Published in the Proc. of ICIP-95, Vol. II, pp.361-365, IEEE Computer Society Press, 1994.

A MULTIRESOLUTION FRAMEWORK FOR STEREOSCOPIC IMAGE SEQUENCE COMPRESSION[1]

Sriram Sethuraman, M.W.Siegel, Angel G. Jordan,

Department of Electrical and Computer Engineering,
The Robotics Institute at the School of Computer Science,
Carnegie Mellon University, Pittsburgh, PA 15213

ABSTRACT

Stereoscopic sequence compression typically involves the exploitation of the spatial redundancy between the left and right streams to achieve higher compressions than are possible with the independent compression of the two streams. In this paper the psychophysical property of the human visual system, that only one high resolution image in a stereo image pair is sufficient for satisfactory depth perception, has been used to further reduce the bit rates. Thus, one of the streams is independently coded along the lines of the MPEG standards, while the other stream is estimated at a lower resolution from this stream. A multiresolution framework has been adopted to facilitate such an estimation of motion and disparity vectors at different resolutions. Experimental results on typical sequences indicate that the additional stream can be compressed to about one-fifth of a highly compressed independently coded stream, without any significant loss in depth perception or perceived image quality.

I INTRODUCTION

Stereoscopic image display is a simple and compact means of portraying depth information on a 2-D screen. The binocular parallax or *disparity* between two images of the same scene, shot from two nearby points-of-view, contains information about the relative depths of the objects in the scene. This relative depth can be deduced by humans, when each eye is presented with its corresponding image. Thus, stereoscopic transmission requires twice the conventional monocular transmission bandwidth. However, several schemes [1], [2], [3], [4] have been developed, that exploit the disparity relation to achieve compression ratios higher than are possible by the independent compression of the two streams.

Psychophysical experiments [5],[6] have shown that a stereo image pair with one high resolution image and one lower resolution image are sufficient to provide good stereoscopic depth perception. For compatibility with existing monocular transmission schemes, one image stream can be compressed by motion compensated prediction and interpolation, as recommended by the MPEG standards. The other stream can be estimated from this stream, at a lower resolution using the disparity relation. To facilitate the estimations at different resolutions and to reduce the computational complexity of the search process, a multiresolutional approach was proposed in [7] by us to compress a ‘still’ stereo image pair. This paper extends the same concept to fit into a motion sequence compression framework.

The paper is organized as follows. Section 2 introduces the concepts behind a low resolution disparity estimation for stereo image compression. Section 3 describes the proposed scheme and how the disparity estimation described in [7] is adapted to a sequence compression context. Section 4 evaluates the compression ratios possible with the proposed scheme. Section 5 discusses the subjective and objective evaluations over a typical stereoscopic image sequence. Section 6 outlines the conclusions and some possible extensions to the scheme.

2. MULTIRESOLUTIONAL DISPARITY ESTIMATION

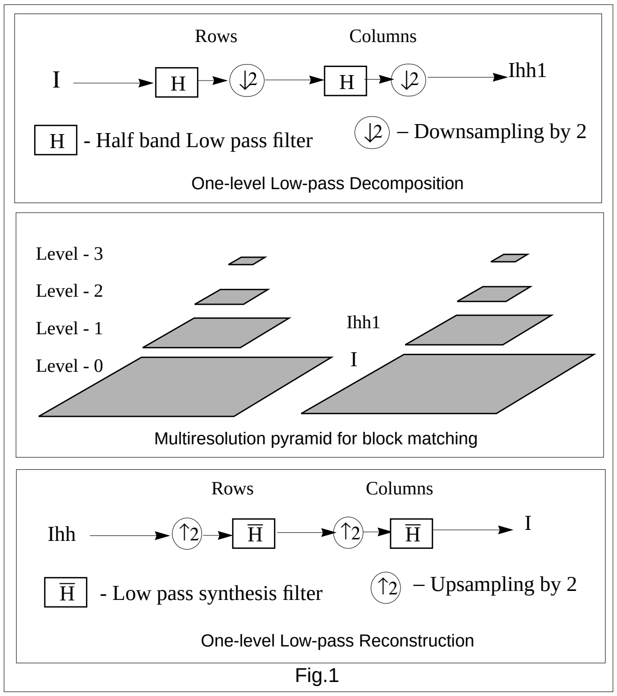
2.1 Disparity estimation

Disparity is the vectorial distance between the two points of a superposed stereo pair that correspond to the same point in the 3-D scene. Estimation of disparity is analogous to displacement estimation between two image frames that are offset temporally. However, the binocular imaging geometry constrains the corresponding points to lie on *epipolar* lines [2]. Thus, the search for the corresponding point is one dimensional. For the imaging geometry considered in this paper, where the camera axes are parallel, the epipolar lines become the corresponding horizontal scan lines. Thus, the disparity becomes a scalar and its estimation requires only a 1-D search.

2.2 Multiresolution based block matching

The image that will be estimated and the image it will be estimated from are both decomposed using a multiresolution

[1] This research was supported by the Advanced Research Projects Agency under ARPA Grant No.MDA 972-92-J-1010.



One-level Low-pass Decomposition

Multiresolution pyramid for block matching

One-level Low-pass Reconstruction

Fig.1

decomposition scheme as in [3],[8]. The motion / disparity over a block (of pixels) is assumed a constant. The sizes of the blocks at the different levels of resolution are fixed. The motion or disparity estimation begins at the coarsest resolution level. A block at resolution level-(j+1) is divided into 4 blocks at the level-j. The estimation proceeds in a coarse-to-fine fashion, with the estimates at level-(j+1) being refined at level-j. The error criterion used for finding the best matching block is the minimum absolute difference (MAD).

The advantage of such a scheme is that the complexity of the search becomes insensitive to the range of the search neighborhood. A larger search neighborhood requires only a small increase in the search neighborhood at the coarsest resolution at which the computational complexity is small due to the smaller number of blocks at that level. Also, the multilevel estimates present a decorrelated representation, thus requiring fewer bits to represent a block’s motion or disparity.

2.3 Low resolution disparity estimation

Consider a scheme wherein the ‘left’ image is intra-coded and the ‘right’ image is estimated from the left using disparity estimation at a lower resolution. The multiresolution based disparity estimation proceeds from the coarsest level up to level-1. The level-1 low pass subimage is estimated using the disparity vectors from the level-1 ‘left’ subimage. A full size (in pixels) reconstruction is obtained by upsampling by a factor of two and reconstructing with the synthesis low pass filter (Refer to Fig.1). Since the number of blocks at level-1 is one-fourth that at level-0, fewer block disparities need to be coded. This leads to compression ratios of about 170 for the right image [7]. The estimation errors at object boundaries due to larger block sizes are not noticeable due to the removal of the high frequencies.

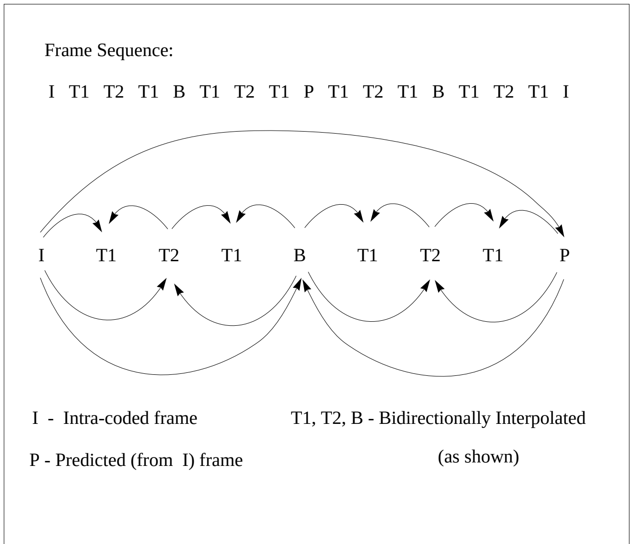

Frame Sequence:
I T1 T2 T1 B T1 T2 T1 P T1 T2 T1 B T1 T2 T1 I
I
T1
T2
T1
B
T1
T2
T1
P
I - Intra-coded frame
P - Predicted (from I) frame
T1, T2, B - Bidirectionally Interpolated
(as shown)

Fig.2: MPEG-type motion compensated prediction & interpolation of left frames

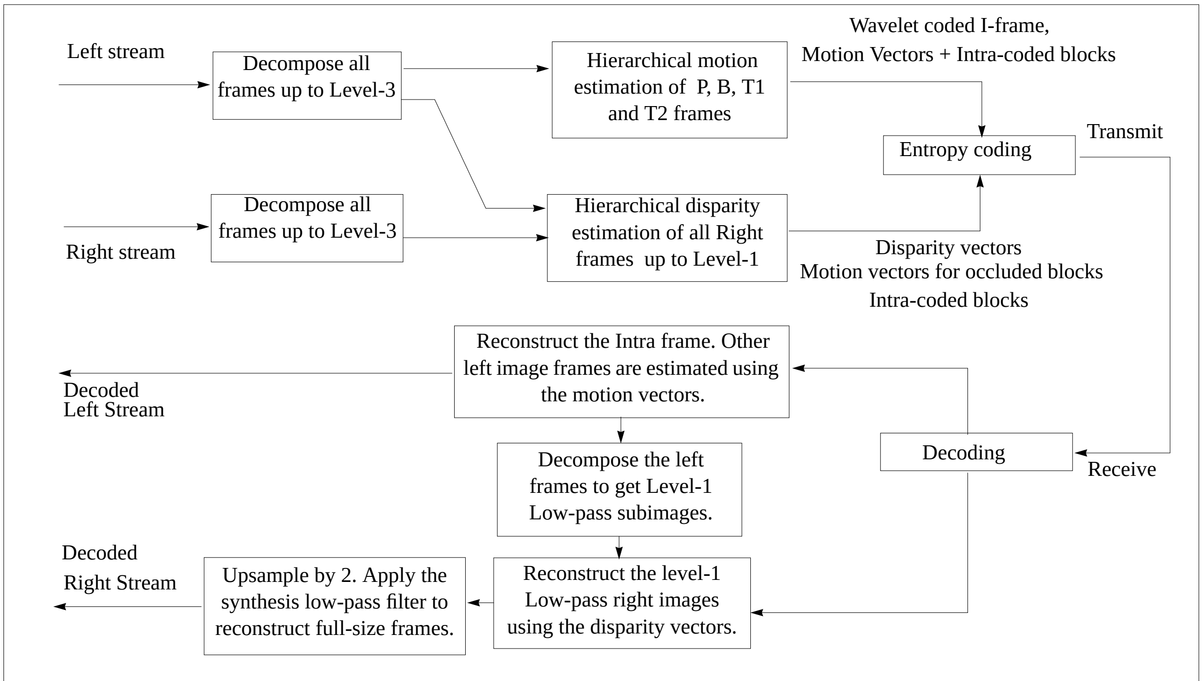
3. PROPOSED COMPRESSION SCHEME

3.1 Coding the left image stream

The left image stream is compressed independent of the right stream using MPEG-type intra-coded frames (I), predicted frames (P) and bidirectionally predicted frames (B). In order to take advantage of the multiresolution scheme in estimating larger motion vectors at the same computational complexity, a larger group-of-pictures (GOP=16) is chosen. Two additional bidirectionally predicted frames, T1 and T2, are introduced. The dependencies of each frame is shown in Fig.2. The motion vector estimation is carried out on the multiresolution pyramid. A half pixel resolution matching is carried out at level-0. The blocks with large residuals are DCT coded. The I-frame is coded by subband coding techniques.

3.2 Coding the right image stream

All the right stream frames are estimated from their corresponding left stream frames, using the low resolution disparity estimation procedure described in section 2.3. A half pixel match is performed at level-1 to improve the estimation accuracy. The occluded regions (regions present only in one view) cannot be estimated by the above procedure. To efficiently code these regions, a temporal prediction is made from the past or future right frames. The same dependency as the left stream is used for this. The blocks that have large residuals are intra-coded. A 2 bit prefix is needed for each block to specify whether it was disparity estimated, forward / backward motion predicted or intra coded. This overhead is offset by the significant reduction in the



Left stream
Decompose all frames up to Level-3
Hierarchical motion estimation of P, B, T1 and T2 frames
Wavelet coded I-frame,
Motion Vectors + Intra-coded blocks
Entropy coding
Transmit
Decompose all frames up to Level-3
Right stream
Hierarchical disparity estimation of all Right frames up to Level-1
Disparity vectors
Motion vectors for occluded blocks
Intra-coded blocks
Reconstruct the Intra frame. Other left image frames are estimated using the motion vectors.
Decoded
Left Stream
Decompose the left frames to get Level-1 Low-pass subimages.
Decoding
Receive
Decoded
Right Stream
Upsample by 2. Apply the synthesis low-pass filter to reconstruct full-size frames.
Reconstruct the level-1 Low-pass right images using the disparity vectors.

Fig.3: Encoder-Decoder structure for the proposed stereoscopic image sequence compression scheme

number of blocks to be intra-coded, due to temporal prediction. Figure 3 illustrates the encoder-decoder structure for the proposed scheme.

4. BIT-RATE CALCULATIONS

Compared to independent compression of the two streams, additional compression in the proposed scheme stems from three factors:

- The low resolution disparity estimation results in one-fourth the number of block disparities needed at full resolution.
- Disparity is a scalar for the camera geometry considered. So, fewer bits are required to represent it than the vector motion [1].
- The I and P frames, which contain more intra-coded blocks, are also disparity estimated.

All intra-coded blocks are assumed to be coded at 1 bit per pixel (bpp). A constant of 8 bits/block is assumed for representing the motion vectors for all T1, T2, B and P frames[2]. The different intra-coding percentages assumed for the different frames are shown in the table below.

Frames	I	P	B	T2	T1
% of intra-coded blocks	100	20	6	2	<1

With these assumptions, ignoring any entropy / runlength / DPCM coding of the motion vectors, the total number of bits to represent a ‘left’ GOP is given by:

$$\overset{I}{MxN} + \overset{P}{21.2B} + \overset{B}{2x(13.8B)} + \overset{T2}{4x(9.8B)} + \overset{T1}{8x(9.8B)}\ \text{bits}^{3}$$

where, MxN is the size of the image (in pixels) and B is the number of blocks per frame = (MxN/block size)

For block size = 8x8, this translates to *3.6MN bits* for 16 frames, resulting in a compression ratio of approximately 35. For 640x480 NTSC resolution at 30 frames-per-second, the bit rate becomes 2Mbits/s. This includes only the luminance signal. Coding of the chrominance components may require half as many bits due to the spatial subsampling (4:2:2) of those components.

For disparity coding, a rate of 3 bits/level/block is assumed. Let the percentage of temporally predicted and intra coded blocks be 20 and 5, respectively for a right frame. The number of bits to represent a ‘right’ GOP is *0.7MN bits*, for an 8x8 block size at level-1. Thus, the right stream can be compressed to about *one-fifth* of the left stream.

5. EXPERIMENTAL RESULTS

Several sequences generated using a fixed 3-D camera were compressed using this stereoscopic compression scheme. The objective results are shown in Figures 4 and 5. It can be seen that the minimum PSNR for the left stream frames is around 30dB and that for the right stream is around 25dB. The disparity compensation improves the uncompensated SNR by about 6dB.

[1] Multiresolution coding is essentially insensitive to the search neighborhoods of the motion and disparity estimations.

[2] The total number of bits required is almost a constant over a large range of neighborhoods because, only the number of bits allocated at the coarsest level (level that has a very small number of blocks) varies with the neighborhood.

[3] This includes the bit fields required to encode whether a block was disparity estimated, forward or backward predicted, or intra-coded.

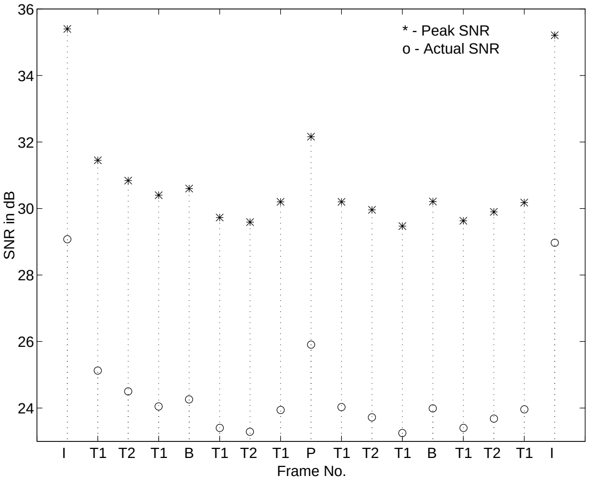


Fig.4: SNR for the left sequence

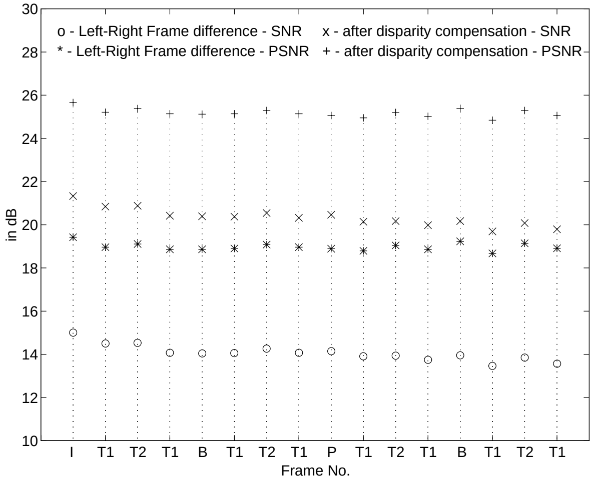


Fig.5: SNR for the right sequence

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right) \qquad SNR = 10\log_{10}\left(\frac{E\,[I^2]}{MSE}\right) \qquad MSE = E\left[(I-\hat{I})^2\right]$$

Figure 6 shows the intra-coded left frame, a disparity estimated right frame, the disparity between the left and right frames and the error after low resolution disparity compensation. When viewed stereoscopically, the decompressed stereo stream gives exceptionally good depth perception. Since the binocular occlusion is minimal in the scene shown, the temporal interpolation is very infrequently used, resulting in lower bit rates. Figure 7 shows the effectiveness of the motion compensation using bidirectional motion prediction for a T2 frame.

6. CONCLUSIONS & FUTURE WORK

Using this compression scheme, a stereoscopic sequence can be compressed to fit into a monocular transmission bandwidth. There is very little loss in the quality of one of the streams. Thus this scheme offers the flexibility to transmit the additional sequence as a marginal option, without seriously affecting the existing monocular transmission schemes.

Currently, we are pursuing a disparity based segmentation scheme to achieve very low bit-rates by efficient segmentation of constant disparity patches. By properly identifying the relations between camera motion and its effects on disparity (for instance, the motion and binocular parallax become the same in the case of a camera panning a still scene), the present scheme can be extended to include camera motions like ‘pan’ and ‘zoom’ and yet achieve very high compressions.

References:

[1] M.G.Perkins, ‘Data compression of stereopairs', IEEE Transactions on Communications, Vol.40, No.4, pp.684-696, April 1992.

[2] A.Tamtaoui, C.Labit, ‘Constrained disparity and motion estimators for 3DTV image sequence coding', Signal Processing: Image Communication, vol.4, 1991.

[3] D.Tzovaras, et al., ‘Evaluation of multiresolution block matching techniques for motion and disparity estimation’, Signal Processing: Image communication, Vol.6, 1994.

[4] R.E.H.Franich, R.L.Lagendijk, J.Biemond, ‘Stereo-enhanced displacement estimation by genetic block matching', SPIE Visual Communications and Image Processing, Vol.2094, pp.362-371, 1993

[5] I.Dinstein et al., ‘Compression of stereo images and the evaluation of its effects on 3-D perception’, Proc. of IEE Applications of Digital Image Processing XII, 1989.

[6] Tetsuo Mitsuhashi, ‘Subjective image position in stereoscopic TV systems - Considerations on comfortable stereoscopic images’, pp. 259-265, SPIE Vol.2179, 1994.

[7] S.Sethuraman, M.W.Siegel, A.G.Jordan, ‘Multiresolution based hierarchical disparity estimation for stereo image pair compression’, Proc. of the symposium on Application of subbands and wavelets, Newark, NJ, 1994.

[8] Stephane G.Mallat, ‘A theory for multiresolution signal decomposition: The wavelet representation’, IEEE Trans. on PAMI’, Vo.II, No.7, July 1989.

[9] M.Uz, M.Vetterli, D.J.LeGall, ‘Interpolative multiresolution coding of advanced television with compatible subchannels’, IEEE Trans. on circuits and systems for video tech., Vol.1, No.1, March 1991.