October 25, 1988

Segmentation-Based Image Coding In A Packet-Switched Network Environment

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

Segmentation-Based Image Coding in a Packet-Switched Network Environment

Sarah A. Rajala and Wonrae M. Lee

North Carolina State University, Electrical and Computer Engineering Department
Box 7911, Raleigh, NC 27695-7911

ABSTRACT

In this paper, a segmentation-based image coding technique incorporating properties of the human visual system is presented. The approach taken in this research is to first segment an image into regions with spatial similarity and then to define an efficient method for encoding the segmented image data. Furthermore, specific attention is given in this paper to defining the requirements for using a segmentation-based image coder for video transmission over a packet-switched network.

1. INTRODUCTION

Predictive and transform coding techniques have shared the mainstream of the image coding area for about 25 years. In general, these techniques attempt to find a representation of the image data such that the mean-squared-error between the original image at the transmitter and the reconstructed image at the receiver is minimized. However, recent progress in the study of the human visual system has opened new vistas in image coding. Development of methods for representing the directional sensitivity of the visual cortex combined with separate processing of contours and textures has lead to a new generation of image coding techniques [1].

The work presented in this paper is an extension of the work presented in [2,3]. In [2], we defined a new segmentation-based image coding algorithm which utilized properties of the human visual system for performing the the segmentation. In [3], we extended the previous work to permit coding of image sequences and presented a modified adaptive 3-D arithmetic coding scheme for efficient encoding of the segmented image data. Here we determine the feasibility of using such a segmentation-based image coding technique in a packet-switched transmission environment.

2. PACKET VIDEO COMMUNICATIONS

With the availability of integrated services digital networks (ISDN's), it is expected that there will be an expansion of the available digital video services. It appears that such an expansion may overload both the communication lines and the switches [4]. As a result, there is an interest in developing a complete understanding of the packet-switched network environment and in developing video compression algorithms that are particularly suited to this environment [4-9]. Because of the fixed bandwidth constraints of circuit-switched networks, most of the video coding techniques that have been developed output data at a fixed rate. With the use of packet-switched networks, it is no longer necessary to constrain the transmission rate to a fixed value. Hence, it is desirable to develop new coding techniques which can produce reconstructed images at the receiver with a fixed image quality.

Variable bit rate compression is a new application area for the field of image coding. In the following I briefly summarize some of the contributions in this field. The first published work appeared in 1986. Lazar and White [6] present results of simulating packetized video and the related protocols using the Columbia University MAGNET local area network testbed. Verbiest [7] reported on the use of asynchronous time division (ATD) video multiplexing to generate a variable rate coding technique. In 1987, Karlsson and Vetterli [8] introduced a sub-band coding technique adapted for use in a packet-switched transmission environment. In this work, priority is given to the bands that contain the most significant image information in order to minimize the effect of packet loss. In 1988, Daly and Hsing [9] proposed a technique for designing a variable bit rate coder using vector quantization.

In the remainder of this paper we present our segmentation-based image coding technique and then discussed how it can be used for packet-switched transmission.

3. SEGMENTATION-BASED IMAGE CODING

The primary goal of our segmentation-based image coding research is to develop an efficient representation of the information in an image assuming that the final viewer of the reconstructed image is a human being. Hence we are interested in determining how we can incorporate properties of the human visual system in the coding strategy. In the following subsections we describe our segmentation-based image coder.

3.1 Pre-processing

In the first stage of the algorithm shown in Figure 1, the input image is filtered to remove granularity. Edge preservation and image quality degradation are considered in order to improve the accuracy of segmentation. The goal of this stage is to suppress the granularity while keeping the same visual quality as the original.

3.2 Segmentation Using Properties of the Human Visual System

We adapt Stockham's HVS model [10] to the process of segmentation. Stockham's model consists of three major units; saturation effect, logarithmic sensitivity, and linear processing. The proposed solution for a link between these visual parameters and the segmentation technique is given as follows. The first stage of the segmentation is applied to each image frame in the spatial domain and utilizes the saturation effect and the logarithmic sensitivity. The saturation effect represents the insensitivity of the human visual system to extreme brightness and extreme darkness. Note, however, the displayed image is a mapping of gray levels onto a phosphor surface, the specific relationship between the gray levels and the dynamic range settings may differ from device to device.

Logarithmic sensitivity, given by the *just-noticeable-difference* (JND) experiment, has been implemented as a gray level JND experiment. In the conventional JND experiment, results were obtained by comparing two illuminated objects. Similarly, the JND observed on a display device needs to be measured by the amount of noticeable gray scale difference. Note, the appropriate JND data is based on well defined viewing conditions, such as brightness adjustment of the display device, room luminance and viewing distance. The gray level JND furnishes the similarity criterion (threshold) for the segmenter. The segmenter is a centroid linkage-based segmenter and the threshold is used to control the number of segments that are produced in each frame. This threshold is the first of two parameters that can be used to control quality.

The second stage of the segmentation entails spatio-temporal linear filtering. The filtering is performed to remove the spatio-temporally irrelevant segments with respect to the perceptual energy visible to the human eye. When the perceptual energy of a segment is lower than a certain threshold level, the segment is considered an irrelevant one. Otherwise, it is considered as a visible segment to the human viewer. An energy model based on the spatio-temporal impulse response of the human visual system is utilized. The threshold used in this filtering stage becomes one of two parameters that can be used to control quality.

3.3 Binary Image Generation and Segmented Image Regeneration

Further compression is obtained using an efficient representation of the segmented image data. We segment the image into constant intensity regions, which are then decomposed into two sources: a mean intensity of each segment and its boundary. In order to achieve a successful reconstruction with a minimum number of boundary pixels, a method of decomposition and reconstruction is developed. A minimal boundary consists of the set of pixels just inside the boundary of the segment. A binary image is then created with a 1 used to denote a boundary pixel and a zero elsewhere. The segment intensity data is represented by 5-bit quantization in binary form. After quantization, the intensity vector is represented by a 5-by-N binary array, where N is the number of intensities. The task of this stage is to convert two sources into binary data for use by the binary encoder.

3.4 Encoder/Decoder

In order to exactly reconstruct the boundary image at the decoder, the source coding method should be noiseless under the assumption of no channel error. Furthermore to achieve a high compression, the coder should adaptively compensate for the nonstationary statistics within a single frame of the image data. One possible candidate for this requirement is the arithmetic code introduced by Langdon and Rissanen [11]. We propose a modified version of the adaptive arithmetic code (MAAC) to overcome a generic difficulty in estimating the conditional probability for small data sets. The modified version of the arithmetic code with adaptation to nonstationarity improves coding efficiency from 3 to 14 percent over the modified stationary model [11]. A 5-by-N quantized intensity array is fed into the two-dimensional modified adaptive arithmetic-code (MMAC) and a boundary image is encoded by the three-dimensional MAAC which utilizes the spatial and the interframe correlations for estimating the conditional probability used in the encoding stage.

3.5 POST-PROCESSING

The original image has been segmented with the guidance of human visual properties. However, the segmented image still has two major difficulties. First, because we assume constant intensity segments we loose texture information inside some segments and second, the image has a blocky appearance along the boundary of the segments. Synthetic texture generation is proposed to replace some of the lost texture. A synthetic texture is generated using a two-dimensional recursive Gaussian Markov pattern which can be extended to the temporal domain. Mach band compensation is used as a solution for the blocky boundary. The deterioration related to the blocky boundaries is compensated for by removing the Mach band overshoots from the segmented image.

4. SEGMENTATION-BASED CODING IN A PACKET SWITCHED ENVIRONMENT

Much of the early work on image coding emphasized the design of a coder/decoder pair that would produce the minimum number of bits needed to represent and reconstruct a faithful replica of the original image given a particular bandwidth constraint. In a packet-switched communications environment the goal is somewhat different. Here, the goal might be posed as follows: Reduce the cost of the image transmission while retaining a fixed image quality.

Keeping in mind the new goal and the fact that an image coder will require a varying channel capacity over time, the packet-switched network must still provide channel capacity depending on the total load of the network. Therefore, an optimal solution to the transmission of variable rate image data over a variable capacity channel requires an interaction between the encoder, network and the decoder. The image codec and the network must be treated as a whole. In the following we list a set of requirements that need to be considered when designing a codec.

Requirements for Designing a Codec

(1) The bit rate should be proportional to the activity level, that is, the codec should transmit only new information.

(2) The codec should be robust to packet loss and packet delay. The image codec should be able to recover from a lost packet without disruption of the video session and be able to maintain good image quality.

(3) There should be good flow control. The image codec should get some warning of congestion from the network so it can adjust its output bit rate to help ease the congestion while maintaining image quality.

(4) The image codec should be able to provide various levels of image quality to the user, for both normal and congested operation. The network could also use this information to accept/deny access to the network based on available capacity.

(5) The system should be able to be implemented in real-time.

The question that still remains is Can a segmentation-based image codec be adapted to a packet-switched transmission environment? We believe the answer is yes for the following reasons. First, the bit rate should be proportional to the activity level. For our segmentation-based codec this can be accomplished if the interframe redundancy is taken into account. In its current implementation, this is accomplished to some degree in the 3-D modified adaptive arithmetic coder. That is the bit rate is not fixed. However, further improvements can be made by incorporating some form of motion compensation in the segmentation-based codec. Second, the codec needs to be robust to packet loss and delay. We believe a good way to make the segmentation-based coding methods more robust is to include motion compensation. When a packet is lost or delayed, the decoder can use the motion compensation algorithm to generate the missing information.

In summary, we believe that segmentation-based image coding algorithms can be adapted to a packet-switched transmission environment. However, much work remains to be done in determining an optimal coding strategy.

5. ACKNOWLEDGEMENTS

The authors wish to thank Day-Fann Shen of North Carolina State University for his contributions to this work. The research leading to this paper was supported in part by NSF Grant No. RII-8620257 and the Industry/University Center for Communications and Signal Processing at North Carolina State University.

6. REFERENCES

1. M. Kunt, A. Ikonomopoulos, and M. Kocher, "Second-Generation Image Coding Techniques," *Proc. IEEE*, Vol. 73(4), pp. 549-574, April 1985.

2. S.A. Rajala, M.R. Civanlar and W.M. Lee, "A Second Generation Image Coding Technique Using Human Visual System Based Segmentation," *Proc. of ICASSP'87*, pp. 32.2.1-32.2.4, 1987.

3. S.A. Rajala, M.R. Civanlar and W.M. Lee, "Video Data Compression Using Three-Dimensional Segmentation Based on HVS Properties," *Proc. of ICASSP'88*, pp. 1092-1095, 1988.

4. B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson and J.D. Robbins, "Performance Models of Statistical Multiplexing in Packet Video Communications," *IEEE Trans. on Communications*, Vol. 36(7), pp. 834-844, July 1988.

5. J.S. Turner, "Design of a Broadcast Packet Switching Network," *IEEE Trans. on Communications*, Vol. 36(6), pp. 734-743, June 1988.

6. A.A. Lazar and J.S. White, "Packetized Video on MAGNET," *Proc. SPIE*, Vol. 707, Sept. 1986.

7. W. Verbiest, "Video Coding in an ATD Environment" *Proc. of 3rd Int. Conf. New Syst. Services Telecomm.*, Belgium, November 1986.

8. G. Karlsson and M. Vetterli, "Three Dimensional Sub-band Coding of Video," *Proc. of ICASSP'88*, pp. 1100-1103, 1988.

9. E. Daly and T.R. Hsing, "Variable Bit Rate Vector Quantization of Video Images For Packet-Switched Networks," *Proc. of ICASSP'88*, pp. 1160-1163, 1988.

10. T.G. Stockham, "Image Processing in the Context of a Visual Model," *Proc. IEEE*, Vol. 60(6), pp. 828-842, July 1972.

11. G.G. Langdon, Jr. and J. Rissanen, "Compression of Black-White Images with Arithmetic Coding," *IEEE Trans. on Communications*, Vol. COM-29(6), pp. 858-867, June 1981.

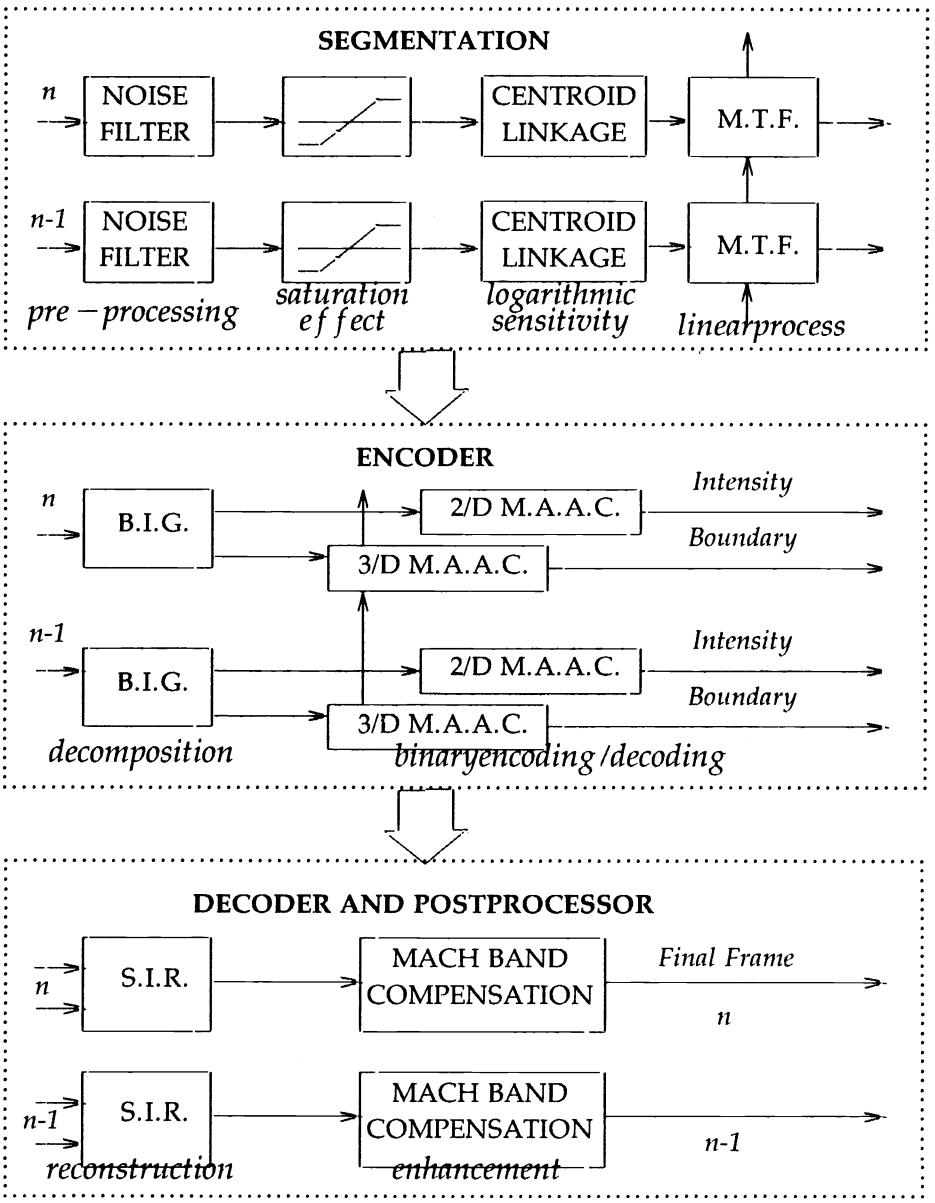

Figure 1. *Overall Block Diagram of a Proposed Compression System*