Optimized Progressive Coding of Stereo Images Using Discrete Wavelet Transform

Torsten Palfner, Alexander Mali and Erika Müller
Institute of Telecommunications and Information Technology, University of Rostock
Rostock, 18119, Germany

ABSTRACT

In this paper, a compression algorithm is introduced which allows the efficient storage and transmission of stereo images. The coder uses a block-based disparity estimation/ compensation technique to decorrelate the image pair. To code both images progressively, we have adapted the well-known SPIHT coder to stereo images. The results presented in this paper are better than any other results published so far.

KEY WORDS
Progressive Coding, Stereo Images, Discrete Wavelet Transform

1. INTRODUCTION

The advantage of stereo images over mono images is, that they make available all the information the human vision system needs. One image for each eye from two different perspectives. The human brain extracts from these two images the depth information. Current display devices ignore this fact, since they are only capable of displaying mono images. As soon as more sophisticated autostereoscopic devices become available, stereo images will become much more popular. Unfortunately, the amount of data, which is necessary to store these stereo images, is twice as much as for normal images. To avoid this, compression has to be used.

One of the first researchers who worked on this topic of stereo image compression was Perkins [1, 2]. Newer results come from Moellenhoff [3] und Woo [4, 5]. Since all these approaches use the old JPEG algorithm [6], which is based on the Discrete Cosine-Transform (DCT), block artifacts appear in the reconstructed image at low bit rates. These block artifacts are unwanted by the viewer, because they affect the subjective quality as well as the depth perception. Current approaches, which use non-linear filter banks [7] or wavelet filter banks [8, 9], have not got this drawback.

In this paper, we present a coding scheme which is also based on the Discrete Wavelet Transform (DWT). But in contrast to the algorithms presented in [7, 9], our coding algorithm exploits the redundancy of the stereo image much more efficiently. And on top of that, it allows progressive transmission.

2. ALGORITHM

The flow chart of our coder is shown in Figure 1. At first, the right image is estimated from the left image. This step is called disparity estimation/compensation (DS). After the substraction, the left image and the right error image are transformed. Finally, these images are coded.

Disparity Estimation and Compensation

Stereo images are highly correlated. As you can see for instance in Fig. 3, the left and the right image differ only slightly. To take advantage of this fact, the right image is divided into blocks of the same size and the disparity is estimated by using a simple full-search block matching algorithm, i. e. the algorithm checks all blocks in the search radius and returns the displacement vector (disparity vector) of the block which has the smallest error energy. The result is a field of disparity vectors as shown in Fig. 2, whereby the length of the disparity vector is inversely proportional to the distance of the object from the camera.

If the setup of the camera taking the stereo image fulfils the parallel axes geometry [2], the search space is reduced considerably. Instead of $M_\mathrm{S} \times N_\mathrm{S}$ comparison operations, where M_S is the number of rows and N_S the number of columns of the search square, only $N_\mathrm{S}/2$ comparison operations are necessary. All vectors point to one direction. Only the length of the disparity vectors is different. Since the calculation of the error energy is quite time consuming, the reduction of the search space speeds up the search algorithm significantly.

The disparity vector field is used to compensate the disparity of both images. The disparity compensated left image is subtracted from the right image. The result of this stage (see Fig. 1) is the left image, the error image and the disparity vector field.

The disparity compensated left image and the error image are shown in Fig. 4. Errors are visible at object boundaries as well as on edges.

DWT

The 2-dimensional Discrete Wavelet Transform (2d-DWT) is applied to the left image and the error image. This transform is preferred, because unlike the DCT there are no block artifacts for high quantization factors. The image is extended symmetrically before it is transformed. The

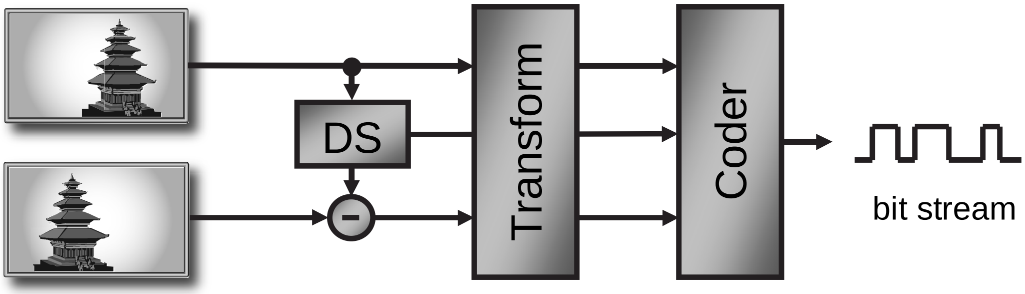


Figure 1. Flow chart of a common stereo coder.

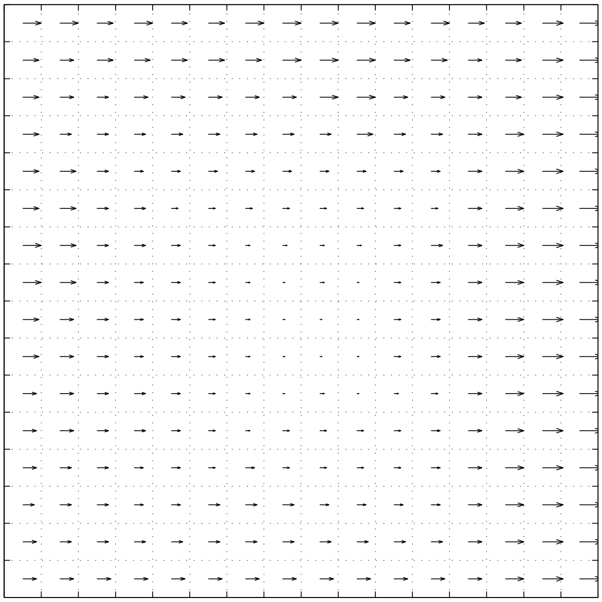


Figure 2. Disparity vector field for blocks of size 16×16.

transform uses the biorthogonal 9/7-wavelet filters implemented in the SPIHT coder [11] or the lossy mode of the new JPEG2000 standard[12].

Coding

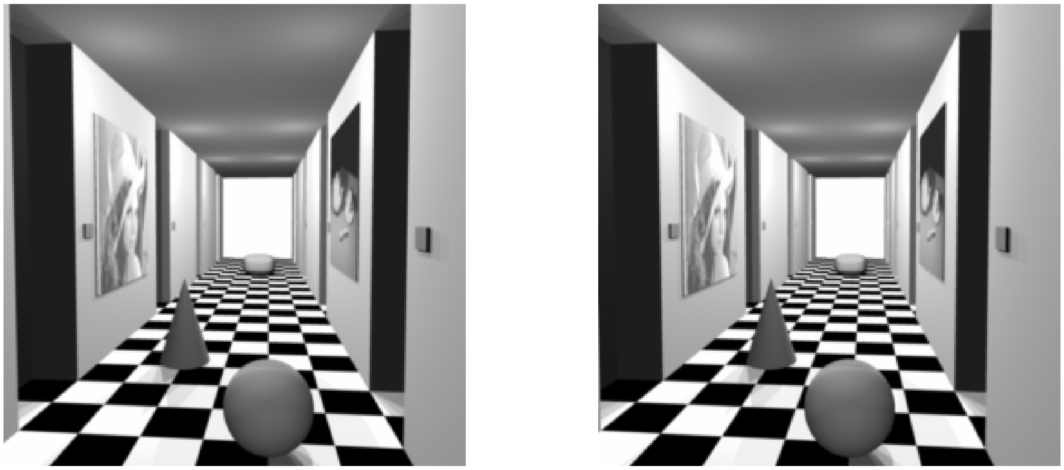
The disparity vector field has to be coded losslessly. This is done by a DPCM coder. The size of the compressed disparity vector field is especially critical at low bit rates. At high bit rates, it becomes less important.

The previous transform stage makes the job of the following coding stage much easier (see Fig. 1). The energy of the 2d-DWT-transformed images is concentrated mainly in the lowpass band and the upper subbands. Most of the coefficients are either zero or very small. To code both images, the SPIHT coder [11] was chosen, because it is much more efficient than the zerotree coder developed by Shapiro [13] and has a similar performance as the JPEG2000 coder [12]. Since the SPIHT coder works only for mono images, it has to be adapted to stereo images here. Modifications are necessary to get a good peak signal-to-noise ratio

$$PSNR = 10 \cdot \log_{10} \frac{255^2}{\left(D_{\mathrm{L}} + D_{\mathrm{R}}\right)/2} \quad \mathrm{dB}$$

for any bit rate, whereby D_{L} and D_{R} are the reconstruction error (mean square error) of the left image and the right image respectively.

The coder by Shapiro [13] was the first coder for 2d-DWT-transformed images which did not only achieve high compression ratios but was also quite simple. In his paper, he introduced the so-called zerotree which is a set of related zeros coded by one ZT-symbol (see Fig. 5.). As stated in [13], it is based on the hypothesis that if a wavelet coefficient is insignificant, i. e. its absolute value lies below a certain threshold T, then all wavelet coefficients of the same orientation in the same spatial location at finer scales are likely to be insignificant with respect to the threshold T. Since this is very likely for a 2d-DWT-transformed im-



Left Image

Right Image

Figure 3. Stereo image ,,Room", Size: 256×256.

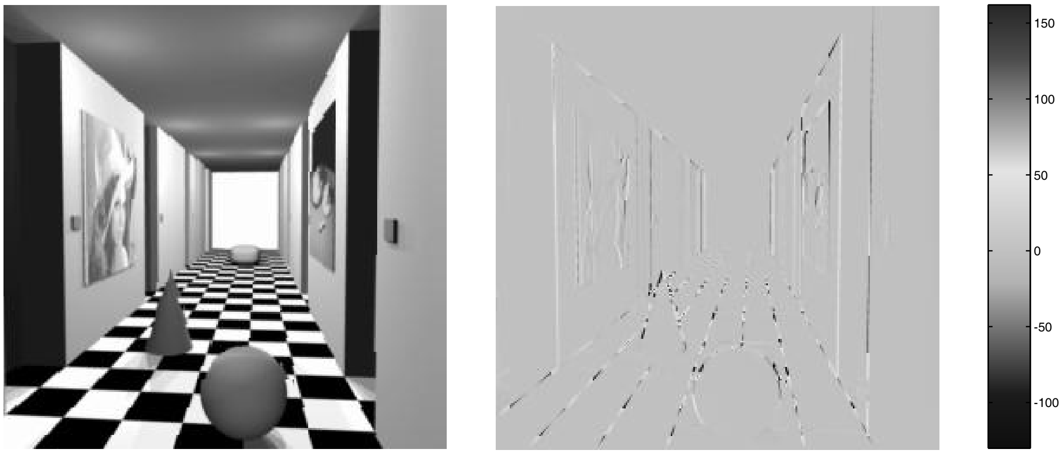
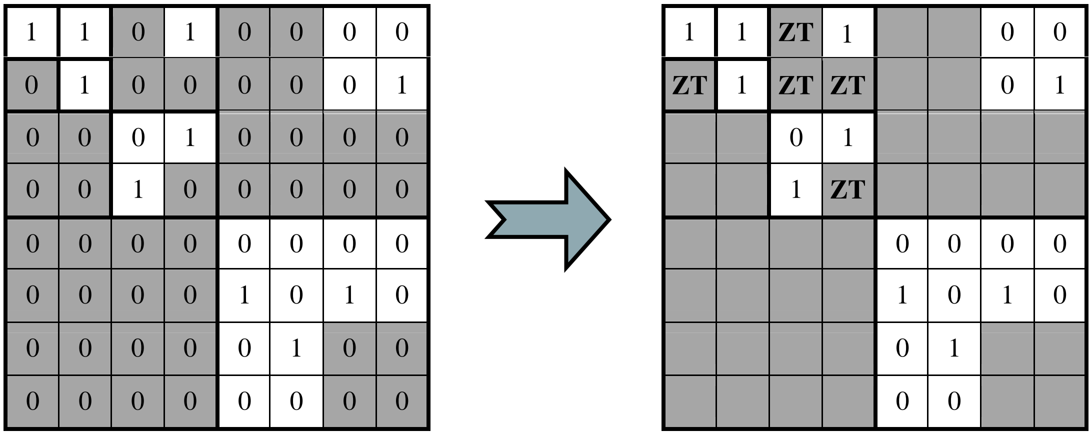


Figure 4. Disparity compensated image (left) and error image (right). Size of blocks: 8×8 pixels.



1	1	0	1	0	0	0	0
0	1	0	0	0	0	0	1
0	0	0	1	0	0	0	0
0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	1	0	1	0
0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0

1	1	**ZT**	1			0	0
ZT	1	**ZT**	**ZT**			0	1
		0	1				
		1	**ZT**				
				0	0	0	0
				1	0	1	0
				0	1		
				0	0		

Figure 5. Example of zerotree (ZT). Size of 2d-DWT-transformed image: 8×8 pixels, three decomposition levels.

age, the number of symbols which has to be coded is reduced dramatically as long as the threshold is high.

Progressive coding is achieved by reducing the threshold. This process is similar to describing a real number by a limited number of bits. For example the real number

$$3{,}33 \approx 3{,}3125 = 0\ 0\ 0\ \underset{\nearrow\ \text{significant bit}}{1}\ \underbrace{1\ 0\ 1\ 0\ 1\ \ .\ \ .}_{\text{refinement bits}}$$

The first threshold is $T_1 = 16$. The number becomes the first time significant for $T_4 = 2$. For threshold $T_5 = 1$ the reconstructed value is 3. By adding two more bits the reconstructed value is $3,25$. The reconstruction error becomes smaller with every bit.

Shapiro breaks up the bit plane into a dominant list and a subordinate list. While the dominant list contains the wavelet coefficients which have not been significant yet, the subordinate list contains the wavelet coefficients which have been significant already for previous thresholds. The dominant list is coded using four symbols (zerotree root, isolated zero, positive significant, negative significant). These symbols are sent to the arithmetic coder before the refinement symbols are coded by the arithmetic coder.

A further development of the zerotree coder is the SPIHT coder by Said and Pearlman [11]. This coder is also based on the zerotree. It gets even better results than Shapiro's coder without getting more complex. It does not use complex symbols any more, only binary numbers. So if speed is a concern, arithmetic coding can be switched off and only the binary symbols are stored. But yet the performance of the SPIHT coder is still better than Shapiro's coder. If arithmetic coding is used, the bits are coded by the arithmetic coder in groups of 2×2.

In contrast to Shapiro's algorithm, the tree roots start in the lowpass band.

The SPIHT coder looks simplistically after the modification as follows:

1. **Initialization:** Start at the highest bit plane, i. e. $n = \lfloor log_2(\max(|\mathbf{W}|)) \rfloor$, and initialize all lists.

2. **Sorting Pass:** Code the bits of the coefficients of the current bit plane n from the left image and the right image in alternating order, which have not been significant before this pass.

3. **Refinement Pass:** Code the bits of the coefficients of the current bit plane n from the left image and the right image in alternating order, which have been significant in previous passes.

4. **Quantization-step update:** decrement n by 1 and go to step 2.

where $\mathbf{W}$ is the 2d-DWT-transformed image. In contrast to the 3d-SPIHT proprosed in [10], this stereo image coder does not use the three-dimensional zerotree.

Due to the structure of the stereo image coder further modifications are necessary. On account of the subtraction operation (see Fig. 1), the bits of the left image reduce the distortion more than the bits of the right image. By shifting the bit planes of the right image down by one, we take advantage of this fact. This step is not only a simple step but also an efficient one. It can be easily implemented.

3. RESULTS

To check the performance of our new coder, we use the test stereo image „Room" of size 256×256 shown in Fig. 3.

Our results are shown for this stereo image in Fig. 6. If you look at this diagram, the overall improvement is highly visible. The FSC published in [9] denoted by „Boulgouris'01" is the only algorithm which comes close to the results of our coder at low bit rates. But at high bit rates they are all worse, not to mention the algorithm ESC [7] denoted by „Boulgouris'00" and the algorithm by Woo [4] denoted by „Woo'99". One disadvantage of the FSC algorithm is, that it is not suited for progressive transmission because it uses feedback. The target bit rate on the decoder side has to be known by the encoder. We would like to point out, that at high bit rates they are even worse than the independent coding using no disparity compensation.

The effect of our adaptation step is also well visible. Instead of a wavy curve (dotted curve), we get the solid curve after the modification.

4. CONCLUSIONS

In this paper, we have introduced an SPIHT coder for stereo images. To be able to apply this coder to stereo images, we had to adapt it. Modifications were necessary to get a good rate-distortion-ratio for any target bit rate. We showed that our coder's performance is much better than the best solution presented by other researchers so far.

To achieve higher compression ratios, methods like interpolation or overlapped blocks may be implemented. These improvements will be published in our next papers.

5. REFERENCES

[1] M. G. Perkins, *Data Compression of Stereopairs*, Ph.D. thesis, Standford University, Feb. 1988.

[2] Michael G. Perkins, "Data compression of stereopairs," *IEEE Transactions on Communications*, vol. 40, no. 4, pp. 684–696, April 1992.

[3] M. S. Moellenhoff and M. W. Maier, "Transform coding of stereo image residuals," *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 804–812, 1998.

[4] W. Woo and A. Ortega, "Optimal blockwise dependent quantization for stero image coding," *IEEE*

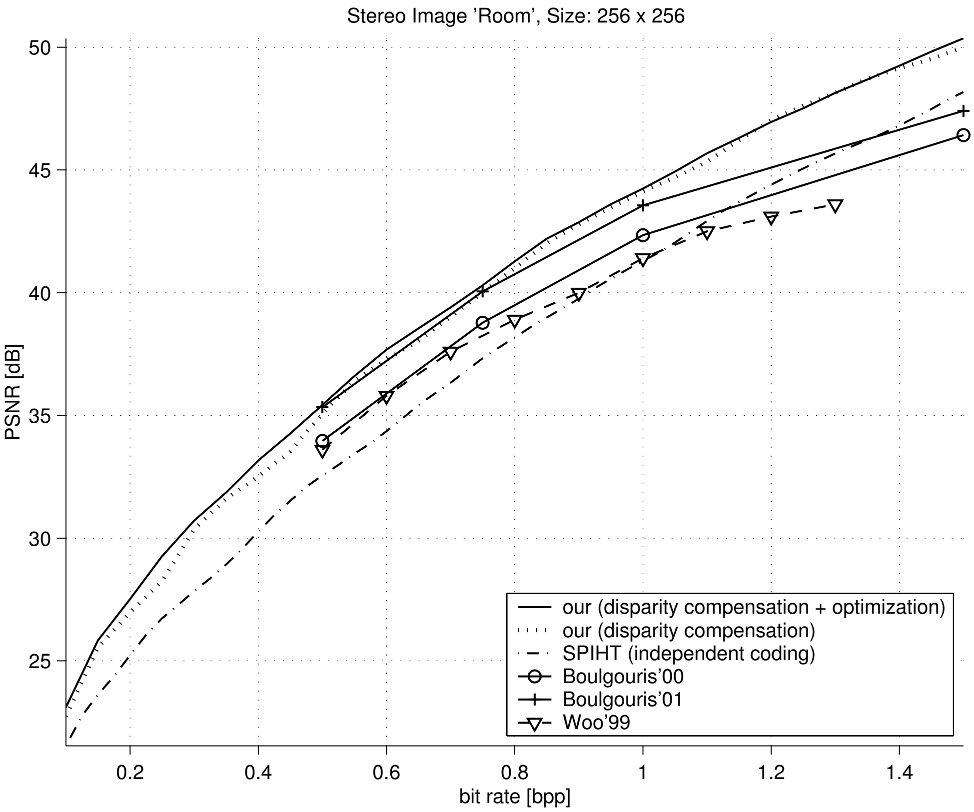


Figure 6. Bitrate-PSNR-diagram for stereo image „Room".

Transactions on Circuits and Systems for Video Technology, vol. 9, no. 6, pp. 861–867, 1999.

[5] W. Woo and A. Ortega, "Overlapped block disparity compensation with adaptive windows for stereo image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 194–200, March 2000.

[6] *Digital Compression and Coding of Continuous Tone Still Images, Part 1, Requirements and Guidelines, ISO/IEC IS 10918-1*, 1992.

[7] Nikolaos V. Boulgouris and Michael G. Strintzis, "Embedded coding of stereo images," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Sept. 2000, pp. 640–643.

[8] William D. Reynolds and Robert V. Kenyon, "The wavelet transform and the suppression theory of binocular vision for stereo image compression," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 1996, pp. 557–560.

[9] Nikolaos V. Boulgouris and Michael G. Strintzis, "Disparity-compensated wavelet coding of stereo images using feedback," *International Conference on Augmented, Virtual Environments and 3D Imaging (ICAV3D)*, pp. 77–80, 2001.

[10] B.-J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (spiht)," in *Proceedings of IEEE Data Compression Conference*, 1997, pp. 252–260.

[11] Amir Said and William A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, June 1996.

[12] "JPEG 2000 part 1, ISO/IEC 15444 part 1," December 2000.

[13] Jerome M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transaction on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.

[14] Woontack Woo, *Rate-Distortion Based Dependent Coding for Stereo Image/Video: Disparity Estimation and Dependent Bit Allocation*, Ph.D. thesis, University of Southern California, LA, CA, USA, 1998.

[15] M. G. Strintzis and S. Malssiotis, "Object-based coding of stereoscopic and 3d image sequences: A review," *IEEE Signal Processing Magazine*, vol. 16, no. 3, pp. 14–28, 1999.