Progressive Image Coding from a Spanning Tree Image Representation

M.J. Biggar[†] BE (Elec), MEngSc, MIREE
A.G. Constantinides BSc (Eng), PhD, CEng, MIEE, SMIEEE

Department of Electrical Engineering
Imperial College of Science and Technology
London, SW7 2BT, ENGLAND

RESUME

Le codage progressif d'image permet l'amélioration progressive de la qualité de l'image visualisée au fur et à mesure que l'information est reçue. Cette technique s'avère être tout à fait utile dans le cas d'accèss interactif avec une base de données d'images par canaux de transmission à faible capacité. Un tel codage s'effectue par décomposition en formes, definies a priori, par filtrage passe-bas ou par des techniques de transformations. La méthode presentée ici est basée sur une décomposition beaucoup moins contraignante. Grâce à l'obtention d'une représentation hiérarchique, les contours irréguliers de l'image peuvent être transmis par ordre d'importance décroissante. L'image est ainsi progressivement restituée en details, et les contours principaux sont prèservés, sans être distordus géométriquement ni rendus flous comme dans d'autres techniques. Les mots source nècessaires à la description de chaque contour supplémentaire ont été mis en évidence et les techniques de codage qui permettent de les representer de façon adéquate ont été développées.

SUMMARY

Progressive image coding involves the gradual improvement in the quality of a displayed image as more information is received. Such a system is useful for interactive access to an image database over low capacity transmission channels. Approaches to this coding task have included the use of a priori fixed decomposition patterns, low-pass filtering and transform domain techniques. Here, a method is presented which is based on decomposition, but without such stringent constraints. A hierarchical image representation is obtained, which permits irregularly shaped region borders to be transmitted in order of decreasing importance. Image detail is built up gradually, in a way which allows the important, global features to be interpreted early in the reconstruction. The perceptually important edges are maintained, neither being geometrically distorted nor blurred as in other techniques. The source words necessary to describe each additional border, and coding techniques appropriate to represent them, are identified. Resulting images are presented, and the advantages and disadvantages of the method are discussed.

1. Introduction

Low capacity transmission channels present considerable challenges to image coding schemes. In recent years, progressive image transmission has been proposed as a means of providing a user with a worthwhile, interpretable image as soon as possible in the specific situation where he/she is interactively interrogating an image database over a low capacity transmission channel, such as the Public Switched Telephone Network (PSTN). This allows a decision to be made as to whether to continue with a more detailed reconstruction, or to terminate it (and, perhaps, go on to the next of a series of images to be reviewed). A particular application is "photographic videotex", in which high-quality images may be embedded in normal videotex pages. Other large digital image databases, such as those emerging in the medical environment, could benefit from progressive coding [1].

Common image coding schemes, such as DPCM, are not well suited to these applications, because the image is built up slowly at full resolution; the subjects of interest may not be revealed until well into the reconstruction [2]. With progressive transmission, the image is represented in a pyramidal, hierarchical fashion, and information is transmitted from broader levels of the pyramid as time progresses.

Previous schemes have generally followed one of three different approaches. The earliest work [2,3] involved regular decomposition of the image. The entire image is initially represented by a small number of very large regular blocks. These are successively decomposed in a regular way to reveal an increasing amount of detail in the image. A second approach is to low-pass filter the image and then subsample it in a series of stages. This is the basis of the "Laplacian pyramid" of [4]. The third approach utilises the hierarchical nature of the coefficients of a transformed image. For instance, the low frequency, or sequency, coefficients contain information about the large, important features in the image. This method has been discussed in [1,5], among others.

The most important features of an image are the large objects and sharp edges which define their limits [6]. However, all of the above coding techniques cause severe distortion of these features in the early stages of reconstruction. Either an artificial structure is imposed, which forces the boundaries onto the edges of large rectangular blocks, or blurring occurs due to low-pass filtering (either directly or through the limitation on the number of transform coefficients). We propose that it may be possible to convey the most important information earlier in the reconstruction by accurately representing a small number of these features. The progressive coding method to be described is a spatial domain approach in which regions of the image are successively split into two in a hierarchical fashion. However, control of the boundaries and the order in which the regions are divided is determined by the properties of the image rather than an externally-imposed regular mechanism. Boundaries are irregularly shaped, but represent the edges of image features much more accurately. The increased number of bits necessary to define an irregular boundary are traded for the gain in perceptual quality resulting from accurate representation of features.

2. Hierarchical Segmentation Procedure

The hierarchical representation of the image is obtained from the region merging procedure described in [7] for the application of image segmentation. Pixels in the image are progressively merged with their neighbours (here, attention is restricted to the 4-connected case) according to a cost function which defines how "similar" the pixels, or groups of pixels, are. Merged pixels are assigned the mean intensity of all pixels in the group, and the cost function between it and its neighbouring groups is re-evaluated after each merger. The links between pixels, and their order of selection, are stored and the resulting list defines a spanning tree of a graph representing the original image. Removal of the last N links from the list defines a spanning forest of N+1 regions. The method therefore permits the definition of as many of the most significant regions in the image as desired, once the spanning tree has been generated.

[†] M.J. Biggar is working under a Study Award from Telecom Australia

In [7], the absolute difference in mean intensity was taken as the measure of linking cost. However, a cost function which seeks to minimise the Sum Squared Error (SSE) over the image gives better results from the subjective and coding rate points of view. The adoption of this cost function means that the segmentation method is similar to the one employed in [8], but treated in a graph theoretic context. While this cost function seems to give good results, it is not necessarily optimal in terms of subjective image assessment. Other cost functions are under investigation.

The image representation resulting from the above method is well suited to a progressive coding scheme. The structure is hierarchical in that the removal of an extra link will split an existing region into two others. A boundary is added within a region, but no existing boundaries are affected.

3. Source Coding

Division of an existing image region into two new regions involves a description of the border and the new intensities, and assignment of codewords to the resulting source words.

3.1 Boundary Description

Additional region boundaries are defined by tracing the line dividing the pixels in either region rather than the pixels on one side or the other of it. This means that there are 3 instead of 4 possible relative directions in which to turn from one line element to the next. The source code set is therefore smaller. Also, the chosen method of representing the sequence of line elements is simplified by not being able to reverse direction.

Before tracing the border, a starting address must be specified. Because the pixel edges are being defined, the operation must be carried out on the array of pixel corners rather than pixel centres. For an image of dimensions X pixels by Y pixels, this array is of size (X+1).(Y+1) points. The 4 image corner points cannot be starting addresses, so there are XY+X+Y-3 valid starting addresses.

An example boundary line to be specified is shown in figure 1. For relatively large regions describing familiar objects, the boundaries are not likely to change direction randomly. In fact, there is considerable redundancy to be exploited in lines describing straight or gently curved objects. A study of thin-line coding techniques appropriate to this application has revealed that a good compromise coding scheme is one in which the line is segmented into portions consisting of elements in one of two adjacent directions from the set of four. The line elements are then run length coded. The method, illustrated in figure 1, is a modification of that described by Kaneko and Okudaira [9]. Each line segment is represented by the number of run lengths in the segment, followed by the sequence of run lengths. This is repeated for each segment as necessary to completely define the line. There is no ambiguity in the change of direction from segment to segment. There are three possible relative directions to move in tracing along a pixel edge from any point. If the current set of runs is complete, then the line segment to follow must not belong to a known two of these. Therefore, it must start off in the third. By rotating the valid pair of coordinates by one, the need to explicitly state the new direction is avoided. An explicit direction need only be stated once at the start of each line (after the starting address has been specified), and this will require 2 bits (since there are 4 possible starting directions). In our implementation we assume a default for the other of the first segment's pair of valid directions. Additional look-ahead processing could be used to find and transmit the correct one at an additional cost of 1 bit per boundary.

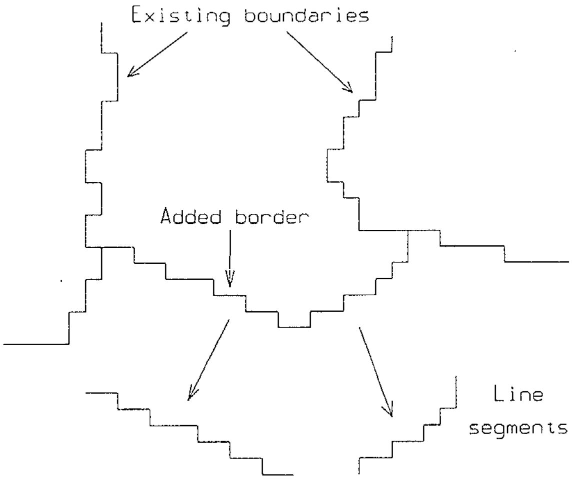

Figure 1. Addition of a new border, showing its representation as two line segments, one having 11 runs and one having 9.

The line terminates when it encounters another boundary, the image border or its own starting point. There is therefore no requirement for additional codewords to specify these conditions.

The boundary line is defined by codewords from two source sets, representing the number of run lengths and the run lengths themselves. Though most of the numbers and runs are likely to be quite small, the coding scheme must be capable of dealing with any possible situation. Run lengths may be anything from 1 to the maximum dimension of the image D_{max}, so the run lengths R will be from the source set $\mathbf{R}$ defined by

$$R \in \mathbf{R} = \{1, 2, \ldots, D_{max}\}, \quad D_{max} \triangleq \max(X, Y). \quad (1)$$

The greatest possible number of consecutive runs in a line segment would be the result of a diagonal region border, spanning the smallest image dimension. It can be shown readily that such a border could have up to $2.D_{min}-1$ run lengths in it, so the run count source word C is from the set $\mathbf{C}$ defined by

$$C \in \mathbf{C} = \{1, 2, \ldots, 2.D_{min}-1\}, \quad D_{min} \triangleq \min(X, Y) \quad (2)$$

3.2 Intensity Description

Assume that a particular region in the reconstructed image, of intensity I_0 and consisting of N_0 pixels, is divided into two new regions, with intensities I_1 and I_2 and pixel counts N_1 and N_2. There are two properties in particular that may be exploited for coding purposes.

1. High correlation may be expected between the divided and original region intensities. That is, only a small intensity change is expected upon region division.

   $\left|I_0 - I_1\right|$ and $\left|I_0 - I_2\right|$ should both be small.

2. To within integer round-off, I_0 is a weighted mean of I_1 and I_2;

$$N_0.I_0 = N_1.I_1+N_2.I_2 \quad \text{and} \quad N_0 = N_1+N_2. \quad (3)$$

The first property suggests that a differential coding technique would be an effective approach, so the quantities to be transmitted are

$$I_{d1} \triangleq I_0-I_1 \quad \text{and} \quad I_{d2} \triangleq I_0-I_2 \quad (4)$$

Assume that I_{d1} is the first transmitted, which defines I_1, then property 2 suggests a method by which the second intensity might be estimated. It is clear from equations 3 and 4 that

$$I_{d1}>0 \Rightarrow I_{d2}<0, \qquad I_{d1}<0 \Rightarrow I_{d2}>0$$

and $I_{d1}=0 \Rightarrow I_{d2}=0$

In practice, integer rounding means that these conditions must be relaxed slightly;

$I_{d1}>0 \Rightarrow I_{d2}\leq 0$, $\quad I_{d1}<0 \Rightarrow I_{d2}\geq 0$

and $I_{d1}=0$ does not imply the sign of I_{d2}.

As long as the first differential intensity received by the decoder is not zero, its sign alone is sufficient to deduce the sign of the second differential intensity. About one bit can therefore be saved in this way for each new boundary defined. This technique was used to code the intensities and obtain the results to be presented in Section 4.

There are, therefore, two additional source word sets needed to code the intensities. For 8-bit monochrome pictures, the range of differential intensities could be from -255 to +255. The first differential intensity code word for each region division must be from this double-sided distribution, and so the codeword I_{d1} is from the set

$$I_{d1} \in \mathrm{ID} = \{-255, -254, \ldots, +254, +255\} \quad (5)$$

The source set for the second received intensity code word I_{d2} then depends upon the value of I_{d1}.

$$I_{d2} \in \mathrm{ID} \quad \text{if} \quad I_{d1} = 0,$$

$$I_{d2} \in \mathrm{IS} = \{0, 1, \ldots, 254, 255\} \quad \text{otherwise.} \quad (6)$$

Note that more sophisticated prediction methods are quite possible. In particular, equation 3 permits both sign and magnitude of the second intensity to be estimated. In order to do this, however, counts of the numbers of pixels in each region must be made and stored as each is created. This would mean a considerable increase in complexity and memory requirements of the decoder. One of the aims of any progressive coding scheme will be to minimise decoder complexity. Since the contribution to overall rate from the intensity components is relatively small anyway (see Section 4), it was decided not to incorporate this feature. A practical realisation may not even incorporate the above-described simple sign prediction procedure.

3.3 Codeword Assignment

As the added boundaries may commence anywhere in the image, and in general there is little reason to expect correlation from one address to the next, there is little potential for redundancy reduction of this source by codeword assignment. A simple fixed length word was assumed in simulations, and the number of address bits required may be calculated simply by $\lceil \log_2(XY+X+Y-3) \rceil$ bits, where $\lceil x \rceil$ means the smallest integer greater than x.

The other four source word sets R, C, ID and IS all exhibit sharply peaked distributions, and therefore offer potential savings by variable length coding. While the minimum bit rate is achievable by the use of Huffman coding, the large cardinalities and very peaked probabilities of these source sets mean that very long words would have to be accommodated. This would lead to a very complex word decoder, though most of the words would almost never be encountered. The solution adopted is the use of a truncated Huffman code, where the most probable words are allocated variable length codes, and one extra code word acts as an indicator that a word outside of these is to be sent. The word is then sent (and expected by the decoder) in simple fixed length binary form.

To illustrate, assume that V variable length words are to be assigned to represent V-1 most probable source words in a source set S plus one word used as the "marker", indicating that a fixed length word is to follow. An incoming variable length word w_i will be interpreted as follows;

$i < V$: Find source word from lookup table.
$i = V$: Expect B bit word to follow.

The B bit word represents all remaining less probable source words. B may easily be found from the number of words remaining; $B = \lceil \log_2(\, C(S)-(V-1)\,) \rceil$, where $C(S)$ is the cardinality of set S.

Codeword assignments based on this scheme have been performed for each of the above source word sets, using the accumulated statistics from several images. Comparison of the average rate per word with the entropy for each source set showed that there is very little penalty to pay for use of this method with quite a small number of variable length words. The number of variable length code words chosen was 16 for R, 16 for C, 32 for IS and 64 for ID. With this arrangement, no code word was longer than 11 bits.

4. Results

The coding scheme described in the previous Section was used to simulate the results attainable in a real implementation. The time taken to transmit a certain number of borders was calculated by dividing the total number of bits by the transmission rate. This does not take into account additional line coding which would be necessary to give the system a degree of error tolerance, since a single bit error could corrupt all future reconstructions in the described system. A "border end" code word, with the facility for requested retransmission, may be sufficient error protection. The simulation also takes no account of processing delay, but this is expected to be small compared with the transmission time. The results, therefore, reflect the time taken to transmit the raw data at the given data rate.

A test image of 256x256 pixels and 8 bit/pixel resolution is shown in figure 2a. Figures 2b and 2c show the coded image after 5 and 30 seconds transmission time at the modest transmission rate of 1200 bit/s. This rate is easily attainable by use of modems over the PSTN, and is commonly used for videotex services. It can be seen that after only 5 seconds, there is sufficient detail in the image for identification of the major features. After 30 seconds, all of the principal features of the original have emerged. These results may be compared with a PCM-coded alternative, in which transmission of the entire image would require 450 seconds at the same rate.

It should be noted that the system becomes less efficient as the number of regions increases due to the addressing overheads. The proportion of the total rate which is due to each component is plotted in figure 3 for the test image. It can be seen that the fixed overhead becomes a significant proportion of the total as the transmission progresses. So, although the method reveals the principal features in an image very rapidly, and a useful representation is obtained after only a few seconds, it is not efficient if an exact reproduction of the image is required. In this case, a dual mode of operation would be a solution. If transmission has not been terminated after a certain number of regions (of the order of several hundred) have been displayed, the system could switch to a differential coding mode. The error between the coded and original image could then be transmitted in a simple serial scan fashion.

5. Practical Considerations

Practical implementation of the system described above requires the generation of the hierarchical structure of Section 2. It is unlikely that it would be practical to do this each time an image is displayed because of the computation required; an additional, unacceptable delay would almost certainly be involved. The system is much better suited to an archival approach, where the linked spanning tree is generated and stored along with the image when it is added to the archive. Each link requires a pixel index plus an extra bit to indicate whether it is a


(a) (b) (c)

Figure 2. Original 256x256 pixel image (a), with reconstructions after 5 (b) and 30 (c) seconds at 1200 bit/s using the proposed progressive coding scheme.

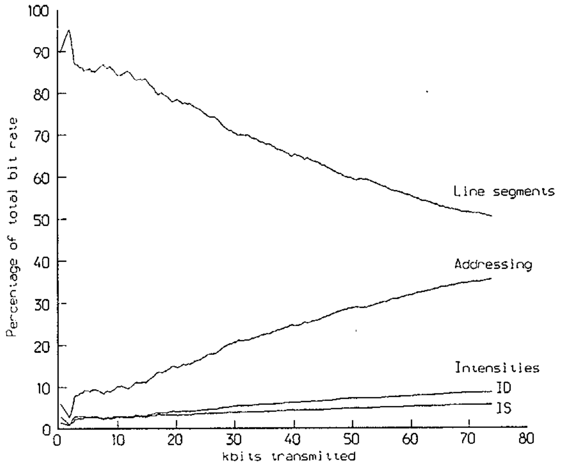

Figure 3. Proportion of total rate due to coding of line segments (sources **C** and **R**), addressing overhead and intensities. Full scale on the horizontal axis corresponds to about 60 seconds at 1200 bit/s.

horizontal or vertical link. It can readily be shown that $(XY-1).(\lceil \log_2 (XY) \rceil + 1)$ bits will be required for storage of the whole structure. For a 256x256 pixel image, this will be about twice as much again as needed for the original image. Such storage requirements may not, however, be a limitation in the light of recent advances in both magnetic and optical mass storage devices.

An alternative, much more efficient, method of storage is possible if only a limited number of regions (say, 1000, for instance) are ever to be displayed, as in the dual-mode system proposed in Section 4. In this case, only the actual transmitted information for these regions need be stored, and this would be a considerable saving not only of storage, but also of the processing burden upon transmission. For a 256x256 pixel image, the data requirement would typically be only about 10 kbytes for 1000 regions.

The main requirement of a receiver is a flexibly addressable frame store, with an additional 4 bit/pixel to define connectivity with other pixels in a region. Other than this, most operations involve straightforward counting and incrementing. Source word lookup tables are also needed. No floating point operations are required. The main complexity, involving the hierarchical image representation, storage and boundary tracing, has been confined to the central databank. By the method proposed above, storage could be minimised and heavy processing could be performed during periods of little use, so a conventional multiuser computer would be a suitable host.

6. Conclusion

The challenge of transmitting pictorial information over low capacity channels has been discussed in this paper. Progressive coding is a method which minimises the effect of this restriction, by presenting limited, but important, information to the user as quickly as possible. Previous approaches to this method were briefly described before a new method, involving irregular spatial decomposition of the image, was presented. This method presents the user with the perceptually most important features of the image without distorting outlines. A coding method has been described, resulting images presented and practical considerations for implementation discussed.

7. Acknowledgement

M.J. Biggar wishes to thank the Director, Research, Telecom Australia, for permission to publish his contribution to this research.

REFERENCES

[1] S.E. Elnahas, K-H. Tzou, J.R. Cox, R.L. Hill, and R.G. Jost, "Progressive Coding and Transmission of Digital Diagnostic Pictures", *IEEE Trans. Med. Imag.*, Vol. MI-5, No. 2, pp. 73-83 (Jun. 1986).

[2] K. Knowlton, "Progressive Transmission of Grey-Scale and Binary Pictures by Simple, Efficient, and Lossless Encoding Schemes.", *Proc. IEEE*, Vol. 68, No. 7, pp. 885-896 (Jul. 1980).

[3] S.L. Tanimoto, "Image Transmission with Gross Information First", *Comp. Graphics and Image. Process.*, Vol. 9, pp. 72-76 (Jan. 1979).

[4] P.J. Burt and E.H. Adelson, "The Laplacian Pyramid as a Compact Image Code", *IEEE Trans. Comm.*, Vol. COM-31, No. 4, pp. 532-540 (Apr. 1983).

[5] H. Lohscheller, "A Subjectively Adapted Image Communication System", *IEEE Trans. Commun.*, Vol. COM-32, No. 12, pp. 1316-1322 (Dec. 1984).

[6] D. Marr, *Vision*, W.H. Freeman (1982).

[7] O.J. Morris, M.de J. Lee, and A.G. Constantinides, "Graph Theory For Image Analysis: An Approach Based On The Shortest Spanning Tree", *Proc. IEE, Part F*, Vol. 133, No. 2, pp. 146-152 (Apr. 1986).

[8] J.-M. Beaulieu and M. Goldberg, "Step-Wise Optimization For Hierarchical Picture Segmentation", *Proc. CVPR '83: IEEE Comput. Soc. Conf. On Comput. Vision & Patt. Recog.*, Washington DC, pp. 59-64 (1983).

[9] T. Kaneko and M. Okudaira, "Encoding of Arbitrary Curves Based on the Chain Code Representation", *IEEE Trans. Comm.*, Vol. COM-33, No. 7, pp. 697-706 (Jul. 1985).