November 20, 1986

Second Generation Hybrid Image-Coding Techniques

M. Reha Civanlar, *North Carolina State University*
Sarah A. Rajala, *North Carolina State University*
Wonrae M. Lee, *North Carolina State University*

Second generation hybrid image-coding techniques

M. Reha Civanlar   Sarah A. Rajala   Wonrae M. Lee

Center for Communication and Signal Processing
North Carolina State University
Box 7914, Raleigh, N.C. 27695-7914

Abstract

Recently, ways to obtain a new generation of image-coding techniques have been proposed. The incorporation of the human visual system (HVS) models and tools of the image analysis, such as segmentation, are two defining features of these techniques. In this paper, an application of the new approach to the classical linear predictive coding (LPC) of images and an HVS based segmentation technique for the second generation coders will be discussed. In the case of LPC, the error image is encoded using an image decomposition approach and binary image coding. This improves the compression ratio keeping the quality nearly the same. The new segmentation technique can be used in single frame image coding applications to obtain acceptable images at extremely high compression rates.

Introduction

Until recently, image analysis and image coding have been considered to be two unrelated fields of image processing.[1, page 469] The aim of image analysis is to extract certain information from a picture, a process of eliminating the unrelated information while enhancing the related information in a picture. That is, although their ultimate products may be different, information lossy image coding and image analysis share very similar goals. In certain applications, such as the LANDSAT, where the relevant information in a picture differs from user to user, it is not allowed to alter the picture for coding purposes. However, in other applications such as video teleconferencing or archival of human pictures, where images with certain characteristics are to be viewed by human observers under more or less well defined conditions, the pictures can be simplified by the removal of irrelevant information. In such applications, the characteristics of the human visual system (HVS) can be used in determining the relevancy of information and an easily codable image containing only the relevant information may be obtained using image analysis tools. The coding methods based on this new approach are called second generation image-coding techniques.[2]

In this paper, two applications of the new technique will be presented. In the first application, a classical image coding technique, linear predictive coding (LPC) is modified using a second generation approach to improve its performance. This is made possible by using a prominent property of the LPC error image which constitutes the major part of the information to be transmitted. This application results in a hybrid technique in that, although it uses tools from second generation techniques, it is based upon the classical LPC and it uses mathematical distance measures for fidelity.

The second application is called HVS based segmentation for image coding and it is a genuine example of the second generation approach. In this technique, the squared coding error may be high, however, it is possible to obtain acceptable images at extremely high compression ratios. The primary reason for this is the use of the HVS model in segmenting the original image. The segmented image can be transmitted using a bandwidth much smaller than that required for the original image. Determination of a fidelity measure for this application is a continuing research area.

LPC with second generation techniques

The importance of the LPC error

The success of the LPC in speech coding has been the driving force for the efforts to develop a two-dimensional (2D) version of the algorithm. In speech coding, the source, i.e. the vocal tract, can be approximated by a causal linear system. In images, it is extremely difficult to define a source model and the causality does not apply. However, in many parts of a picture there is correlation between the neighboring pixels. The classical LPC of images is based on estimating the intensity of a pixel using the intensities of the "previous" pixels which are inside the estimation window. The coefficients used in the linear estimation process and the estimation error constitute the information to be encoded.

In contrary to LPC of speech, in LPC of images, the transmission of the LPC error is much more important than transmission of the coefficients. The effective size of the estimation window, i.e. the number of the neighboring pixels used in estimating the intensity of a pixel, is picture dependent. However, using windows larger than 3 X 3, generally, does not result in a significant improvement.[3] Especially in coarsely sampled images, even the improvement obtained using a 3 X 3 window over a 2 X 2 (previous three pixels) does not justify the increased computational complexity. For the 2 X 2 window, only three coefficients per block have to be transmitted. On the other hand, using a common block size of 16 X 16, the transmission of the LPC error involves encoding of 256 pixels per block.

The reason for using blocks in LPC analysis is the spatially varying correlation properties of images. While this variation is an important factor in the LPC of textured areas, the effect of it is not significant in other areas. Based on this argument, it has been suggested that a standard set of coefficients for the entire image be used.[4] In such an approach, the LPC error picture becomes the only information to be transmitted.

Properties of the LPC error

In fig. 1 and 2 the original image and the LPC error image obtained by a coder with a fixed coefficient set and 2 X 2 estimation window is demonstrated. The LPC error picture is quantized using an eight level quantizer. As expected, the effective dynamic range of the error image is less than the original. The classical approach to the problem is to make use of this reduction in the dynamic range by using appropriate quantizers for the LPC error image. Compression rates as low as 0.7 bits per pixel (bpp) have been obtained using adaptive quantizers.[3]

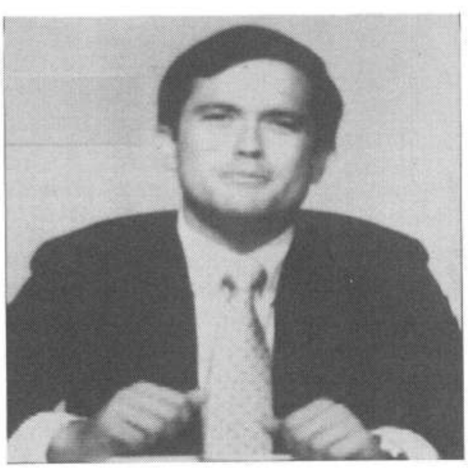

Figure 1 - Original Image

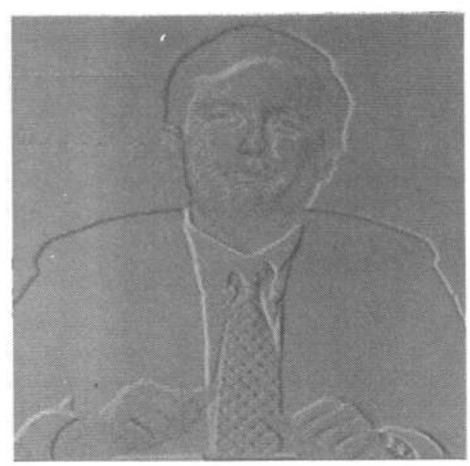

Figure 2 - LPC Error

A very important property of LPC error images which cannot be utilized in straightforward quantization is the localization of the errors around the edges of original images. This is because it is impossible to predict an edge using the pixels located only on one side of it. In fact, this property has been used in edge detectors.[5,6,7] In [7], it is reported that the LPC error is particularly efficient in detecting those edges perpendicular to the direction of recursion and the union of linear predictors starting from four corners of an image is used to capture all of the edges. Clearly, in coding applications detection of all edges is not required. Adaptive quantization can be considered as an indirect way of making use of the localization of the LPC error. However, this property is so prominent that a direct method of exploiting it can easily be developed.

A coder using the localization of LPC errors

Decomposition of the original image One way of improving the performance of LPC is to decompose the original image into two images so that one of these contains the abrupt intensity changes and the other one is easily codable with a LPC. The decomposition requires segmenting the original image and transmitting the segmented image while using the LPC for encoding the difference between the original and the segmented images. If the segmentation algorithm is successful in removing all of the significant edges, the difference picture will not contain any abrupt intensity transition and the LPC error on the difference picture may be negligible. The main problem with this approach is the edges which remain in the difference picture because of segmentation errors. It is possible to improve the performance of the segmentation algorithm to eliminate such errors, however, the quality of the refined segmented picture is so high that transmission of the difference picture becomes unnecessary. This approach will be discussed in the HVS based segmentation section.

An encoder for the LPC error In the LPC error image of fig. 2, only 8% of the pixels have non-zero intensities. These pixels are demonstrated in fig. 3 which is called an LPC error mask.

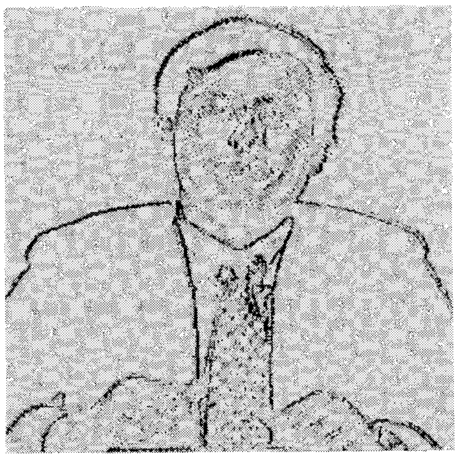

Figure 3 - LPC error mask

Although, the majority of the significant errors are located around edges, the picture in fig. 2 is not suitable for contour coding. This is because of the lack of continuation and varying intensity levels of the pixels located on the possible contours. The modification of the intensities to obtain continuous and homogeneous contours causes strip artifacts in the direction of recursion on the reconstructed image. On the other hand, if fig. 3 is transmitted as a binary image and the pixel values inside the nonzero regions are transmitted separately, modification of the contours becomes possible. In [8] a similar approach has been used to encode the interframe difference pictures of video teleconferencing sequences.

There are numerous techniques for encoding binary images.[9] These are developed for facsimile applications, however, most of them are applicable to encoding of the LPC error mask. In this work, the modified READ code [10] and the arithmetic code [11] have been tested. The modified READ code uses the similarity between the successive lines of a binary image for data compression. Arithmetic coding makes use of local statistics and it is very effective when the majority of the pixels have the same values. Mask pictures have both of these properties. However, the arithmetic coder performs better because of the existence of small isolated regions which cannot be encoded efficiently using the READ code. This is consistent with the observation about the superiority of the arithmetic code in facsimile applications reported in [11] by Langdon and Rissanen.

Once the mask is available, the nonzero pixel values can be transmitted in a specified order, such as a raster scan. Further compression can be obtained by using a level plane coding procedure for the single dimensional vector of nonzero pixel values. It has been observed that, as low as eight quantization levels for the nonzero pixels are sufficient for an acceptable reconstruction of the test image used in this paper.

Results

In fig. 4 the result obtained by using arithmetic code to transmit the LPC error mask and a level-plane run-length code for the nonzero pixel values is demonstrated. In this application the coding rate is 0.41 bpp. In order to code the same image using a straightforward quantization a coding rate of 3 bpp is required.

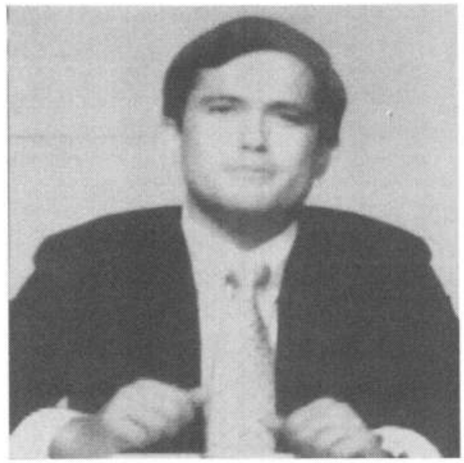

Figure 4 - Coded picture using the new technique

HVS based segmentation

Segmentation of images for coding purposes has been studied by many researchers.[2] As stated in the previous section, it is possible to obtain very high compression rates using this approach. The coding methods using segmentation vary in the criteria they use to determine the homogeneity of the segments and the method they use to encode the boundaries and the insides of the segments. The homogeneity definition is particularly important because it specifies the number of segments which determines the compression rate and the quality of the coded image. In the trivial case, where each pixel is considered to be a segment, there is no compression and the coded image is the same with the original.

In image analysis, the main use of the segmentation is in automatic object recognition. This necessitates tailoring the segmentation algorithms to obtain a one to one correspondence between the segments and various parts of the objects in an image. In coding applications, there is no need for such a correspondence and the visibilitv of the intensitv differences should be the main segmentation criterion. If the end result is intended for a human observer, it is natural to use the characteristics of the HVS to determine the visibility. In this work, the incorporation of the HVS model in the segmentation algorithm is accomplished in two steps: preprocessing the image and determining the thresholds of the segmentation algorithm.

Preprocessing

Two fundamental properties of the HVS used in this application are Weber's law and the modulation transfer function (MTF) of the eye.[12] Weber's law states that the ratio of the intensity difference between a visible region and its background to the intensity of the background is a constant for medium intensity levels (Weber's region). For low and high intensity levels, Weber's law does not hold, and the sensitivity of the human eye to intensity differences reduces a great deal. This property is used in reducing the dynamic range of the original image without loosing visible details for standard viewing conditions. The determination of the Weber's range is accomplished by performing subjective experiments on a set of test images. The upper and lower intensity bounds depend on the equipment used.

The MTF of the eye indicates that the eye has a bandpass response to the spatial frequencies. That is, the visibility of very slow and very fast intensity transitions is low. During the preprocessing stage, this property together with the Weber's law is used in removing small regions which may increase the number of segments. The decision threshold in removing of a region is based on its size and the intensity difference between the region and its neighbors.

Segmentation

A variation of the centroid linkage region growing segmentation algorithm [13] is used to segment the image. In this approach, the image is scanned using a conventional raster scan. The value of each pixel is compared to the value of an already constructed neighbor segment. If the intensity difference is less than the HVS based visibility threshold, which is a function of the intensity of the neighboring segment, the pixel is joined to the existing segment. If the intensity differences between the pixel and its neighbor segments are larger than the corresponding thresholds, a new segment is started.

Experimentally determined thresholds based on HVS are published in various sources.[12] However, the display device and the viewing conditions make considerable difference on the threshold values. Experiments with the display system are invaluable in determining this extremely important function.

Results

In fig. 5, the HVS based segmentation result of the original image of fig. 1 is demonstrated. This image has 205 segments, and using five bits to encode the intensity values inside the segments and contour coding the boundaries, 0.15 bpp is a sufficient rate to transmit this image. A problem with the segmentation approach is the visibility of the segment boundaries in non-edge regions. Postprocessing can be performed to reduce this artifact by rolling off the segment boundaries in such regions. The mathematical difference between the coded image and the original is not small, however, most of the important features of the original are preserved in the coding process.

Figure 5 - HVS based segmentation result

Conclusion

The results obtained using the image analysis approach to image coding are encouraging. However, in order to have a generally applicable codec based on such an approach, there are many questions to be answered. Identification and encoding of the textured areas, and extension of the techniques to make use of the previous frames in image sequence transmission are two examples for these questions.

As in all other information lossy image coding techniques, the image analysis based coders suffer from the lack of a meaningful fidelity criterion. A possible way of comparing the segmented images may be to use an image description language, however, the details of such an approach are still under investigation.

References

[1] W. K. Pratt, *Digital Image Processing,* John Wiley & Sons, New York 1978.

[2] M. Kunt, A. Ikonomopoulos, M. Kocher, "Second-Generation Image-Coding Techniques," *IEEE Proceedings,* vol. 73, no. 4, pp. 549-574, April 1985.

[3] P. A. Maragos, R. W. Schafer, R. M. Mersereau, "Two-Dimensional Linear Prediction and Its Application to Adaptive Predictive Coding of Images," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. ASSP-32, no. 6, pp. 1213-1229, December 1984.

[4] D. Anastassiou, W. B. Pennebaker, J. L. Mitchell, "Gray-Scale Image Coding for Freeze-Frame Videoconferencing," *IEEE Transactions on Communications,* vol. COM-34, no. 4, pp. 382-394, April 1986.

[5] T.F. Quatieri, "Object Detection by Two-Dimensional Linear Prediction," *Proc. IEEE Intl. Conf. Acoustics, Speech, and Signal Processing,* Boston, April 1983.

[6] M.D. Richard, C.W. Therrien, J.S. Lim, "Image Segmentation Using Spatial Linear Prediction," *Proc. IEEE Intl. Conf. Acoustics, Speech, and Signal Processing,* pp. 32.3.1-32.3.4, San Diego, March 1984.

[7] Y. Zhou, R. Challappa, V. Venkateswar, "Edge Detection Using Zero Crossings of Directional Derivatives of a Random Field Model," *Proc. IEEE Intl. Conf. Acoustics, Speech, and Signal Processing,* Tokyo, April 1986.

[8] J. C. Candy, M. A. Franke, B. G. Haskell, F. W. Mounts, "Transmitting Television as Clusters of Frame-to-Frame Differences," *The Bell System Technical Journal,* vol. 50, no. 6, pp. 1889-1917, 1971.

[9] Y. Yasuda, Y. Yamazaki, T. Kamae, K. Kobayashi, "Advances in FAX," *Proceedings of the IEEE,* vol. 73, no. 4, pp. 706-730, April 1985.

[10] R. Hunter, A. H. Robinson, "International Digital Facsimile Coding Standards," *Proceedings of the IEEE,* vol. 68, no. 7, pp. 854-867, July 1980.

[11] G. G. Langdon, J. Rissanen, "Compression of Black-White Images with Arithmetic Coding," *IEEE Transactions on Communications,* vol. COM-29, no. 6, pp. 858-867, June 1981.

[12] J. G. Snodgrass, et al., *Human Experimental Psychology,* Oxford Univ. Press, 1985.

[13] R. M. Haralick, L. G. Shapiro, "Image Segmentation Techniques," *Computer Vision, Graphics, and Image Processing,* vol. 29, pp. 100-132, 1985.