IMAGE CODING WITH MIXED REPRESENTATIONS AND VISUAL MASKING

Bin Zhu, Ahmed H. Tewfik

Department of Electrical Engineering
University of Minnesota
Minneapolis, MN, 55455
e-mail: binzhu, tewfik@ee.umn.edu

Ömer Nezih Gerek

Department of Electrical Engineering
Bilkent University, Bilkent
Ankara 06533, Turkey
e-mail: gerek@ee.bilkent.edu.tr

ABSTRACT

In this paper, we propose a novel approach for low bit rate perceptually transparent image compression. It exploits both frequency and spatial visual masking effects and uses a combination of Fourier and wavelet transforms to encode different bands. Frequency domain masking is computed by using a fine to coarse analysis step. Spatial domain masking is computed either by using Girod's model or a coarse to fine analysis step that accurately computes local contrast. A discrete cosine transform is used in conjunction with frequency domain masking to encode the low frequency bands. The medium and high frequency bands are encoded using spatial domain masking and a wavelet transform. The encoding of these bands is based on a recursive selection of the important edges in each band. It uses cross-band prediction to minimize bit rate. Experiments show the approach can achieve very high quality to nearly transparent compression at bit rates of 0.2 to 0.4 bits/pixel.

1. INTRODUCTION

Image compression with little or no visual distortion is desirable in many applications. It is of considerable importance in these applications to reduce the bitrate as much as possible. To achieve this goal, a coding scheme should exploit both redundancy and irrelevance in an image. Unfortunately most coding schemes proposed in the literature exploit only redundancy in an image, due to the complexity and limited knowledge of the human visual system. We believe the application of visual masking effects in the coding design process is critical to achieve a high level of performance. Recently Safranek et al proposed an empirical visual metric to remove irrelevance in images [1]. We propose here a direct application of both spatial and frequency visual masking effects in coding design. The masking models come directly from psychophysical data and have been validated experimentally. The wavelet transform is used here due to its similarity to the tuning mechanism in the human visual system[2] and to its close approximation to the Karhunen-Loeve transform[7].

Our technique is based on the observation that low frequency images fit the set-up of the experiments that researchers have used to determine frequency domain masking effects. The images that correspond to medium and high frequency bands are edge images. Therefore, masking in these bands is dominated by spatial domain masking. We use a fine-to-coarse analysis step to determine segments of the image over which particular frequency domain masking thresholds hold. A coarse-to-fine analysis combined with the results of [4], [5] or the model of [3] is used to determine spatial masking thresholds. To match the two visual masking models, two representations are employed to encode the subimages. Subimages of medium and high frequencies are encoded in the spatial domain representation using a wavelet representation. Subimages corresponding to low frequencies are encoded in the frequency domain by applying a discrete cosine transform (DCT). Scalar quantization of all transform coefficients is used and designed such that the quantization error at each pixel does not exceed a pixel dependent tolerable error level. It is followed by entropy coding.

2. THE WAVELET TRANSFORM

The wavelet transform provides a compact multiresolution representation of the image. The wavelet transform used in this paper is a tensor product of two identical one-dimensional wavelet filters. The mean of an image is first calculated and removed from the image; then the wavelet transform is used to build a multiresolution representation of the image. The number of decomposition levels is generally of 5 to 6 for images of size 512 by 512. The mean is quantized with 10 bits and passed to the decoder as side information.

Short wavelet filters with large vanishing moments are used for the decomposition. We avoided long wavelet filters to have a better control on the spatial spread of coding and quantization errors. The particular wavelet filters that we have used are based on those of [8]. With these filters, the low-pass subimage is close to a down-sampled version of the original image. This is desirable when we calculate tolerable error level for each pixel (see Sect 3, as well as for the cross-band prediction (see Sect. 5).

3. VISUAL MASKING MODELS

We use a two step procedure to calculate the tolerable error level (TEL) at each pixel of each subimage. The first step is based on Girod's vision model [3]. This model accurately predicts the masking effects around luminance edges as well as masking by uniform background. It can be directly used

0-7803-2431-5/95 $4.00 © 1995 IEEE

in the design of coding systems. After quantization, the model is applied to check if distortions in the quantized image are visible. If any are not acceptable, we re-perform the quantization step. The procedure is then repeated until no distortion is visible. This process is time-consuming. We therefore modify the model so that the tolerable error level can be found for each pixel of each input subimage directly.

To find TEL for each pixel of the original image, we use the linearized model of [3] under the assumption that all errors in the coded image are small. This is a reasonable assumption for nearly transparent image coding. Given the thresholds computed by the model at each pixel, we conservatively calculate the TEL at each pixel of the coded image by neglecting the optical blur implicit in the formation of a retinal image and by using the simplified w-model of [3]. To find the tolerable error level for wavelet decomposed subimages, we can either use the same method as we used above to find the TEL for the original image, or simply use Gaussian filters to process the TEL of the original image, with a spatial spread comparable to the wavelet filter's and coefficients scaled to the visual contrast sensitivity at the central frequency of that subimage. The second method is chosen in this paper since the coding gain predicted by Girod's model is small (see below). The parameters in Girod's model are adjusted to match our actual testing environment.

Note that it is also possible to replace the model of [3] by the spatial domain masking effects predicted by local contrast studies [4], [5]. In this case, we use a coarse-to-fine step that starts with the low frequency subimages and proceeds to higher frequency images to estimate spatial domain masking in the various medium and high frequency bands.

Girod's model predicts very small coding gain, typically around 0.5 bits per pixel [3]. This has lead us to supplement the masking results predicted by this model with those predicted by experiments based on sinusoidal patterns (frequency domain masking) [6]. We use the outputs of the medium and high pass filters in a filter bank decomposition of the image to segment the image into regions over which one can safely apply frequency domain masking based on the results of [6].

4. CODING OF SUBIMAGES

Given the TEL that can be tolerated at each pixel, we proceed to encode the subimages corresponding to the different frequency bands. Recall that these subimages are obtained using a wavelet decomposition based on the wavelets of [8].

Masking in the low frequency images is mostly a frequency domain effect. Therefore, we could encode these images using the same techniques used in [7]. To reduce the complexity, we instead transform subimages of low frequencies into the frequency domain by using the DCT. Next, we use frequency domain masking to determine the TEL in each frequency bin. This then determines the amount of quantization of each frequency coefficient. This information is encoded using a vector quantization approach and sent as side information. This works since the subimages of low frequencies have good resolution in the frequency domain.

In order to encode the medium and high frequency bands (the LH, HL and HH images) at different scales, we proceed to select recursively the important edges in each subimage.

We use a modification of the technique described in [11] for extracting the important extrema of the high frequency sub images. The proposed method is iterative, and can retain as many of the extrema as required. The selection of the important edges and the amount of error that can be tolerated in their representation are both based on the TEL information extracted during the computation of spatial masking in the different bands. To reduce the amount of information transmitted, we exploit the fact that the impulse responses of the wavelet analysis filters are known and that the edges in the various bands are due to discontinuities in the intensity of the image or some of its higher derivatives. At any iteration, we localize the position of the important edges in the highest scale LH, HL and HH subimages. We encode these edge contours using the technique described in the next section for transmitting outlines of regions with similar TEL levels. Next, we predict the position and magnitude of the important edges in lower scale images. Note that the decoder can also implement this operation once it has decoded the highest scale subimages. Finally, we determine if we need to also encode the deviations from these predicted values.

5. THE QUANTIZATION STAGE

Several methods for quantization have been investigated and are still under investigation. They are all designed in such a way that at any pixel of any subimage, the quantization error does not exceed that of the tolerable error level found by masking models for that pixel. A simple way is to use uniform quantization for the subimages. First, set all the values smaller than their TEL to 0. Next, select a quantization step that is twice the value of the smallest TEL of the remaining pixels in that subimage. That ensures that no quantization error is larger than the corresponding TEL. The quantization step of each subimage is represented in a floating point format with 10 bits devoted for mantissa and 2 bits for the exponent. It is sent to the decoder as side information. The optimum dividing point between spatial and frequency domain representations of subimages is found experimentally.

While this method is simple, it does not exploit the large variation in the tolerable error level due to its uniform quantization. This can be avoided if we use adaptive quantization. This is not straight forward since the decoder does not know the tolerable error level at each point. We first segment the TEL image into several regions. Within each region pixels have similar tolerable error level. Each region is uniformly quantized. The quantization step for each region is passed to the decoder. The contours of the regions are either coded by contour coding in [9] or by run length coding, and are sent to the decoder as side information. For run length coding of the contours, we use two end points to represent the contour points in the line connecting the two end points. Since coding of the contours is expensive, we do this for the subimage at the second or third level. For subimages at other scales, the contours are either down-sampled or up-sampled and then extended to be continuous in the sense of 8-connectivity neighborhood, depending on the subimage sizes. The subimages in frequency domain representation do not use this segmentation since the DCT requires rectangular shaped images, and this requirement would offset any saving by doing segmentation, due to the small subimage sizes. A method to do the above segmentation from the subimage of lowest frequency with gradually refined as subimages of high frequencies are considered is currently under investigation. This will save us from coding the contours which is expensive.

6. ENTROPY CODING

After quantization, the two-dimensional array of coefficients is organized into a one-dimensional array. subimages that correspond to a spatial domain representation are scanned as follows. Each LH is scanned horizontally: even numbered rows are scanned from left to right and odd numbered rows are scanned from right to left. Each HL is scanned vertically in a similar manner. For the HH subimages and subimages in frequency domain representation, a zigzag scan same as that in JPEG is used.

The one-dimensional array integers are then translated into symbols. There are two ways to do so. The first way is to use run length to code number of zeros. The run-length zeros are coded by a zero mark followed by an integer which is the number of run-length zeros. Each integer is coded with two symbols, the first is the size of the integer, the second is the code of the integer which is similar to that used in JPEG[10]. .

For subimages of low frequency and those quantized in the frequency domain, we use similar method as the JPEG to form symbols.

The symbols are coded with Huffman coding, which provides further compression.

Figure 1: The original image.

7. EXPERIMENTAL RESULTS

We have tested the performance of the proposed coding scheme on a SUN sparc 2 workstation. Fig. 1 shows part of the original 512 × 512 Lena. This part is more critical for perceptual quality than other parts. Fig. 2 and Fig. 3 show the corresponding part of the decompressed images with coding rates of 0.38 bits/pixel and 0.25 bits/pixel, respectively. The image was decomposed using six levels. All subimages except the low frequency one were quantized in the spatial domain. The LL subimage was coded in the frequency domain. The TEL image was divided into 8 regions and the contour of the regions were coded by run length coding. The run length coding is not as efficient as the edge coding method proposed by Kunt et al. [9], and a better segmentation and contour coding are still currently under investigation. While we have not tried hard to optimize any parameter, Fig 2 is nearly transparent to the eye when looked at from any distance. Fig 3 shows some perceptual distortion when looked at closely, mostly around the shoulder. This is understandable since visual masking around the shoulder is smallest.

8. CONCLUSIONS

We have presented an image compression method to exploit irrelevance in an image by applying spatial and frequency masking effects directly in the design of coding and quantization. Both spatial and frequency domain representations are used to represent the wavelet-decomposed subimages, in conjunction with the two visual masking models. The method achieves low bitrate image compression with nearly transparent quality.

9. REFERENCES

[1] R. J. Safranek and J. D. Johnston, "A Perceptually Tuned Sub-band Image Coder with Image Dependent Quantization and Post-quantization Data Compression," *ICASSP'89*, PP 1945-1948.

[2] S. G. Mallat, "Multifrequency Channel Decompositions of Images and Wavelet Models," *IEEE Trans. ASSP*, Vol. 37, No. 12, PP 2091-2110, 1989.

[3] Bernd Girod, "The Information Theoretical Significance of Spatial and Temporal Masking in Video Signals," *SPIE Vol. 1077 Human Vision, Visual Processing, and Digital Display*, PP 178-187, 1989.

[4] Eli Peli, "Contrast in Complex Images," , *J. Opt. Soc. Am. A.*, vol. 7, no. 10, pp. 2032-2040, 1990.

[5] A. Saadane, H. Senane and D. Barba, "Masking and Quantization Laws in a Visual Subband Coding

Figure 2: The decoded images with 0.38 bits/pixel.

Scheme," *Proc. 1994 IEEE Int. Conf. on Image Proc.*, vol. 2, pp. 869-873, 1994.

[6] G. E. Legge and J. M. Foley, "Contrast Masking in Human Vision," *J. Opt. Soc. Am.*, Vol. 70, No 12, PP 1458-1471, 1980.

[7] D. Sinha and A. H. Tewfik, "Low Bit Rate Transparent Audio Compression Using Adapted Wavelets," *IEEE Trans. Signal Processing*, Vol. 41, No. 12, PP 3463-3479, 1993.

[8] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast Wavelet Transforms and Numerical Algorithms I," *Research Report YALEU/DCS/RR-696*, 1989.

[9] M Kunt, A. Ikonomopoulos, and M. Kocher, "Second Generation IMage Coding Techniques," *Proc. IEEE*, Vol. 74, PP 549-575, 1985.

[10] G. K. Wallace, "The JPEG Still Picture Compression Standard," *Communications ACM*, Vol. 34, No. 4, PP 31-44, 1991.

[11] Y. Xu, J. B. Weaver, D. M. Healy, Jr., and J. Lu, Wavelet Transform Domain Filters: A spatially Selective Noise Filtration Technique, IEEE Trans. Image Processing, Vol. 3, No. 6, Nov. 1994.

Figure 3: The decoded images with 0.25 bits/pixel.