Monograph:
Baghai-Ravary, L., Beet, S.W. and Tokhi, M.O. (1995) Image Coding by Multi-Step, Adaptive Flux Interpolation. Research Report. ACSE Research Report 593 . Department of Automatic Control and Systems Engineering

Department of
AUTOMATIC
CONTROL
and SYSTEMS
ENGINEERING
UNIVERSITY of SHEFFIELD

IMAGE CODING BY MULTI-STEP, ADAPTIVE FLUX INTERPOLATION

L Baghai-Ravary [*], S W Beet [**] and M O Tokhi [*]

[*] Department of Automatic Control and Systems Engineering,
[**] Department of Electronic and Electrical Engineering,
The University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK.

Tel: + 44 (0)114 2825136.
Fax: + 44 (0)114 2731 729.
E-mail: O.Tokhi@sheffield.ac.uk.

Research Report No. 593

July 1995

Abstract

This paper describes and discusses a new technique, the multi-step adaptive flux interpolation (MAFI) and its application to image data for coding. The output of MAFI, when applied to an image, is still in an image form but has a more uniform feature density. This is because the original image has been warped by removing those rows and columns which contain mostly redundant pixels. It is also greatly reduced in size and the side information is minimal. The MAFI output can be further compressed using conventional coders, making its compression ratio even higher. Because of its warped nature, the MAFI output's statistics are also more consistent with the properties assumed by block-based discrete cosine transform (DCT) methods.

Key words: adaptive flux interpolation, image coding, image compression.

CONTENTS

LIST OF TABLES AND FIGURES

1 Introduction

Image coders based on the two-dimensional discrete cosine transform (DCT) (ISO/TC97/SG2/WG8-ADCTG, 1988), such as that specified by JPEG (Joint Photographic Image Expert Group) (Wallace, 1991), produce outputs which are different in form from that of their inputs and cannot be viewed as an image, directly. DCT methods also generally use the same vector quantisation process for all parts of the image. Thus, they inherently assume that the data statistics are uniform, but in many cases this is not true as different parts of the image may exhibit very different feature densities. However, if the image were warped before being encoded, the assumption of uniformity could be made more realistic. An arbitrary two-dimensional warping function could be applied, but this would require something of the order of at least one bit per pixel to specify the proximity of each encoded pixel to its neighbours in each direction. Such a high level of side information would be prohibitive. A more practical solution is to identify and remove the whole of any columns or rows which are redundant.

The redundant columns and rows in a (two dimensional) image contain data which could be interpolated from their neighbours. Provided the interpolation algorithm is sophisticated enough, the neighbours need not be very similar to one another, or to the data being interpolated. Image compression can be achieved by removing this redundant data. Removal of complete columns and rows leaves the data in the form of a reduced dimensionality image (with minimal side information) which can be further compressed by DCT and other algorithms (Cicconi, et.al, 1992; Matsuda, et.al, 1994). This technique improves on the compression ratio and modifies the image statistics, producing a more uniform distribution of spatial frequencies over the image.

2 Compression process

Adaptive Flux Interpolation (AFI) is a powerful method for interpolating a vector (Baghai-Ravary, et. al., 1995a,b). Multi-step AFI (MAFI) interpolates a sequence of vectors and can be used to identify those vectors which are redundant. This algorithm is discussed further in the next section. By treating the image as a matrix with its columns as vector inputs to MAFI, the redundant columns can be removed. The output of the MAFI process is in a compressed form which consists of the remaining columns concatenated, together with a set of numbers recording interpolated-block-sizes to allow reconstruction of the image. The image is, therefore, of a reduced horizontal dimensionality while still maintaining an image form. This reduced image matrix can be transposed, and then the whole process re-applied to remove the redundant rows. Further redundancy can be removed by repeating this process one or more times (see Figure 1).

3 MAFI modus operandi and algorithm

Multi-step Adaptive Flux Interpolation (MAFI) is analogous to optical flow (Adiv, 1985; Bergholm and Carlsson, 1991), as both techniques are used for tracking, but MAFI functions in a two dimensional space whereas optical flow is used for image sequences which are effectively three dimensional.

It is assumed that there are lines of flux within an image which join pixels of similar values. These lines of flux together form the image. Depending on the direction in which they propagate, different sets of these lines are able to represent the same image. As an example, Figure 2(a) show a picture of a box and Figures 2(b) and (c) show the flux lines which are present in that picture (propagating from 'left to right' and from 'top to bottom', respectively).

MAFI identifies the lines of flux in the data and *interpolates* along those lines to achieve an accurate estimation of the redundant data. Where the flux lines bend, the MAFI algorithm *adapts* to cope with those changes. Interpolation along lines of flux is less prone to error because the variation in value along them is small. To take advantage of alternative explanations of the image (such as that in Figures 2(b) and (c)), the data may be transposed and resubmitted to remove further redundancy. Given two vectors, N_1 and N_2, vector N (where $N_1 < N < N_2$) is estimated using MAFI. The likelihood of each pair of elements (pixels) within the vectors being linked is assumed to be given by a zero-mean Gaussian distribution of the change in data value from one end of the link to the other.

Consider a line of flux which joins element u of vector N_1 to element v of N_2, as shown in Figure 3, linking i in vector n-1 to j in vector n. A local distance matrix, Γ, containing the likelihood of every potential flux line, is calculated to give all possible links between the elements of vectors n-1 and n. Each value of this local distance matrix, Γ, is expressed in a normalised log-likelihood form:

$$\gamma_{i,j} = \left| x_{N_1,u} - x_{N_2,v} \right|^2$$

where $x_{N_1,u}$ is the value of element u in vector N_1, at the point of incidence of the flux line. $x_{N_2,v}$ corresponds to the other (extrapolated) end of the link, where it intercepts vector N_2.

Dynamic programming is used to find the set of flux lines which give the maximum total log likelihood over all possible combinations of lines. The use of dynamic programming to identify this mathematically optimal set of links also allows the imposition of arbitrary constraints to ensure that any interpolation is performed in a visually acceptable fashion. In particular, very oblique lines are visually distracting, so these are disallowed by setting the respective elements of the local distance matrix to infinity.

The redundant vectors (columns or rows) are identified by successively attempting to encode blocks of increasing numbers of vectors, until a normalised measure of the root mean square error exceeds a threshold. At this point, the last vector in the block, N_2, is transmitted, together with a number giving the size of the block, N_2-N_1. The threshold is chosen to give the desired image quality.

In the box example, there are only 36 pixels, plus a minimum of side information (in the order of 10^{-4} bits/pixel) which carry all the necessary material for reconstruction of that image. This is independent of the shape or the size of the box, or the size of the picture as a whole, although the side information very slightly increases with the width and height of the picture. This picture is simple and very geometrical, thus after compression, it can be reconstructed with negligible error.

4 Coding considerations

The degree of compression possible with MAFI depends on several factors:

The acceptable level of image distortion The value of the threshold, used to distinguish between the critical and the predictable vectors in the image, determines the level of maximum distortion which is allowed to be present in the data after reconstruction. The higher the threshold, the more data is perceived as being redundant and, thus, the higher the compression ratio.

The nature of the image An image with a low average feature density gives a higher compression ratio than an image dominated by a higher feature density. Fine details contain more critical information and are therefore less predictable than gross detailed ones. However, the shape and structure of the features also influences the final compression. For example, an image containing only combed hair is compressed more than one containing very curly or fuzzy hair.

The form of "visual acceptability" constraints applied during dynamic programming As mentioned previously, dynamic programming (DP) is used to match and link the similar pixels of the columns together, to yield the flux lines. A limit is set in the DP routine so that this procedure is carried out reasonably and unrealistic links, such as a very sharp one from one end of one column to the other end of the next, cannot be made. This type of link is not usually visually acceptable in an image and is therefore disallowed.

The freedom of the feature to match and link from one column to another is thus fixed by a limit parameter. If this limit is set too high the reconstructed data may contain unrealistic and visually unacceptable distortions. This is especially noticeable if the known frames are widely separated, and in practice, the further apart these frames are, the smaller the limit must be to obtain reliable interpolation. Conversely, if the limit is set too low, it restricts the interpolation and may result in unnecessary distortion because the pixels may not be able to be matched correctly (although the distortion is not likely to be of the same severity). In both cases, the compression ratio will generally decrease as less of the data is able to be estimated well.

The number of times MAFI is applied and its direction on each application MAFI can be applied several times without lowering the standard of allowable error in the reconstructed data (set by the threshold) if each application is at an angle to the others. This, in fact, further increases the compression ratio on each application.

In this paper, and for the sake of simplicity, MAFI is oriented orthogonally on each application to remove the redundant rows and columns in the data. However, it is not necessary that the adjustment should be orthogonal, as long as it is at a different angle to its previous application. Figure 4 shows another method for utilising MAFI. Here, MAFI is applied in four directions, each 45^{o} to the other (from Figures 4(a) to (d) respectively: the arrow shows the direction of MAFI application). This technique will not only remove the predictable rows and columns, but also any predictable diagonal vectors. The only drawback to this approach is that the rectangular image format is lost and so the MAFI output would be difficult to encode further.

It is not practical to apply MAFI in the same direction consecutively. This would have a similar effect to an increase in the threshold on the former application and could give incorrect results (see Figure 5).

The compression ratio increases with each application of MAFI, although if it is applied too many times successive increases become smaller and eventually reach zero when all redundant pixels have been removed. The number of times MAFI is usefully applied depends on the nature of the image, the threshold and the DP limit settings (that is on the quality and desired characteristics of the reconstructed image).

Any changes in the MAFI parameters between applications This includes threshold, DP limit, and the direction of application. The structure and the nature of the output of each MAFI application are different from those of its input. For example, the output would have a higher feature density, where redundancy has been detected, and dimensionally it is reduced in size. Consequently better results could be obtained by changing the MAFI parameters for each application to match the properties of the respective input images, and achieve an optimal result. However, in practice, it is the characteristics of the original image which have the greatest effect on the choice of parameters. It is possible to achieve good levels of performance with just a single set of parameters for multiple applications, provided the direction of application is changed each time.

5 Experiments

A number of experiments have been conducted to evaluate the optimum parameters and to observe how critical their values are to the resulting image quality. MAFI has been applied a number of times in orthogonal directions and the signal-to-noise ratio (SNR) of the reconstructed image evaluated. The results of these experiments are summarised in Figures 6 and 7.

Figure 6 shows the variation in compression ratio with the value of the DP limit (for a range of coding error thresholds and numbers of MAFI applications). The value of the limit parameter is numerically equal to the gradient of the steepest line of flux plus one. It is clear that there is an optimum limit value of 2 (a maximum gradient of ± 1), but that this value is much more critical when only one MAFI application is used. Multiple applications provide broadly comparable levels of compression over a range of limit values. The optimum limit is still the same, but it is less critical to the compression. This is because although the shorter blocks, which result from allowing unrealistic flux linkages during the first application, leave a higher level of redundancy in the image, that redundancy is in a form which can be removed easily in the orthogonal direction. That is to say that any features with a gradient greater than ± 1 in the original direction will have a gradient less than that in the orthogonal direction.

Figure 7 shows the relationship between compression ratio and SNR for different numbers of MAFI applications, with the limit set to 2. This graph is, however, virtually identical for other values of the limit. The coding error threshold has been varied to achieve the different SNRs and compression ratios.

Finally, several examples of an image analysed and coded with MAFI are presented. The image in question is the MATLAB clown, which contains large amounts of fine detail. Thus this is quite a difficult image, chosen to illustrate the nature of the distortion introduced by MAFI coding, rather than typical values for its performance. Even so, MAFI can remove around two thirds of the data with four applications. The compression achieved after each MAFI application is shown in Table 1. This shows both the compression achieved with MAFI alone, and that when MAFI is followed by JPEG coding. The JPEG algorithm was from the Graphic Workshop 1.1 package, with a quality factor of 75. This was used as an example to show that further compression is possible with conventional coders.

Figure 8(a) shows the original image of the MATLAB clown and Figures 8(b) and (c), the reconstructed versions after applying MAFI two and four times respectively. Figure 8(d) shows the final encoded image, after all redundant rows and columns have been removed. It is clear that the largest part of the compression has occurred on the left and the lower halves of the image. The upper right hand corner, which contains the hair and is much more finely detailed, has not be so heavily compressed. Thus the spatial frequencies in the MAFI-compressed image have become more uniform.

Another item of interest is the directions of the flux linkages in a complex image of this sort. It is difficult to represent so much data graphically, so Figures 9(a) and (b) show a small section of the image (including an eye and a part of the nose), and the lines of flux, calculated horizontally, with the darkness of the lines of flux denoting the darkness of the image at the respective points. The flux generally follows intuitively meaningful contours in the image, adapting to the shapes of the features within the image.

6 Conclusion

MAFI, even for detailed images such as the one presented here, can remove more than 70% of the data without introducing unacceptable levels of distortion. The SNR ratio at this level of compression is little more than 10dB, but the distortion is not of an obviously structured form and it is visually acceptable.

In practice, two applications of MAFI, one in each direction, are sufficient to obtain the very nearly the best compression ratio for any given SNR. They can remove the bulk of the redundant rows and columns, although some slight improvement can be obtained by further applications. However, the main advantage of using multiple orthogonal applications is not to achieve greater compression, but to make selection of optimum coding parameters less critical. This allows a single set of parameters to be applied successfully to a wide range of image types. Furthermore, MAFI can be used before applying other coders to reduce the size of the encoded data still further. Its output (although in a compressed form) still has an image structure and is viewable and intelligible as an image in its own right. When used in conjunction with JPEG, a reduction in encoded data by a factor of 2.7 is obtained (relative to JPEG alone).

7 References

ADIV, G. (1985). Determining three-dimensional motion and structure from optical flow generated by several moving objects, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **7**, July, pp. 384-401.

BAGHAI-RAVARY, L., BEET, S. W., and TOKHI, M. O. (1995a). *Variable frame-rate speech coding by adaptive-flux interpolation.* Research Report No. 592, Department of Automatic Control and Systems Engineering, The University of Sheffield, UK.

BAGHAI-RAVARY, L., BEET, S. W., and TOKHI, M. O. (1995b). Adaptive flux interpolation, flow-based prediction, delta or delta-delta: which is best?, *Proceedings of Eurospeech'95: 4th European Conference on Speech Communication and Technology*, Madrid, 18-21 September 1995. (To appear).

BERGHOLM, F., and CARLSSON, S. (1991). A theory of optical flow, *CVGIP: Graphic models and image processing*, **53**, (2), pp. 171-188.

CICCONI, P., LEONARDI, R. and KUNT, M. (1992). Symmetry-based image coding, *Proceedings of VCIP-92 Conference*, Boston, SPIE-1818(3), November 1992, pp. 1312-1323.

ISO/TC97/SG2/WG8-ADCTG. (1988). ISO adaptive discrete cosine transform coding scheme for still image telecommunication services, *ISO adct-88*, N640, January 1988.

MATSUDA, I., ITOH, S. and UTSUNOMIYA, T. (1994). Adaptive transform image coding based on variable-shape-blocks and directional autocorrelation models, *Signal Processing VII: Theories and Applications*, September, pp. 163-166.

WALLACE G. K. (1991). The JPEG still picture compression standard, *Communications of the ACM*, 34, (4), pp. 31-44.

Table 1: Compression achieved by successive applications of MAFI.

Number of MAFI applications	Compression		Side information (bits/pixel)
	MAFI	MAFI + JPEG	
0	0.0%	67.8%	0.000
1	59.0%	84.9%	0.006
2	68.9%	87.7%	0.010
3	69.8%	88.1%	0.011
4	70.5%	88.2%	0.012

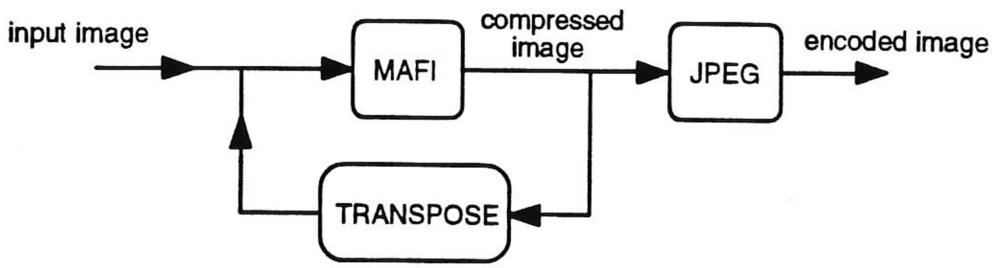
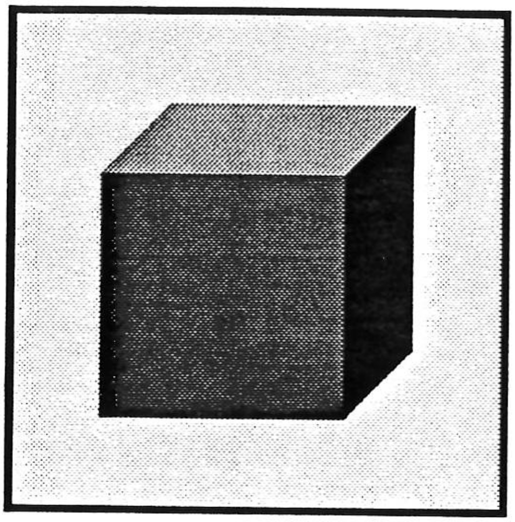
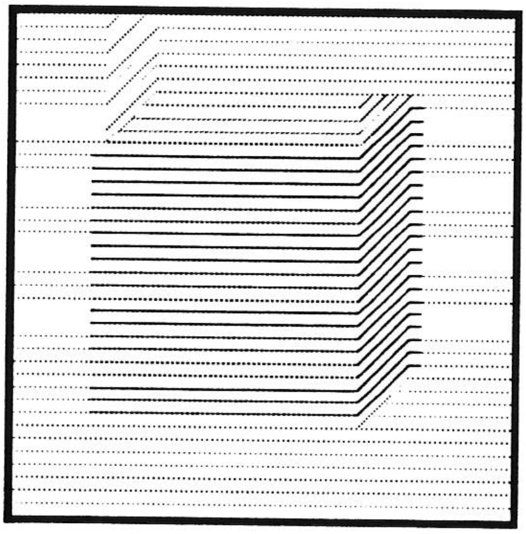


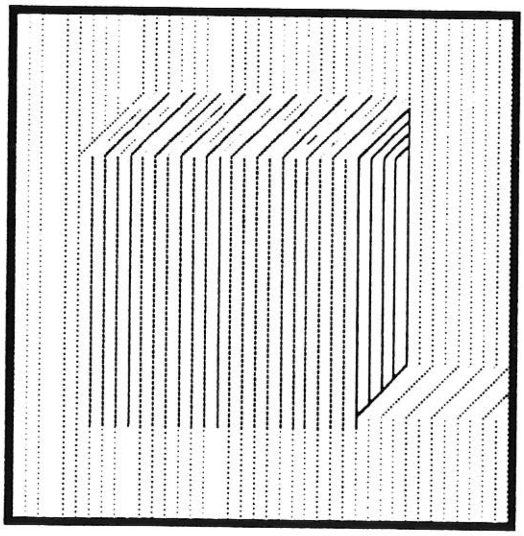
Figure 1: Encoding system.



(a) Image of a box.



(b) Horizontal flux lines.



(c) Vertical flux lines

Figure 2

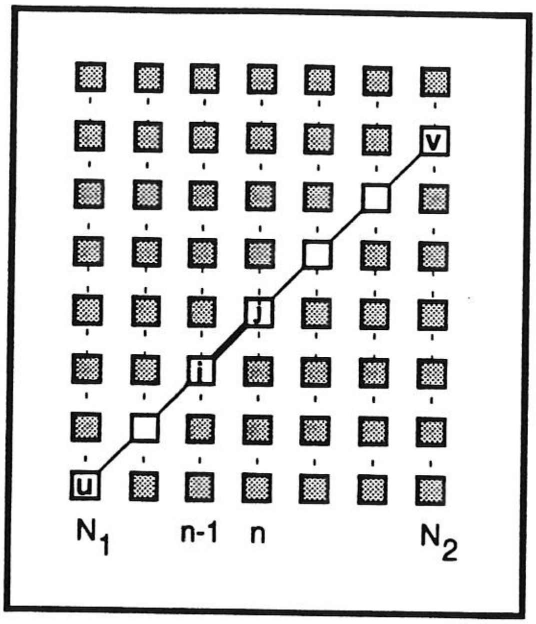


Figure 3: Hypothetical link between two consecutive vectors.

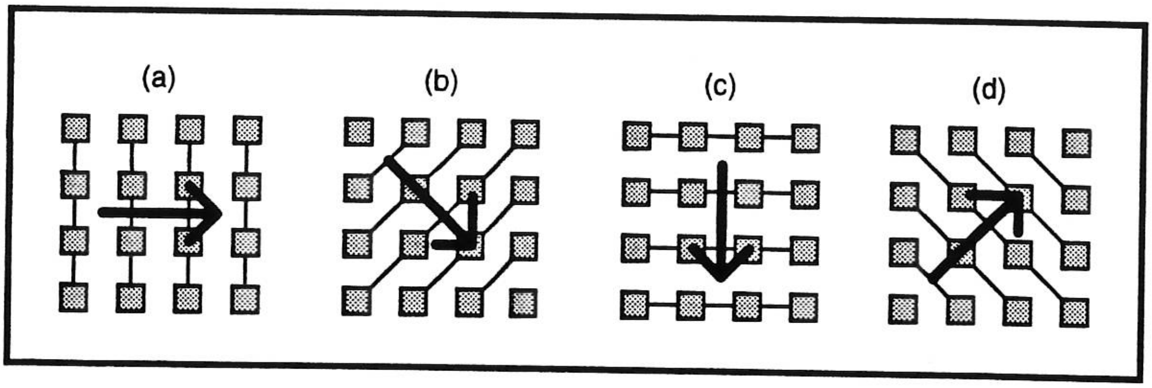


Figure 4: MAFI applied in four non-orthogonal directions.

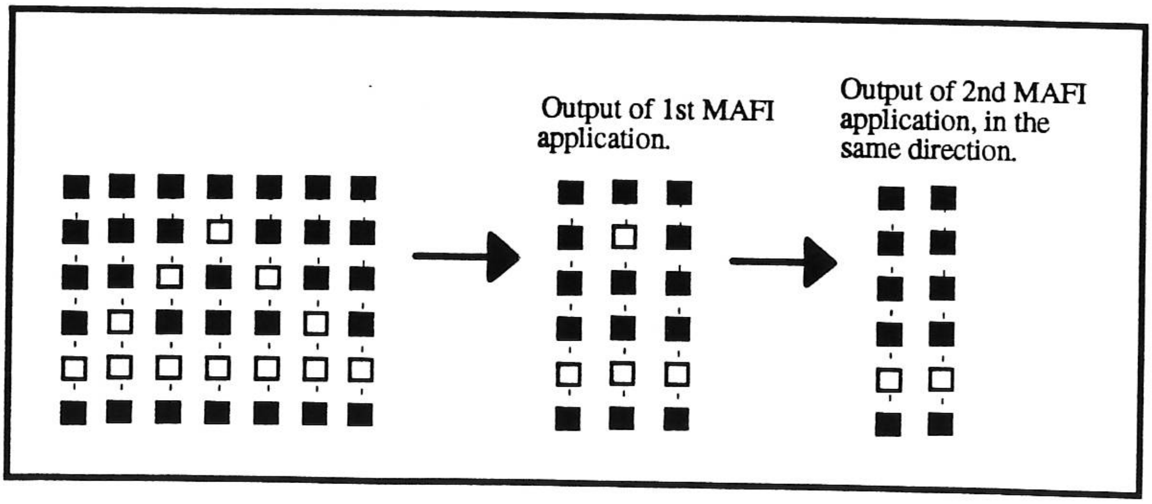


Figure 5: The consequences of applying MAFI in the same direction twice.

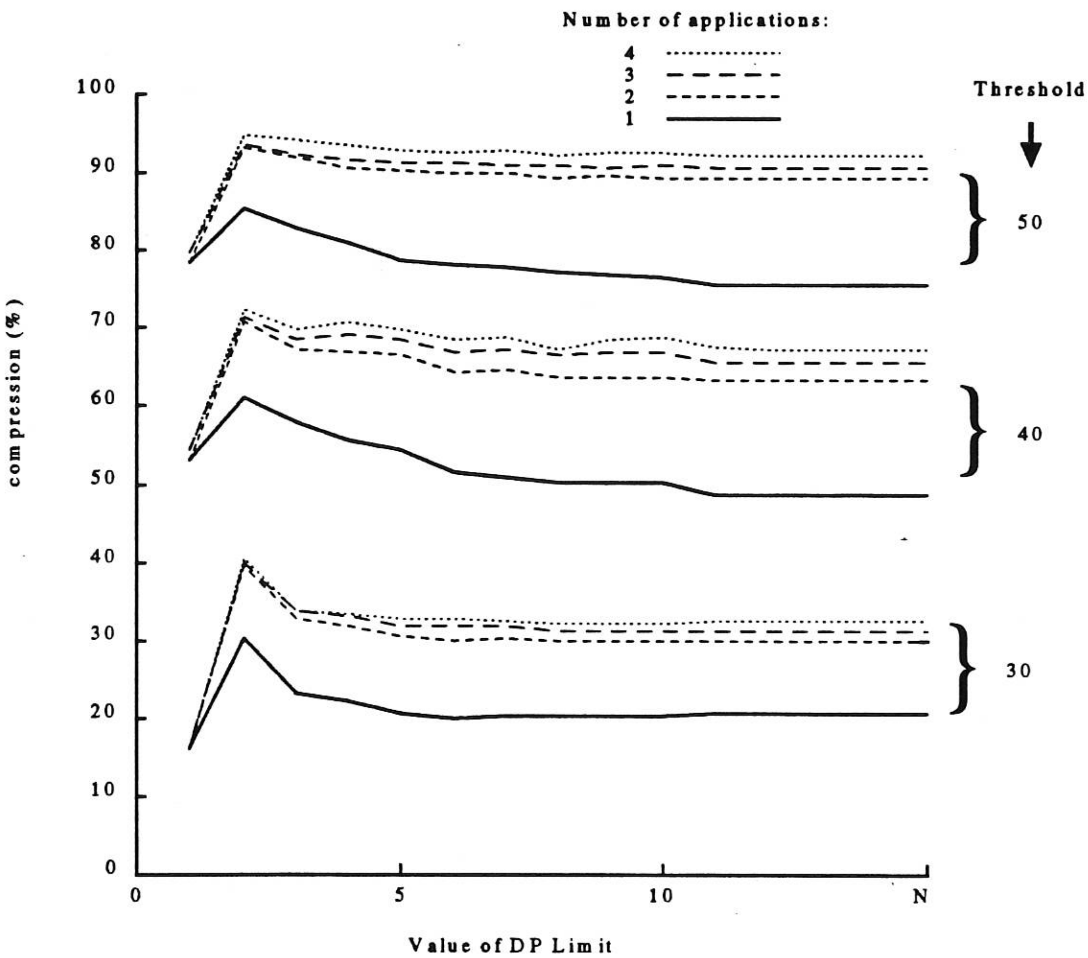


Figure 6: Compression ratio as a function of DP gradient limit.

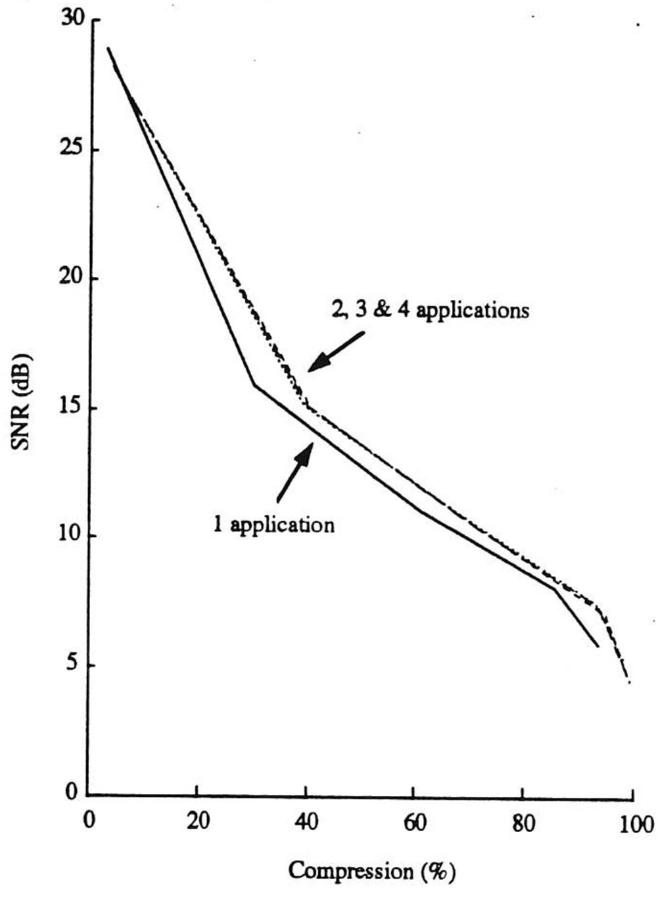


Figure 7: Signal-to-noise ratio as a function of compression ratio.

Figure 8(a): Original image.

Figure 8(b): Reconstructed image after two applications of MAFI.

Figure 8(c): Reconstructed image after four applications of MAFI.

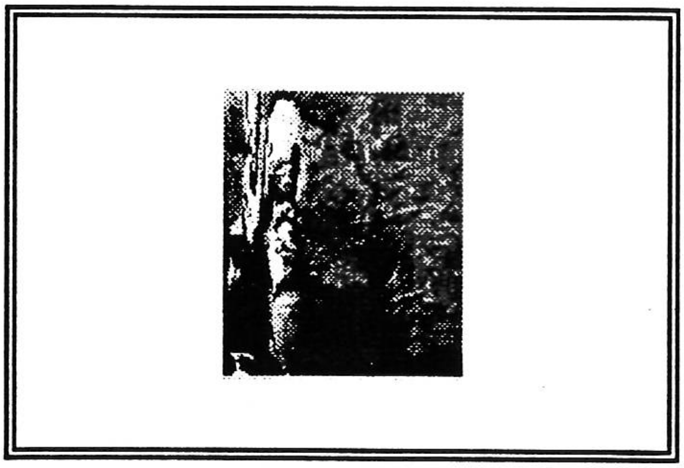


Figure 8(d): Transmitted data after four applications of MAFI.

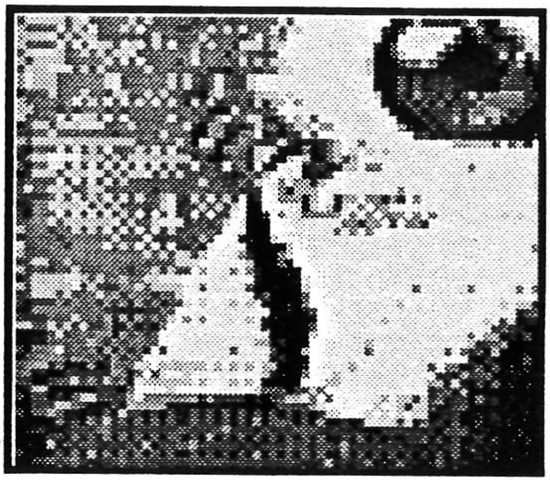


Figure 9(a): A detailed form of Figure 8(a).

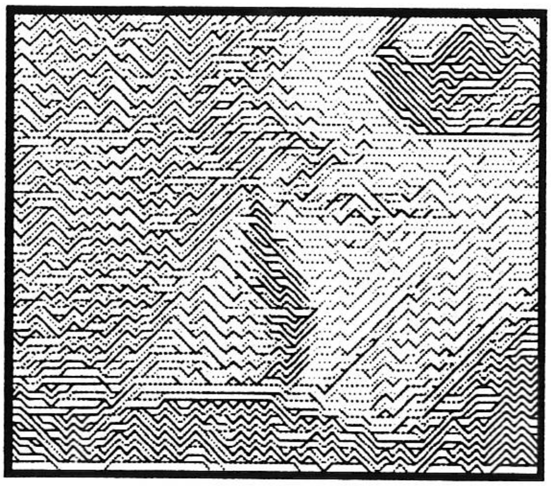


Figure 9(b): Adapted horizontal flux lines.