Enhanced Predictive Block-based Encoding for Stereo Image Compression

Alexander A. Krupev[1] Antoaneta A. Popova[2] and Ivo R. Draganov[3]

Abstract – **The block-based predictive encoding exploits both inter and intra-frame correlation to compress stereoscopic image pairs, similarly to the MPEG standard. Two variations of the method are reviewed and tested in this paper – conventional disparity estimation technique and pioneering block-based encoding. The later does not require extra overhead bits for disparity estimation - it is done in the decoder using information from previously received and decoded blocks. Reported is an enhancement of the method using a bicubic resizing block for extra compression and speed.**

Keywords – **predictive encoding, stereo image compression, DCT, disparity estimation, disparity compensation**

I. Introduction

Stereo pairs consist of two similar images intended for each eye in order to achieve depth perception. This information redundancy is not well exploited by the conventional compression standards like JPEG, GIF and others [1], [3].

A common approach is to compress one image independently (called reference image), then to compress the second image using the knowledge about the first image. For this purpose we use predictive encoding by searching for similar 8x8 blocks between the two images and transmitting their residuals along with the reference image. The DCT transformed and quantized residual blocks are usually sparse matrices (if the stereo matching is performed correctly) and the few non-zero coefficients are compacted, which is a prerequisite for efficient RLE lossless coding.

The pioneering block-based version of the algorithm, first proposed by Jiang et al. [1], uses neighboring block search instead of direct matching in order to avoid sending disparity vectors to the decoder. As we shall see, this is possible because we assume that a standard stereo setup is used (parallel camera optical axes) - the corresponding pixels/blocks lie in the same row. Thus the disparity search range is 1-dimensional. To accelerate the matching block search and achieve extra compression, we added a resize module to the scheme. Human vision is able to compensate the lack of higher frequency information in one of the images if it is present in the other [4], [5]. This allows us to downsize one of the images in the process of encoding and upsize it to its initial resolution again before stereoscopic visualization.

Our goal is to suggest a predictive algorithm for stereo image compression and simulate its impact on stereo perception. The evaluation of the reconstructed compressed images is still an open problem. We employ an objective measure (PSNR), but it doesn't take into account the peculiarities of the human visual system [5]. Subjective experiments are also performed.

II. Predictive Block-based Encoding

The general scheme for block-based predictive encoding with a bicubic resize module, proposed by us, is given below:

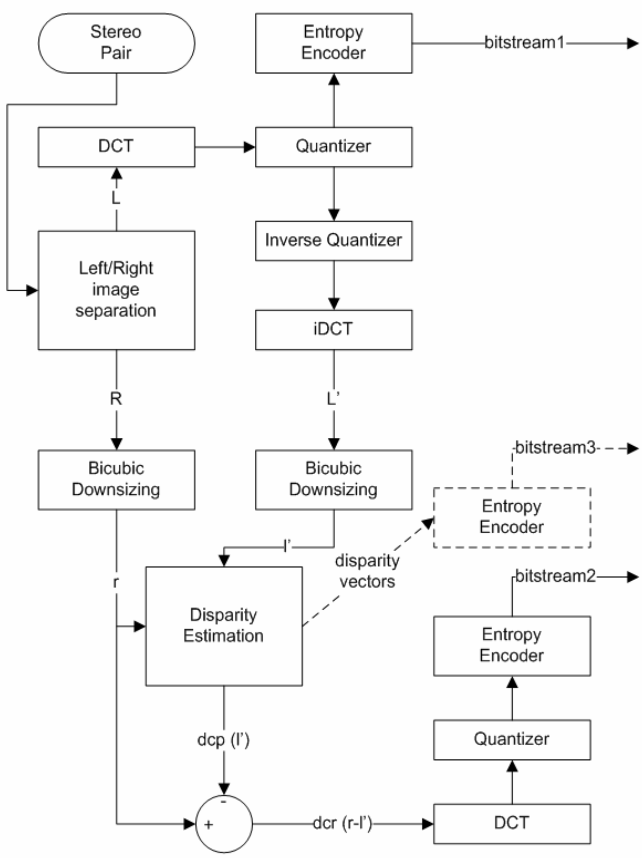

Fig. 1. Disparity compensated DCT encoder for stereo image compression

First the stereo image is separated into left (L) and right (R) images for each eye respectively. L is chosen as a reference image. It is separated into 8x8 blocks, which are DCT transformed and quantized. The intra-frame quantization matrix used as per the MPEG standard [7] is shown in Eq.1.

[1]Alexander A. Krupev is with the Faculty of Communications, Technical University - Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: asfalot@abv.bg

[2]Antoaneta A.. Popova is with the Faculty of Communications, Technical University – Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: antoaneta.popova@tu-sofia.bg

[3]Ivo R. Draganov is with the Faculty of Communications, Technical University - Sofia, Kliment Ohridski 8, 1000 Sofia, Bulgaria, e-mail: idraganov@abv.bg

$$Q = \frac{1}{S}\begin{bmatrix} 8 & 16 & 19 & 22 & 26 & 27 & 29 & 34 \\ 16 & 16 & 22 & 24 & 27 & 29 & 34 & 37 \\ 19 & 22 & 26 & 27 & 29 & 34 & 34 & 38 \\ 22 & 22 & 26 & 27 & 29 & 34 & 34 & 38 \\ 22 & 26 & 27 & 29 & 32 & 35 & 40 & 48 \\ 26 & 27 & 29 & 32 & 35 & 40 & 48 & 58 \\ 26 & 27 & 29 & 34 & 38 & 46 & 56 & 69 \\ 27 & 29 & 35 & 38 & 46 & 56 & 69 & 83 \end{bmatrix}$$
$$S \in [1,\ 16] \tag{1}$$

S is the scale factor – the higher it is, the smaller the quantization steps are, which leads to higher quality and size of the output image. The fraction values are rounded to integers. The quantized values are sent to the lossless entropy encoder (zig-zag scanning, RLE and Huffman coding). The DC coefficients are DPCM coded separately before being sent to the entropy encoder as per JPG standard [6].

To minimize the prediction error, the left image blocks are dequantized and inverse DCT transformed in the encoder. Thus we will be working with the same reference image L' at the encoder and the decoder. Before disparity estimation R and L' can be downsized to increase compression. This will affect the quality of the right image only. Disparity estimation is performed on the downsized left an right images – l' and r. The right image is separated into 8x8 blocks. Let the row/column indices of a current block from r be m and n. The best match search in the left image is performed in the same row – m, on pixel basis. The minimal SSD (Sum of Square Differences) is chosen as a matching criterion (Eq.2):

$$ssd_{m,n} = \sum_{a=0}^{7}\sum_{b=0}^{7}(r_{8m+a,8n+b} - l'_{8m+a,8n+b+k})^2 \tag{2}$$
$$\text{where} \quad k \in \left[-\lceil 0.5N \rceil, \lceil 0.5N \rceil\ \right]$$

The image is composed of $M \times N$ blocks, k represents the disparity vector. It is chosen to be within 1/8 of the image width interval and thus defines the search area. It's unlikely the disparity to be higher for a correctly composed stereo pair. The search area in this case belongs to the interval:

$$search\ \ area_{m,n} \in [8n - \lceil 0.5N \rceil,\ \ 8n + \lceil 0.5N \rceil + 7] \tag{3}$$

After the best match is found, the disparity compensated residual (dcr) $r - l'$ is formed (Eq. 4)

$$dcr_{m,n} = r_{m,n} - l'_{m,n'} \tag{4}$$

n' denotes the disparity relative to n. The residuals are DCT transformed, quantized, losslessly encoded and sent to the decoder. The quantization is uniform with default step value of 16 (inter-frame quantization, [7]), which can be scaled down also, like in Eq. 1. The disparity estimation block also sends the disparity vectors to the decoder, being DPCM encoded first. That's useful because clusters of blocks, belonging to the same objects, are displaced equally [7].

In the pioneering block-based version of the encoder, the disparity vectors (bitstream3) are not sent to the decoder. They are restored there using information from previously decoded blocks. That's why the disparity estimation process uses neighboring blocks to $r_{m,n}$, namely $r_{m-1,n}$ and $r_{m,n-1}$, because they will be already present at the decoder and the vectors could be reproduced there. The pioneering block based disparity estimation, as proposed by Jiang et al. is shown on Fig. 2.

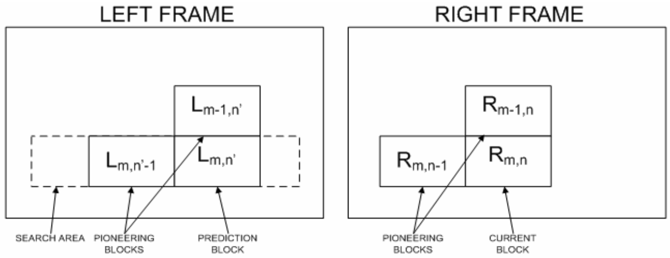

Fig. 2. Pioneering block-based disparity estimation

In our implementation we use the three neighbouring blocks to $r_{m,n}$, to improve precision (Eq. 5).

$$PR_{m,n} = \frac{r_{m-1,n} + r_{m-1,n-1} + r_{m,n-1}}{3} \tag{5}$$

This pioneering block $PR_{m,n}$ is used for disparity estimation, instead of the current block $r_{m,n}$. Analogically, a pioneering block $PL_{m,n'}$ is formed in the left image l', within the search window (Eq. 3) and SSD criterion is used again (Eq. 6) to find the best match.

$$ssd_{m,n} = \sum_{a=0}^{7}\sum_{b=0}^{7}\Big(\frac{r_{8(m-1)+a,8n+b} + r_{8(m-1)+a,8(n-1)+b} + r_{8m+a,8(n-1)+b}}{3} - \frac{l'_{8(m-1)+a,8n+b+k} + l'_{8(m-1)+a,8(n-1)+b+k} + l'_{8m+a,8(n-1)+b+k}}{3}\Big)^2 \tag{6}$$

In the decoder (Fig.3), the left reference image L' is restored after inverse quantization and inverse DCT. It's ready to be displayed, but is also used in the restoration of the right image. L' is downsized to the same size as in the coder and l' is sent to the disparity estimation block as well as the restored residuals (dcr'). The disparity is estimated the same way as in the encoder. The effect of the quantization error of the left image on the estimation process is reduced, because L' (not L) was used in the encoder for prediction. But the matching process can be affected by the quantization error in the residual image blocks (and thus indirectly by L'). There is also an error distribution effect when a disparity vector is not correctly restored. The dependence of the correctly restored vectors on the quantization step (through the scaling coefficients s for intra and z for inter-frame quantization of the residuals) is shown in the experimental results.

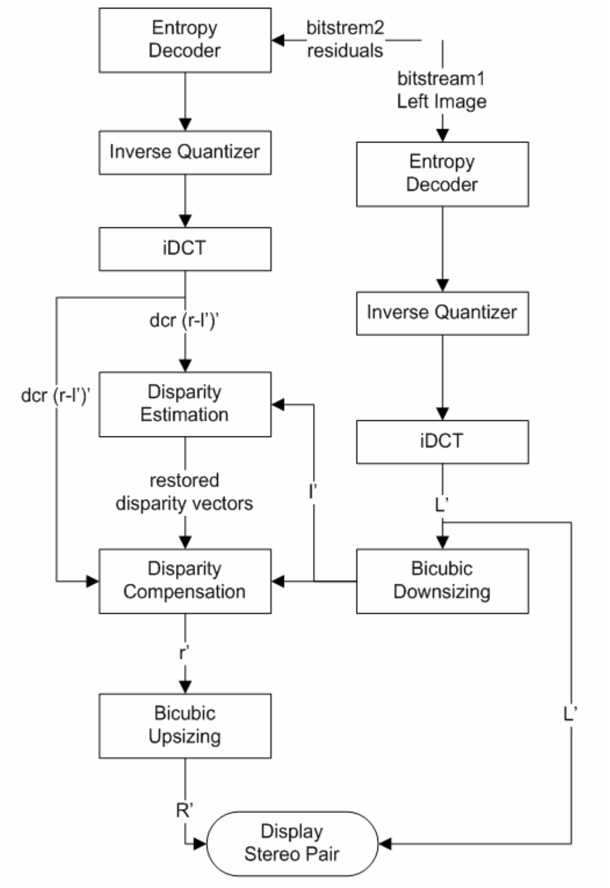

Fig. 3 Pioneering block-based disparity decoder for stereo image compression

Using the restored disparity vectors, the restored disparity compensated residuals dcr' and the downsized left image l', the disparity compensation block restores the right image r' (Eq. 7). It's upsized before being displayed as R'.

$$r'_{m,n} = l'_{m,n'} + dcr'_{m,n} = l'_{m,n'} + (r_{m,n} - l'_{m,n'})' \tag{7}$$

The conventional non-pioneering predictive decoder is a simpler version of the decoder from Fig. 3. Since the vectors are sent from the coder, the disparity estimation block is omitted. There is no risk to apply mismatched residuals in Eq.7. The compression is slightly lower, of course.

III. Experimental Results

The schemes from Figs. 2 and 3 are implemented in Matlab 7.0.3 working environment. The images are processed in RGB color space and grayscale intensity values are used for disparity estimation. The lossless entropy encoders have not been implemented since our main goal is to evaluate quality.

A PSNR evaluation module is added, comparing the processed images R' and L' to the original R and L images according to Eq. 8:

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)[dB] \tag{8}$$

$$MAX_I = 255$$

Here MSE is:

$$MSE = \frac{1}{pq}\sum_{i=0}^{p-1}\sum_{j=0}^{q-1}[I(i,j) - \hat{I}(i,j)]^2 \tag{9}$$

p, q are the dimensions of the images, I and $\hat{I}$ - pixel intensity values before/after processing, ranging from 0 to 255. The images are turned to grayscale before evaluation.

The stereo pair used for the experiments is shown on Fig. 4:


Fig. 4 Mountain stereo pair, used for the experiments

For the first experiment, pioneering encoding is applied without resizing. The disparity vector tables at the encoding and the decoding ends are compared for different quantization scale coefficients (greater scale coefficient gives finer quantization - Eq.1). z affects residual quantization, s – the left/reference image quantization. The results are shown in Table 1.

Table I

Mismatched Block Percentages

Inter scaling coefficient *(z)*	*z=1*	*z=8*	*z=16*
Intra scaling coefficient, *s=1*	21.2%	12.14%	3.93%
Intra scaling coefficient, *s=4*	20.35%	5.98%	0.97%
Intra scaling coefficient, *s=8*	17.39%	6.34%	2.17%
Intra scaling coefficient, *s=12*	16.79%	9.24%	4.11%
Intra scaling coefficient, *s=16*	21.27%	14.13%	7.42%

The correctly restored vectors percentage always increases with the decrease of the inter-frame quantization step (increased *z*). The effect of the inter-frame reference image quantization on the matching process is content-based/random though, since l' participates in the quantized residuals. For coarse residual quantization ($z = 1$), it is dominated by the quantization error of the residual, though. Averagely about 1/7th of the blocks are mismatched due to incorrectly restored vectors. This may manifest itself as blockish artifacts in the image (Fig. 5, Fig 7-d). It doesn't hamper stereo perception, but could be unpleasant for the viewer.

Fig. 5 Restored stereo pair for s, z=1, pioneering encoding, no bicubic resizing of the right frame

The second experiment involves measuring PSNR as a function of the quantization step, which is controlled by the coefficients s and z. The results are shown on Fig. 6. For single images optimal quality/compression is achieved for PSNR between 30-60dB [6]. This applies for the reference image (left) of the stereo pair, but the requirements for the right image are lower. That's because human vision can compensate high frequency loss from the other image.

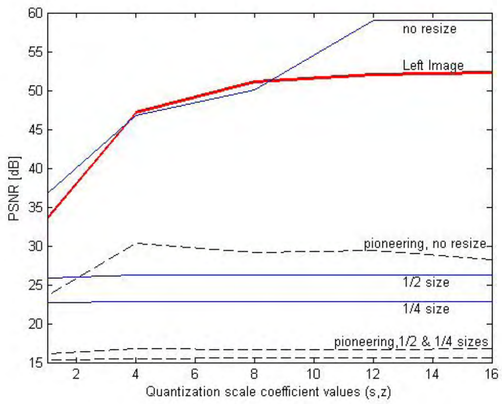

Fig. 6 The dependence of PSNR on the quantization scale coefficients s and z. Assumed is s=z.

The left image is always identical for all stereo pairs. Evidently pioneering block-based encoding does not give satisfactory results in combination with bicubic resize. For $s, z \in (4,6)$ without resizing, though, its PSNR is above 30dB. But without resizing, conventional disparity estimation technique yields 45dB in this case! The compression might be slightly lower because the vectors have to be transmitted, but it's worth the few extra bits. It even surpasses the reference image in quality, when smaller quantization steps are used. The error, induced from resizing, clearly outweighs the quantization errors – the graphs are almost flat lines for ½ and ¼ resizing. Significant compression is achieved this way.

For the purposes of subjective evaluation the processed stereo pairs are visualized using the anaglyph method, based on color separation [2]. Some of the visualized right images are shown on Fig. 7. In our opinion for ½ resize, s=4, z=1, optimal results are achieved (considering not only perceived quality, but compression). For ½ downsizing and pioneering encoding the stereo perception is preserved, but there are lots of visible annoying artifacts. Using ¼ resize yields significant granular noise perception. The stereo pair is very hard to focus on and eye strain occurs fast.

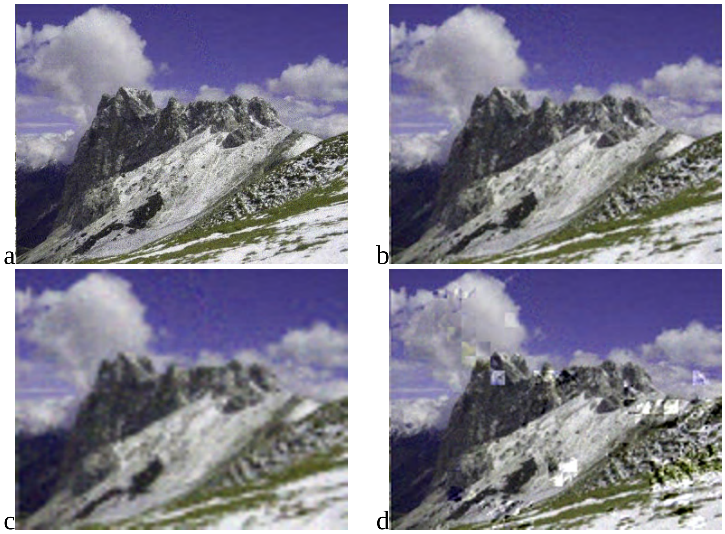

Fig. 7 Right images for s,z=1; a) no resize; b) ½ resize; c) ¼ resize; d) pioneering encoding, ½ resize

IV. Conclusion

Until further improved, the authors would rather employ the conventional predictive encoding over the pioneering block-based prediction scheme. These algorithms can be applied together with other standard compression technologies besides DCT based JPEG, like fractals, wavelets etc. to achieve extra compression for all right frames.

Acknowledgement

This paper was supported by the National Science Fund of the Bulgarian Ministry of Education, Youth and Science (Contract – DDVU 02/13 – "Public and Private Multimedia Network Throughput Increase by Creating Methods for Assessment, Control and Traffic Optimization").

References

[1] J. Jiang, E. Edirisinghe, and H. Schroder, "Algorithm for compression of stereo image pairs" *Electronics Letters*, 33(12):1034-1035, September 1991.

[2] A. Krupev and A. Popova, "Flickering Reduction Algorithm in Anaglyph Stereoscopic Images", ICEST'08 Proceedings of Papers, Vol. 1, pp.125-128, Nis, Serbia, 2008.

[3] M. Bax and A. Vitus, "Stereo Image Compression", http://www.stanford.edu/~mbax/ee392c/report.html, Stanford, Dec. 1997.

[4] I. Dinstein, M.G. Kim, A. Henik, and J. Tzelgov, "Compression of stereo images using subsampling transform coding," Optical Engineering, vol. 30, no. 9, pp. 1359-1364, Sept. 1991.

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

[6] W. Pennebaker and J. Mitchell, "*JPEG still image data compression standard (3rd ed.)"*, Springer. , 1993.

[7] J. Mitchell, W. Pennebaker, C. Fogg, D. LeGall ,"MPEG video compression standard", Kluwer Academic Publishers, 2002