Segmenting Images Corrupted by Correlated Noise

Thomas C.M. Lee, *Member, IEEE*

Abstract—Image segmentation is fundamental to many image analysis problems. It aims to partition a digital image into a set of nonoverlapping homogeneous regions. The main contribution of this paper is the development of a new segmentation procedure which is designed to segment images corrupted by *correlated* noise. This new segmentation procedure is based on Rissanen's minimum description length (MDL) principle and consists of two components: i) an MDL-based criterion in which the "best" segmentation is defined as its minimizer and ii) a merging algorithm which attempts to locate this minimizer. The performance of this procedure is illustrated via a simulation study, with promising results.

Index Terms—Correlated noise, image segmentation, merging algorithm, minimum description length.

1 Introduction

Image segmentation is fundamental to many image analysis problems. It aims to partition a digital image into a set of nonoverlapping regions, in such a way that these segmented regions possess the following properties:

1) pixels within the same region are homogeneous with respect to some characteristic (e.g., gray value or texture); and
2) pixels of neighboring regions are significantly different with respect to the same characteristic (see [11, Chapter 10] and [10, Chapter 4] for references and some general discussion of these issues).

The image characteristic as the basis for segmentation with which this paper is concerned is gray value.

One popular approach to segmenting an image based on gray value is to approximate the image by a two-dimensional (2D) piecewise smooth function and define the segmentation by the discontinuity points of the 2D piecewise smooth function. For example, see [3], [4], [5], [7], [9], [14], [15], [17], [18], [19], [24], and [26]. In this paper, images that can be well modelled by 2D piecewise constant functions (PCFs) corrupted by additive Gaussian noise will receive primary attention. The main contribution here is the development of a new segmentation procedure which is designed to segment such images when the noise is *correlated*. The proposed segmentation procedure is based on Rissanen's minimum description length (MDL) principle [22] and consists of two components:

1) an MDL-based criterion in which the "best" segmentation is defined as its minimizer and
2) a merging algorithm which attempts to locate this minimizer.

To the best of our knowledge, this is the first time that, for the piecewise smooth function modelling approach to the problem of image segmentation, the effect of noise correlation has been taken into account.

In some recent papers, the MDL principle has been applied to the problem of image segmentation, e.g., see [18], [15], [27], [19] and references given therein. The key difference here is that, all such previous work only considered *independent* noise. A more thorough comparison is given in Section 5.

The rest of this paper is organized as follows. Section 2 defines the problem that we consider. Section 3 briefly describes some previous relevant results. Section 4 presents the proposed segmentation procedure. Section 5 compares the proposed procedure with other MDL-based segmentation procedures. Section 6 reports the results of a simulation study, and Section 7 discusses possible extensions to the proposed procedure.

2 Problem Formulation

As suggested before, we model images by 2D PCFs corrupted by additive Gaussian noise. Let $\boldsymbol{f} = (f_1, \ldots, f_n)^T$ be a discrete version of an arbitrary 2D PCF with k constant regions and n pixels. Denote the gray value of the jth region as μ_j, $j = 1, \ldots, k$, and write $\mu = (\mu_1, \ldots, \mu_k)^T$. Then $\boldsymbol{f}$ can be represented by:

$$f_i = \sum_{j=1}^{k} \mu_j I_{\{i \in r_j\}}, \qquad i = 1, \ldots, n$$

where "$i \in r_j$" means "the ith pixel is in the jth region," and I_E is the indicator function for the event E. Notice that we have chosen to use single indexing rather than double indexing for labelling pixel coordinates. Also notice that in our formulation for $\boldsymbol{f}$, region boundaries are composed of horizontal and vertical "edges" between pixels. For convenience, we define $\Omega = \{r_1, \ldots, r_k\}$ as the set of all region boundaries, i.e., Ω defines a partition of the image.

• *The author is with the Department of Statistics, The University of Chicago, 5734 S. University Avenue, Chicago, IL 60637. E-mail: tlee@galton.uchicago.edu.*

Manuscript received 3 Mar. 1997; revised 31 Mar. 1998. Recommended for acceptance by H.R. Keshavan.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 106708.

Our aim is, given an observed image $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ satisfying the model

$$y_i = f_i + \epsilon_i, \qquad i = 1, \ldots, n,$$

where the ϵ_is are zero mean Gaussian random errors, to obtain an estimate $\hat{\boldsymbol{f}}$ for $\boldsymbol{f}$, and define our segmentation of the image by the boundaries separating all constant regions (i.e., all discontinuity points) of $\hat{\boldsymbol{f}}$. We will apply the MDL principle to obtain such an estimated/fitted 2D PCF $\hat{\boldsymbol{f}}$.

Briefly, the MDL principle defines the best-fitted model as the one that enables the best compression/encoding of the data, so that the data can be transmitted in the most economical way. In the present situation, the data is the image to be segmented and a fitted model is simply a fitted 2D PCF (i.e., a segmentation) of the image. That is, the best fitted model/segmentation is the one that produces the shortest code length of the data/image. A typical code length formula usually consists of two parts:

1) the code length for specifying a fitted model plus
2) the code length for specifying the data "conditioning on" the fitted model (i.e., the residuals).

3 MDL-Based Segmentation for Independent Noise

The correlated noise segmentation procedure to be proposed in Section 4 is a modification of the segmentation procedure proposed in [19]. That procedure has been constructed for segmenting images corrupted by independent and identical Gaussian noise ϵ_is with variance σ^2. For convenience, we state the relevant results in [19].

3.1 MDL Criterion for Independent Noise

We first state the result regarding the code length formula for encoding an image (PCF) corrupted by independent noise. We need some notation to proceed. Suppose a fitted 2D PCF is given. Let $\hat{k}$ be its estimated number of regions, $\hat{\Omega} = \{\hat{r}_1, \ldots, \hat{r}_{\hat{k}}\}$ be the set of all its estimated region boundaries (i.e., discontinuity points), and $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_{\hat{k}})^T$ be its estimated region gray values. If a_j and b_j are, respectively, the area (in terms of number of pixels) and perimeter (in terms of number of pixel edges) of the jth region of $\hat{\Omega}$, then it is shown in [19] that

$$\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right) = \hat{k} \log n + \frac{\log 3}{2} \sum_{j=1}^{\hat{k}} b_j + \frac{1}{2} \sum_{j=1}^{\hat{k}} \log a_j + \frac{n}{2} \log\left(RSS_{\hat{k}}/n\right) \qquad (1)$$

is an approximation to the desirable code length formula. Here $RSS_{\hat{k}} = \sum_{i=1}^{n} \left(y_i - \hat{f}_i\right)^2$ is the residual sum of squares with $\hat{f}_i$ defined by

$$\hat{f}_i = \sum_{j=1}^{\hat{k}} \hat{\mu}_j I_{\{i \in \hat{r}_j\}}, \quad \hat{\mu}_j = \frac{1}{a_j} \sum_{s \in \hat{r}_j} y_s, \qquad i = 1, \ldots, n. \qquad (2)$$

It was proposed in [19] to select the minimizer of $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right)$ as the "best" fitted model, or segmentation, of an image corrupted by *independent* noise.

3.2 Merging Algorithm for Independent Noise

Global minimization of $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right)$ is infeasible for images of a reasonable size. In [19] a fast merging algorithm similar to those used in [6] and [15] is proposed. This merging algorithm may miss the global minimum but is guaranteed to find a local minimum of $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right)$.

The general idea of the merging algorithm is as follows. It starts with computing the $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right)$ value of an *oversegmentation* of the image being segmented. Then at each time step it chooses two neighboring regions and merges them to form a new region. These two neighboring regions are chosen in such a way that, when they are merged, it provides the largest reduction in the current value of $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right)$ amongst all other possible merges. Such a merging strategy is often called the "greedy" strategy. The merging algorithm continues until there is only one region left.

If there are K initial regions in the original oversegmentation, then, when the algorithm finishes, K hierarchically fitted models/segmentations are produced. The one that has the smallest $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right)$ value will be chosen as the best fitted model/segmentation.

Also given in [19] are some fast updating formulae for recursively computing $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right)$ as k decreases, and a practical method for obtaining a reasonable oversegmentation.

4 MDL-Based Segmentation for Correlated Noise

This section presents the main contribution of the paper: We modify the independent noise MDL criterion $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right)$ (1) to handle images corrupted by *correlated* noise. We also develop a merging algorithm which does *not* use the greedy merging strategy to approximate the corresponding minimum. We concentrate on the case when the noise can be well modelled by some zero-mean *stationary Gaussian random field* (SGRF). That is, we assume an observed image is a 2D PCF with an SGRF superimposed.

The main modification to $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right)$ is to replace the last term $\frac{n}{2}\log\left(RSS_{\hat{k}}/n\right)$ by $\frac{1}{2}\log\left|\hat{\Sigma}\right|$, which corresponds to the code length for encoding the data conditioning on a fitted model (i.e., the residuals). Here $\hat{\Sigma}$ is the maximum likelihood estimate of the autocovariance matrix Σ of the correlated noise (which we model by a SGRF) and $\left|\hat{\Sigma}\right|$ is its determinant. The second modification to $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right)$ is to add the code length that is required to specify $\hat{\Sigma}$. If we restrict all possible 2D autocovariance functions which give rise to Σ to have the same parametric form, $\hat{\Sigma}$ can be com-

pletely specified by the estimates of the unknown parameters of that parametric form. In this case, as the number of unknown parameters is fixed, the code length for specifying $\hat{\Sigma}$ is constant and so can be ignored. In fact this is the route that we follow below, and the new MDL criterion constructed for handling correlated noise is

$$\mathrm{MDL}_{\mathrm{COR}}\left(\hat{k}, \hat{\Omega}\right) = \hat{k} \log n + \frac{\log 3}{2} \sum_{j=1}^{\hat{k}} b_j + \frac{1}{2} \sum_{j=1}^{\hat{k}} \log a_j + \frac{1}{2} \log \left|\hat{\Sigma}\right| \quad (3)$$

with $\hat{\Sigma}$ to be specified below. In Section 7.4 we will briefly discuss the case when the autocovariance functions can possess different parametric forms. Since the SGRF is assumed to have a zero mean, estimates of the μ_js and f_js are also given by (2).

4.1 Choice of Parametric Form for Autocovariance Functions

It is desirable that one can arbitrarily specify the parametric form of the possible autocovariance functions to handle the noise patterns/textures of the images being segmented. Unfortunately this is infeasible because of various computational difficulties. (For related work on texture classification, consult, for example, [21] and [25].)

Finding or approximating the minimum of the correlated MDL criterion (3) requires repeated computations of $\log\left|\hat{\Sigma}\right|$, which means fast methods for computing $\log\left|\hat{\Sigma}\right|$ are necessary. This is made difficult by the huge size of $\hat{\Sigma}$: for an image of size $n_r \times n_c$, $\hat{\Sigma}$ is of size $n_r n_c \times n_r n_c$. Even the real memory space of a reasonable workstation cannot store the whole $\hat{\Sigma}$ for an image of a reasonable size. Also, unlike $\log\left(RSS_{\hat{k}}\right)$ in the independent case, generally no fast updating formulae exist for the computation of $\hat{\Sigma}$ and $\log\left|\hat{\Sigma}\right|$. Thus, we see two main constraints restricting the choice of a parametric form for the autocovariance:

1) fast methods for estimating the parameters of the parametric form should exist, i.e., a reasonable estimate of $\hat{\Sigma}$ should be obtained quickly; and
2) fast and accurate approximations for $\log\left|\hat{\Sigma}\right|$ should also exist.

Here we propose using the following three-parameter *separable anisotropic exponential* autocovariance parametric form:

$$\mathrm{acv}(r, c) = A \exp\left(-\alpha_r \Delta_r |r| - \alpha_c \Delta_c |c|\right), \quad r = \frac{-n_r}{2}, \ldots, \frac{n_r}{2} - 1,$$
$$c = \frac{-n_c}{2}, \ldots, \frac{n_c}{2} - 1, \quad (4)$$

where Δ_r and Δ_c are the sampling intervals in the row and column directions respectively, $n_r \times n_c$ is the dimension of the image (i.e., $n_r n_c = n$), and $A > 0$, $\alpha_r > 0$, and $\alpha_c > 0$ are parameters. This sort of separable exponential autocovariance function was commonly used in the early days for image representation and compression (e.g., see [13, Chapter 6]). In this paper we assume both n_r and n_c are even, and images are observed in the square $[-1, 1) \times [-1, 1)$. Thus $\Delta_r = 2/n_r$ and $\Delta_c = 2/n_c$.

4.2 Parameter Estimation

The MDL principle suggests that the parameters A, α_r, and α_c should be estimated by the method of maximum likelihood. However, such maximum likelihood estimates often require iterative methods to compute, and these iterative methods typically take a long time to converge. This would make our segmentation method impractical. Below we suggest fast methods for estimating A, α_r, and α_c.

Denote the residuals by $\hat{e}_i = y_i - \hat{f}_i$ for $i = 1, \ldots, n$. We estimate A by

$$\hat{A} = \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i^2 = RSS_{\hat{k}}/n.$$

Our estimators for α_r and α_c are approximate method-of-moments estimators. We first use the approximation $\exp(x) \approx 1 + x$ for small x to write

$$\mathrm{acv}(r, c) = A \exp\left(-\alpha_r \Delta_r |r| - \alpha_c \Delta_c |c|\right)$$
$$\approx A - A\alpha_r \Delta_r |r| - A\alpha_c \Delta_c |c|$$

for small $|r|$ and $|c|$. Define the *variogram* $v(r, c)$ as:

$$v(r, c) = 2(\mathrm{acv}(0, 0) - \mathrm{acv}(r, c))$$
$$\approx 2A(\alpha_r \Delta_r |r| + \alpha_c \Delta_c |c|).$$

Then we have

$$\alpha_r \approx \frac{v(1, 0)}{2A\Delta_r} \quad \text{and} \quad \alpha_c \approx \frac{v(0, 1)}{2A\Delta_c}$$

and, hence, one can estimate α_r and α_c by first estimating $v(1, 0)$ and $v(0, 1)$. Define the following two sets

$S_r = \{(i, j) : \text{pixel } j \text{ is the immediate right neighbor of pixel } i\}$,

$S_c = \{(i, j) : \text{pixel } j \text{ is the immediate bottom neighbor of pixel } i\}$

and let $|S_r|$ and $|S_c|$ be the number of elements in S_r and S_c, respectively. Then $v(1, 0)$ and $v(0, 1)$ can be estimated rapidly by

$$\hat{v}(1, 0) = |S_r|^{-1} \sum_{(i,j) \in S_r} \left(\hat{e}_i - \hat{e}_j\right)^2 \text{ and } \hat{v}(0, 1) = |S_c|^{-1} \sum_{(i,j) \in S_c} \left(\hat{e}_i - \hat{e}_j\right)^2$$

and so α_r and α_c can be estimated by

$$\hat{\alpha}_r = \frac{\hat{v}(1, 0)}{2\hat{A}\Delta_r} \quad \text{and} \quad \hat{\alpha}_c = \frac{\hat{v}(0, 1)}{2\hat{A}\Delta_c}$$

Notice that fast updating formulae for computing $\hat{v}(1, 0)$ and $\hat{v}(0, 1)$ exist.

4.3 Fast Approximation for $\log\left|\hat{\Sigma}\right|$

In this subsection, we develop a fast approximation for $\log\left|\hat{\Sigma}\right|$ when the underlying autocovariance function is of the parametric form (4). We first consider the 1D autocovariance function

$$\gamma_u = \gamma(\Delta|u|) = A\exp(-\alpha\Delta|u|), \quad u = 0, \pm 1, \ldots,$$

where A and α are parameters, and Δ is the sampling interval. Let

$$F(\omega) = \frac{1}{2\pi}\sum_{u=-\infty}^{\infty}\gamma_u \exp(-iu\omega), \quad \omega \in [0, 2\pi)$$

be the corresponding spectrum (in this subsection $i = \sqrt{-1}$), and let τ^2 be the innovation variance. Because $\gamma(t)$ is the autocovariance of an AR(1) (autoregressive of order 1) process, we have

$$\gamma_0 = A, \gamma_0 = \frac{\tau^2}{1-\exp(-2\alpha\Delta)} \Rightarrow \tau^2 = A(1-\exp(-2\alpha\Delta))$$

and

$$F(\omega) = \frac{\tau^2}{2\pi|1-\exp(-\alpha\Delta)\exp(-i\omega)|^2} \Rightarrow$$

$$\int_0^{2\pi}\log F(\omega)d\omega = 2\pi\log\tau^2$$

Now using the results that the determinant of a matrix equals the product of all its eigenvalues (denote the eigenvalues by λ_ks), and that the eigenvalues of the autocovariance matrix Σ of a stationary process with n observations are approximately given by $\lambda_k = F(2\pi k/n)$, $k = 0, \ldots, n-1$, we have

$$\begin{aligned}\log|\Sigma| &= \log\prod_{k=0}^{n-1}\lambda_k \\ &= \sum_{k=0}^{n-1}\log\lambda_k \\ &\approx \frac{n}{2\pi}\int_0^{2\pi}\log F(\omega)d\omega \\ &= \frac{n}{2\pi}2\pi\log\tau^2 \\ &= n\log A(1-\exp(-2\alpha\Delta)).\end{aligned}$$

The separable parametric form (4) is simply a product of γ_us and, thus, by similar arguments one can derive the following approximation:

$$\log|\hat{\Sigma}| \approx n_r n_c \log\left[\hat{A}(1-\exp(-2\hat{\alpha}_r\Delta_r))(1-\exp(-2\hat{\alpha}_c\Delta_c))\right] \quad (5)$$

4.4 MDL Criterion for Correlated Noise

Combining (3) and (5) we obtain the following correlated noise MDL criterion specialized to the autocovariance parametric form (4):

$$\begin{aligned}\mathrm{MDL}_{\mathrm{COR}}(\hat{k},\hat{\Omega}) &= \hat{k}\log n + \frac{\log 3}{2}\sum_{j=1}^{\hat{k}}b_j + \frac{1}{2}\sum_{j=1}^{\hat{k}}\log a_j \\ &+ \frac{n}{2}\log\left[\hat{A}(1-\exp(-2\hat{\alpha}_r\Delta_r))(1-\exp(-2\hat{\alpha}_c\Delta_c))\right] \quad (6)\end{aligned}$$

We propose to choose our best fitted model for images corrupted by correlated noise as the minimizer of $\mathrm{MDL}_{\mathrm{COR}}(\hat{k},\hat{\Omega})$, and define our segmentation by its discontinuity points (i.e., boundaries separating all constant regions). Notice that $\mathrm{MDL}_{\mathrm{COR}}(\hat{k},\hat{\Omega})$ converges to the independent noise criterion $\mathrm{MDL}_{\mathrm{IND}}(\hat{k},\hat{\Omega})$ (1) when both $\hat{\alpha}_r$ and $\hat{\alpha}_c$ approach infinity.

If the noise structure of the image to be segmented is unknown, an obvious question to ask is: When should the independent criterion $\mathrm{MDL}_{\mathrm{IND}}(\hat{k},\hat{\Omega})$ be used and when should the correlated $\mathrm{MDL}_{\mathrm{COR}}(\hat{k},\hat{\Omega})$ be used? We will address this issue in Section 7.6.

4.5 Modified Merging Algorithm

A modified merging algorithm is also developed to approximate the minimizer $\mathrm{MDL}_{\mathrm{COR}}(\hat{k},\hat{\Omega})$. This correlated noise merging algorithm also begins with an oversegmentation of the image being segmented. However, at each time step, it does *not* adopt the natural idea to merge the pair of neighboring regions that produces the largest reduction in $\mathrm{MDL}_{\mathrm{COR}}(\hat{k},\hat{\Omega})$. Instead, it merges the pair that produces the largest reduction in $\mathrm{MDL}_{\mathrm{IND}}(\hat{k},\hat{\Omega})$. That is, the correlated noise merging algorithm uses the *independent greedy merging strategy*, but not the *correlated greedy merging strategy*. Thus it gives the same sequence of nested segmentations as the independent case algorithm. The main difference is that, it picks the segmentation from this nested sequence that has the smallest value of $\mathrm{MDL}_{\mathrm{COR}}(\hat{k},\hat{\Omega})$ as the final segmentation.

There are two strong reasons for not using the correlated greedy merging strategy. Firstly the time required for computing $\log|\hat{\Sigma}|$ is (relatively) much longer than for computing $\log RSS_{\hat{k}}$, even though we have a fast approximation for $\log|\hat{\Sigma}|$. Thus at each time step finding the region pair that gives the best reduction in $\mathrm{MDL}_{\mathrm{IND}}(\hat{k},\hat{\Omega})$ is much quicker than finding the pair giving the best reduction in $\mathrm{MDL}_{\mathrm{COR}}(\hat{k},\hat{\Omega})$, especially when the number of possible pairs to be compared is large. This greatly improves the speed performance of the correlated noise merging algorithm. Our empirical experience suggests that, if the image model assumption is correct (i.e., a 2D piecewise constant function with a SGRF generated by the autocovariance function (4) superimposed), both the correlated and independent greedy merging strategies often produce nearly identical sequences of nested segmentations. Hence there is not much gain in segmentation performance by using the correlated greedy merging strategy, but the gain in speed resulting from using the independent greedy merging strategy is enormous.

The second reason is as follows. When the segmentation sequences produced by the two merging strategies are indeed different, empirical observation suggests that the sequence resulting from the correlated greedy strategy would

be most likely worse than the one resulting from the independent greedy strategy. This can be attributed to the fact that in some step(s) of the merging algorithm, the correlated greedy strategy would choose to merge a pair of two small neighboring regions belonging to two different true regions even if there are other more appropriate pairs available for merging. The independent greedy strategy seldom does this. These situations are more common when the image model assumption is wrong. We believe the reason for this is as follows. Merging two regions belonging to two different true regions would make all the residuals of one region shifted up by a constant and the residuals of the other region shifted down by another constant. As a result of these shifts, residuals of the same region are (much) more correlated and hence give smaller values of $\hat{\alpha}_r$ and $\hat{\alpha}_c$. The value of $\hat{A}$ would also be (slightly) inflated by such a mismerge. When the effect of having small values of $\hat{\alpha}_r$ and $\hat{\alpha}_c$ is dominant in the computation of $\log\left|\hat{\Sigma}\right| = \log\left[\hat{A}\left(1-\exp\left(-2\hat{\alpha}_r\Delta_r\right)\right)\left(1-\exp\left(-2\hat{\alpha}_c\Delta_c\right)\right)\right]$, a potentially largest reduction in $\mathrm{MDL}_{\mathrm{COR}}\left(\hat{k},\hat{\Omega}\right)$ may result, which would suggest a bad merge. Section 7.5 provides some timing figures.

5 Comparison With Closely Related Work

Leclerc [18] and Kanungo et al. [15] have applied the MDL principle to the problem of image segmentation, with the latter work being closer to ours. Both Leclerc and Kanungo et al. only consider the case of independent Gaussian noise, but allow the underlying true image to be piecewise polynomial and different regions to have different noise variances (i.e., spatial varying noise). They also claim that a maximal degree of two (i.e., quadratic) is sufficient for piecewise polynomial surfaces to approximate most images.

To find or approximate the minimum of their MDL criteria, Leclerc uses a minimization procedure which is continuous in nature while Kanungo et al. use a stepwise region-merging algorithm similar to ours. Leclerc's procedure is less likely to miss the global minimum but is more time consuming.

These authors also considered other extensions. Leclerc modified his MDL criterion to handle images which are blurred by a *known* point spread function. Kanungo et al. developed an MDL-based segmentation procedure for *multiband images*: a multiband image can be treated as a stack of possibly correlated grayscale images. The multiband segmentation procedure of Kanungo et al. assumes the noise is independent within bands but is correlated across bands. Our correlated noise segmentation procedure can also be extended to handle multiband images; see [20].

Recently, Zhu and Yuille [27] proposed a new two-stage algorithm, called *region competition*, for approximating the minimum of an MDL or a Bayesian based criterion for image segmentation. These authors claim that their region competition algorithm generalizes other segmentation algorithms such as region growing, snake and balloon methods, and in particular, they constructed such a region competition algorithm for a modified version of Leclerc's MDL criterion. They also considered the problem of segmenting multiband images.

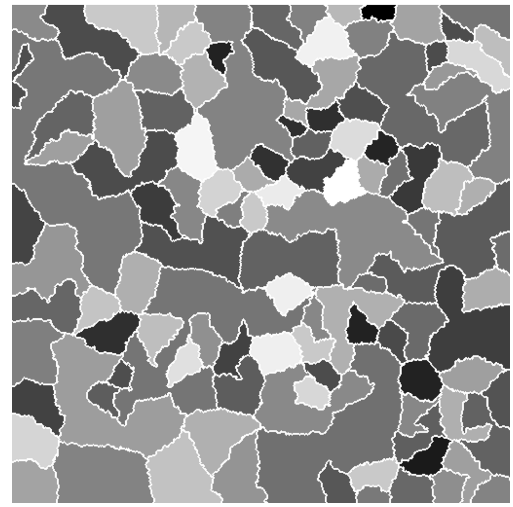


Fig. 1. Test image with true region boundaries (in white) overlaid. Since in our formulation boundaries are composed of horizontal and vertical "edges" between pixels, the actual boundaries cannot be exactly displayed in images. To deal with this problem, we has displayed all pixels which have at least one of their four edges as part of a boundary.

6 Simulation Study

This section reports results of a simulation study which was conducted to evaluate the performance of the proposed correlated noise segmentation procedures.

6.1 Settings

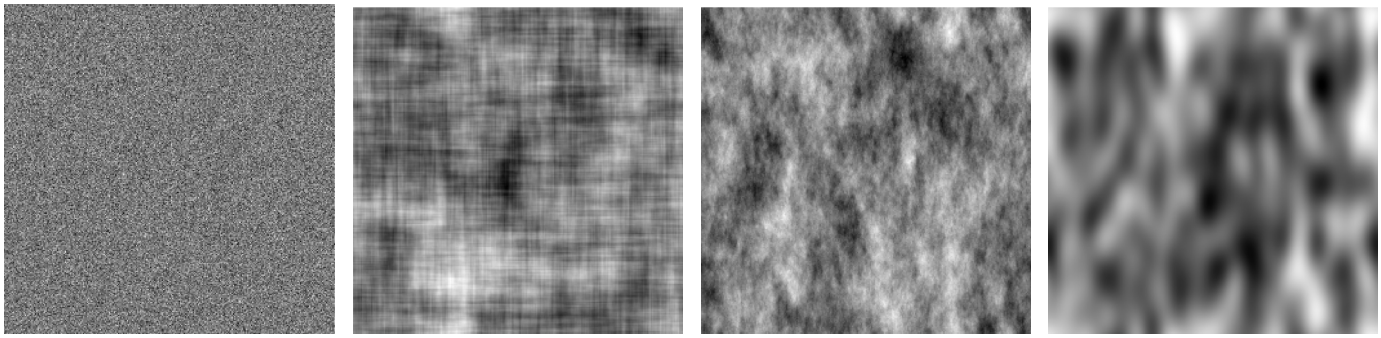
The test image (i.e., the true but unknown image) used throughout the whole simulation study is displayed in Fig. 1. It is composed of 111 distinct regions and is of dimension 512 × 512. It is a realization generated from a model constructed for real aluminium grain images; see [20] for details.

Our simulated noisy images (i.e., the observed images) were then obtained by adding various Gaussian noise patterns / textures to the test image. Fig. 2 displays some typical noise patterns / textures that we used.

In order to evaluate different segmentation results, we will use Baddeley's [2] binary image measure Δ_w^p to rank these results. Also, throughout the whole simulation study we used the *seeded region growing* [1] oversegmentation approach suggested by [19] to obtain initial oversegmentations with which to start the merging algorithm.

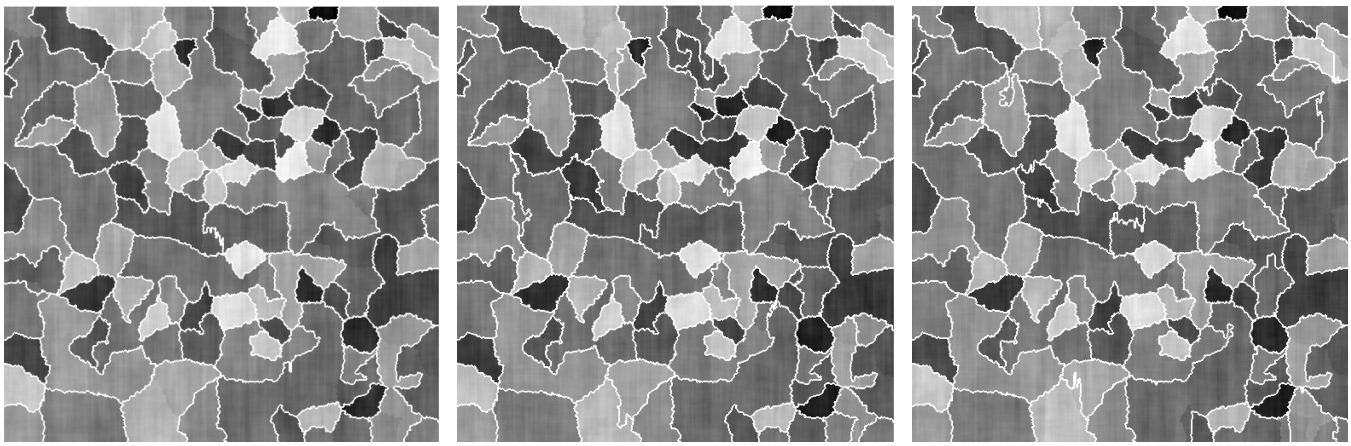
6.2 Correct Autocovariance Assumption

We tested the correlated noise segmentation procedure (i.e., the merging algorithm aims to minimize $\mathrm{MDL}_{\mathrm{COR}}\left(\hat{k},\hat{\Omega}\right)$ with images corrupted by correlated noise generated from the autocovariance given by (4) (see Fig. 2b for an example). We used three different "levels" of noise, which can be characterized by different combinations of autocovariance parameters (see Table 1). Also listed in Table 1 are the *characteristic lengths*, ρ_r and ρ_c, in the *r*- and *c*-directions, respectively. Here a characteristic length is defined to be the dis-



(a) (b) (c) (d)

Fig. 2. Various noise patterns/textures. (a) Independent noise. (b) Correlated noise with separable exponential autocovariance; see Section 6.2 combination 2 for exact formulation. (c) Correlated noise with nonseparable exponential autocovariance; see Section 6.3 for exact formulation. (d) Correlated noise with bivariate Gaussian autocovariance; see Section 6.3 for exact formulation.



(a) (b) (c)

Fig. 3. Simulation results for autocovariance parameter combination 1. (a) Observed image corresponds to good segmentation with segmented region boundaries (in white) overlaid. (b) Similar to (a) but for medium segmentation. (c) Similar to (a) but for poor segmentation.

tance, in terms of number of pixels, that is required for the effect of the noise correlation to drop by 50 percent.

For each parameter combination, we generated 100 noisy observed images, and segmented these observed images with the proposed correlated noise segmentation procedure. We then computed, and for each combination, sorted the values of Δ_w^p in ascending order.

Fig. 3 displays the segmentation results corresponding to the 10th, the 51st, and the 90th smallest Δ_w^p values for parameter combination 1. These results are taken as typical examples of good, medium and poor segmentations, respectively. Results for the remaining two parameter combinations are displayed in a similar fashion in Fig. 4 and Fig. 5, respectively. One can see that the proposed correlated noise segmentation procedure generally performed well in our simulations *when the autocovariance assumption is correct*.

TABLE 1
LEVELS OF NOISE

	A	α_r	α_c	ρ_r	ρ_c
combination 1	20.0	5.0	20.0	35.5	9.0
combination 2	30.0	5.0	5.0	35.5	35.5
combination 3	10.0	10.0	5.0	17.7	35.5

6.3 Incorrect Autocovariance Assumption

In the previous subsection the correlated noise was generated from autocovariance functions satisfying our prior assumption, that is, those autocovariance functions were of the parametric form (4). It is interesting to know how well (or bad) our correlated noise segmentation procedure performs when the underlying noise autocovariance function is *not* of the assumed parametric form (4). To investigate this, we further tested our correlated noise segmentation procedure with images corrupted by correlated Gaussian noise generated from the following two autocovariance functions:

1) *nonseparable exponential* (see Fig. 2c for an example):

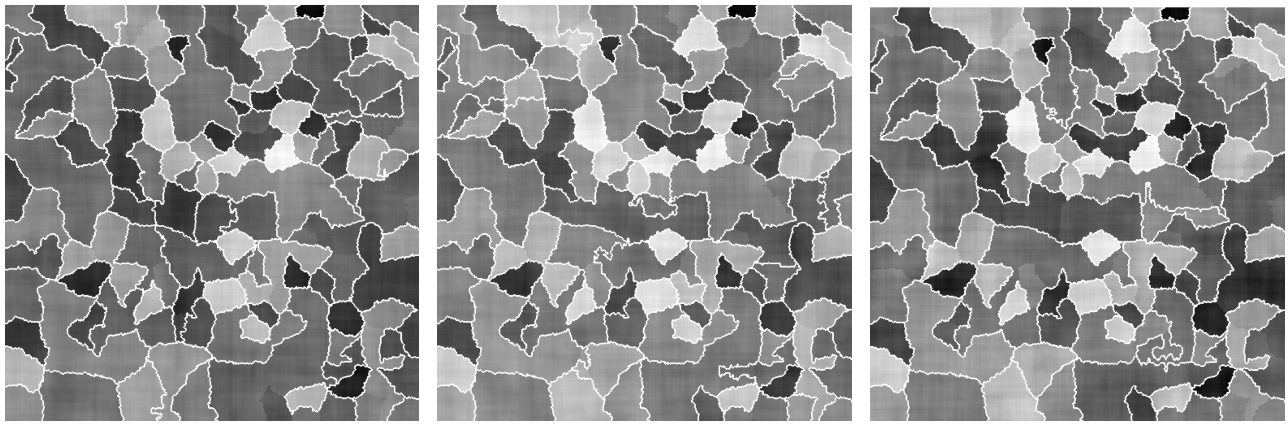
$$\mathrm{acv}(r,c) = 10\exp\left[-5\sqrt{(\Delta_r r)^2 + 4(\Delta_c c)^2}\right]$$

with characteristic lengths $\rho_r = 35.5$ and $\rho_c = 17.7$, and

2) *bivariate Gaussian* (see Fig. 2d for an example):

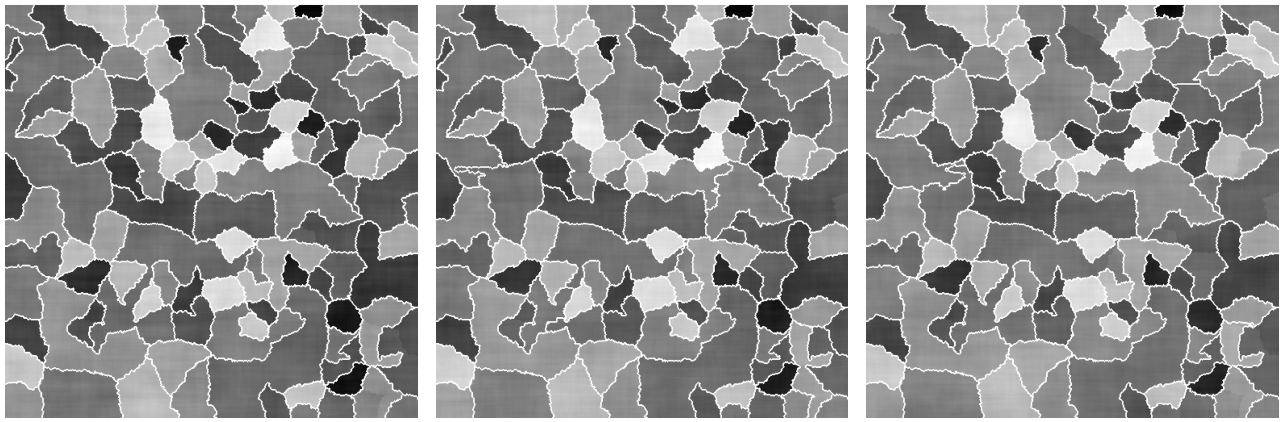
$$\mathrm{acv}(r,c) = 20\exp\left[-20\left\{(\Delta_r r)^2 + 4(\Delta_c c)^2\right\}\right]$$

with characteristic lengths $\rho_r = 47.7$ and $\rho_c = 23.8$,



(a) (b) (c)

Fig. 4. Similar to Fig. 3 except for correlated noise with autocovariance parameter combination 2.



(a) (b) (c)

Fig. 5. Similar to Fig. 3 except for correlated noise with autocovariance parameter combination 3.

where Δ_r and Δ_c are the sampling intervals in the row and column directions, respectively, $n_r \times n_c$ is the dimension of the image (i.e., $n_r n_c = n$), and r and c are indices: $r = -n_r/2, \ldots, n_r/2 - 1$, $c = -n_c/2, \ldots, n_c/2 - 1$.

As before, for each autocovariance function, 100 simulated segmentations were performed, and the corresponding values of Δ_w^p were computed and sorted. We displayed the relevant good, medium and poor segmentations in Figs. 6 and 7.

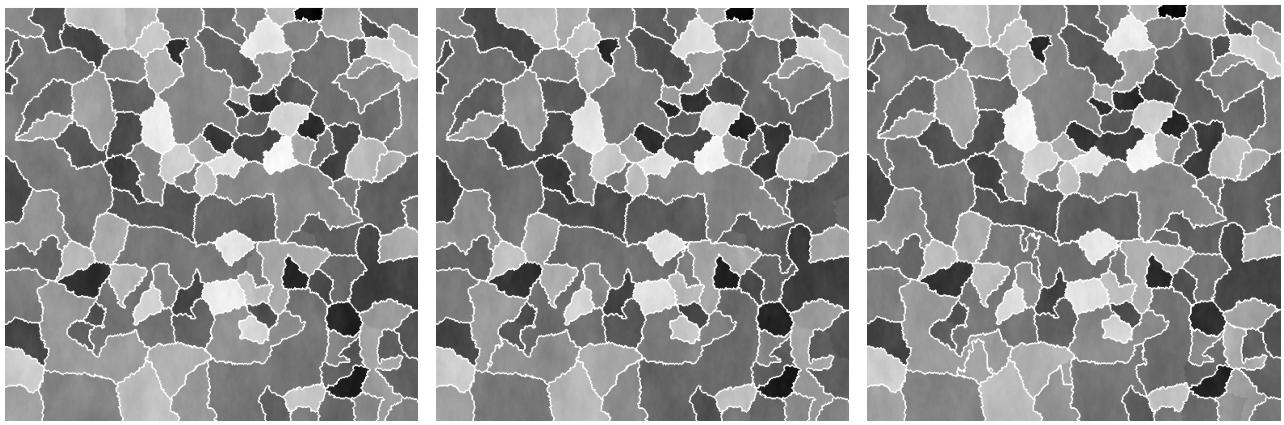
Despite the fact that the noise autocovariance assumption was wrong, our correlated noise segmentation procedure generally produces satisfactory results (although, in the case of the bivariate Gaussian autocovariance function, some obviously spurious boundaries have been added). We suspect the reason is as follows. Although the above nonseparable exponential and bivariate Gaussian autocovariance functions are different to the separable exponential autocovariance (4), their low frequency spectral behaviours are similar. In the context of nonparametric curve estimation with correlated noise, many asymptotic results demonstrate that the important information carried by the observed data is concentrated in the low frequency region of the spectrum (e.g., [12] and references given therein). Thus if such low frequency information can be well captured, one would expect to have a good estimation performance. This phenomenon appears to carry over to the present situation, and hence explains why our correlated noise segmentation procedure has produced satisfactory results even though the autocovariance assumption is wrong.

6.4 Correlated Noise With Independence Assumption and Vice Versa

Two important questions that one may ask are:

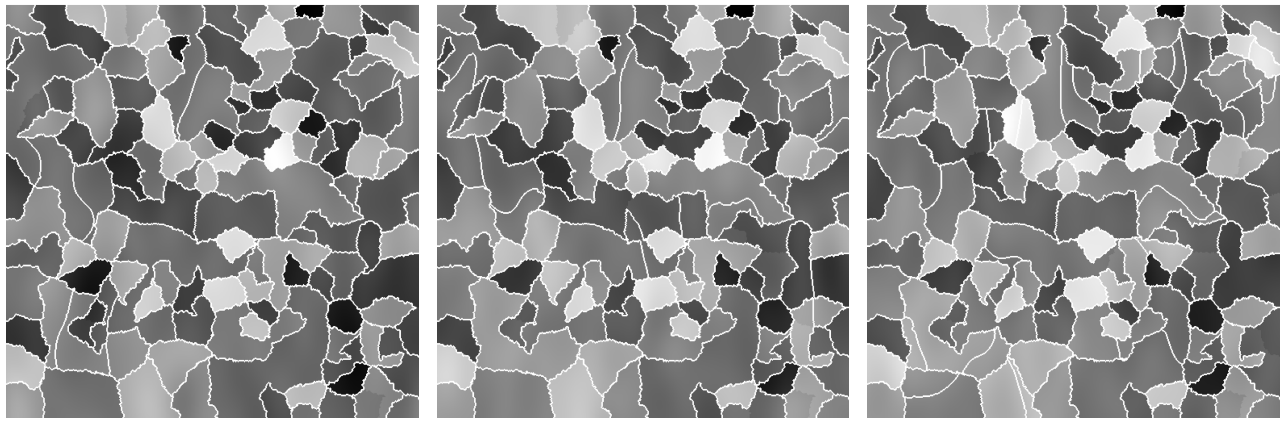
1) how does the independent noise segmentation procedure of [19] (the one that aims to minimize $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right)$ perform if the noise is in fact correlated, and
2) how does the proposed correlated noise segmentation procedure (the one that aims to minimize $\mathrm{MDL}_{\mathrm{COR}}\left(\hat{k}, \hat{\Omega}\right)$ perform if the noise is in fact independent?

In an attempt to answer these two questions, we applied both the independent and the correlated noise segmentation procedures to simultaneously segment four noisy images. The results are displayed in Figs. 8 and 9. The first noisy image is the test image displayed in Fig. 3a with in-



(a) (b) (c)

Fig. 6. Similar to Fig. 3 except for correlated noise with nonseparable exponential autocovariance function.



(a) (b) (c)

Fig. 7. Similar to Fig. 3 except for correlated noise with bivariate Gaussian autocovariance function.

dependent Gaussian noise superimposed. The signal-to-noise ratio (snr) was 4.4 (we define snr = $var(\boldsymbol{f})/\sigma^2$). The remaining three noisy images were selected from those generated in the previous two subsections. These three selected noisy images were those giving rise to the medium segmentations of:

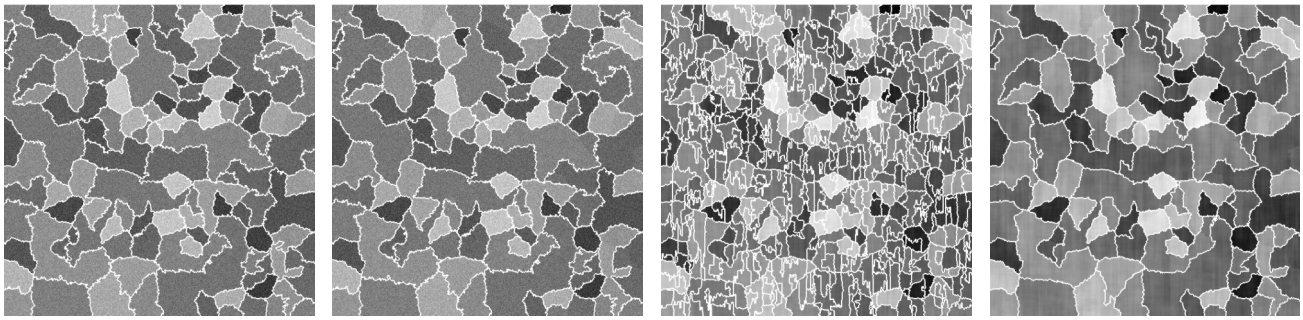
1) correlated noise with autocovariance parameter combination 1 (see Fig. 3c);
2) correlated noise with nonseparable exponential autocovariance (see Fig. 6c); and
3) correlated noise with bivariate Gaussian autocovariance (see Fig. 7c).

Recall that both the independent and the correlated merging algorithms produce the same sequence of nested segmentations, and that for each of these nested segmentations, the values of the relevant MDL criterion (either $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k},\hat{\Omega}\right)$ or $\mathrm{MDL}_{\mathrm{COR}}\left(\hat{k},\hat{\Omega}\right)$) are computed. For comparative purposes, we plotted the MDL values corresponding to the four above-selected images in Fig. 10. For convenience, we call $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k},\hat{\Omega}\right)$ as a function of k an *MDL independent curve*, and we define an *MDL correlated curve* in an analogous manner.

Figs. 8 to 10 provide strong evidence that ignoring the presence of positively correlated noise results in a serious oversegmentation (of course this is not necessarily true if the correlation is so weak that the noise can be treated as independent). Figs. 8 to 10 also suggest that one should usually use the correlated noise segmentation procedure (unless it is known a priori that the noise is spatially independent), as it did not produce any very poor results in our simulation study, while the independent noise segmentation procedure only performed well when the noise was independent (a more thorough simulation study for the independent noise procedure is given in [19]). However there are other situations when the independent noise segmentation procedure is preferred. We defer our discussion of this issue to Section 7.5.

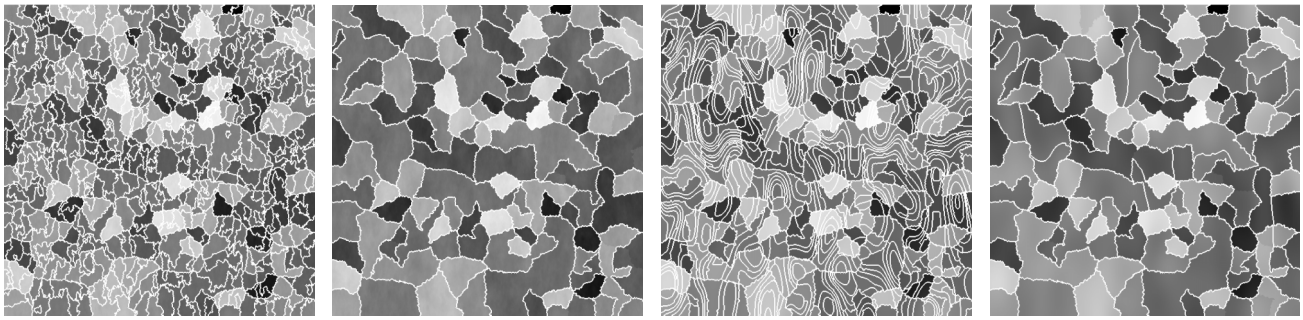
7 Possible Extensions and Discussion

We have already mentioned some possible extensions to the MDL segmentation approach that have been suggested by other authors: Perform piecewise polynomial fitting rather than piecewise constant fitting; allow spatial varying independent noise; account for the effect of blurring if the point spread function is known; and extend the principle to tackle multiband images. In this section, we discuss, and further extend if possible, such extensions to our methods.



(a) (b) (c) (d)

Fig. 8. Comparing independent and correlated noise segmentation procedures. (a) Independent noise image using independent procedure. (b) Independent noise image using correlated procedure. (c) Correlated noise image with autocovariance parameter combination 1 using independent procedure. (d) Correlated noise image with autocovariance parameter combination 1 using correlated procedure.



(a) (b) (c) (d)

Fig. 9. Comparing independent and correlated noise segmentation procedures. (a) Correlated noise image with nonseparable exponential autocovariance using independent procedure. (b) Correlated noise image with nonseparable exponential autocovariance using correlated procedure. (c) Correlated noise image with bivariate Gaussian autocovariance using independent procedure. (d) Correlated noise image with bivariate Gaussian autocovariance using correlated procedure.

We also discuss other extensions and issues in relation to our correlated noise procedure.

7.1 Piecewise Polynomial

Constructing MDL criteria for piecewise polynomial fitting is straightforward: One just needs to derive the additional code length for encoding estimates of the polynomial coefficients. When using a merging algorithm to find a local minimum, fast updating formulae exist for computing polynomial coefficient estimates *if the noise is independent* (e.g., see [15]). However, for the case of correlated noise, such fast updating formulae may not exist. Besides, sometimes it is hard to distinguish between a smooth surface and some "smoothly correlated noise": They confound each other. Thus extending our image model from piecewise constant to piecewise polynomial may not produce any practical MDL-based segmentation procedures.

7.2 Spatially Varying/Inhomogeneous Noise

Let us first consider the case of independent noise. Previous work assumed that the noise within the same region is homogeneous (i.e., of constant variance). This assumption has the nice property that it facilitates the development of fast merging algorithms for finding minima. We can relax this assumption at the expense of longer computation times. One possible way to achieve this is that, at each time step of a new merging algorithm, we segment the "residual image" (that is, the image consists of $\hat{e}_i = y_i - \hat{f}_i$, $i = 1, \ldots, n$) based on the variance characteristic but not the gray scale, and define the current "noise homogeneous zones" by the corresponding variance segmentation of the residual image. Of course, such a new merging algorithm needs more investigation and refinements to make it practical, and it would probably require a long computation time to converge to a local minimum.

The above idea can also be applied to the case of correlated noise. That is, the noise is modelled by a *nonstationary* Gaussian random field. In this case we partition the residual image into different noise homogeneous zones such that each noise homogeneous zone can be well modelled by a *stationary* Gaussian random field. These noise homogeneous zones are probably restricted to having square or rectangular shapes for fast implementation.

7.3 Blurring

Suppose that the observed image $\boldsymbol{y}$ is blurred by a *known* point spread function K: It satisfies

$$y_i = \sum_{j \in S_i} K_j f_{i+j} + \epsilon_i, \qquad i = 1, \ldots, n,$$

where the set $\{K_j\}$ is a discrete version of K and S_i defines the spatial support of $\{K_j\}$ for pixel i. Following Leclerc, we can modify the independent noise criterion $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k}, \hat{\Omega}\right)$ to handle images blurred by K. This is done by replacing

$$RSS_{\hat{k}} = \sum_{i=1}^{n} \left(y_i - \hat{f}_i\right)^2 \quad \text{with} \quad \sum_{i=1}^{n} \left(y_i - \sum_{j \in S_i} K_j \hat{f}_{i+j} \right)^2 .$$

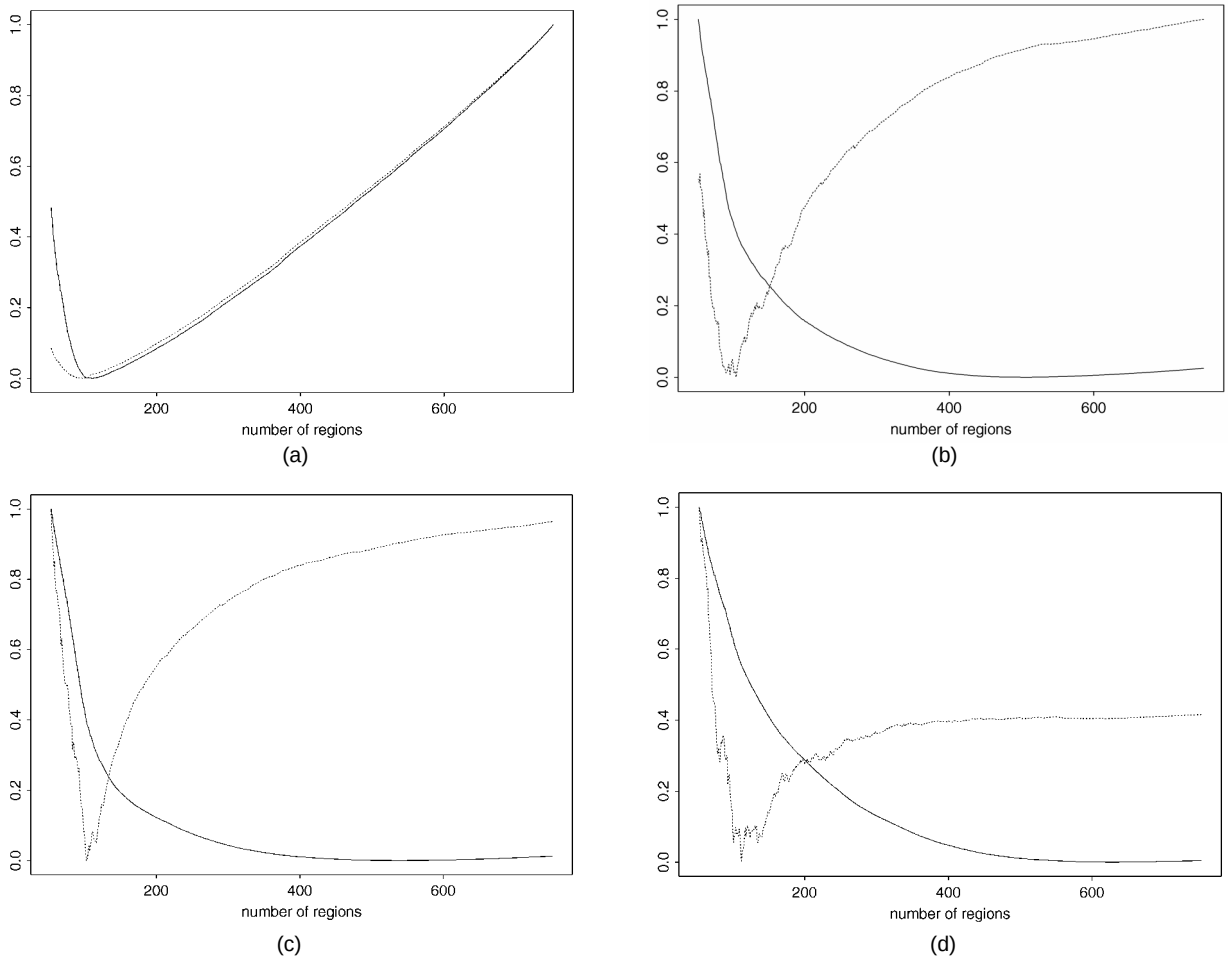


Fig. 10. Plots of MDL independent curves (solid lines) and MDL correlated curves (dotted lines) for the cases of: (a) independent noise; (b) correlated noise with separable exponential autocovariance; (c) correlated noise with nonseparable exponential autocovariance; and (d) correlated noise with bivariate Gaussian autocovariance. The x-axes represent the number of regions in the nested segmentations. Recall that there are 111 regions in the true image. Different curves have been scaled differently to improve readability.

Also, $\hat{f}_i$ should be defined in a slightly different way to (2) to account for the effect of the blurring (Leclerc did *not* address this issue). Nevertheless, we can ignore this issue if it is assumed that the size of K is small compared to the typical sizes of the true regions (i.e., the "range" of blurring is small).

Leclerc's idea can be extended to handle the case when K is only known to have a fixed parametric form. This means K can be completely specified by a (vector-valued) parameter β, and so is written as $K(\beta)$. In this situation $RSS_{\hat{k}}$ should be replaced by

$$\sum_{i=1}^{n}\left(y_i - \sum_{j\in S_i} K_j\left(\hat{\beta}\right)\hat{f}_{i+j}\right)^2,$$

where $\hat{\beta}$ is an estimate (preferably the maximum likelihood estimate) for β. However, one may have to restrict the parametric form of $K(\beta)$ in order to achieve fast computation of $\hat{\beta}$ and the resulting MDL criterion. This is in fact in a similar spirit of the choice of the autocovariance parametric form that we discussed in Section 4.1.

7.4 Other Forms of Correlated Noise

So far we have limited the noise autocovariance function to have the parametric form (4). Ignoring computational issues, there is no difficulty in extending the principle to other autocovariance functions. The only change is to the form of $\log\left|\hat{\Sigma}\right|$ (see (3)). We can also extend the principle to allowing the noise autocovariance function to have different parametric forms. For example, we can model the noise by 2D ARMA (autoregressive moving-average) models (e.g., see [16] and [8]) with variable orders (the autocovariance function (4) considered above is a separable 2D AR model), or we can use the cepstrum modeling approach proposed in [23]. The central issue now is the choice of the orders, but this can be tackled naturally by applying the MDL principle again: one just needs to add the additional (variable) code length for encoding the noise model parameters. However, such ARMA or cepstrum modelling approaches are unlikely to be practical until sufficiently fast computers become available.

7.5 Comparison of Independent and Correlated Procedures

We mentioned that $\mathrm{MDL}_{\mathrm{COR}}\left(\hat{k},\hat{\Omega}\right)$ converges to $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k},\hat{\Omega}\right)$ when both $\hat{\alpha}_r$ and $\hat{\alpha}_c$ approach infinity. Therefore one may treat the independent noise segmentation procedure as a "special case" of the correlated noise segmentation procedure, and hence always use the correlated procedure. Simulation results of Section 6.4 also support this. However, there are reasons for using the independent noise segmentation procedure, when the noise is independent or the correlation is weak.

Firstly, the independent procedure is much faster. For example, for an image of dimension 512×512 with an initial oversegmentation of about 800 initial regions, our implementation of the independent procedure usually finishes in less than 20 seconds, while the correlated procedure usually takes four minutes on a Sparc–10 machine, i.e., the independent procedure is at least 12 times faster than the correlated procedure (faster speed for both procedures is also possible as we have not fully optimized our codes).

Secondly, if the Gaussian noise is independent or "weakly correlated" (i.e., when both autocovariance parameters α_r and α_c are large), we noticed that usually the corresponding MDL correlated curve (see Section 6.4 for its definition) has a sharp local minimum near the correct place, but sometimes also has a *global* minimum at $k = 0$. That is, it suggests the underlying true image consists of one single region. We believe the reason for this is as follows. For the cases of independent or weakly correlated noise, the parameters α_r and α_c are extremely large and are always underestimated by our estimation method (note that $\alpha_r = \alpha_c = \infty$ for independent noise). This would result in inaccurate computations of $\mathrm{MDL}_{\mathrm{COR}}\left(\hat{k},\hat{\Omega}\right)$, and hence a badly behaved MDL correlated curve. This also explains that, even though $\mathrm{MDL}_{\mathrm{COR}}\left(\hat{k},\hat{\Omega}\right)$ converges *theoretically* to $\mathrm{MDL}_{\mathrm{IND}}\left(\hat{k},\hat{\Omega}\right)$ when both $\hat{\alpha}_r$ and $\hat{\alpha}_c$ approach infinity, *in practice* $\mathrm{MDL}_{\mathrm{COR}}\left(\hat{k},\hat{\Omega}\right)$ never does so, as we never have $\hat{\alpha}_r = \hat{\alpha}_c = \infty$.

We have also observed that when the noise is weakly correlated, the independent procedure produces surprisingly good segmentation results, while the correlated procedure sometimes suggests that the true image is itself a single region.

The above observations suggest that, if it is known a priori that the noise is independent or weakly correlated, the independent procedure should be used.

7.6 Independent or Correlated?

A sensible question to ask is: Should one use the independent or the correlated segmentation procedure if the noise structure is unknown? In the present situation, the most consistent way of selecting an appropriate procedure is to compare the *complete* code length for encoding the observed image assuming independent noise with the *complete* code length for encoding the observed image assuming correlated noise, and select the one that gives the smaller code length. When calculating the complete code length for the correlated case, we have to include the additional code length for $\hat{A}$, $\hat{\alpha}_r$, and $\hat{\alpha}_c$. However, these three parameter estimates are themselves correlated, and so an encoding method has to be constructed specifically for them. This is planned as a future research project, and currently we do not have an MDL-based answer to the above question.

Nevertheless, we suggest an *ad hoc* method for choosing an appropriate segmentation procedure. We will make use of the initial estimates of α_r and α_c (i.e., estimates obtained from the starting oversegmentation): If they are both larger than a prechosen critical value α_0, we use the independent procedure, as we have evidence that the noise is either independent or weakly correlated. Otherwise, we use the correlated procedure. The success of this method is highly dependent on the choice of α_0. Our numerical experience suggests $\alpha_0 = 50$.

ACKNOWLEDGMENT

This work was completed while the author was a PhD student at Macquarie University and CSIRO Mathematical and Information Sciences. He would like to express his gratitude to his supervisors, Mark Berman and Vic Solo, for their continuous encouragement, guidance and patience. He would also like to thank Byron Dom for providing a preprint of [15]. The work described in Section 4.3 is mostly due to Vic Solo.

REFERENCES

[1] R. Adams and L. Bischof, "Seeded Region Growing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, pp. 641-647, 1994.

[2] A.J. Baddeley, "Errors in Binary Images and an L^p Version of the Hausdorff Metric," *Nieuw Archief voor Wiskunde*, vol. 10, pp. 157-183, 1992.

[3] J.-M. Beaulieu and M. Goldberg, "Hierarchy in Picture Segmentation: A Stepwise Optimization Approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, pp. 150-163, 1989.

[4] P.J. Besl and R.C. Jain, "Segmentation Through Variable-Order Surface Fitting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, pp. 167-192, 1988.

[5] S. Bose and F. O'Sullivan, "A Region Based Image Segmentation Method for Multi-Channel Data," *J. Am. Statistical Assoc.*, vol. 92, pp. 92-106, 1997.

[6] M.A. Cameron, "An Automatic Non-Parametric Spectrum Estimator," *J. Time Series Analysis*, vol. 8, pp. 379-387, 1987.

[7] Y.-L. Chang and X. Li, "Adaptive Image Region-Growing," *IEEE Trans. Image Processing*, vol. 3, pp. 868-872, 1994.

[8] R. Chellappa and R.L. Kashyap, "Texture Synthesis Using 2-D Noncausal Autoregressive Models," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 33, pp. 194-203, 1985.

[9] S.-Y. Chen, W.-C. Lin, and C.-T. Chen, "Split-and-Merge Image Segmentation Based on Localized Feature Analysis and Statistical Tests," *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 457-475, 1991.

[10] C.A. Glasbey and G.W. Horgan, *Image Analysis for the Biological Sciences*. John Wiley & Sons Ltd, 1995.

[11] R.M. Haralick and L.G. Shapiro, *Computer and Robot Vision*. Addison-Wesley Publishing Company, 1992.

[12] J.D. Hart, "Kernel Regression Estimation With Time Series Errors," *J. Royal Statistical Soc. Series B*, vol. 53, pp. 173-187, 1991.

[13] A.K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[14] V.E. Johnson, "A Model for Segmentation and Analysis of Noisy Images," *J. Am. Statistical Assoc.*, vol. 89, pp. 230-241, 1994.
[15] T. Kanungo, B. Dom, W. Niblack, D. Steele, and J. Sheinvald, "MDL-Based Multi-Band Image Segmentation Using a Fast Region Merging Scheme," Technical Report RJ 9960 (87919), IBM Research Division, 1995.
[16] R.L. Kashyap, "Characterization and Estimation of Two-Dimensional ARMA Models," *IEEE Trans. Information Theory*, vol. 30, pp. 736-745, 1984.
[17] S.M. LaValle and S.A. Hutchinson, "A Bayesian Segmentation Methodology for Parametric Image Models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, pp. 211-217, 1995.
[18] Y.G. Leclerc, "Constructing Simple Stable Descriptions for Image Partitioning," *Int'l J. Computer Vision*, vol. 3, pp. 73-102, 1989.
[19] T.C.M. Lee, "A Minimum Description Length Based Image Segmentation Procedure, and Its Comparison With a Cross–Validation Based Segmentation Procedure," 1997. Unpublished manuscript.
[20] T.C.M. Lee, "Some Models and Methods in Image Segmentation," PhD thesis, Macquarie Univ., Sydney, Australia, 1997.
[21] J.W. Modestino, R.W. Fries, and A.L. Vickers, "Texture Discrimination Based Upon an Assumed Stochastic Texture Model," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 3, pp. 557-580, 1981.
[22] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*. World Scientific, Singapore, 1989.
[23] V. Solo, "Modeling of Two-Dimensional Random Fields by Parametric Cepstrum," *IEEE Trans. Information Theory*, vol. 32, pp. 743-750, 1986.
[24] G. Taubin, "Estimation of Planar Curves, Surfaces, and Nonplanar Space Curves Defined by Implicit Equations With Applications to Edge and Range Image Segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, pp. 1,115-1,138, 1991.
[25] A.L. Vickers and J.W. Modestino, "A Maximum Likelihood Approach to Texture Classification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 4, pp. 61-68, 1982.
[26] J. Zhang and J.W. Modestino, "A Model–Fitting Approach to Cluster Validation With Application to Stochastic Model-Based Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, pp. 1,009-1,017, 1990.
[27] S.C. Zhu and A. Yuille, "Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, pp. 884-900, 1996.

Thomas C.M. Lee received his BappSc (Math) degree in 1992, and the BSc (Hons) (Math) degree with University Medal in 1993, both from the University of Technology, Sydney, Australia. In 1997, he completed a PhD degree jointly from Macquarie University and CSIRO Mathematical and Information Sciences (CMIS), Sydney, Australia. His thesis title was "Some Models and Methods in Image Segmentation."

During 1995, Dr. Lee was on leave from his PhD program and gained practical experience in attacking real-imaging problems by working as an image analyst at CMIS. Currently, he is a visiting assistant professor at the Department of Statistics at the University of Chicago. His main research interests lie in the interface between statistics and image analysis; he has worked on the problems of image restoration and segmentation, texture synthesis, nonparametric regression, errors-in-variables models, staticstical model selection, and linear feature detection in images.