THE WAVELET TRANSFORM AND THE SUPPRESSION THEORY OF BINOCULAR VISION FOR STEREO IMAGE COMPRESSION

William D. Reynolds, Jr.[*†]

[*]Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, IL 60439
E-mail: wreynold@eecs.uic.edu

Robert V. Kenyon[†]

[†]Department of EECS
University of Illinois at Chicago
Chicago, IL 60607
E-mail: kenyon@eecs.uic.edu

ABSTRACT

In this paper we present a method for compression of stereo images. The proposed scheme is a frequency domain approach based on the suppression theory of binocular vision. By using the information in the frequency domain, complex disparity estimation techniques can be avoided. The wavelet transform is used to obtain a multiresolution analysis of the stereo pair by which the subbands convey the necessary frequency domain information.

1. INTRODUCTION

With the increasing interest in three-dimensional displays, television systems, and virtual environments, the processing of stereo images and image sequences is becoming more important. Since stereo images require twice the bandwidth of conventional monoscopic images, compression of stereo images is one of the foremost issues that must be addressed to enable the widespread use of three-dimensional systems.

We present a method for stereoscopic image compression based on the *suppression theory* of binocular vision [1]. This approach is based on the frequency domain relationship between stereo images and not on the spatial domain (disparity) relationship. With the proposed technique, compression can be achieved without implementing disparity estimation techniques, and thus the complexity of the coding scheme can be reduced.

The suppression theory leads to the following compression scheme: one image of the stereo pair retains the details of the scene, while the second image retains the disparity information. Hence, the second image can be highly compressed without affecting the depth information in the compressed stereo images [2]. This compression scheme has been used in several different methods. Serthuraman and et al. [3] use a multiresolution approach, and Rabany and et al. [4] use a disparity compensation technique. Both algorithms compute the disparity between the stereo images. However, Dinstein and et al. [2] and Perkins [5] provided methods where the disparity is not computed. These are the methods we consider.

In these approaches, discrete cosine transform (DCT)-based coding was used on one image (right). For the second image, a low-pass filtered and subsampled version of the second image (left) was obtained; for example, by using the Gaussian pyramid [2]. The DCT-based coding was then applied to the low-resolution version of second image. At the decoder side, the right image was decoded with the DCT method, and the left image was decoded with the DCT method and then upsampled to the original resolution. For both approaches only low-frequency information was used to represent the disparity information.

In each method, a two-step process was used to obtained a low-resolution image and compress the low resolution-image. We propose to use characteristics of binocular suppression with a wavelet transform approach. This approach offers several advantages. First, it combines the two-steps of the compression scheme into one. Second, it provides a more efficient decomposition of the image spectrum. That is, the inherent multiresolution property of the wavelet transform provides a means for obtaining a low-resolution image while simultaneously providing detailed images at different spatial resolutions. This property allows for the inclusion of high-frequency information in the second image. However, the penalty for adding this information is an increase in the bit rate.

2. BINOCULAR VISION

2.1. Suppression Theory

When the two eyes are presented with two similar images, the result is a single percept of the scene. This phenomenon is known as *binocular fusion* [1]. On the other hand, when the two eyes are presented with two distinctly different images, instead of being combined the two images will be in competition with each other. This is known as *binocular rivalry* [1]. The final percept is an unstable shifting between the patterns of each eye. Because of this unstable combination of the two images, certain sections of one image will dominate certain sections of the other image. This alternation of patterns results from a shift in dominance between each eye. Information in the less dominant image will be *suppressed*, while the information in the dominant image will be visible.

Julesz [6] observed this phenomenon while working with random-dot stereograms and reported that binocular fusion depends on the spatial frequency relationship between the two images. For fusion to occur, either the low- or high-frequency (or both) components must be identical. If the

[*]This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Computational and Technology Research, U.S. Department of Energy, under Contract W-31-109-Eng-38.

DISCLAIMER

low- or high-frequency components differ, binocular rivalry will occur, and the image that contains the high-frequency information in a particular section will dominate the final perception [6]. Therefore, we will have suppression of different frequency bands.

Recall that the methods described in the preceding section used a low-resolution version of the second image. In the frequency domain this amounts to splitting the image spectrum in half and discarding the high-frequency components. Since the other image in the stereo pair contains the high-frequency information in areas where the second images does not, it will dominate the view. As the resolution of the second image is decreased (i.e., increasing levels of the Gaussian pyramid), the interval in the frequency domain where suppression occurs is increased. Hence, there is a reduction in image bandwidth, and a corresponding expansion in the interval of suppression.

By adding high-frequency information back into the low-resolution image, the interval of suppression can be decreased. This not only will enhance the depth perception but also will improve the quality of the second image. Previous methods used to obtain the low-resolution image effectively destroyed all the high-frequency information [5]. The wavelet transform avoids this problem by working on the stereo pair.

2.2. Zone of Suppression

In the next section we provide a method for adding high-frequency information to the low-resolution image. The method is based on the experimental results of Liu and Schor [7]. Liu and Schor measured the suppression zone as a function of spatial frequency and orientation (i.e., horizontal and vertical).

The results of their experiments indicate that both the horizontal and vertical dimensions of the suppression zone decrease with increasing spatial frequency. Furthermore, at low spatial frequencies, the vertical dimension was larger than the horizontal dimension. The shape of the corresponding suppression zone was elliptical, with the major axis in the vertical direction. However, at high spatial frequencies, the opposite occurred. The horizontal extent of the suppression zone was larger than the vertical extent and was also elliptical, with the major axis in the horizontal direction. For intermediate spatial frequencies, the shape of the suppression zone was shown to be circular [7].

3. PROPOSED METHOD

A block diagram of our method is shown in Figure 1. The wavelet transform is applied to each image of the stereo pair independently. One image is selected as the reference image, and the other is selected as the secondary image. The reference image represents the detailed image, and the secondary image represents the lower-resolution image. For our purposes, the right image was chosen as the reference image and the left image as the secondary image.

A bit allocation procedure is applied to the reference image to determine the bit rate for each subband. Based on this allocation of bits and properties of binocular vision, the subbands for the secondary image are selected. The subbands from both images are then quantized and entropy encoded, resulting in the compressed stereo pair. At the decoder end, the bit allocation procedure is repeated, and

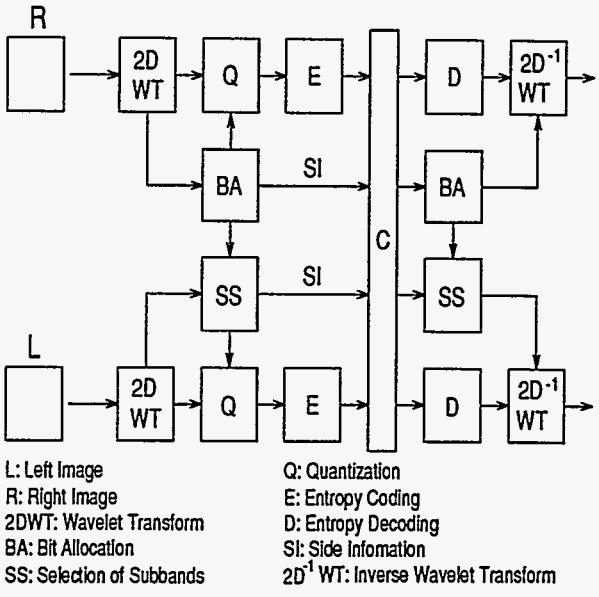

Figure 1: Block diagram of compression algorithm.

the images are reconstructed by using the inverse wavelet transform. Observe that for the bit allocation procedure to be repeated at the decoder, the subband variances of the reference image need to be transmitted. Also, for the secondary image, the subbands that are selected need to be transmitted to the decoder as well. This information is sent as side information.

3.1. Wavelet Transform

The wavelet transform [8] is applied to the left and right images independently. The filtering operations are implemented assuming separability of the image rows and columns. The filters used in this research are 12×4-tap perfect reconstruction, linear phase filters [9]. In order to eliminate any border effects due to the finite extent of the images, symmetric extension of the image borders is used. The number of decomposition levels is three and is constant throughout. This results in ten subbands per image and twenty overall. The template for the wavelet decompositions is shown in Figure 2. Subbands l_1 and r_1 represent the low-frequency subbands, and the remaining subbands l_k and r_k, $k = 2, \ldots, 10$, represent the high-frequency subbands.

3.2. Bit Allocation

After the wavelet decomposition, the bit allocation procedure is applied only to the right image. The objective of the bit allocation is to assign bits to each of the subbands such that the overall distortion or quantization error is minimized for a fixed average bit rate. For the case where the quantization error in each subband can be modeled as $d_k = c_k^2 2^{-2B_k} \sigma_{r_k}^2$, the problem [8] is to

$$min\ D = \sum_{k=1}^{M} d_k \ \ subject\ to\ \ B = \sum_{k=1}^{M} B_k, \qquad (1)$$

where $\sigma_{r_k}^2$ is the variance of the kth subband, B_k is the number of bits allocated to the kth subband, c_k depends on the distribution of the kth subband, B is the average

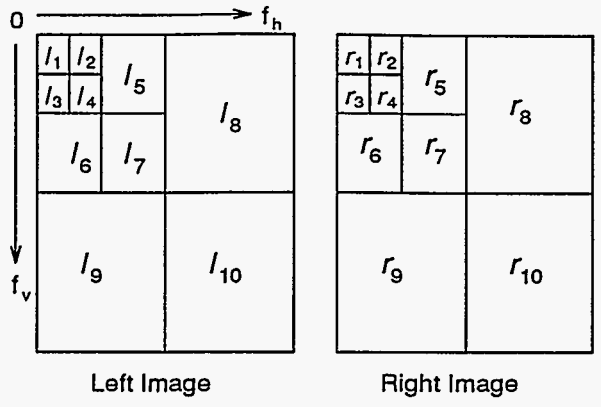

Figure 2: Template for wavelet decomposition.

bit rate, and M is the number of subbands. The solution to this problem is obtained by using the Lagrange multiplier method. For small subband variances, it can result in negative bit rate assignments.

To overcome negative bit rate assignments for B_k, the modified procedure in [10] is used. This procedure results in an iterative method such that the above allocation problem can be rewritten as

$$min\ D = \frac{1}{A}\sum_{k=1}^{M} A_k d_k \ \ subject\ to\ \ B = \frac{1}{A}\sum_{k=1}^{M} A_k B_k, \quad (2)$$

where A is the area of the original image and A_k is the area of the kth subband. Now, we have the additional constraint that B_k is nonnegative [10]. The solution is given by

$$B_k = \frac{A}{A_{s,M}} B + \frac{1}{2}\log_2 \left[\frac{\sigma_{r_k}^2}{\left(\prod_{k=1}^{M} (\sigma_{r_k}^2)^{A_k}\right)^{\frac{1}{A_{s,M}}}} \right], \quad (3)$$

where $A_{s,M} = \sum_{k=1}^{M} A_k$. For a given fixed bit rate, this method converges in a few iterations.

3.3. Selection of Subbands

Section 2 described the spatial frequency relationship between the stereo images. We now use that information to indicate which subbands will be selected in the left image. Recall that at low spatial frequency, vertical edges (which corresponds to horizontal high frequency) are mostly being suppressed, whereas at high spatial frequency, the horizontal edges are mostly being suppressed (which corresponds to vertical high frequency). Therefore, from the left image, subbands $\{l_1, l_2, l_7, l_9\}$ are selected.

Subband l_1 is necessary to represent a coarse description of the disparity information. Since it represents the low-frequency information of the left image, it is also necessary to prevent binocular rivalry [6]. Vertical edges at low spatial frequencies are given by subband l_2. For intermediate frequencies, subband l_7 is chosen. For high spatial frequencies, horizontal edges are mainly being suppressed, so subband l_9 is chosen. The remaining subbands are not transmitted or are set to zero. The bit rate for the selected subbands is determined based on its counterpart in the right image. This maintains the precision of the reconstructed image, at least in the common frequency bands.

3.4. Quantization and Entropy Coding

Once the bit rates have been assigned, the subbands are quantized and entropy encoded. For the low-frequency subbands, differential pulse code modulation is used. The high-frequency subbands, on the other hand, are quantized by using pulse code modulation. The quantization coefficients are Lloyd-Max quantizers and have been generated based on the generalized Gaussian distribution (GGD) [11]. The GGD is given by

$$g(x) = a\exp\{-|bx|^{\alpha}\}, \quad (4)$$

where α is the shape parameter and the parameters a and b are given by

$$a = \frac{b\alpha}{2\Gamma\left(\frac{1}{\alpha}\right)} \quad \text{and} \quad b = \frac{1}{\sigma}\sqrt{\frac{\Gamma\left(\frac{3}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)}}. \quad (5)$$

The parameter σ denotes the standard deviation of the source and is assumed to be unity. The shape parameter, α, for each subband is estimated by using the χ^2-test statistic [12]. A test set consisting of five stereo images with three levels of decomposition is used for estimating α. The results show that for the low-frequency subbands $\alpha = 0.8$ and for the high-frequency subbands $\alpha = 0.6$.

The quantizer outputs are entropy coded for further reduction in bit rate. Huffman coding is applied to the quantizer outputs, which produces a variable length code [13]. The encoded output is then sent to the channel for either transmission or storage.

4. SIMULATIONS

The proposed compression scheme was tested on two different stereo pairs. The stereo images have a spatial resolution of 512 × 512 with 8 bits per pixel (bpp). The images were acquired from the Carnegie Mellon University's Vision and Autonomous Systems Center image database. The images were viewed stereoscopically on a SGI Onyx using CrystalEyes hardware by StereoGraphics.

For comparison of the amount of additional bandwidth or bit rate that is necessary for stereo, assume that the right image has been allocated a bit rate that is suitable for a specific channel. The fractional increase in bit rate, x, for stereo is defined as

$$x = \frac{b_l}{b_r}, \quad (6)$$

where b_r and b_l denote the bit rate (bpp) for the right and left image, respectively. The results of the simulations are shown in Table 1. Figure 3 shows the original and coded TableBlks stereo pair. The coded left image contains high-frequency information at a bit rate of 0.11 bits per pixel (bpp). From the above results, it can be seen that the left image without high-frequency information can be compressed at least 100:1. When high-frequency information is added, the compression ratio decreases to about 20:1. This demonstrates the cost associated with adding high-frequency information. However, perceptually the left image with the high-frequency information is subjectively better, and the depth perception is slightly enhanced.

5. CONCLUSIONS AND FUTURE RESEARCH

We have presented a method for stereo image compression based on the suppression theory of binocular vision. This method is a frequency domain approach that exploits the redundant information in the frequency domain relationship between the stereo images. In contrast to disparity estimation techniques, our method reduces the amount of information presented to the visual system necessary for the computation of depth. The results indicate that when both low- and high-frequency information is present in the low-resolution image, the subjective performance is better than the case when no high-frequency information is present. However, this increase in performance results in an increase in the bit rate for the stereo pair.

Directions for future research include extending this method to compression of stereo video sources and incorporating other perceptual properties into the compression scheme. For example, perceptual weighting in the bit allocation problem has shown to increase performance especially at low bit rates [10]. Also, other perceptual studies focus on providing a better quantitative analysis of the frequency domain relationship between the stereo images.

6. REFERENCES

[1] L. Kaufman, *Sight and Mind: An Introduction to Visual Perception*, Oxford University Press, New York, 1974.

[2] I. Dinstein, M. G. Kim, J. Tzelgov, and A. Henik, "Compression of Stereo Images and the Evaluation of Its Effects on 3-D Perception," *SPIE Applications of Digital Image Processing XII*, Vol. 1153, pp. 522–530, 1989.

[3] S. Sethuraman, M. W. Siegel, and A. G. Jordan, "A Multiresolution Framework for Stereoscopic Image Sequence Compression," *Proc. of ICIP*, 1994.

[4] J. Rabany, I. Dinstein, G. Guy, J. Tzelgov, and A. Henik, "On Stereo Image Coding," *Ninth International Conference on Pattern Recognition*, IEEE Computer Society, pp. 357-359, 1988.

[5] M. G. Perkins, "Data Compression of Stereopairs," *IEEE Trans. on Communications*, Vol. 40, No. 4, pp. 684–696, April 1992.

[6] B. Julesz, *Foundations of Cyclopean Perception*, The University of Chicago Press, Chicago, 1971.

[7] L. Liu and C. M. Schor, "The Spatial Properties of Binocular Suppression Zone," *Vision Research*, Vol. 34, No. 7, pp. 937–947, 1994.

[8] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice Hall, Englewood Cliffs, New Jersey, 1995.

[9] O. Egger and W. Li, "Subband Coding of Images Using Asymmetrical Filter Banks," *IEEE Trans. on Image Processing*, Vol. 4, No. 4, pp. 478–485, April 1995.

[10] M. Balakrishnan and W. A. Pearlman, "Hexagonal Subband Image Coding with Perceptual Weighting," *Optical Engineering*, Vol. 32, No. 7, pp. 1430–1437, July 1993.

Table 1: Simulations results. Here, b_l (bpp) and x represent the case without high-frequency information and b_l' (bpp) and x' represent the case with high-frequency information.

Stereo Pair	b_r	b_l	x	b_l'	x'
TableBlks	0.93	0.073	0.078	0.42	0.45
	0.29	0.073	0.25	0.11	0.38
Pentagon	0.99	0.073	0.073	0.46	0.46
	0.43	0.054	0.13	0.16	0.37

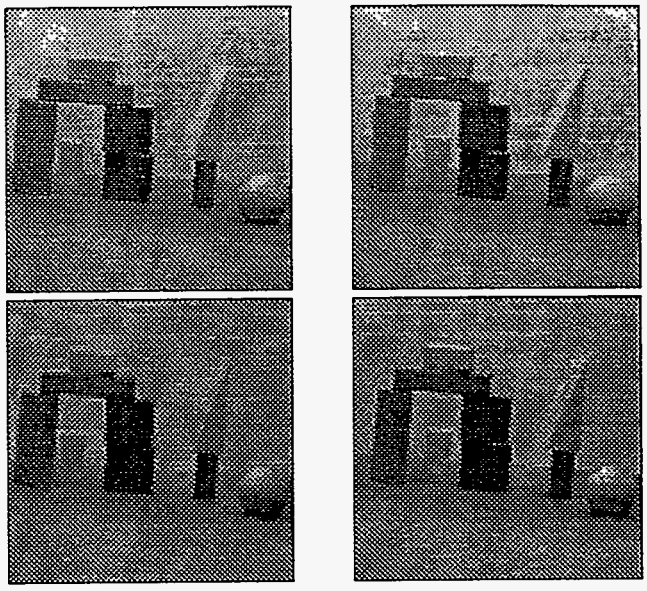

Figure 3: Top Left: Original left; Top Right: Original right; Bottom Left: Coded left at 0.11 bpp; Bottom Right: Coded right at 0.29 bpp.

[11] N. Farvardin and J. W. Modestino, "Optimum Quantization Performance for a Class of Non-Gaussian Memoryless Sources," *IEEE Trans. on Information Theory*, Vol. 30, No. 3, pp. 485–497, May 1984.

[12] W. J. Conover, *Practical Nonparametric Statistics*, 2nd ed., John Wiley & Sons, New York, 1980.

[13] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston, 1992.

7. ACKNOWLEDGMENT

The authors thank Professor Rashid Ansari for his fruitful discussions and for the source code used for generating the Huffman codes.