Progressive Image Coding with Spatially Variable Resolution

Tamás Frajka, P. Greg Sherwood, and Kenneth Zeger

Department of Electrical and Computer Engineering,
University of California, San Diego, La Jolla, CA 92093-0407
email: {frajka,sherwood,zeger}@code.ucsd.edu

Abstract

We present a wavelet-based progressive image coding system that allows the specification of regions of interest to alter the spatial allocation of future transmitted bits. When used in an interactive mode with feedback, the viewer, after seeing a low resolution version of the image, can instruct the encoder to concentrate more effort on coding regions of interest. When operating in independent mode, the encoder can select regions of interest based on automatic feature recognition algorithms or offline human selection. The algorithm is flexible enough to adapt to changes of interest during encoding.

1 Introduction

In many image coding applications it is important that the receiver be able to periodically reconstruct successively refined versions of a transmitted image, in real-time, as more bits arrive from the transmitter. Very elegant "embedded" coders exist which provide continual improvements in the decoded image with each new bit [1, 2]. These "bit progressive" systems allow the viewer to obtain a basic understanding of the transmitted image quickly (e.g. typically at rates as low as 0.05 bits/pixel). Obtaining high resolution in any particular part of the image takes many more bits. However, in certain applications it is beneficial to quickly view a small portion of the image in high detail. For example, based on a low resolution reproduction of an image using a relatively small number of received bits, the viewer might choose to focus attention on one or more interesting features that are recognized in the image, e.g. human faces, written words, other scenery, or perhaps military targets. The potential applications include both video and still image transmission.

The proposed algorithm can be applied in an interactive mode where the viewer can provide feedback and an independent mode where the regions of interest are selected by the encoder based on feature recognition algorithms or human selection. In order to motivate the interactive mode, suppose a user wishes to inspect a 1024 x 1024 pixel grey level image by downloading it over the Internet. During a heavily congested time period, a user might receive on average only hundreds of bytes per second. At a transmission rate of say 300 bytes per second, it would take more than 7 minutes to transmit the image at a resolution of 1.0 bits/pixel. The user might only need to see a high resolution reproduction of one small portion of the image, before being able to stop transmission and move on to inspecting other transmitted images. If a small feedback signal (e.g. via a mouse click) could instruct the transmitter to concentrate on one area of interest, the decision to terminate transmission could be made much sooner. Meanwhile, the independent mode could be used in video conferencing where the background is usually less important for the viewer than the face of the other person. The region of interest could be selected using some kind of feature extractor as in [3].

The proposed algorithm builds upon the wavelet-based algorithm of Said and Pearlman [2] (an enhancement of Shapiro's EZW algorithm [1]) and maintains its progressive (and embedded) property while adding the capability of a spatially adaptive allocation of bits. The bit allocation is based on a general importance function which defines the relative interest of the viewer to the various parts of the image. The algorithm is flexible enough to allow arbitrary changes of "importance" during transmission.

2 Wavelet-based coders

Wavelet transforms provide good spatial and frequency localization that make them very suitable for image processing and compression. Shapiro [1] introduced an embedded coding algorithm based on the concept of zerotrees which takes advantage of dependencies between wavelet coefficients at different scales corresponding to the same spatial locations. This method was further improved by Said and Pearlman [2]. Both algorithms avoid explicitly transmitting the location of significant coefficients (i.e. those whose magnitude is above the current threshold); instead the significance map can be determined by the outcome of comparisons. In the SP algorithm the coefficients are stored on two different lists, one with significant pixels and one with insignificant pixels, which are processed in two distinct passes. The encoder and decoder build identical copies of these lists based on the comparison results. In the sorting pass the insignificant pixels are compared to the current threshold value and are moved to the other list if found significant. Their descendants are checked in the same way. In the refinement pass the values of the significant pixels are further refined to the next bitplane. Since these algorithms transmit the most significant information first, they can stop at any point resulting in an approximation of the image at that rate. The advantage of the encoding is that the target bit rate need not be known ahead of time to design any part of the coder. Of course, low bit rate images are of perceptually poor quality but usually better than the results of DCT based techniques of equal rate.

3 Region Masking

The goal of region masking is to achieve improved perceptual quality in "important" regions (perhaps at the expense of reduced quality in other less important locations) while maintaining the bit rate. One possible solution based on scalar quantization of the coefficients is to use finer quantization in the preferred regions as in [3]. A technique for emphasizing certain predefined regions offline using the EZW algorithm was given in [4].

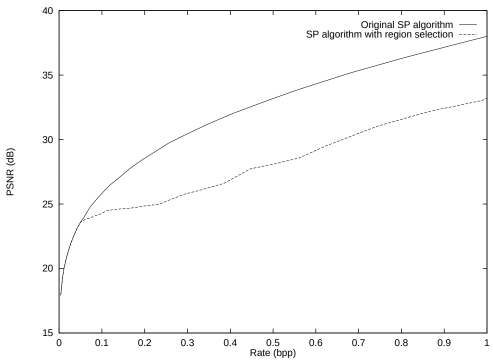
Here we propose a modification of the SP algorithm that idles those coefficients that do not correspond directly to the specified regions. Thus for a few sorting and refinement passes all the bandwidth is assigned to the important areas. For proper communication, both the encoder and the decoder must maintain the same set of active coefficients. The proposed method only requires communication of the region description; everything else is computed locally.

To characterize the relative importance of each pixel location in an image, we define an *importance function* i that assigns a value $0 \leq i(x,y) \leq 1$ to each pixel (x,y) in the spatial domain, such that its sum over the pixels in the image is 1 (i.e. i is a probability mass function). Larger values of $i(x,y)$ are used to indicate that higher resolution is desired for pixel (x,y). Starting with a uniform importance function, its value is updated each time a change in preference occurs.

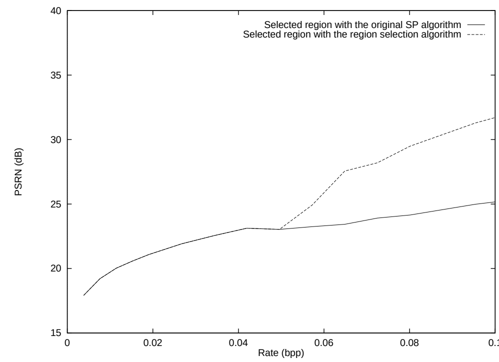
In the SP algorithm, the quantization and coding occurs in the wavelet domain, and hence it is necessary to convert the spatial domain importance description into a bit resolution description for each coordinate (u,v) in the wavelet domain. We define a mapping I, that assigns to each pair (u,v) in the wavelet domain, a number $0 \leq I(u,v) \leq 1$. The mapping I is derived deterministically from the mapping i. As an example, for a given (u,v) the value of $I(u,v)$ might be chosen to be the average value of $i(x,y)$ taken over the set of all pixels (x,y) in the spatial domain, that are mapped to (u,v) via the wavelet transform. Due to the nature of the subband decomposition each subband will contain coefficient(s) corresponding to the selected region(s). Since the SP algorithm does not allocate bits uniformly to each wavelet coefficient, the relative importance of the pixels is not matched exactly by the relative transmission rate assigned to them.

To relate the notion of "importance" to wavelet domain resolution, we delay the processing of wavelet domain pixels associated with unimportant regions in the SP algorithm (i.e. no bits are spent coding the delayed pixels). We introduce a non-negative integer valued *inactivity counter* $S_n(u,v)$ which indicates the number of sorting and refinement passes to wait, after the n^{th} iteration, before resuming processing of the wavelet domain pixel (u,v) in the SP algorithm. Initially we set $S_n(u,v) = 0$ for all (u,v), and after the n^{th} sorting/ refinement pass, we perform the update $S_{n+1}(u,v) = S_n(u,v) - 1$ provided $S_n \neq 0$. If the importance function i has changed during the n^{th} sorting/refinement pass, the values of $S_n(u,v)$ are recomputed. For a newly updated importance function i_n the inactivity counter is set to $S_n(u,v) = \lceil K * (max_{(w,z)} I(w,z) - I(u,v)) \rceil$ where K is a design parameter. Note that this computation can be done by the encoder and decoder independently since they both use the same importance function at any point of the algorithm. The value of K determines the amount of emphasis given to the selected region.

If the importance function changes during encoding, that new information must be provided to the decoder. The importance information needs to be placed in the bit stream carefully to avoid confusion with image data. Two methods of integrating the importance and image data are to place the importance data in the bit stream spaced by a fixed number of image bits or at fixed points in the processing such as the end of a sorting or refinement pass. The encoder can use 1 bit at each such point to indicate whether the next n bytes contain control information or more image information. In the case where a change is prompted by feedback from the decoder, the control information can simply be an acknowledgement of the request. It is important that the encoder does not start to use the new importance function before that control information is sent to the decoder to keep the two consistent.



(a)



(b)

Figure 1: The PSNR comparison between the SP algorithm and the algorithm with region selection for (a) the entire image, (b) for the region selected in the mid-right section.

4 Simulation Results

In this simulation we assumed the following situation: someone is downloading an image and during transmission can select regions using a mouse. A single mouse click is translated into adding a two dimensional Gaussian lobe centered at the mouse click to the importance function. Since the natural synchronization points (end of sorting/refinement pass) are spaced farther apart in the bit stream as the algorithm progresses, we chose to use synchronization points after a predetermined number of information bits for faster feedback.

Figure 2 shows 3 snapshots of both the SP algorithm and of the SP algorithm with region selection. The images obtained with the SP algorithm are shown in the left column (images (a), (c) and (e)). The region in the mid-right part of the image was selected at a bit rate of 0.5 bpp. It can be seen on images (a) and (b) that shortly after the selection (at 0.11 bpp) the head in the selected region is recognizable with the region selection, while still blurry on the other image. The next pair of images shows the situation at 0.23 bpp shortly after the region around the girl in the middle is selected. Finally the last pair shows the result at 1 bpp. The real difference between the two images is the resolution towards the sides of the image. On the image with the region selection, (f), the sides are still a little blurry because less bandwidth was allocated to these parts.

A more quantitative comparison is shown in Figure 1. Part (a) compares the PSNR of the same image encoded with the SP algorithm and with the regions selected as described above. The PSNR with the region selection on the overall image is about 5 dB inferior to the SP algorithm. But as shown in part (b) for only the selected region in the mid right section, the PSNR in that region improves very fast to about 6.3 dB (at 0.1 bpp) over the SP algorithm in that same region.

5 Conclusion

A new region masking technique was presented based on the Said-Pearlman algorithm that can flexibly accommodate the changes in the regions of preference. A modification of the original method was proposed that tries to minimize the overhead information by including many of the computations in both the encoder and decoder based on the information available to both sides. Simulation results show that perceptual image quality in the selected regions is very good even at low bit rates.

Since the algorithm does not require any kind of modification of encoding parameters, it can be used on images already compressed by the SP algorithm. In this case, the algorithm rearranges the bits so that information about the preferred regions is sent first. The final image quality is identical to that achieved by the SP algorithm with essentially the same bit rate (i.e. very small overhead is necessary to specify regions). However the quality in the selected regions improves more rapidly.

6 References

[1] J. M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Trans. on Sig. Proc.*, 41(12):3445–3462, December 1993.

[2] A. Said and W. A. Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Trans. on CSVT*, 6(3):243–250, June 1996.

[3] J. Luo, C. W. Chen, and K. J. Parker. Face location in wavelet-based video compression for high perceptual quality videoconferencing. *IEEE Trans. on CSVT*, 6(4):411–414, August 1996.

[4] J. M. Shapiro. Smart compression using the embedded zerotree wavelet (EZW) algorithm. In *Proc. of the 27th Asilomar Conf. on Signals, Systems and Computers*, volume 1, pages 486–490, November 1993.

Figure 2: The sequence of images at different rates with the SP algorithm (left) and with region selection (right), (a)-(b) 0.11 bpp, (c)-(d) 0.23 bpp, (e)-(f) 1 bpp