Published in SPIE Vol.2419 - Digital Video compression - Algorithms and technologies in 1995

A multiresolutional region based segmentation scheme for stereoscopic image compression

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

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

ABSTRACT

Stereoscopic image sequence transmission over existing monocular digital transmission channels, without seriously affecting the quality of one of the image streams, requires a very low bit-rate coding of the additional stream. Fixed block-size based disparity estimation schemes cannot achieve such low bit-rates without causing severe edge artifacts. Also, textureless regions lead to spurious matches which hampers the efficient coding of block disparities. In this paper, we propose a novel disparity-based segmentation approach, to achieve an efficient partition of the image into regions of more or less fixed disparity. The partitions are edge based, in order to minimize the edge artifacts after disparity compensation. The scheme leads to disparity discontinuity preserving, yet smoother and more accurate disparity fields than fixed block-size based schemes. The smoothness and the reduced number of block disparities lead to efficient coding of one image of a stereo pair given the other.

The segmentation is achieved by performing a quadtree decomposition, with the disparity compensated error as the splitting criterion. The multiresolutional recursive decomposition offers a computationally efficient and non-iterative means of improving the disparity estimates while preserving the disparity discontinuities. The segmented regions can be tracked temporally to achieve very high compression ratios on a stereoscopic image stream.

Keywords: stereoscopic image compression, disparity estimation, quadtree decomposition, segmentation, multiresolution

1. INTRODUCTION

Stereoscopic image display offers a simple and compact way of representing the depth information in a real world scene, on a flat screen. The price for this added realism is the requirement of twice the monocular transmission bandwidth to accommodate the ‘left’ and ‘right’ image streams. However, the two images recorded from two neighboring points of view possess an inherent redundancy, by proper exploitation of which it is possible to compress the two streams and to transmit them over a monocular channel with very little loss in the quality of one of the image streams.

Such high compression ratios (of compressing two streams to the size of one maximally compressed stream) can be achieved only by taking advantage of the structure imposed by the additional viewpoint and the high tolerances of the human visual system. This end can be better realized by resorting to object-oriented compression schemes compared to conventional fixed block-size based schemes. This paper presents a region based scheme that utilizes the binocular disparity information itself to segment the image.

The paper is organized as follows. Section 2 provides the background needed to understand the binocular imaging geometry and how it can be used to compress a pair of stereo images. Section 3 emphasizes the need for variable block size based schemes to achieve higher compression and better disparity fields. Section 4 outlines the proposed approach and justifies its choice. Section 5 elaborates on the implementation details of such a scheme. Section 6 discusses the experimental results that substantiate the compression gains possible using the proposed approach.

2. BINOCULAR IMAGING AND DISPARITY ESTIMATION

Figure 1 shows the geometry of a binocular imaging setup. A point P in the 3D world is projected in perspective, on to P_L and P_R on the two imaging sensors, respectively through the left and right viewing points L and R. The vectorial distance

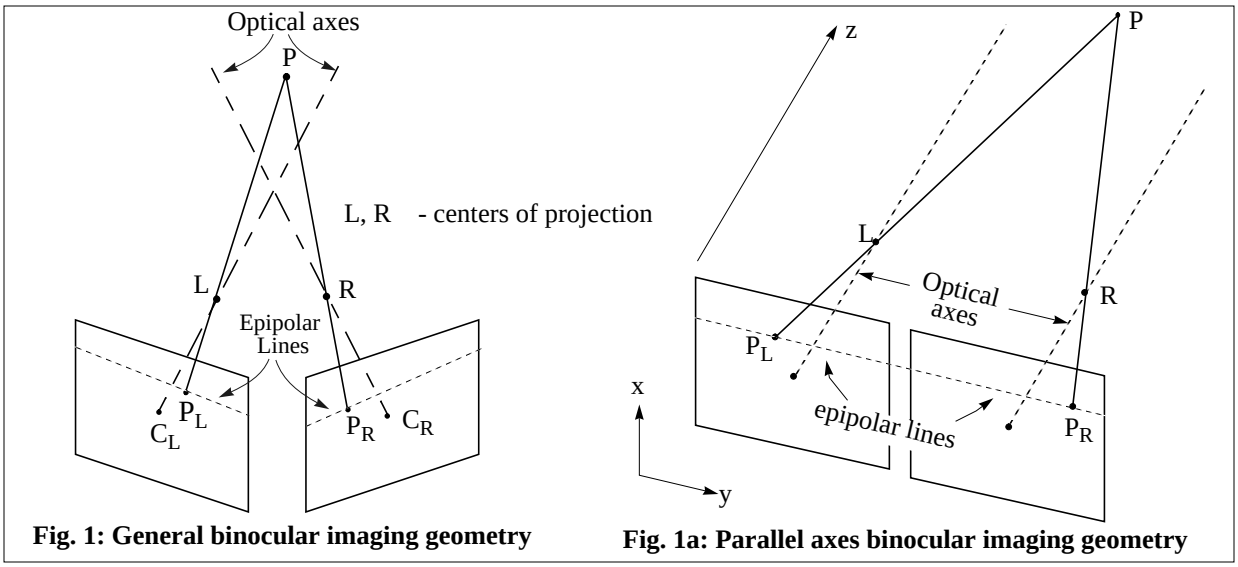


Fig. 1: General binocular imaging geometry

Fig. 1a: Parallel axes binocular imaging geometry

between P_L and P_R, when the two imaging regions are placed one on top of the other, is called the *disparity* of point P. The problem of finding the pair P_L and P_R is commonly known as the *correspondence* or *disparity estimation* problem, and is analogous to the estimation of motion between two temporally separated image frames. However so long as the perpendiculars to the image planes passing through the respective centers of projection (optical axes) lie on the same plane, the geometry of the setup dictates the corresponding points to lie on the *epipolar* lines, which are obtained by the intersection of the image planes with the plane defined by the points (P, L, and R). For the case, where the optical axes of the left and right imaging elements are parallel, the epipolar lines become the horizontal scan lines (Fig.1a). Thus for this special case the disparity estimation for a point in one image is restricted to a search along the corresponding scan line in the other image.

3. LOW BIT RATE STEREO IMAGE COMPRESSION

Several stereo compression techniques have been developed based on the concept of disparity estimation.[1, 2, 3, 4, 5] Most compression techniques partition the image into rectangular regions of fixed size (in pixels) and estimate the block disparity of each block, assuming a constant disparity over the entire block. This fixed block size (FBS) based partition eliminates the need for any coding overhead to specify the locations of the blocks and thus requires only the transmission of block disparities and disparity compensated error. However, the FBS based estimation has certain drawbacks.

The FBS scheme cannot take advantage of large areas of constant disparity that typically make up the background of a scene. These areas get divided into multiple blocks, thus requiring more block disparities to be coded. Also, matching of small featureless blocks leads to spurious matches which reduces the smoothness of the disparity field. Also, blocks falling on objects at different depths will lead to erroneous estimation and large errors after motion/disparity compensation. Transmission of error information to suppress any visible artifacts in these blocks increases the overall bit rate. Accurate disparity fields are needed to perform other tasks such as synthesis of intermediate views. In short, the allocation of bits is not proportional to the disparity detail present in an area. Higher compression and better estimation accuracy can be achieved by adapting the block size to the disparity detail present in that region. A genetic block matching algorithm was proposed in [7] to achieve dense and accurate disparity fields and to partially overcome the disadvantage of FBS schemes.

In this paper, we propose a segmentation scheme that adapts the local block size depending on the disparity detail to achieve a smoother and more accurate disparity field while substantially reducing the total number of block disparities to be transmitted. The coding overhead needed to specify the block locations and the computational overhead needed for the segmentation have both been reduced by incorporating the entire algorithm in a multiresolutional framework.

4. PROPOSED SEGMENTATION SCHEME

A pyramid structure for hierarchical disparity estimation is constructed using a multiresolution decomposition, as shown in Fig. 2. A conventional quadtree decomposition is employed to split the image based on the disparity compensated error of a block. The disparity estimation and splitting are carried out in succession. The disparity estimation commences at the coarsest resolution level. The splitting proceeds recursively, until the disparity compensated error for a block is below a preset threshold or the size of the block drops below a preset minimum. The location for splitting at lower resolutions is decided based on strong intensity edges, which are obtained using a simple scheme outlined in section 5. At higher resolutions, the division occurs at the midpoints of the sides of a block. Processing at each level ends when the recursion stops. The leaf nodes of the tree generated at this level become the root nodes at the next higher resolution level and the estimation and splitting proceeds recursively for these root nodes at that level. This process is repeated until the decomposition is complete at the

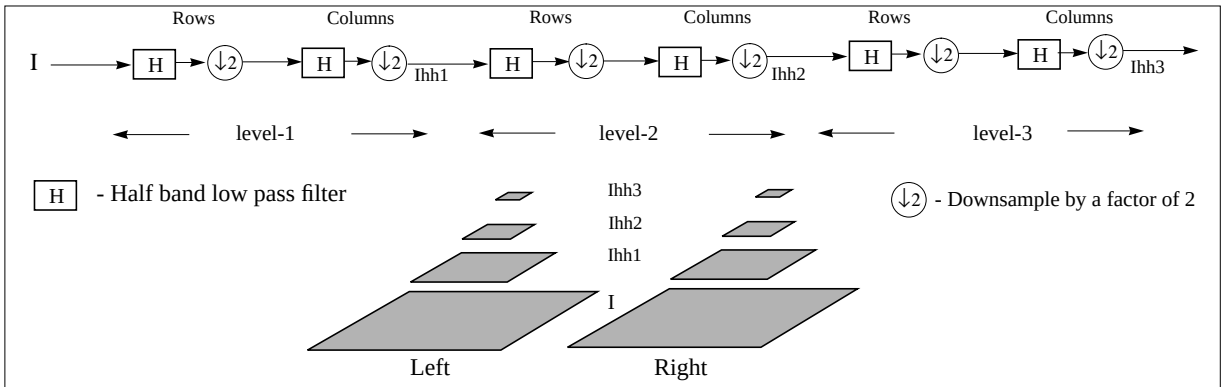


Fig. 2: Multiresolution Decomposition and the stereo pyramid structure

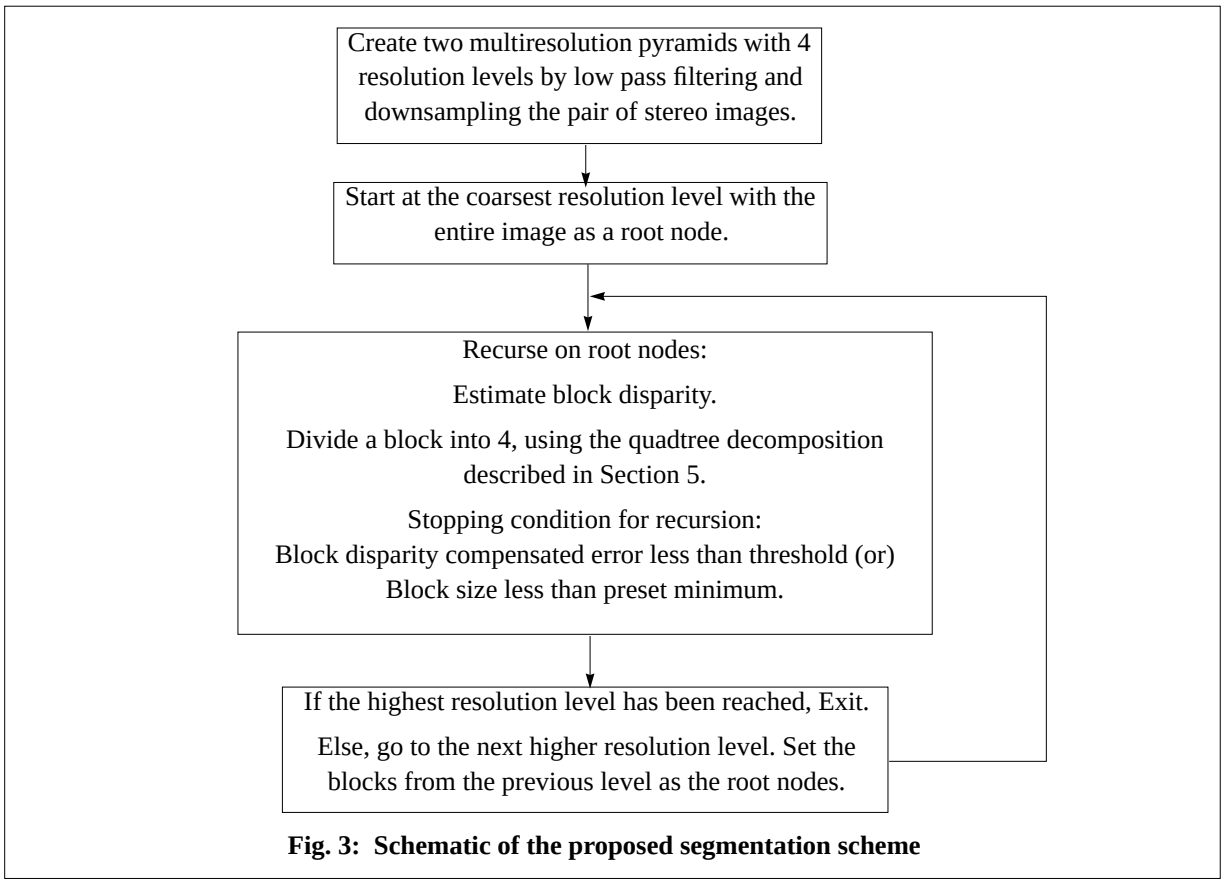


Fig. 3: Schematic of the proposed segmentation scheme

highest resolution level. Figure 3 presents a schematic of the proposed scheme.

4.1 Advantages of quadtree decomposition

A region based segmentation method (quadtree) has been chosen, instead of feature based methods such as contour based segmentation, because the extraction of features is computationally intensive, and establishing correspondences between the extracted features is not simple. Also, propagating the estimated sparse motion/disparity correspondences to the region within the contour is not trivial. On the other hand, region based segmentation relies only on the intensity values that are already available; and since the segmentation itself takes into account the texture of the region, no propagation of correspondences is required. Further, the overhead for coding arbitrarily shaped regions is very high compared to coding the location of rectangular blocks obtained using recursive region splitting.

4.2 Comparison with motion-based segmentation

The novelty of the proposed segmentation is that it is disparity based. A disparity based segmentation is superior to a displacement (motion) based one because a disparity field is a measure of the relative depths of the objects in the scene, while a displacement field is only a measure of the 'change' in the scene. Thus a segmentation based on disparity provides a handle on the real objects in the scene. The ability to represent real objects helps in targeting more bits on desired objects at a specific depth, while allocating fewer bits to other regions (and blurring those regions to suppress any visually distracting artifacts). Such a scheme to separate the person at the foreground from the background has been suggested in [8]. Since entire objects can be segmented, the global velocity of the object can be computed and used to constrain the search for the motion vectors of specific portions of that object.

For a parallel axes camera geometry, which is the correct geometry for stereoscopic display on a single flat screen [11], the disparity field is essentially a scalar field, and thus the estimation of disparity requires only a 1D, one-directional search (along the corresponding scan line, and to the right or left of the pixel of interest); In contrast a 2D, two directional search is required to estimate a displacement vector. A parallel axes configuration thus reduces the search computations and the estimate coding bits considerably compared to motion vector computation and coding.

4.3 Multiresolutional segmentation - a non-iterative solution

Typically, a disparity based segmentation technique requires a good disparity map. But a smooth and accurate disparity map can be obtained only by estimation. Thus the segmentation procedure based on disparity has to be iterative. However the multiresolutional estimation allows us to progressively refine both the partition and the disparity map from the coarse to fine resolution, thus minimizing the computational burden associated with iterative estimation at the highest resolution level. Also, the process of estimating the disparities for large blocks at the top of the tree is computationally expensive. Since the size (in pixels) of the image at coarsest resolution is small, the high computational expense associated with matching large blocks at the top of the tree is considerably reduced. Also, by adopting different splitting strategies at the different resolutions, the coding overhead is reduced. This is explained in detail in the next section.

5. IMPLEMENTATION DETAILS

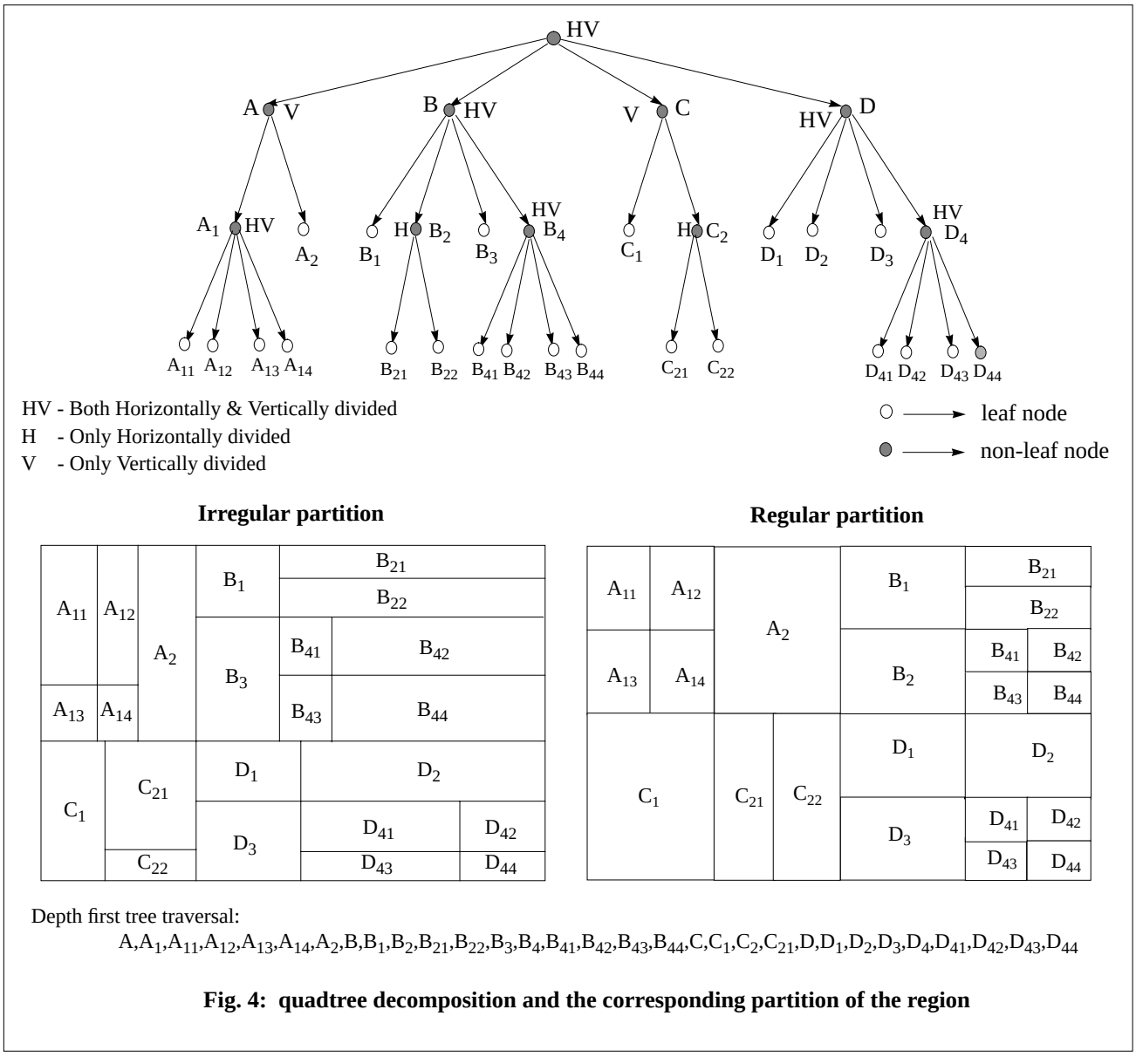
5.1 Generalized quadtree

In a conventionally used *regular* quadtree, a block can be divided only at the midpoint of its sides. Such a decomposition requires very little overhead information (1 bit per node -- bit 1 if divided, bit 0 if not) to specify the tree structure and hence the locations of the blocks. However, as the divisions do not occur near intensity discontinuities, the total number of segmented blocks is higher than necessary. The other extreme would be to divide a block only at locations close to sharp intensity discontinuities. Such an *irregular* quadtree provides the minimum number of blocks. But specifying the location now requires $\log_2$(size of the block) bits per node.

As a trade-off between these two extremes, a *generalized* quadtree can be considered wherein a block can be divided at 2^K-1 locations where K is the number of bits that can be allocated per side per node. The division takes place at the permitted location that lies closest to a sharp intensity discontinuity. The two extreme cases will become special cases of such a scheme. In the multiresolution based scheme, larger K values are used for the coarser resolutions to assure smaller number of blocks. Regular decomposition (K=1) can be used at fine resolutions where the block sizes are small anyway.

5.2 Coding overhead calculations

The coding overhead needed to specify the locations of the blocks is estimated here, assuming that arbitrary partitions (maximum K) are allowed only at the coarsest resolution (level-3) and only a regular partition (K=1) is allowed at all the subsequent levels of resolution. The recursive tree structure and the corresponding block partitions are shown in Fig.4. The locations of the blocks can be uniquely decoded from a string of numbers that specify the x-y pixel offsets from the top left corner of the block that is being divided, with a depth first tree traversal. The number of bits needed to represent the x-y offset is $\log_2$(size of the block). As the blocks get progressively smaller, the number of bits needed to specify the partition also



Depth first tree traversal:

$A,A_1,A_{11},A_{12},A_{13},A_{14},A_2,B,B_1,B_2,B_{21},B_{22},B_3,B_4,B_{41},B_{42},B_{43},B_{44},C,C_1,C_2,C_{21},D,D_1,D_2,D_3,D_4,D_{41},D_{42},D_{43},D_{44}$

Fig. 4: quadtree decomposition and the corresponding partition of the region

decreases. A special escape code can be used to signify a leaf node (an undivided node). Since, the number of blocks at the end of segmentation at level-3 is about only 10% of the total number of blocks (Refer to Table 1), the coding of the block locations at level-3 requires only a few hundred bits for an NTSC resolution image. Due to the regular partition from level-2, the subsequent decomposition requires only 1 bit/node at the remaining resolution levels. Thus, the overall coding overhead constitutes only a small fraction of the total bits needed to code one image of the stereo pair given the other.

5.3 Detecting intensity discontinuities

The sharp intensity discontinuity within a block is detected by picking the maximum of the high pass filtered row and column 'means' of the block. Since the split is made only vertically and/or horizontally, such a simple technique would be sufficient instead of a full scale edge detection within the block. Also, the averaging provides the overall best location to divide, within the block.

For a block of size N_x x N_y starting at (i , j) in image I, let the best matching block in the other view of the stereo pair start at $(i + \delta_x, j + \delta_y)$. If the Mean Absolute Difference (MAD) given by

$$\left(\left(\frac{1}{N_x N_y} \right) \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left| I(i,j) - I_{ref}\left(i + \delta_x, j + \delta_y \right) \right| \right) > threshold\ T$$

then, that block is divided horizontally and vertically at the following locations:

Horizontally: $\underset{i}{Max} \left[g_x(i) = \left| Conv(m_x, f) \right|; \quad 1 < i < N_x \right]$ (Operator Conv denotes convolution)

Vertically: $\underset{j}{Max} \left[g_y(j) = \left| Conv(m_y, f) \right|; \quad 1 < j < N_y \right]$

where, $m_x(i) = \sum_{j=1}^{N_y} I(i,j)$, $m_y(j) = \sum_{i=1}^{N_x} I(i,j)$ and *f* is the symmetric difference filter [-1, 0, 1].

5.4 Smoothness of the disparity field

Since the splitting may remove the strongest edge present in a block, spurious matching might occur due to lack of features. To overcome this, additional rows or columns of pixels around the block are also considered for matching, with the errors weighted by a gaussian window centered at the center of the block. The edges present in the boundaries of the block prevent many spurious matches and a smoother disparity field is achieved.

6. EXPERIMENTAL RESULTS

The variable block size based disparity estimation was tested over several markedly different stereo pairs. The right image of a stereo pair was estimated from the left using the variable block size (VBS) based approach and a fixed block size based hierarchical approach. The signal-to-noise ratio (SNR) of the estimated image has been kept constant in both the approaches. The results are tabulated in Table 1. It can be seen that the VBS results in about *30-65%* of the number of blocks needed using FBS. As can be seen from Table 1, the number of blocks at the end of segmentation at level-3 is fairly small, thus requiring lower coding overhead. This implies that, compared to the FBS scheme, the VBS scheme requires fewer bits to represent the right image, given the left image, at the same image quality (measured by SNR). If the number of blocks is kept a constant, the VBS scheme gives a higher SNR than the FBS scheme. Figure 5 illustrates this for the test stereo pair 'flower garden'. The high fixed block sizes required in this case lead to considerable estimation errors.

Figure 6 illustrates the partition and disparity map obtained using the VBS and FBS schemes for the stereo pair 'juggler'.

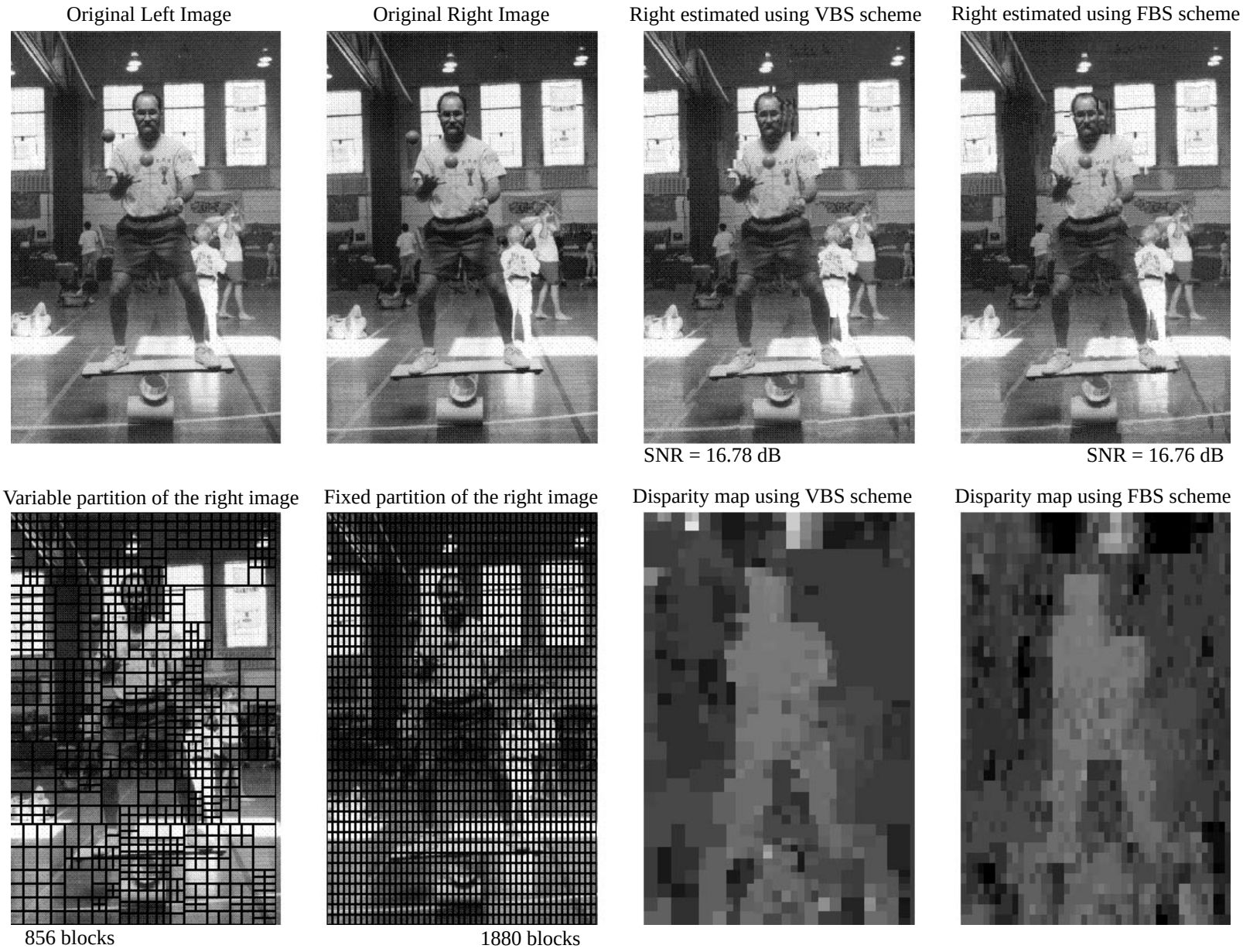


Fig. 6: Comparison between the VBS and FBS schemes - at a fixed SNR

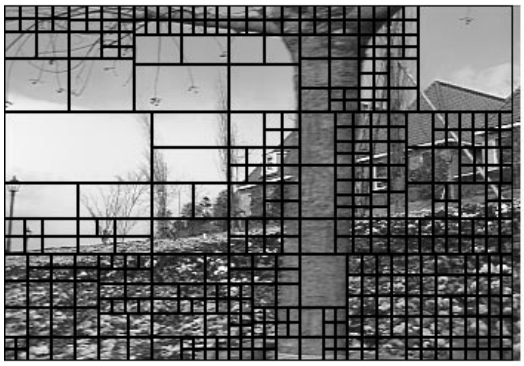
Original Left Image

Original Right Image

Right estimated using the VBS scheme (19.22 dB)

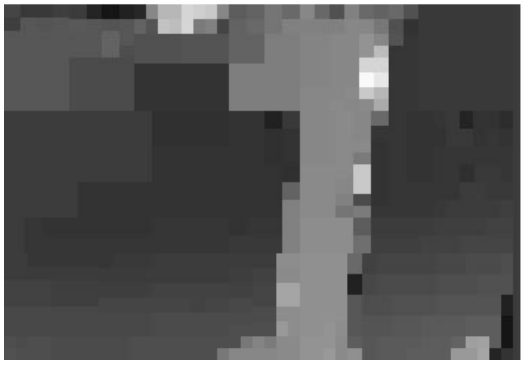
Right estimated using the FBS scheme (17.6 dB)

Variable partition of the right image (403 blocks)



Fixed partition of the right image (408 blocks)

Disparity map using the VBS scheme



Disparity map using the FBS scheme

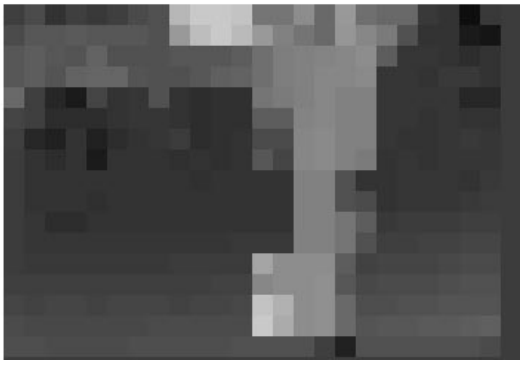


Fig. 5: Comparison between the VBS and FBS schemes - at almost equal block counts

The disparity field estimated using the VBS is more accurate and has fewer spurious matches. It can also be seen that the density of small blocks is high in the occluded areas (areas that lack correspondence); thus this approach can tackle occlusions better than an FBS scheme (which requires intra-coding of erroneous blocks).

TABLE 1: Comparison of block sizes with VBS and FBS coding at a fixed SNR

Stereo image pair	Size in pixels	Variable Block Size Coding					Fixed Block Size Coding		N_0/N
		N_3	N_2	N_1	N_0	SNR (dB)	N	SNR (dB)	%
Juggler	360x240	93	317	671	856	16.78	1880	16.76	45.5
F. Garden	240x350	57	173	531	640	19.94	1990	19.89	32.1
Crowd	420x280	122	347	717	813	18.51	2400	18.54	33.9
Book Sale	640x240	139	376	915	976	18.66	1920	18.65	50.8
Portage	245x370	108	359	999	1022	17.41	1850	17.44	56.1
Baltimore	484x470	87	329	930	959	17.64	1880	17.82	51.0
Karp	277x422	76	415	891	901	16.84	1440	16.73	62.5
Lake	549x329	178	268	418	419	20.35	1060	20.26	39.5
Swing	376x319	121	300	924	931	12.26	1870	12.38	49.7

N_i - Number of blocks at the end of segmentation at Level - i

N - Number of blocks needed to achieve the same SNR using FBS coding

$$SNR = 10 \cdot \log_{10}\left(\frac{\sum_i \sum_j I^2(i,j)}{\sum_i \sum_j [I(i,j) - \hat{I}(i,j)]^2} \right)$$

I - Original image

$\hat{I}$ - Estimated image

7. CONCLUSIONS & FUTURE WORK

A computationally efficient variable block size based scheme that results in high compression ratios and better subjective image quality than fixed block size based schemes is presented in this paper. The resulting disparity fields are smooth, accurate and yet preserve disparity discontinuities and are hence suitable for applications such as the synthesis of intermediate views.

The segmented blocks correspond to portions of objects in the scene each of which may undergo individual displacements over time. By tracking these blocks temporally, the global motion of the blocks can be estimated. These global vectors can be used to constrain the search for the displacements of specific portions of the segmented primitives. Thus the proposed segmentation scheme can eventually reduce the motion estimation complexity and also can minimize the number of motion vectors to be transmitted. Such a depth based motion estimation would also aid in the automatic detection of motion occlusions. Currently, efforts are on to extend this compression technique to a stereoscopic sequence compression framework. In this regard, critical issues such as the frequency of segmentation and coding of occluded regions are being addressed.

8. ACKNOWLEDGEMENTS

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

9. REFERENCES

1. M.G. Perkins, ‘Data compression of stereopairs', IEEE Trans. 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, pp.45-54,1991.
3. A. Tamtaoui, C. Labit, ‘Coherent disparity and motion compensation in 3DTV image sequence coding schemes', Proc. of ICASSP ‘91, Vol.IV, pp.2845-2848, 1991.
4. S.Sethuraman, M.W. Siegel, A.G. Jordan, ‘A multiresolution framework for stereoscopic image sequence compression’, Proc. of ICIP ‘94, Vol. II, pp.361-365, IEEE Computer Society Press, 1994.
5. 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.
6. D. Tzovaras, M.G. Strintzis, H. Sahinoglou, ‘Evaluation of multiresolution block matching techniques for motion and disparity estimation’, Signal Processing: Image Communication, Vol. 6, 59-67, 1994.
7. R. 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.
8. M. Waldowski, ‘A new segmentation algorithm for videophone applications based on stereo image pairs’, IEEE Trans. on Communications, Vol.39, No.12, December 1991.
9. S. Liu, M. Hayes, ‘Segmentation based coding of motion difference and motion field images for low bit-rate video compression’, Proc. of ICASSP ‘92, Vol.III, pp.525-528, 1992.
10. I. Dinstein, et al., Compression of stereo images and the evaluation of its effects on 3-D perception’, SPIE Applications of Digital Image Processing XII, Vol.1153, pp.522-529, 1989.
11. V.S. Grinberg, G. Podnar, M.W. Siegel, ‘Geometry of binocular imaging’, Proc. of the IS&T/SPIE Symp. on Electronic Imaging, Stereoscopic Displays and applications, Vol.2177, 1994.