A Second Generation Image Coding Technique

Based on Human Visual System Segmentation

by

S.A. Rajala
M.R. Civanlar
W.M. Lee

Center for Communications and Signal Processing
Department of Electrical and Computer Engineering
North Carolina State University

February 1987

CCSP-TR-87/1

ABSTRACT

A new segmentation method based on the properties of the human visual system is developed in this work. The segmentation method is part of a new second generation image coder. In this coder, the difference image, between the original and the segmented version, is not transmitted (or stored) for reconstruction of the received signal. The difference image will be referred to as the visual residual and it is assumed that the characteristics of this visual residual are known. If this is so, it is only necessary to transmit the boundaries of the segmented image and the value of each segment inside a boundary.

1. INTRODUCTION

The visual residual is defined as the amount of image intensity information that could be removed without deteriorating the signal for visual perception. At this point in time, no unified human visual system (HVS) model has been established. However, some of the features are widely known, for examples, the nonlinear sensitivity of the receptors and the bandpass filtering characteristics of the human visual system (HVS). The motivation of this work is to incorporate the well known properties of the HVS into the segmentation process for image coding.

Prior to segmentation, preprocessing is performed to flatten the homogeneous regions of the image. A selective filter is used in order to maintain the sharpness of edges and flatten the homogeneous regions simultaneously.

It is a familiar fact that the relationship between the light stimulus and the psychophysical reaction in the human visual system is logarithmic. This simple fact illustrates that a measure like mean square error (MSE) is not suitable criterion for quality assessment. Since most of the segmentation algorithms require such a measure for thresholding, it is important to define a measure based on visual perception.

The binary image representing the contour of the segmented image will be treated as a facsimile image. There are numerous coding techniques available for binary images such as READ coder or arithmetic coder. The arithmetic coder can reach 98 percent of Huffman coding without any coding noise [1]. It can be easily implemented in hardware with shift registers. In some case the run-length coder is more beneficial than the arithmetic coder.

In Fig.(1), the residual image between the original and the segmented image are not transmitted. It has a special correlation to the visual phenomena. The post-processor will recover the structure compensating the residual-image as close as possible. The blocky decoded-image can be enhanced by the application of Mach band phenomenon.

2. PREPROCESSING

The purpose of preprocessing is to smooth the homogeneous regions of an image, while maintaining the sharpness in the edges.

The dynamic range of the receptors in the eye spans nine order of magnitude and can be divided into three different sections. The lower section spans from 0.01 μL to 1 μL is categorized as the infra- (or, absolute-) threshold region, while the upper region above 100 mL is referred to as the supra-threshold region. The middle range between 1 μL and 100 mL nearly follows the Weber-Fechner's law [2]. A plot shows that the middle range has a slope nearly equal to one on log-log coordinates. Over this linear portion, Weber's law is a good approximation, i.e, $\Delta I = K\ I$ and the Weber fraction, K, differs from one stimulus to the next. Weber's law concerns the relationship between the just noticeable difference (JND, ΔI) and the input stimulus,I. It is reasonably accurate when we are examining stimuli in the middle range, although it may not hold for stimuli that are extremely small or extremely large. For this reason, the dynamic range addressed in this paper is restricted to the middle or normal viewing dynamic range. The monitor is adjusted to give rise to a brightness of 100 mL for the highest gray level 255. The Weber fraction is not important to us, but the fact that the JND is proportional to the standard stimulus is an important clue in designing the segmenter. The fraction will vary with the experimental circumstances such as size of an image, room luminance, observation duration, etc. One typical experiment is given in [3, pp.81-85].

The noise contained in an image generally has a higher spatial frequency content than the image scene components. A linear convolution filtering of noise removing is a crude approximation of low-pass filter. The finite mask array will also smooth the edges and eventually it turns out blurred image. A median filter with (-) or (+) shape performs well except there is a chance of removing important small-sized features. For instance, the small size of an eyeball in head and shoulder pictures might be removed by a median filter. Therefore, we selectively filter out the noise based on visual criterion.

The selective filtering scheme reported in [4] solved the edge blurring problem with a nonlinear noise stripping method and used here.

A one-dimensional selective filter is given by

$$I_o' = I_o + \frac{\Delta_{-1} + \Delta_{+1}}{4} \tag{1}$$

where

$$\Delta_m = \begin{cases} \delta_m & \text{if } |\delta_m| \le T_o \\ +T_o & \text{if } \delta_m > +T_o \\ -T_o & \text{if } \delta_m < -T_o \end{cases} \tag{2}$$

and

$$T_o = 4.0 + 0.12301\ I_o \tag{3}$$

I_o' is the filtered value, I_o is the pre-filtered value, and the JND threshold is determined by current pre-filtered value.

By experiment, the absolute threshold is chosen to be around 60 out of 256 dynamic range of this monitor arrangement. If the threshold Δ_m in Eq.(2) set as the visual threshold in which each individual pixel has different just-noticeable-difference (JND) intensity value, then this process adapts better to the visual perception. The value δ_m was clipped to the threshold whenever it exceeded the limit in stead of being set to zero. Eq.(3) was obtained by experiment in the normal viewing conditions. The experimental equation is very closely following the first order polynomial approximation of log-log Weber-Fechner's law.

The cross (+) shaped of two-dimensional array demonstrates that some additional improvement in segmentation work is obtained with the second-stage two nearest-neighbor filter given by

$$I_o'' = I_o' + \frac{\Delta_{-2} + 2\Delta_{-1} + 2\Delta_{+1} + \Delta_{+2}}{8} \tag{4}$$

where the Δ_m's were clipped to $\frac{T_o}{4}$.

The result maintains the sharpness along the edges and the smoothness in the homogeneous area.

3. SEGMENTATION

Haralick and Shapiro [5] present a good survey of the variety if image segmentation techniques in existence. However, achieving all the perfect results is difficult because homogeneous regions are typically full of small holes and have ragged boundaries in an image. In addition, the segmenters does not take into account properties of the HVS.

The modulation transfer function (MTF) of the HVS indicates a bandpass response to spatial frequencies. That is, the visibility of very slow and very fast intensity transitions is low. During the segmentation stage, this property together with Weber's law is used to remove the small holes and to cluster the homogeneous pixels. The decision threshold for removing a small region is based on its size and the contrast difference between its neighbors. A variation of the centroid linkage region growing segmentation algorithm is used.

In Fig.(2), let [O] be the observation pixel, then attempt to merge this pixel with one of the previously segmented regions. The algorithm scans the pixels from top to bottom and from left to right. The centroid linkage window does not take into account the future pixels which have not yet been clustered. The pixels labelled [1],[2],[3], and [4] might be same or different from each other. The segmenter looks at all neighboring pixels, [1],[2],[3] and [4] to find the neighbor which has no visually significant difference. If none of the neighboring pixels match the criterion in Eq.(3), then the algorithm declares a new region. If there is only one good neighboring pixel to which pixel [O] can merge, then the pixel [O] is included into the region with that neighboring pixel. This assures the connectivity of all pixels in the same segment. If there are more than two good neighboring pixels, then situation is more complicated.

A linkage pixel is defined as the pixel posited between separate homogeneous regions. Without this pixel, a merge between regions can not be established. Suppose the observation pixel [O] is positioned on the linkage pixel in Fig.(3) where #23 and #32 segments have nearly same intensity respectively. The two separate regions can be connected and merged through the linkage pixel. Then new intensity value and the number of pixel in the merged segment are updated.

For a normal viewing distance (around 3 feet), a single pixel having a JND value larger than the threshold value in Eq.(3) does not have much of a contribution to the overall perception. However, this single pixel was not removed in the preprocessing stage because its value was larger than the visual threshold established by Weber's law. At this stage, the single pixel is merged with one of its neighbor in the second pass. By imposing a cost function between the size and the intensity difference from neighboring pixels, we attempt to merge this insignificant small-sized region, typically having less than 6 pixels, with the closest neighbor.

The segmenter accurately identifies most of the edges. The exception is wide ramp edges, for which it is also difficult for a human to identify an exact positioning of the edge. If the width of a ramp edge is relatively small (less than 4 pixels wide), then the algorithm replaces this edge with a step edge. Therefore, the maximum segmentation error occurring in ramp edges is 2 pixels. Ramp edge with large width is segmented to a staircase edge.

4. CODEC

In order to reduce the number of boundary pixels, it is necessary for the binary image generator to perform under the following conditions; (1) a minimum number of boundary pixels (1's) should be maintained and (2) the inside of the segments should be represented by the correct intensity in the decoding stage. The correspondence process of representing the exact intensity of a segment is called "painting". The decoded binary segments are filled with the appropriate intensities in sequential order by the painter. The final decoded binary image can be reconstructed exactly under the conditions that the codec is noiseless and the channel is error free.

Several efficient coders exist in the facsimile environment. The modified READ code and the arithmetic code are two examples [6]. It has been observed that the arithmetic coder performs better. One difference between the typical facsimile binary coder and the current application is the number of pixels. Facsimile coding operates over nearly 2.8 million pixels while the number of pixels in this application is fairly small (65.5K pixels) and the correlation among 1's is very low. Therefore, the compression rate in the binary coding stage expects to be lower than for facsimile data compression. It should be noted, however, that the rate in this stage governs overall compression rate.

5. POSTPROCESSING

There are two useful visual effects that should be taken into account: (1) Relatively large error occurring right on the edges are not very noticeable, thus degradation in the edges can be tolerated more than degradation in the homogeneous regions. (2) The Mach band effect should be taken into account, that is, a simultaneous brightness contrast pattern, see Fig.(4-a), having uniform luminance in a strip is perceived to have a nonuniform distribution.

To compensate for the Mach band effect in the segmented image, the postprocessing filter generates a compensation pattern, Fig.(4-c), which is similar to the redundancy image discarded in the segmentation stage. Compensation is only done for ramp edges while the high contrast edges remain untouched.

6. RESULTS

In Fig.(5), the upper left one is the original USC girl and the upper right is the segmented image. The lower left is the residual image. It should be noted that the several black and white dots along the big contrast edges do not affect the subjective quality of the image. The lower right image is a decoded image with compensation. Note that the compensation of the Mach band effect makes the segmented edges more smooth but not in the sharp edges. Small-sized active pixels in hair area of Fig.(5) have been merged into small number of segments. This is mainly due to the cost function of MTF is computed locally. This is one of inherent problem of the centroid linkage segmentation.

The original image in Fig.(5-a) is a 256-by-256, 8 bits per pixel representation. There are 307 segments in the second image, Fig.(5-b). 12473 bits are required to describe the binary image and the intensity vectors. The intensity vectors are nonuniformly quantized with 5 bits. The overall compression rate for the image in Fig.(5-d) is 42:1.

To compare the performance of the visual segmenter, the fixed number of segment 307 is given as a constraint for the fixed threshold segmenter. Same parameters for the preprocessing and the compensation are given for both. The fixed threshold value 21 is applied for the given constraint. Compare Fig.(6-a) to Fig.(6-b).

This technique has been extended to time varying imagery. In that case compression, the quality degradation in the individual frame due to segmentation can be more or less disregarded. Fig.(6-a) shows one of time varying video frame in the teleconferencings, and Fig(6-b) is the decoded frame as a rate of 56:1 compression (9328 bits with 159 segments).

7. REFERENCES

[1] G.G. Langdon and J. Rissanen,"Compression of Black-White Images with Arithmetic Coding," IEEE Trans. Comm., vol. COM-29, No. 6, pp. 858-867, June 1981

[2] D. Jameson and L.M. Hurvich Ed., *Visual Psychophysics,* in Handbook of Sensory Physiology VII/4, Springer-Velag, 1979

[3] T.N. Cornsweet, *Visual Perception,* Academic Press, N.Y., 1970

[4] D. Anastaussiou, et al.,"Gray Scale Image Coding for Free-Frame Videoconferencing," IEEE Trans. on Comm., vol. COM-34, No. 4, pp. 382-394, April 1986

[5] R.M. Haralick and L.G. Shapiro,"SURVEY: Image Segmentation Techniques," Comp. Vision, Graphics, and Image Proc., vol. 29, pp. 100-132, 1985

[6] R. Hunter and A.H. Robinson,"International Digital Facsimile Coding Standards," Proc. of IEEE, vol. 68, No. 7, pp. 854-867, July 1980

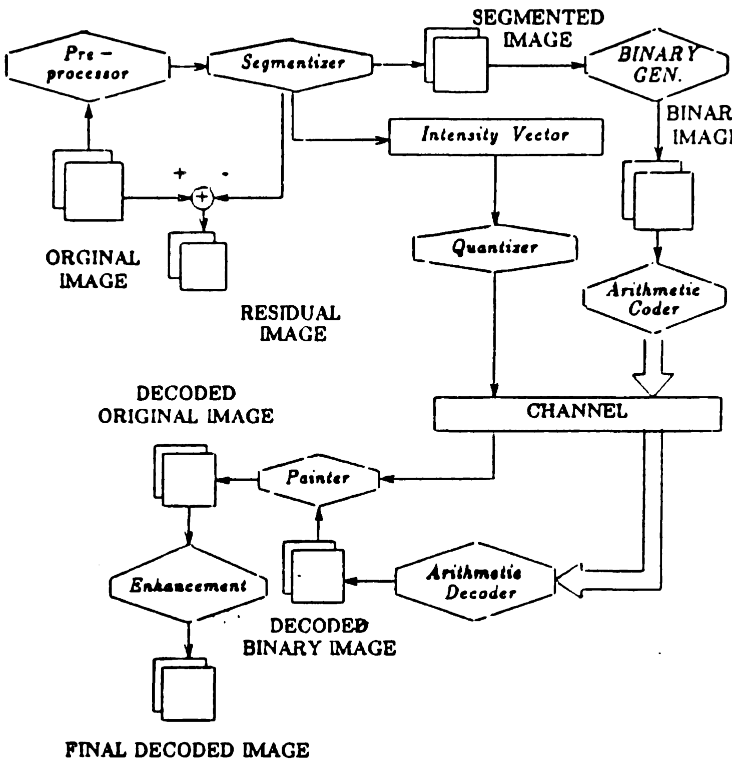

Figure (1): Overall Scheme

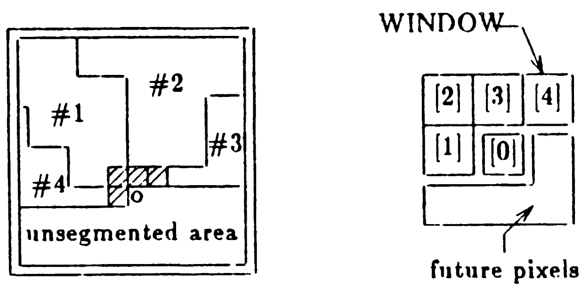

Figure (2): A Centroid Linkage Pixel

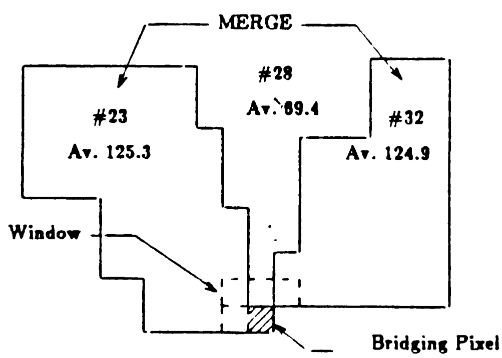

Figure (3): A Bridging Pixel

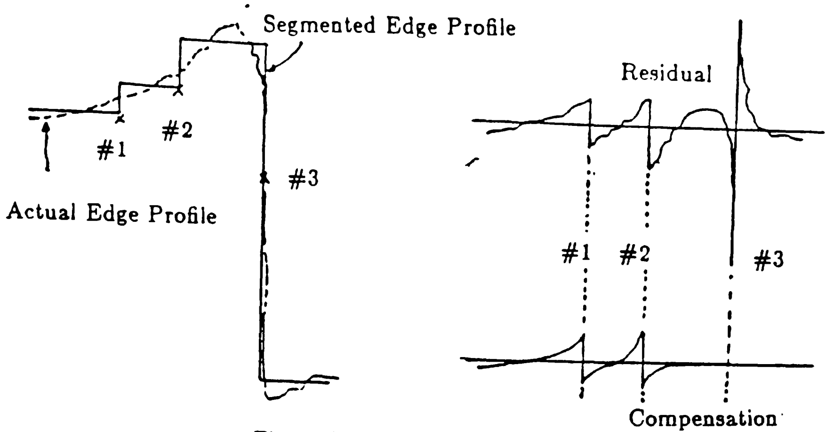

Figure (4): Post Processing

(5-a)	(5-b)
(5-c)	(5-d)

(6-a)	(6-b)
(6-a)	(6-d)