \in Q(\vec{v})} \left\{ -\frac{\mu}{2} \kappa_{k(\vec{v})} \vec{n}_{k(\vec{v})} + \log P(I_{(\vec{v})} | \alpha_k) \vec{n}_{k(\vec{v})} \right\} \quad (10)$$

where $Q(\vec{v}) = \{k \mid \vec{v} \text{ lies on } \Gamma_k\}$, i.e. the summation is done over all regions R_k such that \vec{v} is on Γ_k . $\kappa_{k(\vec{v})}$ is the curvature of Γ_k at point \vec{v} and $\vec{n}_{k(\vec{v})}$ is the unit normal to Γ_k at point \vec{v} . By the direction of each Γ_k defined at the beginning of this section, \vec{n}_k will point out of R_k .

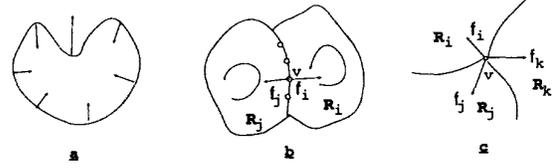


Figure 1: The forces acting on the contour: a. the smoothing forces, b. the statistics force at a boundary point. c. the statistics forces at a junction point.

The right side of equation (10) has a simple intuitive interpretation. It represents two kinds of "forces" acting at point \vec{v} on the contour, both pointing along the normal. The first term, the *smoothing force*, is strongest at points of high curvature. Figure (1.a) shows the smoothing force at points along the region boundary. This force is independent of the direction of the curve and it tries to make the curve as straight as possible. The second term is the *statistics force*, $\vec{f} = \log P(I | \alpha) \vec{n}$. Since $\log P \leq 0$, the statistics forces always compresses the region. The better the point \vec{v} satisfies the homogeneity requirement, the larger $P(I | \alpha)$, and hence the weaker the statistics force.

For example, as shown in Figure (1.b), \vec{v} is a point on the common boundary of region R_i and R_j . Since curves Γ_i and Γ_j have inverse normal vectors at \vec{v} , $\vec{n}_i = -\vec{n}_j$ and $\kappa_i \vec{n}_i = \kappa_j \vec{n}_j$. The motion equation for \vec{v} is:

$$\frac{d\vec{v}}{dt} = -\mu \kappa_{i(\vec{v})} \vec{n}_{i(\vec{v})} + \log \left(\frac{P(I_{(\vec{v})} | \alpha_i)}{P(I_{(\vec{v})} | \alpha_j)} \right) \vec{n}_{i(\vec{v})} \quad (11)$$

It is easy to see that the smoothing term by itself is the *Euclidean geometric heat flow* equation used for curve smoothing and evolution, and it is equal to the following heat diffusion equation.

$$\frac{d\Gamma}{dt} = \frac{\partial^2 \Gamma}{\partial s^2} \quad (12)$$

where s is the arc length of the curve Γ . Detailed discussion about the properties of this equation is found in [13].

Besides the smoothing term, the motion of \vec{v} is determined by the likelihood ratio test. If $P(I_{(\vec{v})}|\alpha_i) > P(I_{(\vec{v})}|\alpha_j)$, i.e. the intensity at \vec{v} fits better to the distribution of region R_i than to that of region R_j , then the boundary will move in the direction of \vec{n}_i . Similarly for points which are shared by several region boundaries, Figure (1.c) show the statistics forces at a junction point. Intuitively, adjacent regions compete for ownership of pixels along their boundaries, subject to the smoothness constraint, and motion of the edges are mainly determined by the probability distributions of each region and the properties of the edge. This is why we call our algorithm *region competition*.

Now we find it easy to unify region growing, the balloon model and MDL in a common statistics framework.

Firstly, the region growing algorithm can be considered to be a degenerate case of region competition where we treat the growing region as R_1 , with $P(I : \alpha_1)$ chosen according to our desired homogeneity criterion, and the background region as R_0 (so that $R_1 \cup R_0 = R$) which has uniform probability distribution P_0 . Then the motion equation for each point \vec{v} along the boundary is:

$$\frac{d\vec{v}}{dt} = (\log P(I_{(\vec{v})}|\alpha_1) - \log P_0)\vec{n}_{(\vec{v})} \quad (13)$$

where \vec{n} is the normal of the region contour. This directly corresponds to region growing where people test the probability $P(I_{(\vec{v})}|\alpha_1)$, and compare it with the absolute threshold P_0 . Fisher's test and the χ^2 test [12] correspond to particular choices of $P(I_{(\vec{v})}|\alpha_1)$.

Secondly, the underlying statistical assumption for the balloon model is even simpler, it treats both the current region R_1 and the background R_0 as uniform distributions P_1, P_0 with $\nu = \log P_1 - \log P_0$. According to the balloon model, taking the derivatives of equation (3), we have the motion equation for \vec{v} on the boundary:

$$\frac{d\vec{\Gamma}_{(s)}}{dt} = -\alpha\Gamma_{ss} + \beta\Gamma_{ssss} + (\log P_1 - \log P_0)\vec{n}_{(s)} + \lambda\vec{\nabla} \left| \vec{\nabla} I \cdot \vec{\nabla} I \right| \quad (14)$$

In this equation, the smoothing force is only slightly different from that in Equation (11), observe that $\kappa = \vec{\Gamma}_{ss} \cdot \vec{n}$ (i.e. the curvature is the projection of Γ_{ss} along the normal). In equation (14), the threshold for deciding where to stop moving the contour is provided by the local edge measurement—the last term. Since the balloon model does not use the statistics inside the regions it often fails to segment images into homogeneous regions.

In summary, we can compare the region growing, balloon models and Bayes/MDL (as well as the filter approaches) in a common three-stage statistics framework: hypotheses, tests, and decisions. The differences between these approaches can be easily found by studying what kind of hypotheses they have made for the input images, how they test these hypotheses, and how they make the decision. The MDL criterion, which we adopt, assumes that images consist of regions with smooth boundaries and homogeneous properties defined by a family of probability distributions. For each region it tests the statistics $P(I : \alpha)$ which can be specified by any test such as Fisher's test, the χ^2 test etc. It decides the position of the boundary by the likelihood ratio

instead of choosing a threshold. Moreover, region competition provides a global criterion for segmenting the entire image, and it is faster than the annealing approaches.

3.2. Generalizing the MDL Criterion for Region Competition

To proceed further we need to specify a family of probability distributions $P(I : \alpha)$. In this paper we will consider Gaussian distributions. Thus we set $\alpha = (\mu, \sigma)$, and

$$P(I_{(x,y)} : (\mu, \sigma)) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(I_{(x,y)} - \mu)^2}{2\sigma^2}\right\}.$$

However, an analysis of the motion Equations (10,11) reveals some underlying disadvantages for the greedy algorithm that we proposed.

First, the statistics force at each boundary point (x, y) will depend on $P(I_{(x,y)}|\mu, \sigma)$, in other words on the probability that $I_{(x,y)}$ on the boundary can be generated by a Gaussian distribution $N(\mu, \sigma^2)$. This force seems plausible but since only a single sample is taken from this distribution there is a reasonable chance that $I_{(x,y)}$ will lie on one of the tails of the distribution. The statistics force therefore may be overly sensitive to fluctuations in the image.

Second, in the extreme case where two distributions have the same mean but different variances, the classification error will be intolerable. One such image is shown in Figure (2.a). In such cases, we need to measure the second order statistics (variance) near each boundary point (x, y) in order to tell whether the point (x, y) should be classified as belonging to one distribution or to the other. As we mentioned earlier, this cannot be done by Fisher's test. Furthermore, if the distribution $P(I|\alpha)$ is more sophisticated than the Gaussian distribution then more complex statistics must be computed in the neighborhood of (x, y) .

At this point we should make it clear that the problems of misclassification are not the fault of the original energy functional (equation (7)). In fact, we can prove even in cases where intensities in two adjacent regions have same mean but different variances, the global minimum of equation (7) still gives the correct segmentation. The problems of misclassification are due to the greediness of the algorithm, and the difficulty of finding the global minimum of equation (7).

Our method for dealing with these problems is to use a window of m pixels around each boundary point (x, y) . We call these pixels the neighbor set of (x, y) denoted by $\mathcal{W}_{(x,y)}$. The effect of this is to replace $P(I_{(x,y)} : \alpha)$ by the joint probability: $\prod_{(u,v) \in \mathcal{W}_{(x,y)}} P(I_{(u,v)} : \alpha)$, divided by the size of window m .

Correspondingly, the energy function becomes:

$$E[\Gamma, \{\alpha_i\}] = \sum_{i=1}^M \left\{ \frac{\mu}{2} \oint_{\partial R_i} ds - \iint_{R_i} \frac{1}{m} \iint_{\mathcal{W}_{(x,y)}} \log P(I_{(u,v)} : \alpha_i) dudvdx dy + (\lambda\mathfrak{J}) \right\}$$

The window \mathcal{W} solves our problems. First we observe that the larger m is, then the bigger the chance that it is representative of the distribution, and the smaller the risk of misclassification. Second with such samples, we can do

higher order statistics over $\{I_{(u,v)} \mid (u,v) \in \mathcal{W}_{(x,y)}\}$ to answer the second problem raised earlier. But the window size m cannot be too big, otherwise the boundary will not be located precisely. Later in section (4) we will further analyze the relationship between the precision of the boundary and the choice of windows.

To interpret equation (15) in the sense of MDL, we observe that each pixel is coded m times, then the final code length is divided by m . Therefore by exchanging the order of integral in the second term, we can rewrite equation (15) as:

$$E[\Gamma, \{\alpha_i\}] = \mu \oint_{\Gamma} ds - \iint_R \sum_{i=1}^M \pi_{i(x,y)} \log P(I_{(x,y)} : \alpha_i) dx dy + \lambda M \quad (16)$$

where $\Gamma = \cup_{i=1}^M \partial R_i$, and $\pi_{i(x,y)} = \frac{\|R_i \cap \mathcal{W}_{(x,y)}\|}{m}$ such that $\sum_{i=1}^M \pi_{i(x,y)} = 1$ for all (x,y) . In other words, each pixel (x,y) in the image is coded by a mixture distribution. Detailed discussion is given in [16].

In the rest of this section, we will explore equation (15) by assuming Gaussian distributions.

Suppose we have window samples $\{I_{(u,v)} \mid (u,v) \in \mathcal{W}_{(x,y)}\}$, with each $I_{(u,v)} \sim N(\mu, \sigma^2)$, and $\bar{I}_{(x,y)}$ and $S_{(x,y)}^2$ are the sample mean, and sample variance for $\{I_{(u,v)}\}$.

Again by applying the variational principle[16], we can rewrite the statistics force generated by a single region R_i at (x,y) in the direction of \vec{n}_i , as the following:

$$\begin{aligned} & \frac{1}{m} \iint_{\mathcal{W}_{(x,y)}} \log P(I_{(u,v)} \mid \alpha_i) du dv \\ &= \frac{1}{m} \log \prod_{(u,v) \in \mathcal{W}_{(x,y)}} P(I_{(u,v)} \mid \mu_i, \sigma_i^2) \\ &= -\frac{1}{2} \left[\log(2\pi\sigma_i^2) + \frac{(\bar{I}_{(x,y)} - \mu_i)^2}{\sigma_i^2} + \frac{S_{(x,y)}^2}{\sigma_i^2} \right] \quad (17) \end{aligned}$$

From equation (17), we observe: (i) The second term tests the mean. Since $\bar{I}_{(x,y)} \sim N(\mu, \sigma^2/m)$, (i.e. the variances have been divided by m), the misclassification risk is greatly reduced. (ii). The third term tests the variance. With the addition of this ‘‘variance’’ force, we can detect two regions with the same means but different variances. Equation (17) will be generalized to higher dimension in the next section (see Equation (20)).

Now we can derive the motion equation for point \vec{v} at the boundary $\Gamma_i \cap \Gamma_j$ by plugging equation (17) into the motion equation (11).

$$\begin{aligned} \frac{d\vec{v}}{dt} &= -\mu \kappa_{i(\vec{v})} \vec{n}_{i(\vec{v})} - \frac{1}{2} \left\{ \log \frac{\sigma_i^2}{\sigma_j^2} \right. \\ &+ \left. \left(\frac{(\bar{I} - \mu_i)^2}{\sigma_i^2} - \frac{(\bar{I} - \mu_j)^2}{\sigma_j^2} \right) + \left(\frac{s_i^2}{\sigma_i^2} - \frac{s_j^2}{\sigma_j^2} \right) \right\} \vec{n}_i \quad (18) \end{aligned}$$

3.3. Simulations on Grey Level Images

In this subsection, we illustrate the region competition algorithm on two grey level images.

As shown in figure (2.a), image 1 is a 100×100 image partitioned into two regions by a S-shaped curve. The intensities in these two regions are generated randomly from two

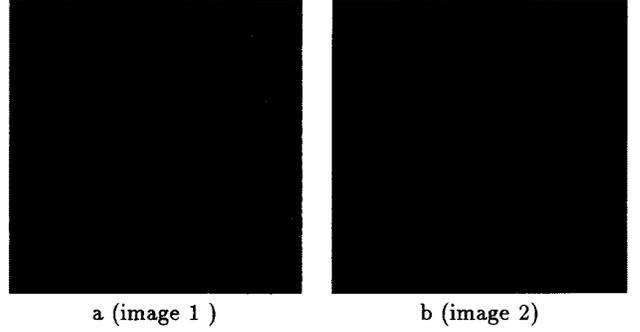


Figure 2: Two typical grey level Images

Gaussian distributions with identical means: $N(128, 10^2)$ and $N(128, 35^2)$. Suppose the image is initially segmented by a slanted straight line, see Figure (3.a). Then the motion of the regions at iteration steps $t = 0, 20, 30, 50$ are shown in Figure (3.a.b.c.d) respectively. Finally Figure (3.e) shows how the boundary moves if we drop the last term in Equation (18). Thus if we do not test the variance then the edge just moves to the low-right corner and disappears. This results in an incorrect segmentation with the image being perceived as a single region. In this image, \mathcal{W} is a circular window which contains 32 pixels.

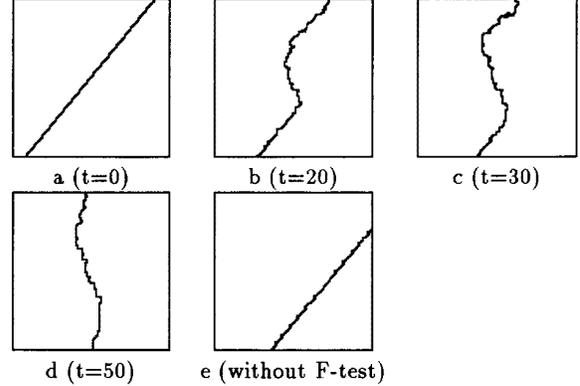


Figure 3: Region competition runs on image 1

As shown in figure (2.b), image 2 is a 150×150 image consists of 4 regions whose intensities are generated randomly from 4 Gaussian distributions: $N(50, 10^2)$, $N(80, 10^2)$, $N(110, 11^2)$, and $N(150, 15^2)$. The segmentation is started by putting nine seed regions (each has 80 pixels) on a grid, with four seeds inside the four ‘‘true’’ regions, four seeds straddling the boundaries, and one seed lying on the intersection point of the four regions. As we will define in the next section, the first four seeds are called *good seeds*, while the later five are *bad seeds*. The background (shown by the shadow) is treated as a single region which has uniform probability distribution P_0 . In fact P_0 provides the forces

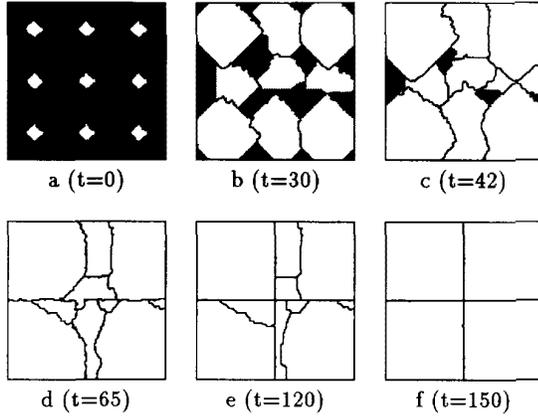


Figure 4: The competition of *good* and *bad* seed regions.

causing the seed regions to grow.

The motion of regions at iteration steps $t = 0, 30, 42, 65, 120, 150$ are shown in Figures (4.a.b.c.d.e.f) respectively. At $t = 30$ we observe that the good seeds grow faster than the bad seeds. Where these regions meet edges are formed, see Figure (4.b). Then the four good seeds keep compressing the bad seeds and taking over the background, see Figure (4.c.d.). In this competition process, some regions may be squeezed out by their neighbors. Finally the competition converges at $t = 120$ in Figure (4.f), where all the bad seed regions are driven into one of the four real regions. Then we merge the pairs of adjacent regions which cause the energy function to decrease the most, and restart region competition. After 30 more iteratives, the algorithm converges at $t = 150$, see Figure (4.g). A detailed description of the algorithm is given in [16].

4. ANALYZING REGION COMPETITION

This section briefly analyzes the performance of the region competition algorithm by discussing the following two aspects. More analysis is given in [16].

I. The precision of the segmentation, and the choice of window. In section (3.2), we argue that a window of size $m \neq 1$ is necessary for robust segmentation. Analytic results can be found if we consider the statistics force $f_{(x,y)}$ as a random variable depending on both the location (x, y) and the size (and shape) of the window \mathcal{W} . The boundary between regions is located at (x, y) where $f_{(x,y)} = 0$ and the probability distribution of this boundary can be studied. There exists an uncertainty interval \mathcal{U} for each boundary point and when the window size m is too small or too big \mathcal{U} will be intolerably large. Thus an optimal m can be chosen. For example, in cases where the two adjacent regions are generated by Gaussians $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, then the optimal window size is defined to be $m = \frac{4\alpha_o^2\sigma^2}{(\mu_1 - \mu_2)^2}$, where α_o is a confidence constant. We see we should set $m = 1$ only if $\|\mu_1 - \mu_2\| \gg \sigma$. This explains why conventional region growing (which uses $m = 1$) works

on images with high contrast (or SNR) regions.

II. The Initial conditions. Like all greedy algorithms, the performance of region competition will depend on the initial condition—more precisely, on the choice of the seeds. Because we assume the image consists of M homogeneous regions, and each region is generated by one of the pre-specified probability distributions, each initial seed is a hypothesis about the *location* of a certain *probability distribution family (or a model)*. Therefore, we define a seed to be “good”, if

1. it is completely inside a ‘true’ region, and
2. it selects the correct probability family (or model).

Failing to satisfy either of these two conditions causes a seed (or a hypothesis) to be ‘bad’. Thus to test whether a seed is good or not is to test how good the sample characteristics (in most case, the empirical distribution function) inside the seed fit a probability model. In statistics, there are many standard methods for testing the goodness of fit, such as the moment test, Chi-square test, and Kolmogorov test.

Empirically we observe that a sufficient condition for an optimal segmentation is that at the beginning each “true” region includes at least one good seed. But this is not a necessary condition. Below we demonstrate an interesting domino effect of transferring bad seeds into good ones.

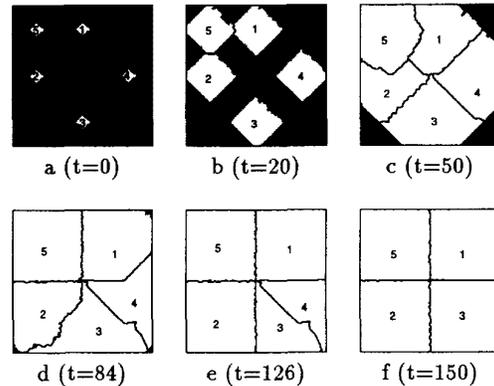


Figure 5: *The Domino effect of seeds transformation*

In figure (5), we show several steps of region competition running on the image shown in figure (2b). Four seeds are bad, straddling the boundaries, while seed 5 which lies in the upper-left region is good. In the first stage, all bad seeds grow across the boundaries (see $t=20$). After they contact each other, seed 5 drives seed 1 and seed 2 out of the upper-left region (see $t=50$). By time $t = 84$ seeds 1 and 2 are completely driven out of the upper-left region and both become good seeds while seed 3 and 4 remain bad seeds. Now seeds 1 and 2 drive seeds 3 and 4 into the lower-right region, at $t = 126$, where they become good seeds and are finally merged into a single region at $t=150$. Such domino effects will continue when more regions are involved, and we think this is a key difference between the region competition algorithm with conventional region growing approaches.

5. MULTIVARIATE MODELS: COLOR

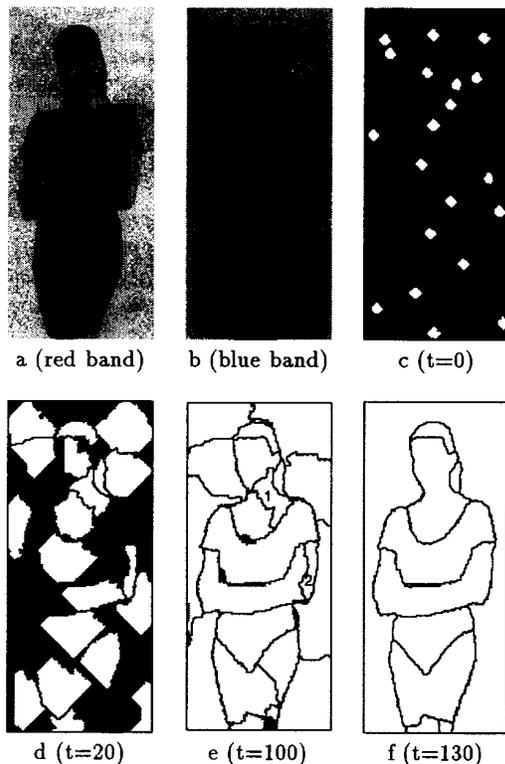


Figure 6: Segmentation of a color image with highlight regions

The previous sections have described segmentation techniques where regions are considered to have homogeneous intensity properties. For many applications image segmentation should be based on the albedo, because it corresponds to properties of the material independent of the lighting conditions or the geometric configuration of the material. For example, if we are looking at a person, the skin or each patch of his(her) clothes will have constant albedo properties but the intensity of the skin or clothes may change dramatically due to the changes of the geometry.

In color images, the probability family for each homogeneous region R_i can be modelled by:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix}_{(x,y)} = \begin{pmatrix} r \\ g \\ b \end{pmatrix} I_{(x,y)} + \begin{pmatrix} r_s \\ g_s \\ b_s \end{pmatrix} I_{S(x,y)} + \begin{pmatrix} e_r \\ e_g \\ e_b \end{pmatrix}_{(x,y)} \quad (19)$$

where $(R, G, B)_{(x,y)}$ is the input color, (r, g, b) is the (assumed) spatially constant body color within each region, (r_s, g_s, b_s) is the (assumed) spatially constant color of the light source, $I_{(x,y)}$ is the illumination, $I_{S(x,y)}$ is the specular component, and $(e_r, e_g, e_b)_{(x,y)}$ are the residues (or noise).

Therefore the homogeneous probability model is defined as $P((R, G, B)_{(x,y)} : (\bar{\mu}_i, \Sigma_i, (r_i, g_i, b_i)))$, where $(\bar{\mu}_i, \Sigma_i)$ are

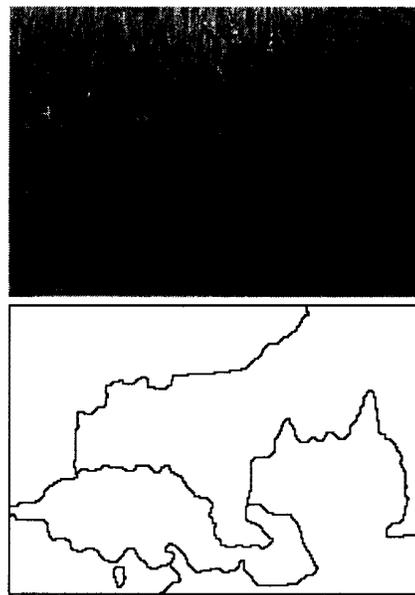


Figure 7: Segmentation of a texture image

the parameters of the Gaussian distribution of the residues. As before the region competition algorithm iterates the following two stages.

In the first stage, we fix the boundary Γ , and compute the best fitting of color (r_i^*, g_i^*, b_i^*) for each region R_i by least square regression. Then the fitting residues lie on a 2D plane β_i , whose normal is (r_i^*, g_i^*, b_i^*) , for each $(x, y) \in R_i$. We calculate $\bar{\mu}_i$ and Σ_i as the sample mean and sample covariance of the residues on the β_i plane in [16], we show that the color (r_i^*, g_i^*, b_i^*) computed by the least square method and the sample mean μ and sample variance Σ maximize the probability $P((R, G, B)_{(x,y)} : (\bar{\mu}_i, \Sigma_i, (r_i, g_i, b_i)))$ and thus minimize the energy functional.

The second stage is to move the boundary Γ by minimizing the MDL criterion.

For a pixel $p = (u, v)$ on the boundary of region R_i we first calculate the projection of $(R, G, B)_{(u,v)}$ onto the (r_i^*, g_i^*, b_i^*) vector, and then we calculate the fitting residue $\bar{Y}_{(u,v)}$ in the β_i plane.

As for the grey level case, we sample a window containing m pixels and obtain a set of measurements $(\bar{Y}_1, \dots, \bar{Y}_m)$. We define the window mean and co-variance by: $\bar{\bar{Y}}$ and S respectively. Thus the statistics force generated by a single R_i at (x, y) is $f_i \cdot \bar{\mathbf{n}}_i$, with:

$$f_i = -\frac{1}{2} [\log(2\pi) + \log \|\Sigma\| + (\bar{\bar{Y}} - \bar{\mu}_i) \Sigma_i^{-1} (\bar{\bar{Y}} - \bar{\mu}_i) + \text{tr}(\Sigma_i^{-1} S)] \quad (20)$$

Observe that Equations (20) generalizes equation (17) to higher dimensions. As for the grey level case we can identify the terms on the right hand side of Equation (20) as the standard generalizations of the T and F tests to higher dimensions.

Figures (6a,b) show the red and blue bands of a 261×116 color image. Observe the noise, shading in the cloth and the highlights on several parts of the skin.

The algorithm was implemented as for the grey level case. Figure (6c) shows the initial seeds at iteration step $t = 0$. Figures (6d,e,f) shows the sequence at steps $t = 20, 100, 135$ respectively. The competition between the seed regions converged at $t = 100$ as shown in Figure (6g) where the small shadow regions are the image spaces which are not occupied by the 20 seeds. In such regions we need to introduce additional seeds which can compete with the initial ones. We observe that the eyes and mouth are included in the face region because they are too small (about 10 25 pixels) and thus are treated as outlier noise.

Observe in Figure (6e) that the algorithm detects the highlights correctly. These highlight regions are: (1) on the neck and shoulder, (2) on the left arm, (3) on the left leg, (4) on the breast. Such highlight regions are detected because the distributions of the fitting residues in the 2D plane for these regions corresponds to ellipses with bigger aspect ratios than those for the non-highlight regions. Figure (6f) shows the final segmentation after spreading additional seeds and merging. We claim that the color model used in this paper has advantages over the existing methods, such as [9] in terms of both color description and highlight detection, detailed analysis is given in [16].

6. TEXTURE SEGMENTATION

We can directly adapt our algorithm to perform texture segmentation. As an example, figure (7.a) shows a 172×247 texture image. At each pixel (x, y) we measure the vector $\vec{X} = \nabla(G_\sigma * I)$ which captures the local orientation of the texture elements (or textons). Then we find that \vec{X} inside each region –such as grass, cheetah and bull are respectively subject to 2D Gaussian distribution $N(\vec{\mu}_i, \Sigma_i)$, while the $\vec{\mu}_i$'s are all close to zero and the covariance Σ_i is the major discriminator. Detailed analysis is given in [16].

Figure (7b) is the final result after merging. Observe that the belly of the cheetah, which is highly shadowed, is merged with the grass under the body. This is due to the similarity of their covariance matrices. This example has only used two texture filtering bands. In general we would prefer to use a larger number of filters geared to detecting different texture patterns. Moreover, with a larger number of filters it would be possible to perform a normalization between texture bands, as we did for the color bands, and reduce the effect of gradients caused by shadows and geometric effects.

7. CONCLUSION

In this paper, we assume that the image consists of M homogeneous regions, each of which is generated by one of K probability models. Therefore, the precondition for the region competition algorithm to work is that it should know *a priori* all the probability models underlying the input images. In other words, if K models are not sufficient for an input image, there is no reason to expect satisfactory segmentation. Such limit is true for all image segmentation algorithms.

We are currently extending this work by integrating grey level, color and texture cues. We are using this algorithm as a front end for our object recognition systems[17].

–The authors apologize for not discussing many highly related papers due to space limitations. More references are given in [16].

8. REFERENCES

- [1] A. Blake and A. Zisserman. **Visual Reconstruction**. Cambridge, MA. the MIT press, 1987.
- [2] J.F. Canny. "A computational approach to edge detection". *IEEE Trans. PAMI*-8(6):679-698. 1986.
- [3] A. Chakraborty et.al. "Deformable Boundary Finding Influenced by Region Homogeneity". *CVPR'94*.
- [4] L.D. Cohen. "On Active Contour Models and Balloons". *CVGIP:Image Und.*, Vol.53:(2), 1991.
- [5] D. Geiger and A.L. Yuille. "A common framework for image segmentation". *IJCV* 6(3), pp 227-243 1991.
- [6] S. Geman and D. Geman. "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images". *IEE Trans. PAMI* 6. pp 721-741. 1984.
- [7] T. Kanungo et al. "A fast algorithm for MDL-based multi-band image segmentation". *Proc. CVPR 1994*.
- [8] M. Kass, A. Witkin and D. Terzopoulos. "Snakes: Active Contour Models". *Proc. 1st ICCV87*.
- [9] G.J. Klinker et al. "A Physical Approach to Color Image Understanding". *IJCV*, 4. 1990.
- [10] . Y.G. Leclerc. "Constructing Simple Stable Descriptions for Image Partitioning". *IJCV*, 3, 1989.
- [11] J.M.Morel and S.Solimini. **Variational Methods in Image Segmentation**. Birkhauser pub.inc. 1995.
- [12] A. Leonardis et al. "Segmentation as the Search for the Best Description of the Image in Terms of Primitives". GRASP Laboratory. U.Penn 1990.
- [13] F.Mokhtarian and A.Mackworth, "A theory of multi-scale, curvature-based shape representation for planar curves", *PAMI* vol.15:(7) 1992.
- [14] D. Mumford and J. Shah. "Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems." *Comm. Pure Appl. Math.*, 42, pp 577-684. 1989.
- [15] A.P. Pentland. "Automatic Extraction of Deformable Part Models". *IJCV* vol.4: pp 107-126. 1990.
- [16] S.C.Zhu. "Region Competition: Unifying" Harvard Robotics Lab. TR-10-94, Version III March'95.
- [17] S.C.Zhu and A.L. Yuille. "FORMS: a Flexible Object Recognition and Modelling System". *Proc. of ICCV'95*. To appear in *IJCV*.