AGGREGATION AND REFINEMENT IN BINARY IMAGE RESTORATION

Mike Jubb and Christopher Jennison
School of Mathematical Sciences
University of Bath
Bath, BA2 7AY
United Kingdom

ABSTRACT

The work reported in this paper grew out of an investigation into the use of "refinement" methods to obtain restorations of an image with detail at a level finer than the pixel grid on which records are observed. A "cascade" algorithm is presented that produces restorations on successively finer pixel grids, starting with a single large pixel and ending with the original grid.

1. Introduction

Recent developments in statistical image restoration use a Bayesian approach. One observes a degraded version of a true scene after the addition of noise and possibly, blurring. If the degradation process and noise distribution are known, the likelihood of the record can be combined with a prior probability model to produce a posterior distribution for the true scene. A common approach is then to seek the *maximum a posteriori* (MAP) estimate of the scene and present this as the restored image.

For computational purposes it is extremely convenient to work with Markov random field (MRF) models. Under a MRF model the scene is divided into pixels, each of which can take a single color or grey level, a neighborhood structure for the pixels is specified and the key property of the model is that the distribution of the coloring of any pixel is conditionally independent of all other pixels, given the coloring of its neighbors.

There are two main approaches to searching for the MAP estimate. Geman & Geman (1984) proposed the method of simulated annealing. They have shown this to be a versatile and effective method although the amount of computation involved is often high. Besag (1986) suggested a computationally simpler method which he refers to as the method of iterated conditional modes (ICM). This method will normally converge to a local rather than global maximum of the *a posteriori* likelihood; however, convergence is rapid and, given the approximate nature of the MRF model, failure to find the global maximum may not be a serious drawback.

Jennison (1986) and Jennison & Jubb (1987) have shown that the same form of MRF model can be used to obtain restorations of an image with detail at a finer level than the pixel grid on which records are observed. In their original examples the noise level was very low. The work reported in this paper grew out of an investigation into the use of "refinement" methods in the presence of greater noise: the main problem in this case is to find a good starting point for the refinement algorithm. In some of our exploratory examples we discovered that the ICM method itself experienced serious difficulties at very high noise levels. One solution to this problem is to increase the signal to noise ratio by aggregating the records of, say, each 2 by 2 block of pixels into a single record: satisfactory results were obtained by applying ICM to the aggregated signal and then using the resulting restoration as the starting point for ICM on the original pixel grid. A natural extension of this idea is a "cascade" algorithm, similar to that of Gidas (1989), which produces restorations on successively finer pixel grids, starting with a single large pixel and ending with the original grid. We have found that this approach provides a simple and efficient way of adapting the ICM method to very noisy data. It also solves the refinement problem, since the end product of this algorithm, or even a restoration based on aggregated data, will provide a good starting point for the refinement process.

Our intention in this paper is to follow the ICM approach as much as possible. There are several places where simulated annealing might be incorporated but it would require substantially more computing, and there is no guarantee that it would provide better results. The main advantage of simulated annealing is that it allows one to escape from a local maximum of the posterior likelihood by a process of trial and error, however, use of the cascade algorithm to choose a good starting point for the deterministic ICM algorithm may be just as effective. We do introduce a version of simulated annealing to implement the refinement method of Section 5. Although this provides a very convenient way of exploring a larger set of restorations, its impact on the final restored image for our example is slight.

Some comment on the role of the prior model for the true scene is called for. Gidas (1989) goes to great lengths to ensure that, in his cascade algorithm, the models at different pixel sizes are mutually consistent. We are not committed to a single model and will be happy as long as the final restoration is a good one. It should also be remembered that all that we require of the end product of one stage of the cascade algorithm is that it should provide a good starting point for the next. We do not assume that we have a global MAP estimate at any stage, nor do we try to make use of such a property.

We shall use a single illustrative example throughout the paper. In the original image the boundaries of objects are smooth in parts but irregular in other places and certain features are extremely difficult to restore given the level of noise in the data. Thus, the example shows both the power of the proposed method and its limitations.

2. Model and notation

We first consider a rectangular region partitioned into pixels labelled $1, 2, \ldots, n$ each pixel is colored black or white and the color of pixel i is denoted by x_i^* which takes the value 0 for white and 1 for black. The x_i^* are unobserved. It is assumed that the conditional density function $f(y_i|x_i^*)$ is known and for the remainder of this paper we shall assume that the records y_i are independently distributed as Gaussian with mean x_i^* and variance σ^2. The set of records is denoted by $y = \{y_i; i = 1, \ldots, n\}$. A coloring of pixel i (not necessarily the true coloring, x_i^*) is denoted by x_i and a specific coloring of the whole region is denoted by $x = \{x_i; i = 1, \ldots, n\}$.

In the MRF model for the true scene we shall use a neighborhood system in which pixels are considered to be first order neighbors if they are horizontally or vertically adjacent to each other and second order neighbors if they are diagonally adjacent. In our model, the prior distribution for the true scene, $p(x)$, is

$$p(x) \propto \exp[-\{\beta_1 Z_1(x) + \beta_z Z_2(x)\}], \tag{2.1}$$

where $Z_1(x)$ is the number of discrepant first order pairs in the scene x, i.e. the number of pairs of first order neighbors which are of opposite color, $Z_2(x)$ is the number of discrepant second order pairs and β_1 and β_2 are fixed positive constants.

The MAP estimate of the true scene is the value of x which maximizes $P(x|y)$, the conditional probability of x given the record y. By Bayes' theorem

$$P(x|y) \propto l(y|x)p(x), \tag{2.2}$$

where $l(y|x)$ is the conditional likelihood of the observed record y, given the true coloring, x, and $p(x)$ is the prior probability of x. Thus, the maximization of $P(x|y)$ corresponds to the minimization of

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i)^2 + [\beta_1 Z_1(x) + \beta_2 Z_2(x)], \tag{2.3}$$

over values of $x = \{x_i; k = 1, \ldots, n\}$.

Besag's (1986) method of iterated conditional modes updates each pixel in turn, choosing for it the most likely color based on its record and the current coloring of its neighbors, i.e., minimizing (2.3) with respect to x_i with all the other pixel colorings fixed. The expression in (2.3) must decrease or remain constant at each updating but convergence will usually be to a local minimum. We shall see later in this paper that the choice of the initial coloring can have a great influence on the accuracy of the final restoration. Throughout this paper, when ICM is applied, a second order neighborhood system will be used with $\beta_2 = \beta_2/\sqrt{2}$; this ratio of β_1 to β_2 minimizes the rotational variance of the second term of (2.3) with respect to the positioning of the pixel grid on a given scene (see Brown, Jennison and Silverman, 1987).

In the above model for the true scene it is assumed that each pixel is colored wholly black or white. This is at best an approximation: more generally, one might expect pixels on the boundary of an object to contain areas of each color, in which case the record y_i will be distributed as Gaussian with variance σ^2 and mean equal to the proportion of pixel i colored black. Although we shall consider problems in which there is a general true scene, we start by considering restorations based on a discrete MRF model in which each pixel has a single color. The refinement method described in Section 5 does, however, allow boundary pixels to be colored partly black and partly white.

3. An example

An example of a binary scene containing two separate objects is shown in Figure 1. A 256 by 256 pixel grid was superimposed on this scene and the proportion, p_i, of black of pixel i as calculated for each pixel. The record y_i was obtained by adding Gaussian noise with variance 4 to this proportion, p_i. Figure 2 shows the closest mean classifier for this record, in which a pixel is colored black if its record is greater than 0.5 and white otherwise. One would not normally hope to recover an image which has been exposed to such a large amount of noise and Figure 3 shows the rather unsatisfactory restoration obtained by applying ICM with $\beta_1 = 4$. The value $\beta_1 = 4$ is unusually high but we found this to give the best results. (Note that even if $\beta_1 \to \infty$ certain configurations of pixels remain unsmoothed.)

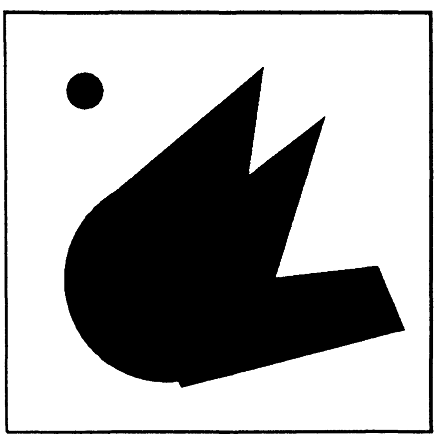


Figure 1. The true scene.

The major problem in our example is the low signal to noise ratio. This ratio may be improved by aggregating the record, i.e., by replacing sets of 2 by 2 pixels by a single large pixel with record equal to the average of the original four. This also corresponds to viewing the original image on a coarser grid. The variance of the new record is one quarter that of the original but the range of the p_i's is still $[0, 1]$; thus there is a substantial increase in the signal to noise ratio. The restoration shown in Figure 5 was obtained by applying ICM to the aggregated record; the prior model for the true scene had the same form as (1.1) but was applied to larger pixels, the value $\beta_1 = 4$ was also used here as it was found to give the best results. The clear superiority of this restoration to that shown in Figure 4 demonstrates the advantage of working with the aggregated record. One explanation of the success of this restoration process is that it allows the ICM algorithm to look further afield when gathering neighbor information; ICM on the original pixel grid can easily be trapped in a local maximum of the *a posteriori* likelihood when only one pixel is allowed to change at a time.

Repeating the aggregation process gives the restoration shown in Figures 5 and 6, which are the restorations at two and three levels of aggregation respectively. These restorations were obtained using $\beta_1 = 1$, a more typical value, which we have found gives good results in cases where the signal to noise ratio is moderate. Note that the computational time and storage requirements for the processing of a 32 by 32 image are approximately $\frac{1}{64}$ times those needed to process a 256 by 256 image.

So far, we have followed Besag's method and used the closest mean classifier as our initial coloring for the 256 by 256 case and this partly responsible for the poor quality of the restoration in Figure 3. A better initial coloring might be the final restoration obtained from an aggregated record. Figure 7 shows that result of using Figure 5 as the initial coloring for ICM on the 256 by 256 grid with $\beta_1 = 4$; a similar result is obtained with $\beta_1 = 1$. The superiority of this restoration to that of Figure 3 demonstrates the influence of the initial coloring on the resulting image.

The method of simulated annealing is less dependent on the initial coloring, since it can progress from one local minimum of (2.3) to another whilst passing through higher intermediate values. Thus, simulated annealing is able to search at least a little further afield than the myopic ICM strategy. An advantage of using an aggregation procedure is that it allows the ICM approach to use more distant neighbor information whilst maintaining its computational speed.

4. The Cascade algorithm

In the previous section we introduced the idea of using the restoration obtained from an aggregated record as the initial coloring for restoration on a finer scale. We now extend this idea to define a "cascade" algorithm in which restorations obtained from 2^m by 2^m grid are used as the initial colorings for restorations on 2^{m+1} by 2^{m+1} grid. A single pixel restoration is obtained by aggregating the record until it is one pixel in size: this is then used as the initial coloring for the ICM method on the 2 by 2 grid. This restoration is in turn used as the initial coloring for ICM on the 4 by 4 grid and we continue in this way, obtaining restorations right up to the level of the original record. The last six in the series of restorations for our example are shown in Figures 8–13; the value $\beta_1 = 1$ was used at each level, though it is interesting to note that using higher values at the 128 and 256 levels made virtually no difference to the image obtained.

The method of Gidas (1989) is very similar to the procedure we have just described. However, Gidas uses a single MRF model defined on the finest pixel grid and employs

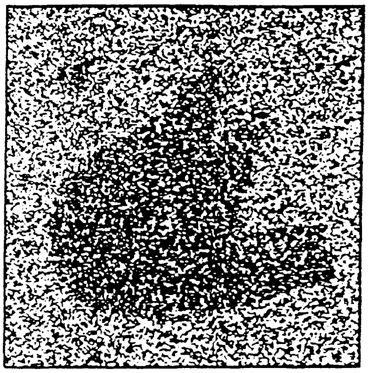


Figure 2.

Figure 3.

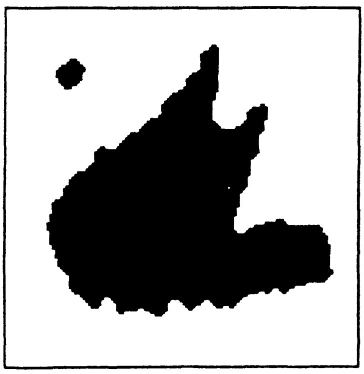


Figure 4.

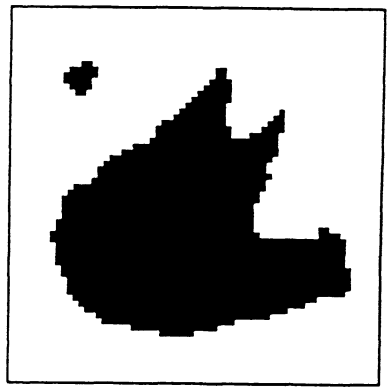


Figure 5.

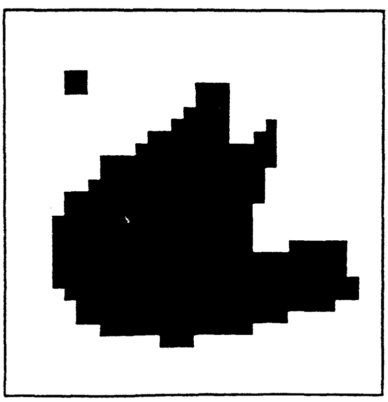


Figure 6.

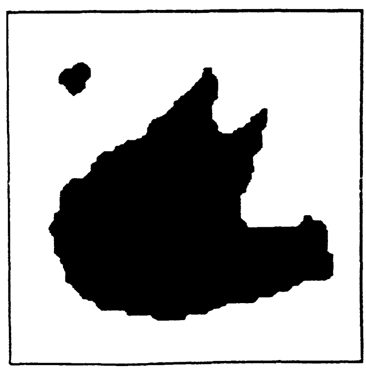


Figure 7.

Figures 2–7.

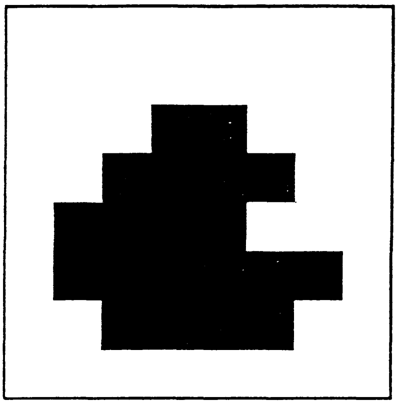


Figure 8.

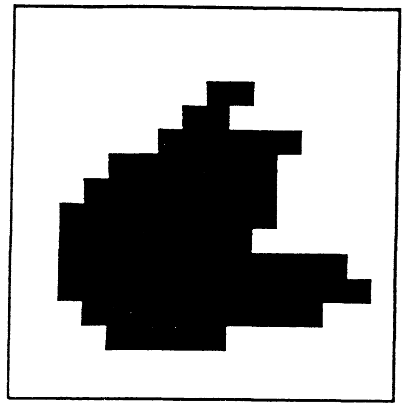


Figure 9.

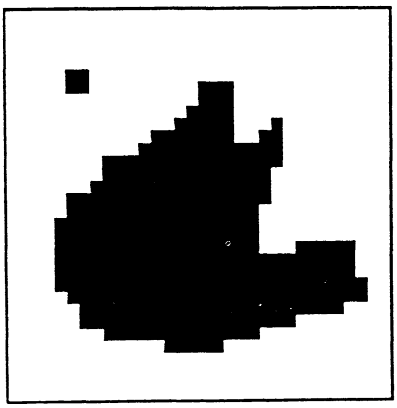


Figure 10.

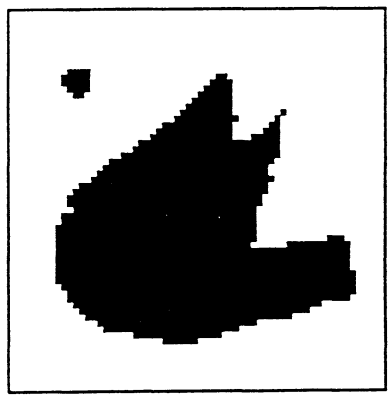


Figure 11.

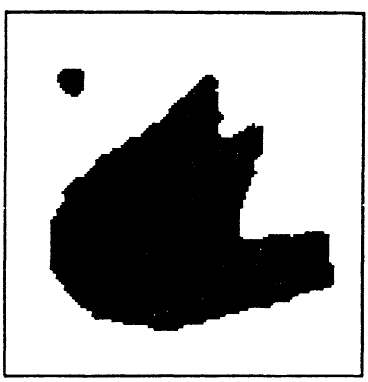


Figure 12.

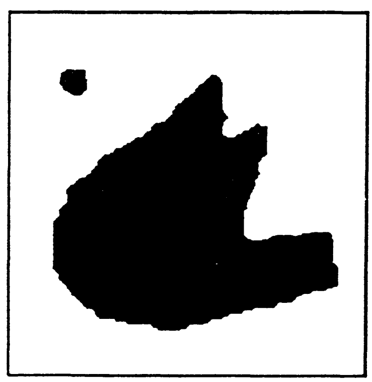


Figure 13.

Figures 8–13.

the "renormalization group" approach to compute the models implied for coarser grids. Both the complexity of the models at the aggregated levels and the use of simulated annealing at each stage makes this a computationally demanding method. We have tried to keep computation to a minimum at the expense of a less rigorous treatment of the prior model: given the approximate nature of this model, we would argue that this is not unreasonable.

One might at least try to develop theoretical arguments to produce a "correct" sequence of values of β_1 for use at different stages of the cascade algorithm. Brown, Jennison and Silverman (1987) interpret the second term of (2.3) as a penalty and suggest that it should be chosen to be approximately independent of the pixel grid superimposed. They suggest that this penalty should approximate a constant multiple of the total boundary length in the image. In our application this would imply that the parameter β_1 be halved as the pixel sizes are quartered but we have not found this to be very successful in practice. Using the same value of β_1 at each stage produced substantially better results.

When processing the larger images we avoid unnecessary computations by storing the coordinates of pixels whose colorings have changed in the current iteration. If the number of these is small, only pixels whose neighbors have changed color in the last iteration are considered for updating in the next iteration. For each of the images shown in Figures 8–13 one complete iteration plus some minor changes was all that was required. Summing a geometric series, we see that the total computation required is approximately equivalent to $1\frac{1}{3}$ iterations of ICM on the finest pixel grid.

We have seen that the restorations obtained on the finer grids have been insensitive to the choice of β_1. This is partly attributable to the high noise level (updating is essentially by the "majority vote rule" at quite high values of β_1) but also suggests that, for a given image, restoration at too fine a pixel level is unnecessary, adding only computation and superfluous detail to what is already a satisfactory restoration. We are able to make a direct comparison of restorations obtained at different levels of aggregation by superimposing the the finer grid on the coarser image and calculating penalties for both, based on the finer record and the MRF model at that level. The coarser image is disadvantaged, since it was chosen when searching for the minimum of a different penalty. We measure the benefit of restoring at the finer level by the percentage decrease in the penalty. The values are tabulated below.

Grid size of coarse restoration	Grid size of fine restoration	percentage reduction in penalty
2×2	4×4	68.1
4×4	8×8	75.8
8×8	16×16	49.6
16×16	32×32	21.5
32×32	64×64	5.2
64×64	128×128	1.6
128×128	256×256	0.6

Analysis of these values is purely subjective but appears to suggest that the 64 by 64 level is satisfactory. Inspection of Figures 8–13 also leads to the same conclusions.

5. Subpixel refinement

So far the restoration techniques we have used have colored each pixel wholly one color, even though pixels on the edges of objects in the true scene may be partly black and partly white. We now consider techniques which allow both colors to appear in a single pixel. Jennison (1986) used a modification of the ICM method to obtain a restoration in which each pixel was divided into a 4 subpixel quarters and a separate color allocated to each subpixel. His method used the ICM restoration at full pixel size as a starting point for restoration at the subpixel level. The success of this technique prompted Jennison and Jubb (1987) to consider the further refinement of pixels.

Since the number of different coloring of a pixel grows exponentially with the number of subpixels, the extension of Jennison's method to a finer subdivision of each pixel is computationally prohibitive. However, the limit of this process, in which an arbitrary coloring of each pixel is allowed, can be made tractable. Rather than specify a MRF model for the true scene we interpret the minimization of (2.3) as a form of penalized maximum likelihood. The second term of (2.3) is, approximately, a multiple of the total boundary length in the image, x. Thus, an analogous penalty for a general restoration, x, is

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i p_i(x))^2 + \beta L(x), \tag{5.1}$$

where $p_i(x)$ denotes the proportion of black in pixel i, $L(x)$ is the total edge length in scene x and β is a fixed constant. For computational simplicity we restrict attention to restorations in which pixels are either of a single color or are separated into areas of different color by a single straight line with the line segments defining such areas in adjacent pixels meeting at a point.

A black and white image can be regarded as a series of line segments separating the two colors. Jennison and Jubb (1987) use the restoration obtained from Jennison's quarter pixel method is used an initial representation for the line segments. The updating process treats pixels in pairs, selecting the best place for two edges to meet, given the current restoration of neighboring pixels. We repeat the details for completeness.

As an example, consider the configuration at pixels i and j shown in Figure 14. The distances a and b are determined by the current coloring of neighboring pixels and treats as constant for the moment. The distance W is chosen to minimize the contribution from pixels i and j to the total penalty (5.1), i.e.

$$g(W) = \frac{1}{2\sigma^2}\sum_{k=i,j}(y_k - p_{kW})^2 + \beta(e_{iW} + e_{jW}), \tag{5.2}$$

where e_{kW} is the length of edge in pixel k when the join is at W and p_{kW} is the proportion of black in pixel k when the join is at W.

For the case shown in Figure 14, this penalty is

$$g_1(W) = \frac{1}{2\sigma^2}\{(y_i - a - \frac{1}{2}(W-a))^2 + (y_j - b - \frac{1}{2}(W-b))^2\} + \beta\{\sqrt{1+(W-a)^2} + \sqrt{1+(W-b)^2}\}.$$

This can not be minimized directly but the form of

$$\frac{dg_1(W)}{dW} = \frac{1}{4\sigma^2}(2W + a - 2y_i + b - 2y_j) + \beta\left[\frac{(W-a)}{\sqrt{1+(W-a)^2}} + \frac{(W-b)}{\sqrt{1+(W-b)^2}}\right]$$

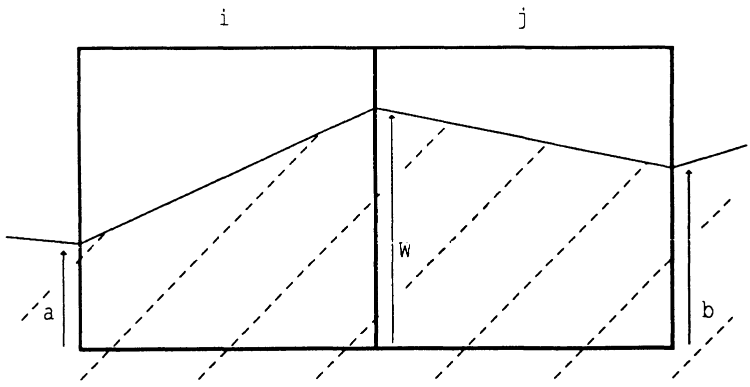


Figure 14. Updating the position of edges in pixels i and j.

suggests an iterative approach. Given an approximate solution W_{s-1} we solve

$$\frac{1}{4\sigma^2}(2W_s + a - 2y_i + b - 2y_j) + \beta \left[\frac{(W_s - a)}{\sqrt{1 + (W_{s-1} - a)^2}} + \frac{(W_s - b)}{\sqrt{1 + (W_{s-1} - b)^2}} \right] = 0$$

to obtain

$$W_s = \frac{4\sigma^2\beta \left[\frac{a}{\sqrt{1+(W_{s-1}-a)^2}} + \frac{b}{\sqrt{1+(W_{s-1}-b)^2}} \right] + (2y_i - a + 2y_j - b)}{2 + 4\sigma^2\beta \left[\frac{1}{\sqrt{1+(W_{s-1}-a)^2}} + \frac{1}{\sqrt{1+(W_{s-1}-b)^2}} \right]}.$$

Starting from any sensible initial value, W_0, accuracy to 3 decimal places was achieved after at most four iterations. In practice we take W_0 to be the value of W prior to this update.

Different forms of (5.2) are possible depending on which neighbors of pixels i and j contain both colors. There are only four distinct cases that may arise and these are shown in Figure 15.

We have shown the method of solution for case (i); cases (ii)–(iv) are solved in a similar way. All other cases can be reduced to one of the above by means of exchanging and/or inverting the pixels and their colours. The edge pixels are updated in turn, following an edge around, completing circuits of the edge until convergence.

The complete restoration algorithm

We can now combine both aggregation and refinement into a three stage algorithm:

Stage 1: Apply the cascade algorithm using ICM on the aggregated records up to a suitable point. The record is now fixed at this level and no further use will be made of the original record. (If the record is still aggregated at this level substantial savings in computation will result).

Stage 2: Iterate Jennison's quarter pixel refinement to convergence. This is very quick and supplies a good starting point for the line fitting process.

Stage 3: Apply the line fitting algorithm to convergence.

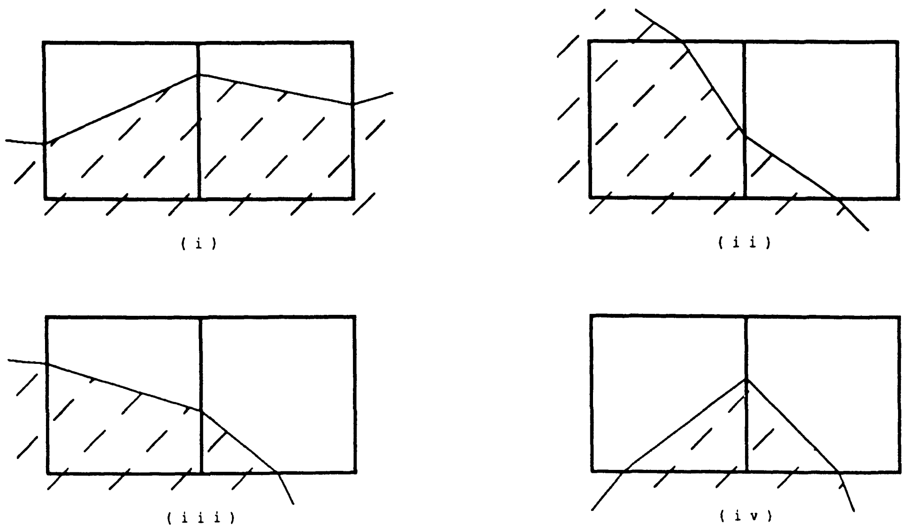


Figure 15. Possible configurations of edges in two neighboring pixels.

A development in the line fitting algorithm

In the line fitting algorithm described by Jennison and Jubb (1987) the route that the lines take through pixel edges is determined once and for all by the restoration obtained at the quarter pixel level.

We have now extended the algorithm to allow changes in this route. Each time the point at which the edge crosses a pixel boundary is updated an alternative route is compared. A number of cases have to be treated separately; three qualitatively different configurations are shown in Figure 16.

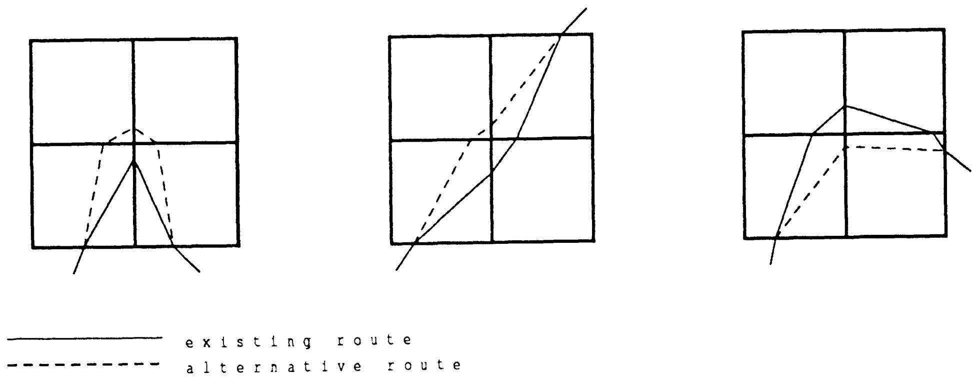
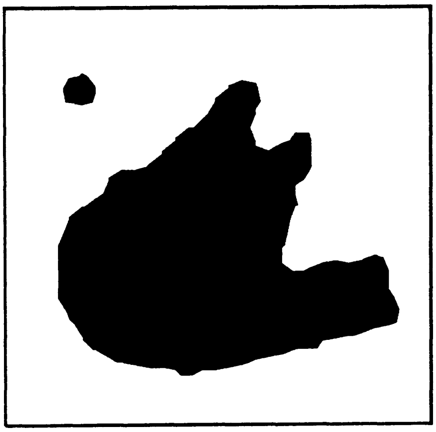
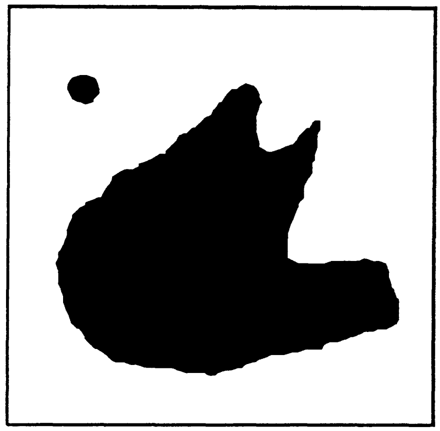


Figure 16. Examples of configurations at which alternative routes are considered.

The contribution to the total penalty from all four pixels is calculated for each of the two routes with line edges chosen optimally for that route. In the basic method, the route which has smallest penalty is then chosen.

Figures 17 and 18 show the restorations obtained from applying the line fitting method to the aggregated record in the example. In Figure 17 the cascade algorithm was applied until the grid size was 32 by 32, and in Figure 18 a 64 by 64 grid was used. In the previous section we suggested that a grid of 64 by 64 would be sufficient and the restoration shown in Figure 18 is indeed satisfactory. In both cases we used $\beta_1 = 1$ at the ICM and quarter pixel levels of restoration and $\beta = 4$ for the line fitting.





Figures 17–18.

The updating process in the above line fitting procedure has the general characteristics of an ICM method: the penalty (5.1) is minimized with respect to one component of the boundary whilst everything else is held fixed. This method will generally yield a local minimum of (5.1) and it is possible that the final restoration could be improved further by making a number of route changes simultaneously. For example, the penalty (5.1) might be reduced by moving a long vertical edge one pixel to the left whereas it would increase initially if only one route change were made at a time.

To allow further exploration of alternative routes we have implemented a form of simulated annealing. This method retains the property that for a given route the point on a pixel edge at which two line segments meet is chosen optimally. However, when comparing the minimum penalties for *different* routes we allow the route with the larger penalty to be chosen with non-zero probability. Suppose two routes, A and B, have minimum penalties pen_A and pen_B then, when the annealing process is at temperature T we select route A and its optimal edges with probability

$$\frac{e^{(-pen_A/T)}}{e^{(-pen_A/T)} + e^{(-pen_B/T)}}$$

otherwise we choose route B. Of course, only the contribution to the total penalty from the four pixels concerned need actually be calculated.

By restricting the random choice to the route alone, we ensure that, effectively, the annealing process is applied to a fairly low dimension problem, the number of variables being of the order of the number of boundary pixels. Theorem B of Geman and Geman (1984) demonstrates the convergence of their simulated annealing method. In its stated form, this theorem does not apply to our hybrid procedure whose iterative steps combine a random choice of route with a deterministic choice of edges given that route the currently fixed end points. Perhaps a sufficiently general result could be proved but this would, presumably, still only apply for gentle cooling schedules. However, we prefer to think of the annealing method simply as a convenient numerical procedure which searches a little further afield than the ICM approach.

We have experimented with a variety of cooling schedules for our example using the aggregated record at both the 32 by 32 and 64 by 64 grid levels. The best results were obtained using a cooling schedule in which T decreased logarithmically from 3.5 to 0.5 over several hundred sweeps and linearly from 0.5 to zero over several hundred more. We then continued to update using $T = 0$ until convergence, which usually required only a few sweeps. This corresponds to ICM and guarantees convergence to a local maximum. Although simulated annealing often produced a lower penalty, the restoration produced was never visually superior to that obtained using the local maximization procedure.

Our conclusion is that the starting point provided by the cascade algorithm was sufficiently good that the deterministic line fitting algorithm was very nearly optimal.

6. Concluding remarks

Combining the line fitting procedure with the cascade algorithm has produced a fast and effective method for obtaining a high quality restoration from noisy data. Further work is required to provide an automatic choice of suitable values of β_1 at different grid levels and a criterion for terminating the cascade algorithm at the most appropriate level of aggregation. Although we have considered only two-color images in this paper, it is clear that the basic ideas are more generally applicable: we hope to continue work on the development of an aggregation and refinement algorithm for grey level images.

References

Besag, J.E. (1986) On the statistical analysis of dirty pictures, *J. Royal Statist. Soc. B,* 148, 259–302.

Brown, T.C., Jennison, C. and Silverman, B.W. (1987) Edge process models for regular and irregular pixels. Submitted for publication.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6, 721–741.

Gidas, B. (1989) A renormalization approach to image processing problems. *IEEE Trans. Pattern Anal. Machine Intell., to appear.*

Jennison, C. (1986) Contribution to discussion of Besag (1986) *J. Royal Statist. Soc. B,* 148, 288–289.

Jennison, C. and Jubb, M. (1987) Statistical image restoration and refinement, *Proc. XII Inf. Proc. in Med. Imaging Conf., to appear.*