

A Useful Bound for Region Merging Algorithms in a Bayesian Model

Trevor C. Tao†‡#

David J. Crisp†‡#

†Department of Applied Mathematics

University of Adelaide, Adelaide SA 5005, Australia.

‡Intelligence, Surveillance & Reconnaissance Division

Defence Science & Technology Organisation,

PO Box 1500, Edinburgh, SA 5111, Australia.

Centre for Sensor Signal and Information Processing,
Mawson Lakes, SA 5095, Australia. Email: ttao@maths.adelaide.edu.au

Email: david.crisp@dsto.defence.gov.au

Abstract

A well known and effective approach to image segmentation is based on the Mumford-Shah model, where finding an optimal segmentation of an image is posed as the minimization of an energy functional. The simplest of these models is mathematically tractable and desirable properties of the resulting segmentations have been rigorously established. An important step in the proofs involves establishing an “inverse isoperimetric” type of bound. In particular, it can be shown that for segmentations where it is impossible to decrease the functional energy by merging any pair of regions, the length of the boundary between two regions must be bounded above by their areas and variation of the image itself. Recent work of one of the authors has extended the Mumford-Shah segmentation model to a Bayesian setting. In this report we show that an “inverse isoperimetric” type of bound also exists in this new setting. The new bound can likewise be used to prove desirable properties of the corresponding segmentations.

Keywords: bound, image segmentation, region merging, Mumford-Shah, Bayesian.

1 Introduction

Image segmentation is a standard problem in the field of image processing with many applications. The aim is to partition an image into a set of regions that satisfy desirable properties such as homogeneity within regions and contrast between adjacent regions according to some predetermined criteria. The value of segmentation lies in the fact that it enables further processing, such as image classification, to be performed using higher-level structures of the image, rather than individual pixels.

Many researchers have considered the problem of image segmentation and a plethora of techniques and algorithms exist. However, there is a lack of work on evaluating and comparing the performance of the various approaches and algorithms. Whilst research is currently active in this area, as seen in, say, (Zhang 2001) and (Martin & Fowlkes & Tal & Malik 2001), few benchmark databases and standard evaluation procedures exist, and selecting the best algorithm for a particular application is usually a matter of trial and error. In light of this dilemma, it is very useful to be able to establish the nature and properties of segmentation algorithms analytically. An analytic evaluation avoids the lack of rigour and

objectivity in the alternative empirical approaches, (Zhang 1996). The main aim of this paper is to present an analytic result for the segmentation algorithm described in (Crisp & Newsam 2000).

The particular approach to segmentation considered in (Crisp & Newsam 2000) is based on the well-known Mumford-Shah energy functional. One of the attractions of the Mumford-Shah functional is that it is mathematically tractable and important properties of its segmentations can be established analytically. A good exposition of the analysis can be found in the book (Morel & Solimini 1995). We will mainly be concerned with the results in Chapter 5 where Morel and Solimini present an in-depth study of the simplest functional, the Mumford and Shah piecewise constant model. This is the functional for which the most complete and comprehensive theory is available. The theory applies to a region merging approach to minimizing the functional. A full description of the algorithm and a demonstration of the quality of segmentations produced is given in (Koepfler & Lopez & Morel 1994).

It was observed by Crisp and Newsam, (Crisp & Newsam 2000), that in a Bayesian setting the minimizer of the Mumford-Shah functional with piecewise constant image model can be re-interpreted as a maximum *a posteriori* (MAP) estimate within a class of simple stochastic models for the image. By extending the class of stochastic models, a new MAP segmentation functional was described.

In this paper, we address the obvious question of the extent to which the analysis in (Morel & Solimini 1995) and (Koepfler & Lopez & Morel 1994) applies to the new MAP functional. Our main interest lies in a particular “inverse isoperimetric” type inequality which is central to much of the Morel and Solimini analysis. Our main result is that this inequality does indeed generalize (with some modifications) to the new functional. We also outline how the new inequality can be used to establish some desirable segmentation properties.

In the next section, we introduce some basic notation, describe the original Mumford-Shah functional, and discuss the relevant “inverse isoperimetric” inequality. In the following section we outline the region merging technique used for approximating the minimizers of the Mumford-Shah functional. In Section 4 and Section 5, we discuss the extension of the Mumford-Shah model to a Bayesian setting and introduce the MAP functional. Though much of these two sections is a repeat of the work in (Crisp & Newsam 2000), we include it here in order to set the notation and context. While the MAP functional itself is well-behaved, the region merging technique for minimising it can have initialisation problems. In Sec-

tion 6, we introduce a “variance offset” and show how it provides a good solution to the region merging initialisation problems. Our main result is presented in Section 7. This is followed by a discussion of its significance in Section 8. We conclude with a section on future work.

2 The Mumford-Shah Model

A formal definition of segmentation is as follows: An image can be described by a vectorial function $g : \Omega \rightarrow R^n$ where Ω is the region of interest and the values in the range represent data such as pixel intensity. The segmentation problem is to determine a “reasonable” pair of values (u, K) where u is a simple approximation of g and K is the boundary set. Here, for clarity of exposition, we only consider scalar intensities where $n = 1$, in which case the function represents a gray-scale image. We also assume Ω is a 2-D rectangular domain.

An important effort in analysing the problem of segmentation was made by Mumford and Shah, (Mumford & Shah 1989). The simplest version of their model is:

$$E(u, K) = \int_{\Omega-K} (u - g)^2 dx + \lambda \ell(K)$$

where the problem is to determine the “best” segmentation of an image by minimizing the functional. Here λ is a scale parameter, u is constrained to be piecewise constant in each of the individual regions, and $\ell(K)$ is the length of the boundary K . The parameter λ is preselected and controls the scale of the resulting segmentations – small values yield fine segmentations while large values yield coarse segmentations.

Many theoretical results have been established. For example, Morel and Solimini (Morel & Solimini 1995) proved an important theorem due to Mumford and Shah:

Theorem 1 *Let g be a measurable bounded function in Ω . Then the minimum of $E(u, K) = \int_{\Omega} (u-g)^2 dx + \lambda \ell(K)$ is attained at some K . Moreover, the minimal boundary sets have the following geometric property: Either the points of K are regular, C^1 and with curvature bounded by $\frac{8 \text{osc}(g)}{\lambda}$, or the singular points are of two types, namely, triple points where three branches meet with 120-degree angles and boundary points where K meets the boundary of Ω at 90 degrees.*

One of the first results needed in their proof is a lemma that determines an upper bound for the length of the common boundary between a pair of regions depending only on properties of the regions in question. More specifically, they showed that if a segmentation is 2-normal¹ then

Lemma:

$$\lambda l(\delta(O, O')) \leq \min(|O|, |O'|) \text{osc}^2(g) \quad (1)$$

where $l(\delta(O, O'))$ is the length of the common boundary between the two regions O and O' . The oscillation $\text{osc}(g)$ is the difference between maximum and minimum gray values of g and is thus independent of the segmentation.

¹A segmentation is 2-normal if it is not possible to merge a pair of regions without increasing the energy functional.

3 Region Merging

The problem of determining the optimal segmentation according to the energy functional is intractable because the search space of all possible segmentations is extremely large. Hence we seek a fast algorithm for producing only a “reasonable” segmentation which we hope will have desirable properties even though it is not optimal. A detailed description of region merging is given in (Morel & Solimini 1995) and we only present a brief discussion here.

Starting with an initial segmentation with very small regions, we repeatedly select a pair of regions to be merged until it is no longer possible to do so without increasing the energy functional. When this happens, the segmentation is, in some sense, locally optimal for a given choice of scale parameter. In this way, a segmentation can be calculated given any value of scale parameter. The simplest choice of initial segmentation is the so called *trivial segmentation*, where each pixel is considered an individual region. Although there are other schemes for initializing the segmentation, we consider only the trivial segmentation here. Before region merging takes place, the relevant equations, derived from the continuous setting, must be discretized. For example, in the Mumford-Shah functional, the term

$$\int_{\Omega-K} (u - g(\mathbf{x}))^2 dx$$

becomes

$$\sum_{\mathbf{x} \in \Omega} (u - g(\mathbf{x}))^2$$

and the boundaries are treated as edges between pixels. Thus, boundary curves can only bend at right angles and their length is computed accordingly.

4 Extension to a Bayesian Model

One of the shortcomings of the Mumford-Shah Model is that it can only handle simple “cartoon-like” images where the essential regions are nearly constant. Said differently, if two adjacent regions had comparable mean gray-values but different higher-order statistics such as Gaussian variance, then a region-merging algorithm will either merge both regions into one or “undersegment” both, depending on the scale parameter λ . This shortcoming was addressed in (Crisp & Newsam 2000) by viewing the Mumford-Shah Model from a Bayesian setting and then extending the image model in an appropriate way. In order to introduce notation and set the context for later sections, we repeat (with variations) the relevant contents of (Crisp & Newsam 2000).

The original Mumford-Shah functional consists of two terms, where the data fidelity competes with the simplicity of the model. In a Bayesian setting, it is natural to consider both the model and the conditional data as random variables. Thus our problem is to find the best segmentation model given the data.

$$\begin{aligned} M_{opt} &= \max \arg_M p(M|g) \\ &= \max \arg_M p(g|M)p(M)/p(g) \\ &= \max \arg_M p(g|M)p(M) \\ &= \min \arg_M \left\{ -\ln(p(g|M)) - \ln(p(M)) \right\}. \end{aligned}$$

Observe the similarity between this model and the Mumford-Shah functional. The former term corresponds to data fidelity and the latter the model complexity. Although there is no scale parameter, it is

implicit in the latter term since we have yet to define a measure for the complexity of the model. Our task is to minimize the functional:

$$-\ln(p(g|M)) - \ln(p(M)).$$

One difficulty with the Mumford-Shah functional is in reconciling both terms with common measures, because the data-fidelity term requires a two-dimensional measure to compute the integral whereas the boundary between regions is one-dimensional. Similar considerations hold for the Bayesian setting, hence we introduce a new parameter and define

$$-\ln(p(M)) = \gamma l(K)$$

where K is the boundary set corresponding to M . This is a natural choice since it equates the likelihood of a model with its complexity and therefore encapsulates an ‘‘Ockam’s Razor’’ type of principle – i.e. the simplest segmentation is the best. Note that this argument can be put on a rigorous footing using the Minimum Description Length (MDL) principle, as in done in (Kanungo & Dom & Niblack & Steele 1994), although a slightly different and more complicated energy functional results.

Now we consider an image model which incorporates known data errors using stochastic methods. For each region R_i in the segmentation, we model the image by

$$g(\mathbf{x}) = \mu_i + \epsilon(\mathbf{x})$$

where $\mathbf{x} = (x, y)$ is the coordinates of a pixel and the noise $\epsilon(\mathbf{x})$ is a zero mean Gaussian random variable. In order to distinguish two regions with similar values of μ_i but different noise variance, we allow the noise variance to vary from region to region. Thus each region R_i has two model parameters, μ_i and σ_i . If we assume the data errors are independent from pixel to pixel, then we have

$$p(g(\mathbf{x})|\mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \frac{-(g(\mathbf{x}) - \mu_i)^2}{2\sigma_i^2}$$

$$p(g|M) = \prod_{i=1}^N \prod_{\mathbf{x} \in R_i} p(g(\mathbf{x})|\mu_i, \sigma_i)$$

where $\mathbf{x} \in R_i$, σ_i and μ_i are the parameter values for R_i , and N is the number of regions. Thus

$$-\ln(p(g|M)) = \sum_{i=1}^N \sum_{\mathbf{x} \in R_i} \left(\ln(\sigma_i \sqrt{2\pi}) + \frac{(g(\mathbf{x}) - \mu_i)^2}{2\sigma_i^2} \right)$$

In fact, it can be shown that this model is a direct generalization of the Mumford-Shah functional. Consider the simplification where the noise variance is constant across the entire image. In other words σ_i is replaced with a constant σ . Then

$$-\ln(p(g|M)) = \sum_{i=1}^N \sum_{\mathbf{x} \in R_i} \left(\ln(\sigma \sqrt{2\pi}) + \frac{(g(\mathbf{x}) - \mu_i)^2}{2\sigma^2} \right)$$

The term $\ln(\sigma \sqrt{2\pi})$ is constant, so does not affect the minimization. Hence, (ignoring the constant)

$$\begin{aligned} & -\ln(p(g|M)) - \ln(p(M)) \\ &= \sum_{i=1}^N \sum_{\mathbf{x} \in R_i} \frac{(g(\mathbf{x}) - \mu_i)^2}{2\sigma^2} + \gamma l(K). \end{aligned}$$

Thus, by setting $\gamma = \lambda/(2\sigma^2)$ we can recover the Mumford-Shah functional.

It is well known that once the boundary set K is determined for Mumford-Shah segmentation, the optimal estimated parameter values μ_i are determined analytically by:

$$\mu_i = \frac{1}{|R_i|} \sum_{\mathbf{x} \in R_i} g(\mathbf{x}).$$

Similarly, it is easily shown that for the Gaussian model the estimated values of μ_i are as given above and the variances σ_i are determined by:

$$\sigma_i^2 = \frac{1}{|R_i|} \sum_{\mathbf{x} \in R_i} (g(\mathbf{x}) - \mu_i)^2. \quad (2)$$

5 Region merging in the Bayesian Model

For region merging algorithms, we wish to calculate the change in energy corresponding to a merge between two regions.

$$\begin{aligned} E(K') - E(K) &= \ln p(g|M) - \ln p(g|M') \\ &\quad + \ln p(M) - \ln p(M') \end{aligned} \quad (3)$$

where K and K' are the boundary sets corresponding to M and M' , respectively. In our notation we will use the boundary set K as input to the energy function and the actual model as input to the probability function p .

Because the negative log-likelihood is calculated as a sum of contributions for each region, we see an important advantage of region merging: Whenever two regions are merged, the difference between the log-likelihoods of the segmentation before and after a merge depends only on the two regions in question. It is easy to see that:

$$\ln p(M) - \ln p(M') = \gamma l(\delta(R_i, R_j)).$$

In other words, the second term of (3) is just the length of the common boundary between two regions multiplied by the scale parameter.

Given the optimal mean and variance parameters above, it can be shown that the negative log-likelihood for a segmentation M satisfies:

$$-\ln(p(g|M)) = \sum_{i=1}^N (\ln(2\pi\sigma_i^2) + 1) \frac{|R_i|}{2}.$$

It follows that:

$$\begin{aligned} & \ln(p(g|M)) - \ln(p(g|M')) \\ &= \frac{1}{2} \left(|R_{ij}| \ln \sigma_{ij}^2 - |R_i| \ln \sigma_i^2 - |R_j| \ln \sigma_j^2 \right). \end{aligned}$$

Therefore the change of energy induced by a merge is

$$\begin{aligned} E(K) - E(K') &= \frac{-1}{2} \left(|R_{ij}| \ln \sigma_{ij}^2 - |R_i| \ln \sigma_i^2 - \right. \\ &\quad \left. |R_j| \ln \sigma_j^2 \right) + \gamma l(\delta(R_i, R_j)). \end{aligned} \quad (4)$$

For region merging algorithms, it is also necessary to be able to update the model parameters after each merge. In other words, if we use R_{ij} to denote the region obtained by merging R_i and R_j then we require update formulae for μ_{ij} and σ_{ij} in terms of

$\mu_i, \mu_j, \sigma_i, \sigma_j$. It can easily be shown that:

$$\sigma_{ij}^2 = \frac{1}{|R_i| + |R_j|} \left(|R_i|(\sigma_i^2 + (\mu_i - \mu_{ij})^2) + |R_j|(\sigma_j^2 + (\mu_j - \mu_{ij})^2) \right) \quad (5)$$

$$\mu_{ij} = \frac{|R_i|\mu_i + |R_j|\mu_j}{|R_i| + |R_j|}. \quad (6)$$

To complete the description of the region merging algorithm, it only remains to specify the order in which the regions are merged. Here, there is an important choice. The usual technique (see (Kanungo & Dom & Niblack & Steele 1994) for instance) is to take a greedy approach and to always merge the pair of regions which yield the largest decrease in the energy functional. Note that, the decrease is easily calculated using equation (4). However, Crisp and Newsam (Crisp & Newsam 2000) use a different technique. They associate with each pair of regions, a critical value of λ where the energy change associated with the merger becomes zero. This critical value of λ is called the *merge cost* of the region pair. They then select region pairs according to the size of the merge costs, not ΔE . The cheapest merges are done first and merging continues until the merge costs exceed the stopping value of λ . They refer to this algorithm as the Full Lambda Schedule Algorithm (FLSA). Note that FLSA was first reported in (Redding & Crisp & Tang & Newsam 1999) and that the original motivation was in the work of Koepfler et al. (Koepfler & Lopez & Morel 1994). Experimental results lead us to suspect that FLSA is superior and it is the algorithm we use.

6 An Offset for the Variance Estimates

It was noted in (Crisp & Newsam 2000) that the region merging algorithm for minimizing the MAP functional presents a problem as it stands. The algorithm is initialized with each pixel being a separate region, implying the initial estimates of the region variances are all zero. However, the change in energy in equation (4) involves taking logarithms of the region variances. While one way around this dilemma is suggested in (Crisp & Newsam 2000), we have since found that better results are obtained by using the simpler approach of merely initializing the variance estimates at some prescribed non-zero value. We denote this value by σ_0^2 and call it the *variance offset*.

The variance offset must be chosen carefully. If it is too small, then experience shows that the segmentations are dominated by a few large regions which absorb the remaining single pixel regions one at a time. On the other hand, if the offset is too large, then the resulting segmentations are similar to those from the Mumford-Shah functional and so region variances are not taken into account at all. It is easy to see why the latter is so. If σ_i^2 and σ_j^2 in (5) are replaced with $\sigma_i^2 + \sigma_0^2$, and $\sigma_j^2 + \sigma_0^2$ then σ_{ij}^2 is likewise replaced by $\sigma_{ij}^2 + \sigma_0^2$. It follows that all region variances are offset by the same amount, namely σ_0^2 . If the offset is large enough, it will dominate the true variances.

We conclude this section by noting that there are other solutions to the initialisation problem. See for instance Kanungo et al., (Kanungo & Dom & Niblack & Steele 1994) and (Kanungo & Dom & Niblack & Steele & Sheinvald 1995), where the problem of poor estimates of variance for small regions was considered and possible solutions suggested. See also earlier work of the authors, (Crisp & Tao 2002) where a variation of FLSA called Synchronous Locally Best

Merging (SLBM) was suggested. The idea of SLBM is to treat the initial merges as local processes operating independently. Other algorithms in the literature look for globally best merges. These algorithms suffer from “region domination” where large regions tend to swallow individual pixels one by one before they have a chance to merge with “correct” neighbours. SLBM automatically avoids these problems by forcing a merged region to wait for all other existing LB merges by placing them in a queue.

7 A New Bound for the Bayesian Model

Recall that Morel and Solimini established a bound given by equation (1) for 2-normal segmentations. Equation (1) says that the length of the common boundary between two regions is bounded above by their areas, choice of scale parameter and oscillation of the entire image itself. In other words, a 2-normal segmentation will not have overly complex boundaries given a reasonable choice of λ and a well-behaved image g . We seek to establish a similar bound for the common boundary length in our Bayesian formulation.

To prove our main result, we will require the region variances to be bounded. In fact, we require

$$0 < \text{off}^2(g) \leq \sigma_i^2 \leq \text{osc}^2(g) \quad (7)$$

for all regions R_i during every step of the segmentation. Note that the RHS inequality is always true. The LHS is also clearly true for our algorithm if we set

$$\text{off}^2(g) = \sigma_0^2$$

because $\sigma_0^2 \leq \sigma_i^2$ holds for the initial trivial segmentation and inspection of the update formula (5) shows that merging a pair of regions can only increase the variance ($\sigma_{ij} \geq \max(\sigma_i, \sigma_j)$). Similarly, (7) holds for any region merging algorithm which begins with regions whose variances are non-zero as long as $\text{off}^2(g)$ is set below the smallest variance.

We are now in a position to state our main result:

Lemma:

$$\gamma l(\delta(R_i, R_j)) \leq \frac{\text{osc}^2(g)}{\text{off}^2(g)} \min(|R_i|, |R_j|). \quad (8)$$

Proof: By 2-normality, we have

$$\begin{aligned} 0 &\geq E(K) - E(K') \\ &= -\left(\ln p(g|M) - \ln p(g|M') \right) \\ &\quad + \gamma l(\delta(R_i, R_j)). \end{aligned} \quad (9)$$

Hence, using the facts that $|R_{ij}| = |R_i| + |R_j|$ and $\ln(x) \leq x - 1$ for all real x , we have

$$\begin{aligned} &2\gamma l(\delta(R_i, R_j)) \\ &\leq 2 \ln p(g|M) - \ln p(g|M') \\ &= \left(|R_{ij}| \ln \sigma_{ij}^2 - |R_i| \ln \sigma_i^2 - |R_j| \ln \sigma_j^2 \right) \\ &= \left(|R_i| \ln \frac{\sigma_{ij}^2}{\sigma_i^2} + |R_j| \ln \frac{\sigma_{ij}^2}{\sigma_j^2} \right) \\ &\leq \left(|R_i| \left(\frac{\sigma_{ij}^2 - \sigma_i^2}{\sigma_i^2} \right) + |R_j| \left(\frac{\sigma_{ij}^2 - \sigma_j^2}{\sigma_j^2} \right) \right). \end{aligned}$$

From the update formulae for the model parameters (5) and (6), we easily obtain

$$\begin{aligned}\sigma_{ij}^2 - \sigma_i^2 &= \frac{1}{|R_i| + |R_j|} \left(|R_j|(\sigma_j^2 - \sigma_i^2) \right. \\ &\quad \left. + |R_i|(\mu_i - \mu_{ij})^2 + |R_j|(\mu_j - \mu_{ij})^2 \right) \\ \mu_{ij} - \mu_i &= \frac{|R_j|(\mu_j - \mu_i)}{|R_i| + |R_j|}\end{aligned}$$

and the analogous equations hold with indices i and j swapped. Therefore:

$$\begin{aligned}& 2\gamma l(\delta(R_i, R_j)) \\ & \leq \left(|R_i| \left(\frac{\sigma_{ij}^2 - \sigma_i^2}{\sigma_i^2} \right) + |R_j| \left(\frac{\sigma_{ij}^2 - \sigma_j^2}{\sigma_j^2} \right) \right) \\ & = \frac{|R_i||R_j|}{|R_i| + |R_j|} \left[(\sigma_j^2 - \sigma_i^2) \left(\frac{1}{\sigma_i^2} - \frac{1}{\sigma_j^2} \right) \right. \\ & \quad \left. + \frac{(\mu_i - \mu_j)^2}{|R_i| + |R_j|} \left(\frac{R_i}{\sigma_i^2} + \frac{R_j}{\sigma_j^2} \right) \right] \\ & = \frac{|R_i||R_j|}{|R_i| + |R_j|} \left[\left(\frac{\sigma_j}{\sigma_i} - \frac{\sigma_i}{\sigma_j} \right)^2 \right. \\ & \quad \left. + \frac{(\mu_i - \mu_j)^2}{|R_i| + |R_j|} \left(\frac{R_i}{\sigma_i^2} + \frac{R_j}{\sigma_j^2} \right) \right] \\ & \leq \frac{|R_i||R_j|}{|R_i| + |R_j|} \left[\left(\frac{\text{osc}(g)}{\text{off}(g)} - \frac{\text{off}(g)}{\text{osc}(g)} \right)^2 \right. \\ & \quad \left. + \frac{\text{osc}^2(g)}{|R_i| + |R_j|} \left(\frac{|R_i| + |R_j|}{\text{off}^2(g)} \right) \right] \\ & < \min(|R_i|, |R_j|) \left(\frac{2\text{osc}^2(g)}{\text{off}^2(g)} \right).\end{aligned}$$

from which (8) follows.

8 Discussion

Our result only differs from (1) by the presence of the offset term in the denominator. The offset represents a trade-off between the tightness of the bound and data fidelity. If the offset is too small, then the bound is meaningless. If the offset is too large, we have recovered the Mumford-Shah formulation. Note also that our bound depends only on the ratio of oscillation and offset, which implies that a linear transformation in gray-scale values will not alter the bound. This property does not hold for the original Mumford-Shah model.

In (Morel & Solimini 1995), the authors deduce from their original bound that for a large scale parameter, small regions will be eliminated when the segmentation becomes 2-normal. More specifically, they establish a lower bound for the area of any region O :

$$\frac{288|\Omega|\text{osc}^4(g)}{C^2\lambda^2} \geq \frac{C\lambda}{\sqrt{|O|\text{osc}^2(g)}}$$

or

$$|O| \geq \frac{C^6\lambda^6}{288^2|\Omega|^2\text{osc}^{12}(g)}$$

where C is the isoperimetric constant in Ω . By following the proof in (Morel & Solimini 1995), we can easily arrive at a similar bound:

Lemma:

$$|O| \geq \frac{C^6\gamma^6\text{off}^{12}(g)}{288^2|\Omega|^2\text{osc}^{12}(g)}$$

This result is useful from an image processing point of view since it establishes that a small cluster of pixels will never be treated as a separate region provided the scale parameter is sufficiently large.

9 Future Work

Now that we have established the bound (8), an obvious question is: Can the bound be used to prove a result for the MAP functional which is analogous to Theorem 1? We believe that such a result does indeed hold by adapting the proofs in (Morel & Solimini 1995). We intend to follow up on this soon.

One significant difficulty with the Mumford-Shah functional is that there is no automatic method for selecting the scale parameter which controls the trade-off between data fidelity and model complexity, and our Bayesian framework fails to address this. At present, the algorithm is run with different values of λ and a human operator must decide the ‘‘best’’ segmentation. Minimum Description Length (MDL) avoids these problems by equating the cost of a segmentation with the number of bits necessary to encode it, see (Kanungo & Dom & Niblack & Steele & Sheinvald 1995), but the resulting functional is different. We believe that the differences do not affect the validity of our algorithm and that MDL offers a sound way of choosing the stopping value of λ . We hope to investigate this further in the future.

Another method of removing the choice of scale-parameter is the method of *L-curves*. This requires that segmentations be computed for all values of λ and then a graph can be drawn showing the relationship between model complexity and data fidelity. The optimum segmentation can be defined as that corresponding to the point with some desirable property such as ‘‘greatest curvature’’ on the L-curve. Obviously this method depends on the ability to calculate segmentations for many values of λ quickly and efficiently. In this respect, a useful property of segmentation algorithms is *causality*, i.e. given two values of parameter $\lambda_1 < \lambda_2$ and corresponding segmentations M_1, M_2 , the set of curves in the segmentation M_2 should be contained in that of M_1 (Morel & Solimini 1995). One of the advantages the FLSA algorithm has over a greedy approach is that causality holds and hence L-curves are easily calculated.

Another question concerns the choice of initial variance σ_0^0 , which should only depend on the data g . As noted before, we must not set the initial variance too large or too small. Clearly we should also avoid the loss of useful information in the image. Hence we enforce the condition that if one pixel has intensity greater than that of a neighbour, then it should be ‘‘highly probable’’ the inequality will not be reversed when the error noise $N(0, \sigma_0)$ is taken into account. If we define the gray-scale resolution ϵ_0 to be the minimum non-zero difference between pixel intensities for a discretized image, then it is reasonable to assume that the initial variance is proportional to the gray-scale resolution with $\sigma_0 = \kappa\epsilon_0$. There is no obvious choice of κ and this needs to be addressed.

References

- Crisp, D.J. & Tao, T.C. (2002), ‘Fast Region Merging Algorithms for Image Segmentation’, in ‘Proceedings of the Fifth Asian Conference on Computer Vision (ACCV2002)’, Melbourne, Australia, 22–25 January 2002, pp.412–417.

- Crisp, D.J. & Newsam, G.N. (2000), 'A Fast, Efficient Segmentation Algorithm Based on Region Merging', in 'Proceedings of Image and Vision Computing New Zealand 2000 (IVCNZ00)', Hamilton, New Zealand, November 2000, pp.180–185.
- Kanungo, T., Byron, D., Niblack, W. & Steele, D. (1994), 'A Fast Algorithm for MDL-based Multi-Band Image Segmentation' in 'IEEE Conference on Computer Vision and Pattern Recognition', Seattle, Washington, June 1994, pp.609–616.
- Kanungo, T., Byron, D., Niblack, W., Steele, D. & Sheinvald, J. (1995), 'MDL-Based Multi-Band Image Segmentation Using a Fast Region Merging Scheme', *IBM Technical Report* Number RJ 9960 (87919), IBM Research Division (1995).
- Koepfler, G., Lopez, C. & Morel, J.-M. (1994), 'A Multi-Scale Algorithm for Image Segmentation by Variational Method', *SIAM Journal of Numerical Analysis* **31**, 282–299.
- Martin, D. & Fowlkes, C. & Tal, D. & Malik, J. (2001), 'A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics', in 'Proceedings of the Eighth IEEE International Conference on Computer Vision', 2, 2001, pp. 416–423.
- Morel, J.-M & Solimini, S. (1995), *Variational Methods in Image Segmentation*, Birkhäuser, Boston.
- Mumford, D. & Shah, J. (1989), 'Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems', *Communications on Pure and Applied Mathematics* **42**(4), 577–685.
- Redding N.J. & Crisp, D.J. & Tang, D. & Newsam G.N. (1999), 'An Efficient Algorithm for Mumford-Shah Segmentation and Its Application to SAR Imagery', in 'Proceedings of the Fifth Conference on Digital Image Computing, techniques, and Applications (DICTA'99)', Perth, Australia, 7–8 December 1999, pp.35–41.
- Zhang, Y.J. (1996), 'A Survey on Evaluation Methods for Image Segmentation', *Pattern Recognition* **29**(8), 1335–46.
- Zhang Y.J., 'A Review of Recent Evaluation Methods for Image Segmentation', in 'International Symposium on Signal Processing and its Applications (ISSPA)', Malaysia, 13–16 August 2001.