

Lossy Compression of Head and Shoulder Images Using Zerotrees of Wavelet Coefficients

David B. Gerhard and W. Kinsner

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba, Canada R3T 5V6

Abstract

This paper presents an image compression system based on the wavelet transform. The compression system includes pre-processing to remove unimportant information, wavelet transform of the image, thresholding of the wavelet coefficients, zerotree coding of the thresholded coefficients, and finally arithmetic coding of the zerotree code. Post-processing is done to remove salt and pepper noise, an artifact of the wavelet transform at high compression levels. The transform procedure is designed for the particular class of images showing head and shoulders of individual people facing front, with a homogeneous background. A group of 12 passport photographs is used as a representative image set for testing purposes. Results obtained using the wavelet transform procedure are compared to the results of compressing the same images using the JPEG grayscale compression standard. The quality of the images is judged using peak signal to noise ratio (PSNR). At compression ratios below 40:1, the JPEG compression system produced better quality images than the wavelet system. At compression ratios above 40:1, the wavelet transform produced images of better quality. The wavelet transform produced images of above 30dB PSNR at compression ratios up to 150:1. These images are acceptable for subject recognition.

1. INTRODUCTION

A passport image or similar photographic identification contains the information necessary to visually recognize an individual. It is desired to code images of this type with high compression ratios for storage, transmission and security while retaining the information necessary for recognition.

General image compression calls for specific techniques. These techniques fall into two categories: lossless and lossy. Lossy techniques work on the principle of removal of information, and include techniques based on the discrete cosine transform (DCT). Lossless techniques work on the principle of redundancy reduction, and do not remove any information. The decompressed file is identical to the original file. The wavelet transform is a lossy technique which provides greater compression than other popular lossy techniques at the same image quality.

Pre-processing the images increases the compression ratio by taking advantage of properties of the images that remain consistent throughout the image set. Such properties must be identified and shown to be consistent throughout the set before they can be used to increase the compression. In the image set used, these properties include the position of the head and the homogeneous background.

2. THEORETICAL BACKGROUND

2.1 Wavelet Theory

The discrete Fourier transform represents a sampled signal as a sum of sinusoids. This method has been used for the compression of data, by recognizing that while a Fourier representation of a signal typically has an infinite number of coefficients, the majority of the information in the data is contained in the first few coefficients. There is an inherent

limitation of Fourier methods: non-periodic signals and signals with discontinuities require many coefficients to represent accurately.

The Joint Photographic Experts Group (JPEG) compression standard is based on the Discrete Cosine Transform (DCT). This is a type of Fourier transform, using cosines as the base function. DCT has the property that the coefficients of high value tend to be centered around one corner of the image. It is this property that lossy DCT compression techniques take advantage of.

We would like to devise a transform using a function other than a sinusoid as a basis, which would represent non-periodic functions in a small number of terms, and represent periodic functions well also. This is the idea behind wavelets and the wavelet transform. A wavelet is a function which is localized in both frequency and time—wavelets decay to zero toward infinity in both domains. Wavelets have the further property that they integrate to zero, which implies that the functions are “balanced” above and below the time axis.

To take a wavelet transform, we first select a wavelet base function, and then correlate each possible scaling and translation of that wavelet with the function to be transformed. The correlation values thus obtained are the coefficients of the wavelet transform. The equation for the scaling and translation of the wavelet function is

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad (1)$$

where $\psi(t)$ is the original, or “mother” wavelet, a is the scale factor, and b is the translation factor. Because wavelets have two variables associated with them, the wavelet “spectrum” of a function is two dimensional. While the Fourier transform is a function of frequency, the wavelet transform is a function of both scale and translation. This allows for much greater time-frequency resolution than the

Fourier transform, which cannot localize in time. The discrete wavelet transform uses wavelets at non-overlapping locations and a finite number of scales, each half as detailed as the last.

There are many wavelets used in current work. One of the most common families of wavelets is the Daubechies family, named after Ingrid Daubechies who has done extensive work in this field [Daub90] [Daub92]. This paper uses the Daubechies 4-tap wavelet.

2.2 Wavelets for Image Compression

The Fourier transform of an image stored as an array of pixel values consists of taking the Fourier transform of each row of data, and then of each column. This can be easily extended to wavelet transforms. The two-dimensional wavelet transform has three associated variables: translation in the x -direction, translation in the y -direction, and scale. The transform is generated by taking the wavelet transform of the rows and of the columns. The order in which the transforms is performed is irrelevant.

The wavelet transform of a digital image makes use of multiresolution analysis to represent the three-dimensional wavelet transform in two dimensions. Multiresolution analysis merely implies working with the data at different scales. There are a finite number of scales at which to view sampled data such as an image. At each progressively coarser resolution, we take a wavelet transform using a wavelet of one scale.

The wavelet transform splits the image into high-frequency (H) and low-frequency (L) components at each scale. The H components at each scale are retained, and the L components are filtered again at the next scale. The splitting is done in both directions, partitioning the image into four sections, as shown in Fig. 1. Here, LL is the section consisting of the low frequency components in both directions. HL is the horizontal high frequency components, LH is the vertical high frequency components, and HH is the high frequency components in both directions.

The process continues, splitting the LL subsection of the image into four smaller subsections in the same way, until the LL subimage is as many pixels wide as the number of taps for the wavelet. The result is a multiresolution pyramid structure containing information about the image at each scale, as in Fig. 1 (d). This is an exact representation of the image data—the inverse transform reproduces the original image.

The compression associated with wavelet transforms comes when we see that many of the wavelet coefficients in these

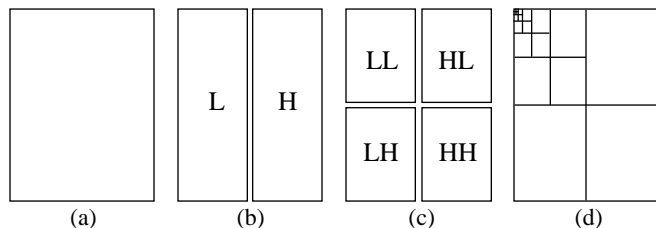


Fig. 1. Multiresolution signal decomposition. a) original. b) after rows are transformed. c) after rows and columns are transformed. d) after entire image is transformed.

decompositions are small. If we remove some of these small coefficients, we reduce the storage size of the image. The human visual perceptual system cannot detect these small errors in the image, and because we are designing a system for the purposes of face recognition, we can remove a great deal of information from the reconstruction and still have an image that is suitable for recognition.

2.3 Zerotree Theory

Zerotree coding is a technique developed by Jerome Shapiro which provides a compact representation of the wavelet coefficients of an image [Shap93]. The theory begins by recognizing that the values and locations of coefficients in different levels of a multiresolution wavelet coefficient representation are similar. Because of this, we can recognize that a zero-valued coefficient in one level often corresponds to four zero-valued coefficients in a lower level. This is shown in Fig. 2. These trees of zero coefficients can be coded by merely coding the root coefficient of the tree. This provides the compression—none of the coefficients below the root need to be coded.

Shapiro considers a coding scheme whereby each coefficient that is not a member of a zerotree is coded by one of 4 symbols. These are POS, NEG, ZTR, and ISZ, corresponding to positive coefficients, negative coefficients, zerotree roots, and isolated zeros, respectively. Zerotree coding is lossless.

Zerotree symbols are assigned to wavelet coefficients as the coefficient file is scanned recursively from the smallest sections to the largest sections. If a coefficient is zero, then corresponding coefficients in lower resolutions are checked. If all of these “children” are zero, then the coefficient is coded as a ZTR. If even one of the children is not zero, the coefficient is coded as an ISZ. The absolute values of the POS and NEG coefficients are coded in a list after the zerotree symbol list.

2.4 Arithmetic Coding Theory

Arithmetic coding is a method of lossless data compression which uses the frequency of data symbols to encode the data. The arithmetic coder used in this system will be coding the zerotree symbols produced by the zerotree coder described above.

The concept behind arithmetic coding is to represent a message as an interval of real numbers between 0 and 1 [Kins91]. Each symbol in the data string is mapped to a range which represents its relative probability of occurrence in the message. As the message is encoded, each symbol reduces the real number range. Once all the symbols in the message have

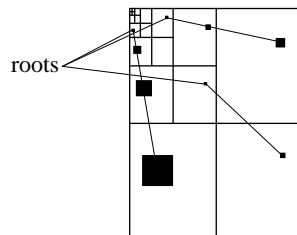


Fig. 2. Zerotrees in a wavelet coefficient representation.

been used, the message is completely represented by the two real numbers that mark the final range.

2.5 The Images

The techniques described above can be used to compress any general image. We are concerned with the use of wavelets to compress head-and-shoulder images, frequently used as identification on a driver's license or passport. Certain constraints to are necessary make the image set homogeneous and suitable for identification purposes:

- The image should contain the subject's face and upper portion of torso.
- The subject should be facing forward, and the head should be relatively level and centered in the image.
- The background should be non-textured, preferably white.
- If the person uses eyeglasses for normal vision, they should be worn. No other artificial elements should appear on the face (e.g. hats).

Constraints on the set of images used to verify the system must also be considered:

- Each image in the set should have the same type of background, the same relative position and orientation of the subject, and should be the same size and shape.
- The subject base should be as diverse as possible.
- The size of the images should facilitate processing—they should contain sufficient detail, but not be unwieldy. The pixel resolution of the images should permit multiple bisections in order to perform the wavelet transform.

A search of the internet was performed, and many image sets were found. However, most of these were produced for the purpose of face recognition, and were therefore intentionally not homogeneous.

It was decided then that we would create a database of face images suitable to our use. A group of volunteers had their passport photographs taken, and these photographs were scanned to 512 x 640 pixels, at a resolution of 300 dpi. The image set consists of 12 subjects, representative of characteristics such as age, gender and race.

3. IMPLEMENTATION CONSIDERATIONS

3.1 The System

The compression system operates by thresholding to zero small coefficients in the wavelet representation of the pre-processed image, and then zerotree coding the thresholded wavelet coefficients and arithmetic coding the zerotree code. The number of coefficients thresholded is selected by percentage.

3.2 Pre-Processing

A smoother image will compress better because more coefficients will be zero after thresholding, and thus larger zerotrees will be found. We smooth portions of the image that are unimportant to the identification of the subject, such as the background. The shoulder line and the face contain the

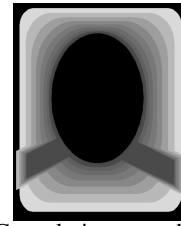


Fig. 3. Convolution smoothing mask.

important information, and so these sections would not be smoothed as much.

The smoothing is performed using an averaging filter of different sizes at different locations in the image, according to a pre-determined mask. The mask used is shown in Fig. 3.

A normal smoothing filter leaves the edges of the image unchanged. The wavelet transform picks up the boundary between the smoothed and unsmoothed areas, as well as the higher frequency information in the unsmoothed areas. This is undesirable. Since the recognition information in our images is concentrated toward the center of the image, the edges can be altered. We replace the edges of the image with pixels of value 255, corresponding to the white background of the image. The filter then gradually smooths from the white of the edge to the image data.

3.3 Post-processing

At high compression, the discrete wavelet transform has the tendency to introduce salt and pepper noise into the reconstructed image. This noise increases as more wavelet coefficients are removed from the image.

The salt and pepper noise is removed by finding regions bounded by high differentials, and filling these regions with a constant level which is an average of surrounding pixels.

The high differentials indicate the noise and not the image data because the result of a high compression discrete wavelet transform is an image with little high frequency information.

4. EXPERIMENTAL RESULTS

4.1 Experiments

The pre-processed images were compressed using the JPEG standard and the wavelet system. The JPEG compressor used contains a parameter called JPEG quality factor, with an integer range from 1 to 100. The wavelet system used a parameter corresponding to the percentage of coefficients thresholded to zero. This parameter has a real number range from 0 to 100.

Each pre-processed image was compressed with the JPEG compression standard at 7 different settings of the quality parameter: 1, 5, 10, 20, 50, 75, and 90. The same images were compressed using the wavelet system at 8 different levels: 50%, 95%, 98%, 99%, 99.5%, 99.7%, and 99.9%. Both of these file sets were further compressed with the arithmetic coder. The sizes of these files were recorded, and the quality was compared. The quality of the wavelet compressed images was

calculated before and after post processing, to show the effect of the post-processor.

The goal was to compare the wavelet transform compression and the JPEG compression, all other parameters being equal. This is why we use the pre-processed images, and compress the result with the arithmetic coder.

4.2 Measures

The compression is measured using bit rate, in bits per pixel (bpp), and is calculated by dividing the number of bits used to store the image (file size $\times 8$) by the number of pixels in the image, $512 \times 640 = 327,680$. An uncompressed image has a compression of 8 bpp. The compression ratio is found by dividing the original file size, including header (327,736 bytes) by the compressed file size. An uncompressed image has a compression ratio of 1:1.

The quality of the image is assessed using peak signal to noise ratio (PSNR) which is calculated as:

$$PSNR = 10 \log \frac{MAX^2}{\frac{1}{w \times h} \sum_{i=1}^w \sum_{j=1}^h (o(i, j) - c(i, j))^2} \quad (2)$$

where w and h are the width and height of the image respectively, o is the original image data, and c is the compressed image data. MAX is the maximum value that a pixel can have, 255. A PSNR of 30 dB or greater is considered to be acceptable for subject recognition.

PSNR is a measure of mathematical similarity between the original image and the compressed image. It is not a direct measure of the perceptual visual quality, I. e. the way the image looks to the human eye. However, PSNR can be used as an indicator of the image quality.

4.3 Results

Fig. 4 shows the results of compressing the image “al” using the JPEG compression standard and the wavelet transform developed in this paper. Presented are both the result of the wavelet transform alone, and the result of the wavelet transform with post-processing to remove salt and pepper noise. No such post-processing was used for the JPEG compressed images because the salt and pepper noise is a result of the wavelet transform.

Comparable results were achieved with all the images, for JPEG and the wavelet system after post-processing. Significant differences were observed in the results of the wavelet system before post-processing. Three classes of images were recognized: images with results similar to those of the image “al”, where noise effects are seen only after about 40:1 compression ratio; images with results similar to the image “jt”, where noise effects produce a consistent degradation in image quality over the entire compression range, and images with results similar to the image “ej”, where noise effects are more pronounced at lower compression ratios and at very high compression ratios. The results for the images “jt” and “ej” are presented in Figs. 5 and 6, respectively. The investigation of the cause of these phenomena is beyond the scope of this paper.

4.5 Discussion of Results

The charts of the relationship between compression ratio and PSNR show that at lower compression ratios (near 20:1), the JPEG compression system produced images of higher quality than the wavelet compression system, for the same compression level. However as the compression ratio was increased, the PSNR of the JPEG compressed images degraded more rapidly than that of the wavelet compressed images. At higher compression ratios (near 100:1), the wavelet system produced images of higher quality.

A PSNR of near 33dB, which is considered acceptable for subject recognition, is achieved by the JPEG system at a compression of 0.092bpp. A comparable PSNR is achieved

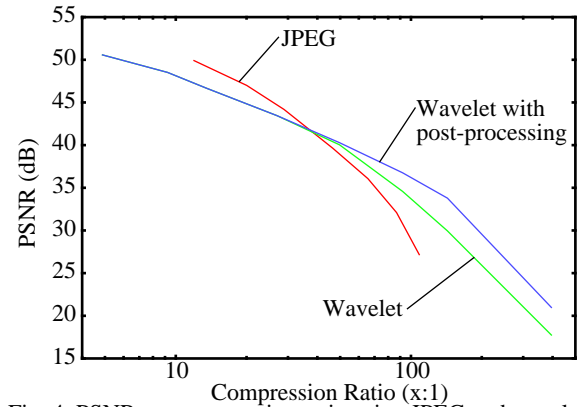


Fig. 4. PSNR vs. compression ratio using JPEG and wavelet compression for the image “al”.

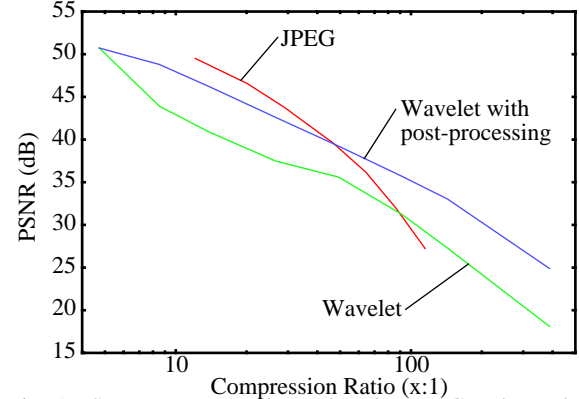


Fig. 5. PSNR vs. compression ratio using JPEG and wavelet compression for the image “jt”.

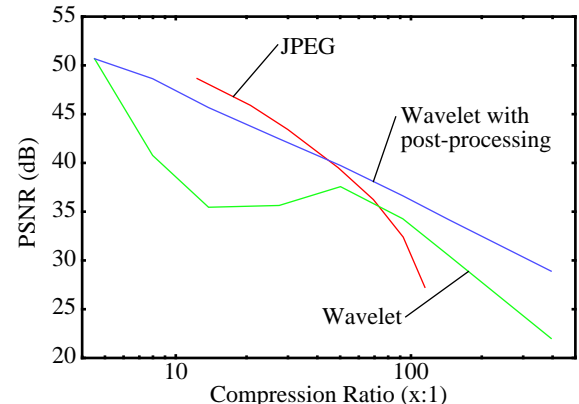


Fig. 6. PSNR vs. compression ratio using JPEG and wavelet compression for the image “ej”.

by the wavelet system at only 0.056bpp. A comparable compression (to 0.088bpp) using the wavelet system has a considerably higher PSNR of 36.76dB.

Apart from PSNR, the visual quality of the images must also be considered. Because the original intent of the system is to compress identification photographs, we can judge visual quality by how readily the images can be used to identify the subject. Fig. 7 shows comparisons of the image “al” compressed using the JPEG system, and using the wavelet system at comparable quality, and comparable compression.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This paper developed a system based on the wavelet transform for compression of head and shoulder images for subject recognition. The general wavelet transform system performed better than the JPEG compression system in terms of compression and resulting image quality, at higher compression ratios. A pre-processing module was developed to remove information unnecessary for identification of the subject. A post-processing module was developed to remove salt and pepper noise, which is an artifact of the wavelet transform at high compression levels.

The head and shoulder images in the set were consistently compressed to near 0.055 bits per pixel at a peak signal to noise ratio of near 33 decibels using this wavelet compression system. JPEG compression of the same images resulted in PSNR of near 27 dB for compression of near 0.070bpp. At comparable compression levels, the wavelet system produced images of higher PSNR, and at comparable image quality, the wavelet system achieved lower bit rates.

5.2 Recommendations for Further Work

There are many areas touched on by this paper that could be expanded to future research work. The wavelet transform used in the system developed here employed the Daubechies 4-coefficient wavelet as a basis. Other wavelets could be used in the system. The relative efficiency of these different wavelets

to compress facial images could be investigated.

The wavelet transform was compared to the JPEG compression standard only. Further comparisons could be made to other image compression techniques, such as neural net codebook compression, or compression schemes based on techniques such as fuzzy logic or genetic algorithms.

It was seen that there are three categories of noise effects in the image set. These categories could be investigated further—What features of the images do these categories represent? Are there more than three categories? This seems to be a problem that would be suited to neural networks.

Acknowledgments

Partial support from the Natural Sciences and Engineering Research Council of Canada is acknowledged.

References

- [CoLR94] T. Cormen, C. Leiserson and R. Rivest. *Introduction to Algorithms*. Toronto: McGraw-Hill, 1994.
- [Daub90] I. Daubechies; “The wavelet transform, time-frequency localization and signal analysis,” *IEEE Trans. Information Theory*; Vol. 36, No. 5, pp. 961-1005, Sep. 1990.
- [Daub92] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.
- [GoWo92] R. Gonzales and R. Woods. *Digital Image Processing*. Don Mills: Addison Wesley, 1992.
- [Gree95] M. Green. *Perceptual Image Compression with Wavelet Transform*. Computer Engineering Undergraduate Thesis, University of Manitoba; 1995.
- [Kins91] W. Kinsner. *Review of Compression Methods*. Technical Report DEL91-1, University of Manitoba, 1991.
- [Lync85] T. Lynch. *Data Compression, Techniques and Applications*. Belmont: Lifetime Learning Publications, 1985.
- [Papo80] A. Papoulis. *Circuits and Systems*. New York: Holt, Rinehart and Winston, 1980.
- [PTVF92] W. Press, S. Teukolsky, W. Vetterling and B. Flannery. *Numerical Recipes in C*. New York: Cambridge University Press, 1992.
- [Shap93] J. Shapiro. “Embedded image coding using zerotrees of wavelet coefficients,” *IEEE Trans. Signal Processing*. Vol. 41, No. 12, pp. 3445-3462, Dec. 1993.



Fig. 7. a) original image “al” after pre-processing. b) “al” compressed to 0.092bpp using JPEG. c) “al” compressed to 0.087bpp using wavelet compression. d) “al” compressed to 0.056bpp using wavelet compression.