RESIDUAL IMAGE CODING FOR STEREO IMAGE COMPRESSION

Tamás Frajka and Kenneth Zeger

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

ABSTRACT

The main focus of research in stereo image coding has been disparity estimation (DE), a technique used to reduce coding rate by taking advantage of the redundancy in a stereo image pair. Significantly less effort has been put into the coding of the residual image. In this paper we propose a new method for the coding of residual images that takes into account the properties of residual images. Particular attention is paid to the effects of occlusion and the correlation properties of residual images that result from block-based disparity estimation. The embedded, progressive nature of our coder allows one to stop decoding at any time. We demonstrate that it is possible to achieve good results with a computationally simple method.

1. INTRODUCTION

Human depth perception is in part based on the difference in the images the left and right eyes send to the brain. By presenting the appropriate image of a stereo pair to the left and right eyes, the viewer perceives scenes in 3 dimensions instead of as a 2-dimensional image. Such binocular visual information is useful in many fields, such as tele-presence style video conferencing, telemedicine, remote sensing, and computer vision.

These applications require the storage or transmission of the stereo pair. Since the images seen with the left and right eye differ only in small areas, techniques that try to exploit the dependency can yield better results than independent coding of the image pair.

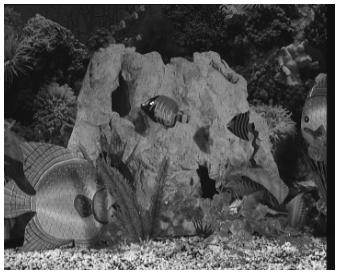
Most successful techniques rely on disparity compensation to achieve good performance. Disparity compensation is similar to the well known motion compensation for video compression. [1][2] [3][4] employ disparity compensation in the spatial domain, while [5] uses the wavelet domain. Disparity compensation can be a computationally complex process. In [6] a wavelet transform based method is used that does not rely on disparity compensation.

Many of the above works use discrete cosine transform (DCT) based coding of the images which uses a rate allocation method to divide the available bandwidth between the two images. Embedded coding techniques based on the wavelet transform [7] provide improved performance for still images when compared with DCT-based methods. A progressive stereo image coding scheme is proposed in [8] that achieves good performance without having to use rate allocation.

With disparity compensation, one image is used as a reference image, and the other is predicted using the reference image. The gain over independent coding comes from compressing the residual image that is obtained as the difference of the original and predicted image. Little attention has been paid to the coding of the residual image. Moellenhoff et al. [9] looked at the properties of disparity compensated residual images and proposed some DCT and wavelet techniques for their improved encoding.

Supported in part by the National Science Foundation and the UCSD Center for Wireless Communications.

a



b

Fig. 1. Original left images of the (a) *Room* (b) *Aqua* stereo pairs.

In this paper we propose a progressive coding technique for the compression of stereo images. The emphasis of our work is on the coding of the residual image. These images exhibit properties different from natural images. Our coding techniques make use of these differences. We propose to use transforms that take into account those differences as well as the block-based nature of disparity estimation. In our coder we treat occluded blocks differently from blocks that are well estimated with the DE process. Multi-Grid Embedding (MGE) by Lan and Tewfik [10] is used as the image coder for its flexibility. All these yield to some significant improvements over other methods in the coding of stereo images.

The outline of the paper is as follows. Section 2 gives an overview of stereo image coding. Our contribution is in Section 3 with experimental results provided in Section 4.

2. STEREO IMAGE CODING

Stereoscopic image pairs represent a view of the same scene from two slightly different positions. When the images are presented to the respective eye the human observer perceives the depth in the scene as in 3 dimensions. One can obtain stereo pairs by taking pictures with two cameras that are placed in parallel 2-3 inches apart. The left image of the stereo pairs used in this work can be seen in Figure 1.

Because of the different perspective, the same point in the object will be mapped to different coordinates in the left and right

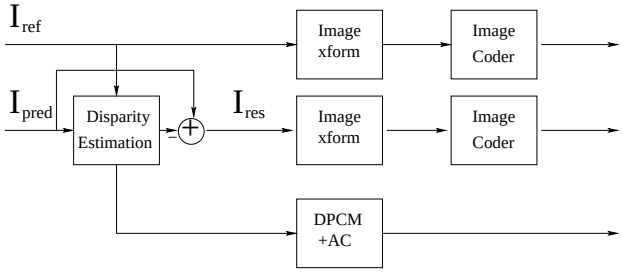


Fig. 2. Stereo image encoder.

images. Let (x_l, y_l) and (x_r, y_r) denote the coordinates of an object point in the left and right images, respectively. The *disparity* is the difference between these coordinates, $\mathbf{d} = (x_l - x_r, y_l - y_r)$. If the cameras are placed in parallel, $y_l - y_r = 0$, and the disparity is limited to the horizontal direction. One image of the pair serves as a reference image, I_{ref}, and the other I_{pred} is disparity estimated with respect to the reference image. A block diagram of the encoder is shown in Figure 2.

The disparity of each object in the image depends on the distance between the object and the camera lenses. (See [11] for more details). The disparity estimation process tries to determine the displacement for each image pixel. Since this process would be quite complex if done for each pixel individually, it is carried out for $k \times k$ blocks instead. Block sizes of $k = 8$ or $k = 16$ provide a good trade-off between accuracy of the estimation and the entropy necessary to encode the disparity vector, $\mathbf{d}$, for each block.

The search for the matching block is carried out in a limited search window. Given the reference image the optimal match could be any $k \times k$ block of this image. This exhaustive search is computationally complex. From the parallel camera axis assumption one can restrict the search to horizontal displacements only. From the camera set-up it is clear that the disparity for objects in the left image with respect to the right image is positive and vice versa. This observation helps further limit the scope of the search.

This estimation process works well for blocks that are present in both images. However, occlusion may result if certain image information is only present in one of the images. Occlusion can happen for two main reasons: finite viewing area and depth discontinuity. Finite viewing area occurs on the left side of the left image and the right side of the right image where the respective eye can see objects that the other eye cannot. Depth discontinuity is due to overlapping objects in the image; certain portions can be covered from one eye on which the other eye has direct sight.

The disparity vectors are usually losslessly transmitted using differential pulse code modulation (DPCM) followed by arithmetic coding [12].

Given the disparity estimate of the image, the residual I_{res} is formed by subtracting the estimate from the original. This residual and the reference image are then encoded. Many proposed techniques use DCT-based block coding methods for the encoding of both images. They also require a bit allocation mechanism to determine the coding rate of each image (this bit allocation is carried out in addition to the bit allocation between the DCT-transformed blocks of each image). For each target bit rate, a separate optimization is used to determine the appropriate bit allocation.

Embedded image coders can be terminated at any bitrate and still yield the best reconstruction to that rate without a priori optimization. Zerotree-style techniques such as Set Partitioning in Hierarchical Trees (SPIHT [7]) by Said and Pearlman offer excellent compression performance for still images. This zerotree technique is extended to stereo images [8] by Boulgouris et al. The bitplane coding is performed on both the residual and reference image at the same time guaranteeing that the most significant information for both images is sent before the less significant information.

The decoding of stereo images is rather simple. Both the residual and reference image are reconstructed. Using the DE information and the reconstructed reference image the decoder can recover the other image of the stereo pair.

3. RESIDUAL IMAGE CODING

The goal of our research was to make stereo image coding more efficient by improving the coding of the residual image. The DE we chose is rather simple, but even with such a simple disparity estimator our proposed coding technique has very good performance.

3.1. Image Coding Method

Embedded coding yields good performance coupled with simplicity of coding due to not having to perform any bit allocation procedure. MGE [10] uses a quadtree structure instead of the zerotrees of SPIHT. It employs the same bitplane coding, starting from the most significant bits of the transform domain image down to the least significant. For each bitplane, the quadtree structure is used to identify the significant coefficients; coefficients whose most significant bit is found on that bitplane. The sorting pass identifies the coefficients that become significant on the current bitplane, while the refinement pass refines those coefficients that have previously become significant.

The way we use MGE for stereo image compression is similar to [8]. For each bitplane first the sorting and refinement pass are executed for the reference image and then for the residual image. The highest magnitude coefficient is usually smaller for the residual image than for the reference image. The residual image contains mostly high frequency information. MGE was chosen over SPIHT because it can encode the above scenarios more efficiently.

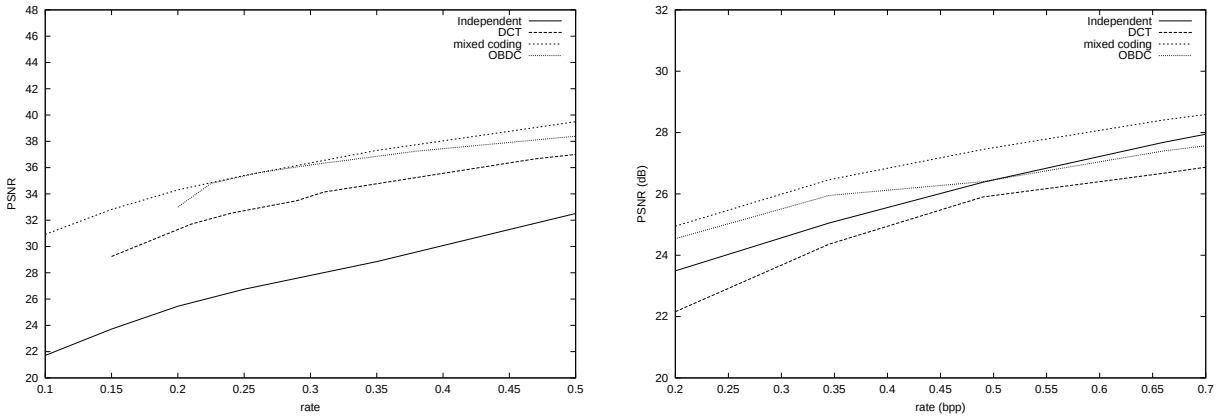
3.2. Occlusion

As noted in Section 2 there are two kinds of occlusion that may occur in DE. A finite viewing area can be overcome in certain cases. In the case of blocks near the image edge where a one directional search could run out of image pixels if we allow the search to continue in the other direction, it may find blocks similar to the one to be estimated.

The residual of those blocks that are occluded because of depth discontinuity displays different characteristics from the other parts of the image. As noted in [9], the occluded blocks are more correlated. We propose to detect such blocks, and code them differently from the rest of the residual image blocks for improved efficiency.

3.3. Image Transform

Moellenhoff's analysis [9] shows that residual images show significantly different characteristics from natural images. They mainly contain edges and other high frequency information. The correlation between neighboring pixels is smaller as well. This suggests that transforms that work well for natural images may not be as useful for residual images.



a b

Fig. 3. Comparison of independent coding, JPEG-style coding, OBDC, and mixed transform coding for the left image residual (with reference image JPEG-coded with quality factor 75) for (a) the Room, (b) the Aqua images.

	1	2	3	4	5	6	7	8
RO	0.93	0.94	0.96	0.96	0.97	0.95	0.94	0.94
RR	0.27	0.38	0.41	0.45	0.44	0.31	0.33	0.03
AO	0.90	0.89	0.89	0.89	0.88	0.88	0.87	0.89
AR	0.24	0.25	0.22	0.23	0.25	0.23	0.26	0.12

Table 1. 1-pixel correlation for columns of 8×8 blocks for RO=Room original, RR=Room residual, AO=Aqua original, AR=Aqua residual, horizontal direction, right images.

In wavelet transform coding one of the most widely used filters is the $9-7$ filter by Daubechies [13]. It is preferred for its regularity and smoothing properties. With the image pixels less correlated in residual images, we propose the use of Haar-filters. These 2-tap filters take the average and difference of two neighboring pixels.

DE uses $k \times k$ size blocks to find the best estimates for the image. There is no reason to expect neighboring blocks to exhibit similar residual properties. For one block the algorithm can find a relatively good match while its neighbor could be harder to predict from the reference image.

Moellenhoff's results only indicate that the pixels of the residual image are less correlated than that of the original image. These results do not reveal much about the local correlation of pixels, namely across the $k \times k$ block boundaries. We propose to look at 1-pixel correlation on a more local scale in both horizontal and vertical direction. Instead of gathering these statistics for the whole image, we only look at the correlation between all pixels in the n^{th} and its immediate neighbor in the $(n+1)^{th}$ column or row of all $k \times k$ blocks for the case of horizontal or vertical correlation, respectively. Note that the correlation between the k^{th} and $(k+1)^{th}$ columns/rows gives the correlation just across the boundary between two neighboring blocks. The 1-pixel correlations in the horizontal direction are shown in Table 1 for the Room and Aqua images. (Vertical correlations show similar trends.) For the original images the 1-pixel correlation is about the same for each position in a block, while for residual images it drops off significantly at the block boundary (column 8). This further supports our assumption that different blocks exhibit different properties in the residual image.

Based on this observation we focus on block-based transforms that could better capture the differences between the blocks than a global transform, such as the wavelet transform, that sweeps across the block boundaries. DCT in practice is performed on $k \times k$ blocks. Its performance is diminished by the JPEG encoding method. However, if the DCT coefficients are regrouped into a wavelet decomposition style subband structure as proposed in [14], and are encoded using an embedded coder, the performance approaches that of wavelet based methods (this method is referred to as Embedded Zerotree DCT (EZDCT)) .

None of the proposed image transforms so far take into account the effect of occlusion. For an occluded block the best match can still be a very distorted one. In those cases not using the estimate for the given block at all could be the best strategy. For each block the estimator should decide if the best match is good enough. If not, the given block is left intact. This process creates a mixed residual image, with some parts having mostly edges and high frequency information, and other parts blocks from the original image. For residual blocks that contain significant high frequency information a uniform band partitioning (such as with DCT) works better than octave-band signal decomposition (see [15]), while octave-band decomposition is desirable in blocks of the original image.

Note that the Haar transform only uses two neighboring pixels to compute the low and high frequency coefficients, then moves on to the next pair. If k, the block size is even, starting at the left edge of the block, the Haar transform can be performed without having to include pixels from outside the block for the computation of Haar-wavelet coefficients for all pixels in the block. Furthermore, this can be repeated up to $\lfloor \log_2 k \rfloor$ levels without affecting coefficients from outside the $k \times k$ block. We propose to use a mixed image transform. This transform consists of a Haar transform of 3 levels for occluded blocks and a DCT for others with the DCT coefficients regrouped into the wavelet subbands to line up with the Haar-transformed coefficients.

4. EXPERIMENTAL RESULTS

In our simulations we used the 256×256 Room and 288×360 Aqua stereo image pairs shown in Figure 1. The reference image was transformed using the $9-7$ filters. For DE a simple scheme

was used with a 64 pixel horizontal search window. Occlusion detection consisted of looking for blocks where the estimation error was above a given threshold.

For stereo images, the Peak Signal-to-Noise Ratio (PSNR) is computed using the average of the mean squared error of the reconstructed left and right images,

$PSNR = 10 \log_{10} \frac{255^2}{(MSE_l + MSE_r)/2}$.

First we show the comparison of different methods for the coding of the disparity estimated left image. The reference image is the uncoded right image. The bitrate figures include the coding of the disparity vector field. In the case of the mixed transform, for each block an extra bit is encoded to signal if that block is considered as occluded. (Clearly, in the case of independent coding there is no need to encode any disparity information.) The PSNR is computed using the MSE for the left image alone. The JPEG-style coder in our comparison uses quantization tables from the MPEG predicted frame coder.

Figures 3 compares independent wavelet coding, JPEG-style coding, overlapped block disparity compensation (OBDC) [4], and mixed transform coding. Mixed transform coding significantly outperforms both independent and JPEG-style coding with a gain of about 3 dB over the JPEG-style encoding. It also performs as well or better than OBDC coding which uses a computationally more complex disparity estimator.

Next we compare our proposed methods and the results from [8]. Good residual image performance alone does not guarantee overall good performance when the entire stereo image is concerned in an embedded coding scenario. Recall that the decoder uses the compressed reference image to recreate the estimate for the other image. If the coding of the residual image takes away bits from the coding of the reference image the overall result may not be as good as the coding of the residual image would suggest.

Figure 4 demonstrates this comparison. In this case the left image is chosen as the reference image. In the comparison, "Boulgouris2" refers to new results (received from the authors) obtained by an improved version of the original Embedded Stereo Coder. It uses a more sophisticated disparity estimator and better wavelet filters. Our proposed method outperforms this improved algorithm as well by $0.70 - 1$ dB.

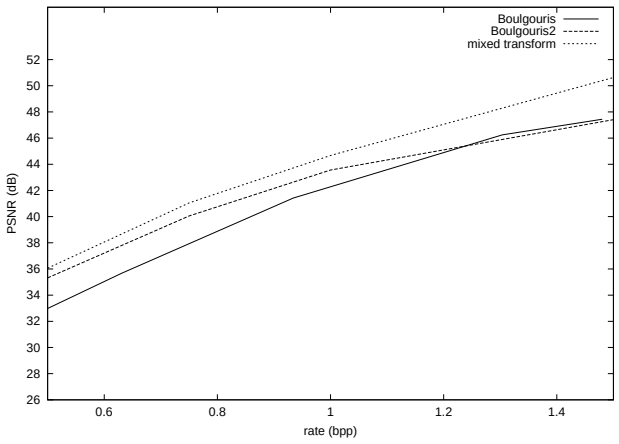


Fig. 4. Comparison of proposed method with the Embedded Stereo Coding scheme from Boulgouris and its improved version for the full stereo pair.

5. ACKNOWLEDGEMENT

We would like to thank Nikolaos Boulgouris for providing data for the performance comparison.

6. REFERENCES

[1] W.-H. Kim and S.-W. Ra, "Fast disparity estimation using geometric properties and selective sample decimation for stereoscopic image coding," *IEEE Trans. on Cons. Elec.*, vol. 45, no. 1, pp. 203–209, Feb. 1999.

[2] C.-W Lin, E.-Y. Fei, and Y.-C. Chen, "Hierarchical disparity estimation using spatial correlation," *IEEE Trans. on Cons. Elec.*, vol. 44, no. 3, pp. 630–637, Aug. 1998.

[3] D. Tzovaras and M. G. Strintzis, "Disparity estimation using rate-distortion theory for stereo image sequence coding," in *Int. Conf. on DSP*, July 1997, vol. 1, pp. 413–416.

[4] W. Woo and A. Ortega, "Overelapped block disparity compensation with adaptive windows for stereo image coding," *IEEE Trans. on Circ. and Sys. for Video Tech.*, vol. 10, no. 2, pp. 861–867, Mar. 2000.

[5] Q. Jiang, J. J. Lee, and III M. H. Hayes, "A wavelet based stereo image coding algorithm," in *ICASSP*, Mar. 1999, vol. 6, pp. 3157–3160.

[6] W. D. Reynolds and R. V. Kenyon, "The wavelet transform and the suppression theory of binocular vision for stereo image compression.," in *ICIP*, Sept. 1996, vol. 2, pp. 557–560.

[7] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circ. and Sys. for Video Tech.*, vol. 6, no. 3, pp. 243–250, June 1996.

[8] N. V. Boulgouris and M. G. Strintzis, "Embedded coding of stereo images," in *ICIP*, Sept. 2000, vol. 3, pp. 640–643.

[9] M. S. Moellenhoff and M. W. Maier, "Characteristics of disparity-compensated stereo image pair residuals," *Sig. Proc.:Image Comm.*, vol. 14, no. 1–2, pp. 55–69, Nov. 1998.

[10] T. Lan and A. H. Tewfik, "Multigrid embedding (MGE) image coding," in *ICIP*, Oct. 1999, vol. 3, pp. 369–373.

[11] H. Yamaguchi, Y. Tatehira, K. Akiyama, and Y. Kobayashi, "Stereoscopic images disparity for predictive coding," in *ICASSP*, May 1989, vol. 3, pp. 1976–1979.

[12] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Communications of the ACM*, vol. 30, no. 6, pp. 520–540, June 1987.

[13] M. Antonioni, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. on Image Proc.*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

[14] Z. Xiong, O. G. Guleryuz, and M. T. Orchard, "A DCT-based embedded image coder," *IEEE Signal Processing Letters*, vol. 3, no. 11, pp. 289–290, Nov. 1996.

[15] T. D. Tran and T. Q. Nguyen, "A progressive transmission image coder using linear phase uniform filterbanks as block transforms," *IEEE Trans. on Image Proc.*, vol. 8, no. 11, pp. 1493–1507, Nov. 1999.