COMPRESSION OF STEREO VIDEO STREAMS

P. D. Gunatilake, M. W. Siegel, A. G. Jordan
Department of Electrical & Computer Engineering and School of Computer Science
Carnegie Mellon University

Direct application of monocular compression schemes to stereo video streams is suboptimal. This is because these techniques do not take advantage of the high correlation between the left and right stereo image pairs. This paper introduces new still and motion stereo compression schemes that are far efficient than monocular compression schemes in compressing stereo image streams.

BACKGROUND

Efficient compression will have to be an integral part of any practical 3-D TV system. Fortunately, the high correlation between stereo image pairs leads us to expect efficient compression to be possible. It is important that compression is not gained at the loss of image quality since image realism is the principal paradigm of 3-D TV. We propose schemes for compressing still and motion stereo that are capable of achieving high compression while retaining image quality.

Monocular compression schemes such as DPCM, transform coding (DCT), JPEG and MPEG achieve compression by decorrelating in the spatial and/or temporal correlation domains[1-4]. Stereo video streams add a new dimension, namely the correlation between the left and right images. Direct application of any monocular compression schemes is suboptimal because the decorrelation is limited to spatial and temporal correlation domains. Several authors have proposed stereo image compression schemes that take advantage of psychophysical aspects of the human vision system[5,6]. Our schemes while benefiting from those artifacts of the visual system, achieves compression by extending the decorrelation process of monocular compression schemes to spatial, temporal and left-right image stream correlation domains.

INTRODUCTION

MODELLING OF THE CORRELATION

Typical stereo image streams have a high degree of similarity (common objects or background) between left and right image streams. This similarity or correlation is formally defined as “worldline correlation[1]”. Worldline correlation describes the correlation between time offset frames of the left and right image streams. A special case of the worldline correlation is the correlation between temporally corresponding stereo image pairs(zero time offset). This correlation is named “cross image correlation” to emphasize the temporal correspondence between the left and right stereo image pair.

1. “worldline” borrowed from the field of special relativity, means the path of an object in 4-dimensional space - time.

The consideration of the worldline correlation lead to efficient compression. The general modelling of the worldline correlation involves identifying corresponding similar areas of left and right time off set[2] stereo images. There are several widely known area matching algorithms [7-9] that identify similar areas of a stereo image pair. By matching, we mean the process of finding, for a given block of pixels of the left[3] image, the block of the right image that is most nearly the same image fragment. Similar image fragments are related to each other by their relative displacement vector with respect to a common origin in the image coordinate plane. Hence an image fragment is completely described by its corresponding similar image fragment and the associated relative displacement vector. We will define the following terms and algorithms before describing the details of our compression schemes.

PIXEL INTENSITY DEFINITION

A pixel intensity denoted by x positioned at coordinate (i,j) is defined as x(i,j). Another notation for x(i,j) is $x(\vec{\Gamma})$ where $\vec{\Gamma} = \begin{bmatrix} i & j \end{bmatrix}$. They are equivalent since a pixel coordinate can be represented as an ordered pair (i,j) or as a vector $\begin{bmatrix} i & j \end{bmatrix}$

BLOCK DEFINTION

A block of pixels in an image is described by its upper left hand pixel coordinate together with its pixel height and length. Define

$$block\ \ R(\vec{\alpha}_0) = (R(i,j) \mid (i_0 \le i \le i_0 + m), (j_0 \le j \le j_0 + n)) \qquad \text{(EQ 1)}$$

where $block\ \ R(\vec{\alpha}_0)$ is a block of height m+1pixels and length n+1 pixels of the right image and $\vec{\alpha}_0 = \begin{bmatrix} i_0 & j_0 \end{bmatrix}$. Note that the upper left hand pixel of the block is at pixel coordinate (i_0,j_0). $block\ \ L(\vec{\alpha}_0)$ is defined in an analogous manner.

BLOCK MATCHING ALGORITHM (BMA)

Assume that we are required to find the most similar block for $block\ \ L(\vec{\alpha}_0)$. Define the corresponding search area S of the right image to be

$$S = (R(i,j) \mid (i_0 - i_1 \le i \le i_0 + i_1 + m), (j_0 - j_1 \le j \le j_0 + j_1 + n)) \qquad \text{(EQ 2)}$$

where i_1 and j_1 are constants that control the size of S

BMA using Mean-Square-Error (MSE)

The MSE between $block\ \ L(\vec{\alpha}_0)$ and a $block\ \ R(\vec{\beta}_0)$ where $\vec{\beta}_0 = \begin{bmatrix} k_0 & l_0 \end{bmatrix}$ in S is defined as

$$M(k_0, l_0) = \frac{\sum_{i=0}^{m} \sum_{j=0}^{n} \{L(i_0 + i, j_0 + j) - R(k_0 + i, l_0 + j)\}}{(m+1) \times (n+1)} \qquad \text{(EQ 3)}$$

2. Includes zero time offset

3. Matching for a block of pixels of the right image achieved similarly with appropriate changes.

Let $M(k_{min},l_{min}) = \min\{M(k_0,l_0)\}$ for all $block\ R(\vec{\beta}_0)$'s in S. The $block\ R(\vec{\beta}_{min})$ corresponding to $M(k_{min},l_{min})$ is the most similar block of the right image that approximates $block\ L(\vec{\alpha}_0)$.

DISPARITY VECTOR

Let $block\ R(\vec{\beta}_0)$ where $\vec{\beta}_0 = \begin{bmatrix} k_0 & l_0 \end{bmatrix}$ be the most similar block found using a block matching criterion for $block\ L(\vec{\alpha}_0)$. The disparity vector of $block\ L(\vec{\alpha}_0)$ is defined as

$$\delta(\vec{\alpha}_0) = \begin{bmatrix} k_0 - i_0 & l_0 - j_0 \end{bmatrix} \quad \text{(EQ 4)}$$

The matrix of disparity vectors of all blocks in the left image is called the disparity map of the left image with respect to the right image. Because reasonable matching criteria will be symmetric with respect to interchange of left and right images $\delta(\vec{\beta}_0) = -\delta(\vec{\alpha}_0)$

ESTIMATED BLOCK

The $block\ R(\vec{\beta}_0)$ that is the most similar block in the right image that approximates $block\ L(\vec{\alpha}_0)$ is called the $estimated\ block\ L(\vec{\alpha}_0)$. Note that

$$estimated\ block\ L(\vec{\alpha}_0) = block\ R(\vec{\alpha}_0 + \delta(\vec{\alpha}_0)) \quad \text{(EQ 5)}$$

The matrix of estimated left blocks constitute the estimated left image that approximates the left image.

MOTION ESTIMATION

Assume that two frames of a monocular image stream contain essentially similar image fragments in different locations within each frame. This difference is usually the result of scene motion during the time interval between the two frames. Then, given one frame and the set of relative displacement vectors of essentially similar image fragments, the other frame can be usefully approximated. The procedure of finding the set of relative displacement vectors is called motion estimation.

In practice, image fragments are considered to be square arrays of pixels. Consider a block $X_{n+k}(\vec{\alpha}_0)$ of image X_{n+k} where k > 1 and an area S of X_n that contains block $X_n(\vec{\beta}_0)$ that is essentially similar to block $X_{n+k}(\vec{\alpha}_0)$. The displacement vector $\vec{D}_0$ between block $X_{n+k}(\vec{\alpha}_0)$ and block $X_n(\vec{\beta}_0)$ is given by $\vec{D}_0 = \vec{\alpha}_0 - \vec{\beta}_0$

Block $X_{n+k}(\vec{\alpha}_0)$ can be written as

$$block\ X_{n+k}(\vec{\alpha}_0) = block\ X_n(\vec{\alpha}_0 - \vec{D}_0) = block\ X_n(\vec{\beta}_0) \quad \text{(EQ 6)}$$

Let $x_{n+k}(\vec{\Gamma})$ be any pixel at $\vec{\Gamma} = \begin{bmatrix} i & j \end{bmatrix}$ within block $X_{n+k}(\vec{\alpha}_0)$

Then define Displaced Pixel Difference as

$$DPD(x_{n+k}(\vec{\Gamma}), \vec{D}_0(i-1)) = x_{n+k}(\vec{\Gamma}) - x_n(\vec{\Gamma} - \vec{D}_0(i-1)) \quad \text{(EQ 7)}$$

between pixel $x_{n+k}(\vec{\Gamma})$ and a pixel $x_n(\vec{\Gamma} - \vec{D}_0(i-1))$ in block $X_n(\vec{\alpha}_0 - \vec{D}_0(i-1))$. $\vec{D}_0(i-1)$ is an initial estimate of the displacement vector $\vec{D}_0$. Using (EQ 6), (EQ 7) can be written as

$$DPD(x_{n+k}(\vec{\Gamma}), \vec{D}_0(i-1)) = x_{n+k}(\vec{\Gamma}) - x_{n+k}(\vec{\Gamma} - \vec{D}_0(i-1) + \vec{D}_0) \qquad \text{(EQ 8)}$$

Note that if $\vec{D}_0(i-1) = \vec{D}_0$, then $DPD(x_{n+k}(\vec{\Gamma}), \vec{D}_0(i-1)) = 0$

The motion estimation algorithm used in this paper estimates $\vec{D}_0(i)$ from an initial estimate $\vec{D}_0(i-1)$ by recursively minimizing the quantity $DPD(x_{n+k}(\vec{\Gamma}), \vec{D}_0(i-1))$ which measures the difference between the two pixel intensity values $x_{n+k}(\vec{\Gamma})$ and $x_n(\vec{\Gamma} - \vec{D}_0(i-1))$[8]. The minimization is done using a steepest descent technique. It is derived below.

Expanding (EQ 8) using Taylor series

$$DPD(x_{n+k}(\vec{\Gamma}), \vec{D}_0(i-1)) = -(\vec{D}_0(i) - \vec{D}_0(i-1)) \nabla x^T_{n+k}(\vec{\Gamma} + \text{higher order terms} \qquad \text{(EQ 9)}$$

where ∇ is a discrete approximation to the gradient with respect to pixel intensity. (EQ 9) can be approximated by (EQ 10) assuming the higher order terms are negligible.

$$DPD(x_{n+k}(\vec{\Gamma}), \vec{D}_0(i-1)) \cong -(\vec{D}_0(i) - \vec{D}_0(i-1)) \nabla x^T_{n+k}(\vec{\Gamma}) \qquad \text{(EQ 10)}$$

$\nabla x^T_{n+k}(\vec{\Gamma})$ can be estimated as

$$\nabla x^T_{n+k}(\vec{\Gamma}) = \frac{1}{2}\begin{bmatrix} x_{n+k}(\vec{\Gamma} - [1\ 0]) - x_{n+k}(\vec{\Gamma} + [1\ 0]) \\ x_{n+k}(\vec{\Gamma} + [0\ 1]) - x_{n+k}(\vec{\Gamma} - [0\ 1]) \end{bmatrix} \qquad \text{(EQ 11)}$$

Summing (EQ 10) over all pixels $x_{n+k}(\vec{\Gamma})$ in block $X_{n+k}(\vec{\alpha}_0)$ and rearranging

$$\vec{D}_0(i) = \vec{D}_0(i-1) - \frac{\sum_{x_{n+k}(\vec{\Gamma}) \in X_{n+k}(\vec{\alpha}_0)} DPD(x_{n+k}(\vec{\Gamma}), \vec{D}_0(i-1)) \cdot \nabla x_{n+k}(\vec{\Gamma})}{\left[\sum_{x_{n+k}(\vec{\Gamma}) \in X_{n+k}(\vec{\alpha}_0)} \nabla x_{n+k}(\vec{\Gamma}) \cdot \nabla x^T_{n+k}(\vec{\Gamma})\right]} \qquad \text{(EQ 12)}$$

COMPRESSION OF STILL STEREO IMAGES

We have developed two different methods of compressing still stereo images. Both methods exploit the cross image correlation between the left and right images of a stereo pair. The first, which we call “direct integration” compresses one image of a stereo pair using any non-stereo still image compression scheme while the other image is represented by its disparity map. The second which we call “hybrid integration” makes changes to non-stereo still image compression schemes so that they are able to take advantage of the cross image correlation. We will describe direct integration in this paper.

DIRECT INTEGRATION

The left image is divided into blocks of m+1 by n+1 pixels. The search area for $block\ \ L(\vec{\alpha}_0)$ is defined as in (EQ 2)

$$S = (R(i,j) \mid (i_0 - i_1 \le i \le i_0 + i_1 + m), (j_0 - j_1 \le j \le j_0 + j_1 + n))$$

where i_1 and j_1 control the size of the search area. In general, j_1 is much larger than i_1 since the horizontal disparity is much greater than the vertical disparity for most reasonable stereo image pairs. Note that we are searching areas to the left as well as the right of $block\ \ R(\vec{\alpha}_0)$ for the most similar block corresponding to $block\ \ L(\vec{\alpha}_0)$. The disparity vector corresponding to the resulting most similar block is given by (EQ 4). The disparity map associated with the left image is formed by arraying the individual disparity vectors of all the blocks of the left image. The number of bytes required to represent the disparity map is approximately one percent of the number of bytes required to represent the left image as a pixel intensity map for a block size of 8 by 8 pixels.

Next, the right image is compressed using any non-stereo still compression scheme. Compression is gained by reducing the stereo pair to a conventionally compressed right image and a small disparity map from which in conjunction with the right image, the left image can be estimated. The estimated left image consists of blocks of the right image that are appropriately displaced to approximate the left image. When viewed stereoscopically, the pair are of subjectively good quality and exhibit essentially full depth resolution. Figure 1 is a block diagram of the direct integration scheme. The arrow heads indicate the direction of information flow.

The similarity between the estimated left image and the original left image depends on three factors. The most important factor is the underlying assumption of direct integration that left and right images of a stereo pair essentially contain the same image fragments. This assumption is a good approximation for most stereo image pairs. The second factor is the ability of the algorithm to determine the true most similar image block from the right image given a block from the left image. This depends on whether the search area S contain the true most similar block. For most reasonable stereo images i_1=3 pixels and j_1=16 pixels have proved to be sufficient bounds for the horizontal and vertical disparity. Hence a search area S that covers these bounds would have a high probability of finding the correct most similar block. The last factor is the block size. It is reasonable to expect the similarity between the estimated left image and the original left image to increase as the block size gets smaller. Large block sizes lead to larger differences since the probability of left and right images having the same large area is small. Figure 2 illustrates the effect of the block size on the quality of the estimated left image for the stereo image pair in Fig 3.

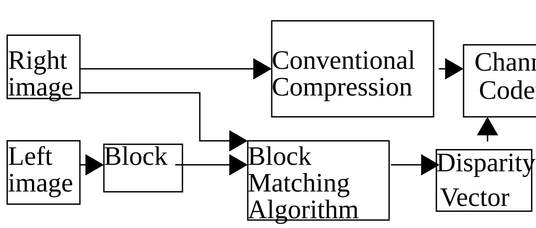


Figure 1
Block diagram of Direct integration

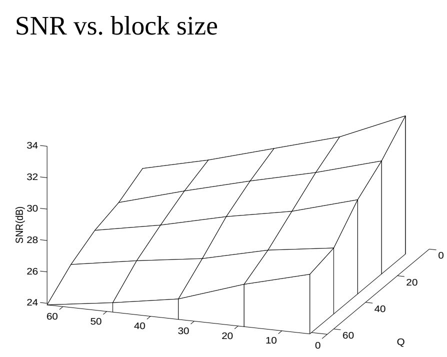


Figure 2.

The metric used to measure the similarity between the estimated left image and the original left image is the Signal -to -Noise Ratio(SNR). P and Q denote the height and length of a block. Figure 3 illustrates the original and the decompressed direct integrated stereo image pairs. The block size used was 8 by 8 pixels

Original left image **Decompressed left image**

Original right image **Decompressed right image**

Figure 3.

COMPRESSION OF MOTION STEREO STREAMS

A monocular image stream has considerable redundant[4] image content between temporally offset image frames. This redundancy is defined as temporal correlation. A stereo stream, in addition to the temporal correlation, has significant redundancy between time offset[5] left and right image frames. We call this redundancy the worldline correlation. Motion compression schemes identify and efficiently code the temporal redundancy. Efficient coding involves coding the location information of the redundant image parts. This leads to significant compression since the bandwidth of the location information is far less than the bandwidth of the redundant image part.

Monocular motion compression schemes such as MPEG will compress a stereo stream assuming the left and right image streams to be independent of each other. The decoupling

4. Redundancy defined according to some similarity criterion.
5. Zero offset included.

of the left and right image streams leads to the discarding of the worldline correlation. Therefore information in one image stream may appear in the other image stream. This is inefficient from a compression standpoint. We propose a new stereo image compression scheme that considers the left and right image streams to be correlated. Hence it is far more efficient than monocular motion compression schemes in compressing stereo streams.

The proposed stereo image compression scheme can be thought of as a process that preselects a few non consecutive image frames from a given segment of the original stereo stream. Then each of the remaining image frames of the original segment are approximated by combining image sections of the preselected image frames. The preselected image frames and the approximated image frames form the compressed stereo image stream. A compressed stereo image segment is shown in Fig 4. We call the preselected image frames “Main group” frames while the approximated frames are called the “Sub group” frames. The subgroup frames are close approximations of their corresponding original frames due to the high temporal and worldline correlations between frames in a stereo stream. The white frames in Fig 4 are the main group frames while the shaded frames are the subgroup frames. Compression is gained since only a few frames, namely the main group frames, contain image data that require high bandwidth. The remaining frames, the sub group frames, require very little bandwidth since they are constructed using image data from the main group frames. The bandwidth of a sub group frame is the bandwidth required by the correlation information that is needed for its construction from the main group frames. Details of the main and sub group frames are presented below.

MAIN GROUP FRAMES

Main group frames are frames that contain image data. There are two types of frames that belong to the main group. They are intraframe coded frame (I-frame) and motion estimated and error compensated frame (M-frame).

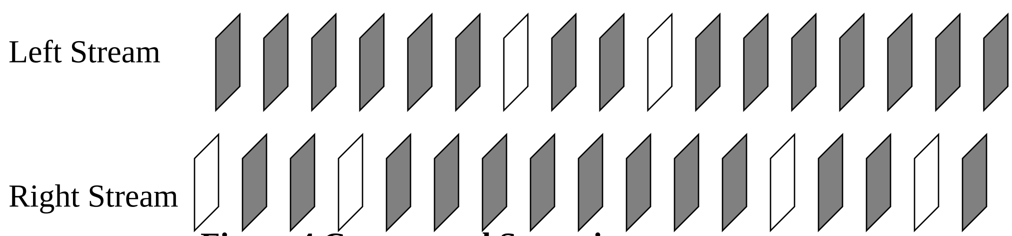


Figure 4 Compressed Stereo image stream

INTRAFRAME CODED FRAMES(I-Frame)

An I-frame is a periodically chosen frame from either the left or right image stream of a stereo stream. An I-frame is subjected to intraframe coding. Widely used intraframe coding schemes are JPEG or DPCM. The I-frame is similar to the I-frame used in MPEG[4].

MOTION ESTIMATED & ERROR COMPENSATED FRAME (M-Frame)

Let X_n and X_{n+k} be two frames in a monocular image stream that are k frames apart, where X denotes a left or right image frame. Figure 5 shows the process of generating a M-frame, M_{n+k} that approximates X_{n+k} from X_n and X_{n+k}. The first step of this process is to estimate $\overline{X}_{n+k}$ that approximates X_{n+k} from X_n using motion estimation[8,9]. The error image E_{n+k} is generated by subtracting $\overline{X}_{n+k}$ from X_{n+k}. The error image E_{n+k} appears to be due to the fact that X_n and X_{n+k} do not contain the same image content. E_{n+k} is quantized to $\overline{E}_{n+k}$. Now M_{n+k} is the sum of $\overline{X}_{n+k}$ and $\overline{E}_{n+k}$. Representing X_{n+k} by M_{n+k} leads to savings of bandwidth. This is because $\overline{X}_{n+k}$ is represented by a set of displacement vectors and $\overline{E}_{n+k}$ is limited to a finite number of levels. Figure 6 displays the SNR as a function of k for a fixed quantizer. The SNR decreases with increasing k. This is reasonable since as k increases, the temporal correlation between X_n and X_{n+k} decreases leading to a more significant E_{n+k}. Figure 7 displays the original image, the estimated image $X_{n+k}$and the resulting M-frame for k=8

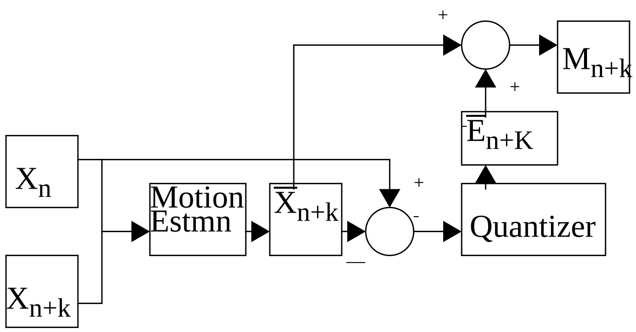


Figure 5
Construction of a M-frame

SNR vs. k for a M-frame

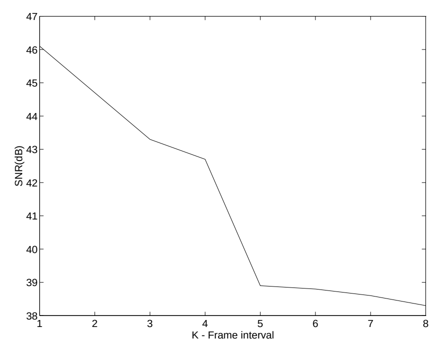


Figure 6

Original image M - frame Error image

Figure 7.

SUB GROUP FRAMES

A sub group frame in the compressed stereo stream is constructed using image data from the main group frames so that it approximates the corresponding frame in the original stereo stream. The bandwidth of a sub group frame is the bandwidth of the correlation information that is required for its construction from the main group. There are two different types of frames in the sub group. They are the bidirectional interpolated frame (B-frame) and the worldline frame (W-frame).

BIDIRECTIONALLY- INTERPOLATED FRAME (B-Frame)

Figure 8 displays the process of constructing a B-frame, B_{n+k} that approximates an original image frame X_{n+k} using frames X_n and X_{n+m} where $1 \le k < m$. The underlying assumption of this technique is that the image content of X_{n+k} lies in the union of the image content of X_n and X_{n+m}. B_{n+k} is generated using (EQ 13) where PX_{n+k} and FX_{n+k} are the motion estimated frames from X_n and X_{n+m} respectively.

$$B_{n+k} = PX_{n+k} + \left(\frac{k}{m}\right)\left(FX_{n+k} - PX_{n+k}\right) \qquad \text{(EQ 13)}$$

The image quality of a B-frame depends on how well the underlying assumption is satisfied. Figure 9 illustrates the original image frame, the corresponding B-frame and the resulting error image. The B-frame approximation appears to be a low-pass version of the original image. This is indicated by the error image which contains high frequency information. A B-frame in the compressed stereo stream is represented by two sets of displacement vectors.

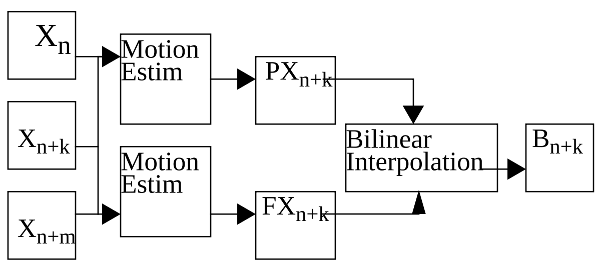


Figure 8 - Construction of a B-frame

Original image B-frame Error image

Figure 9

WORLDLINE FRAME (W-Frame)

A worldline frame(W-frame) is responsible for the consideration of worldline correlation in the proposed stereo compression scheme. The consideration of worldline correlation is significant in motion stereo image compression since it may easily be higher than either spatial or temporal correlations in a stereo scene with motion. It is the consideration of worldline correlation that makes the proposed motion stereo compression scheme more efficient than conventional compression schemes in compressing stereo image streams.

A W-frame, W_n approximates an original image frame X_n using original image frames X_{n-k}, Y_n, Y_{n-1} and Y_{n-2}. X and Y denote left/right or right/left image frames. A W-frame is constructed in the following manner.

1. Divide X_n into blocks of pixels.

2. For a given block of pixels in X_n, say $block\ \ X_n(\vec{\alpha})$, find the most similar blocks from original image frames X_{n-k}, Y_n, Y_{n-1} and Y_{n-2}. Let the corresponding four blocks be

   $block\ \ X_{n-k}(\vec{\alpha}1)$, $block\ \ Y_n(\vec{\alpha}2)$, $block\ \ Y_{n-1}(\vec{\alpha}3)$, $block\ \ Y_{n-2}(\vec{\alpha}4)$.

3. Select the most similar block to $block\ \ X_n(\vec{\alpha})$ from $block\ \ X_{n-k}(\vec{\alpha}1)$, $block\ \ Y_n(\vec{\alpha}2)$, $block\ \ Y_{n-1}(\vec{\alpha}3)$ and $block\ \ Y_{n-2}(\vec{\alpha}4)$. This block is $block\ \ W_n(\vec{\alpha})$.

4. Execute steps 2 and 3 for all $block\ \ X_n(\vec{\alpha})$'s in X_n to find corresponding $block\ \ W_n(\vec{\alpha})$'s that constitute W_n.

In step 2, $block\ \ X_{n-k}(\vec{\alpha}1)$ is found using motion estimation between X_{n-k} and X_n. $block\ \ Y_n(\vec{\alpha}2)$ is found using disparity estimation on Y_n and X_n as described in direct integration. Estimation of $block\ \ Y_{n-1}(\vec{\alpha}3)$ and $block\ \ Y_{n-2}(\vec{\alpha}4)$ are carried out in the following manner. We first note that for $block\ \ X_n(\vec{\alpha})$ and its stereo match block $block\ \ Y_n(\vec{\alpha}2)$

$$\vec{\delta}(n-k) + \vec{D}_x(n-k) = \vec{\delta}(n) + \vec{D}_y(n-k) \quad \text{(EQ 14)}$$

where $\vec{\delta}\ \ and\ \ \vec{D}$ are the disparity and displacement vectors as shown in Fig 10. Note that $\vec{\delta}(n) = \vec{\alpha}2 - \vec{\alpha}\ \ and\ \ \vec{D}_x(n-k) = \vec{\alpha}1 - \vec{\alpha}$. We further note that the disparity vector $\vec{\delta}(n)$ changes slowly over time. Hence we assume $\vec{\delta}(n) = \vec{\delta}(n-k)$ for small k. Substituting into (EQ 14), we have $\vec{D}_x(n-k) = \vec{D}_y(n-k)$ that is the X and Y displacement vectors are equal. Hence

$$block\ \ Y_{n-1}(\vec{\alpha}3) = block\ \ Y_{n-1}(\vec{\alpha} + \vec{\delta}(n) + \vec{D}_x(n-1)) \quad \text{(EQ 15)}$$

$$block\ \ Y_{n-2}(\vec{\alpha}4) = block\ \ Y_{n-2}(\vec{\alpha} + \vec{\delta}(n) + \vec{D}_x(n-2)) \quad \text{(EQ 16)}$$

Note that $\vec{\alpha}\ \ and\ \ \vec{\delta}(n)$ are known vectors. $\vec{D}_x(n-1)\ \ and\ \ \vec{D}_x(n-2)$ in (EQ 15) and (EQ 16) is directly estimated from $\vec{D}_x(n-k)$ as

$$\vec{D}_x(n-p) = \left(\frac{p}{k}\right)\vec{D}_x(n-k)\ \ where\ \ p = 1\ \ or\ \ 2 \quad \text{(EQ 17)}$$

Note that $\vec{D}_x(n-k)$ is found from the motion estimation carried out to determine $block\ \ X_{n-k}(\vec{\alpha}1)$. The underlying assumption of (EQ 17) is that objects are in uniform motion for a small frame interval k. In practice, to compensate for possible errors due to approximations, a rectangle neighborhood surrounding the blocks of pixels given by (EQ 15) and (EQ 16) is searched for the most similar block to $block\ \ X_n(\vec{\alpha})$ in image frames Y_{n-1} and Y_{n-2}. Figure 11 shows the four most similar blocks and the corresponding vector relationships. Figure 12 illustrates the original image frame, the corresponding W-frame and the resulting error image. As in the case of B-frame, the W-frame appears to be a low-pass approximation of the original frame.

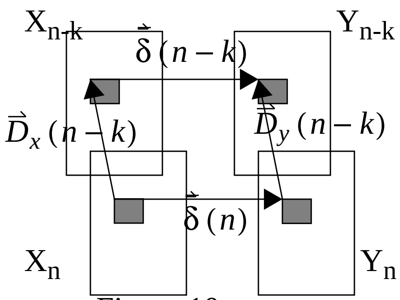


Figure 10
Disparity and Displacement vector diagram

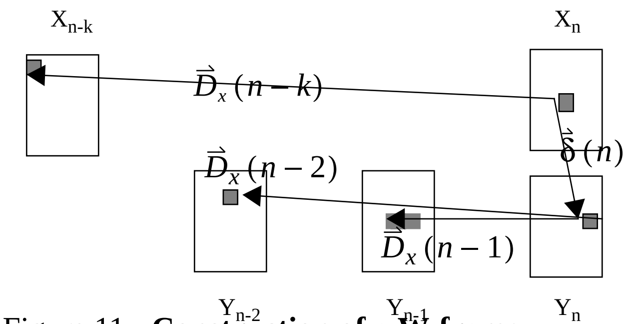


Figure 11 - **Construction of a W-frame**

Figure 12.

COMPRESSED STEREO IMAGE FRAME FORMAT

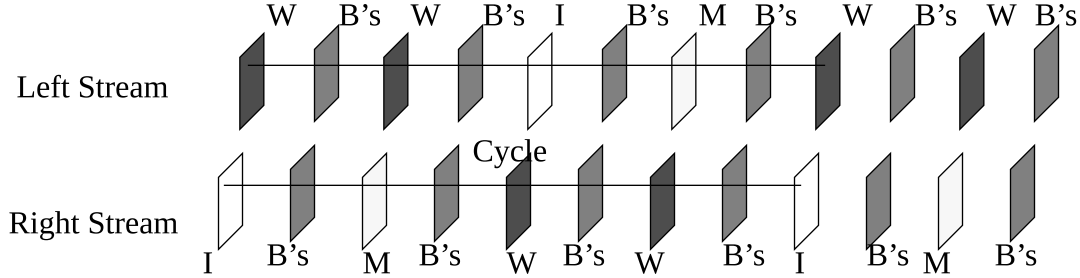
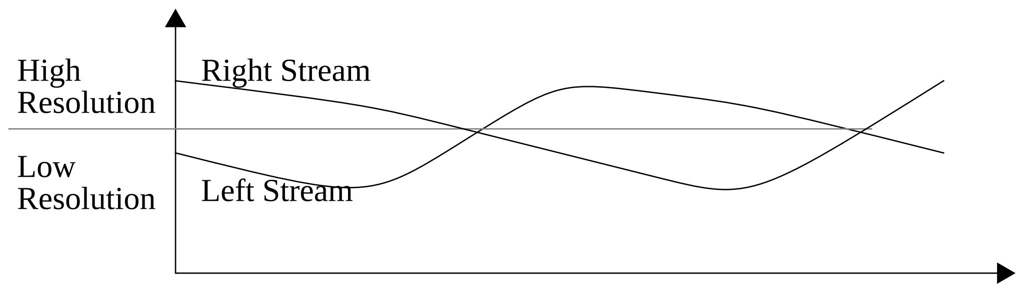


Figure 13
Compressed Stereo Image frame format

Figure 13 displays a segment of the compressed stereo image frame format. Note that the left and right compressed stereo image streams are cyclical. One cycle from each stream is depicted by a solid line in Fig 13. Each cycle has a frame period of 4k +4 frames where 4k is the total number of B frames in a cycle. Note that there are k B-frames between (I and M), (M and W), (W and W) and (W and I). The number of B frames per cycle is a design choice. It is a trade-off between compression gain and image quality. The compressed left stream leads the compressed right stream by half a cycle(2k+2 frames). This ensures that, at any given time, the compressed stereo image stream contains a high resolution(sharp) and a low resolution(low-pass) image pair. Research in stereoscopic vision in the human visual system has indicated that a stereo pair containing high and low resolution images has effectively no loss of image quality or depth perception[5,6]. The proposed stereo image compression scheme takes advantage of this phenomenon. It effectively masks the low resolution images by pairing them with high resolution images. Note that the low resolution images contribute to very high compression. Figure 14 displays this “smart masking” during one cycle period.



Right Stream:I B M B W B W B I

Left Stream; W B W B I B M B W

Figure 14 - Image quality of reconstructed stereo image stream

The compressed stereo image stream has a very high compression gain. This is because only a few frames, namely the main group frames, contain image data that require considerable bandwidth. The remaining frames require only a small fraction of the bandwidth needed to represent the main group frames since they contain only correlation information. The compression gain depends on the length of the image cycle. For a cycle of 15 frames, the compression gain of this scheme on a test image stream was approximately 40% greater than the corresponding compression gain of MPEG for the same reconstructed image quality

CONCLUSIONS AND FUTURE WORK

The application of monocular compression schemes in compressing stereo images is suboptimal since the correlation between left and right images is not considered. This paper describes new still and motion stereo compression schemes that are far more efficient than monocular compression schemes in compressing stereo streams.

The still stereo image compression scheme, direct integration, requires only an addition of 3% to 6% more bandwidth than a single image to represent a still stereo image. A decompressed still stereo image pair is of good stereoscopic image quality and exhibit essentially full depth resolution. A motion stereo image compression scheme is presented. It is unique in the sense that the left-right image correlation along with stereoscopic properties of the human visual system are used to achieve very high compression while retaining overall stereoscopic image quality. The gain of the proposed motion stereo image compression scheme was found to be about 40% greater than the corresponding compression gain of MPEG for a test stereo image sequence

Future efforts are focussed on decreasing the computational load required for the implementation of these compression schemes. Efforts are also underway to evaluate the robustness of the proposed compression schemes in handling different types of still and motion stereo images.

Acknowledgment

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

References.

[1] M. Rabbani and P.W. Jones, "Digital image compression techniques",SPIE Optical Engineering Press, Bellingham, Washington, 1991.

[2] K.R. Rao, P. Yip,. "Discrete Cosine Transform--Algorithms, Advantages, Applications", Academic Press, Inc. London, 1990.

[3] G.K. Wallace, "Overview of the JPEG (ISO/CCITT) still image compressions standard".Proceedings of SPIE, vol. 1244 (Feb. 1990), pp.220-233.

[4] Moving Pictures Expert Group, MPEG-2 Test Model 5, Doc. ISO-IEC/JTC1/SC29/WG11/MPEG93, Mar 1993.

[5] I. Dinstein, M.G. Kim, J. Tzelgov, A. Henik, "Compression of Stereo image and the Evaluation of its Effects on 3-D Perception",SPIE, Vol 1153, 1989.

[6] I. Dinstein, M. Guy, J. Rabany, J. Tzelgov, A. Henik, "On Stereo Image Coding", 9 th International Conference of Pattern Recognition.

[7] Marsha. J. Hannah, "Computer matching of areas in stereo images",STAN-CS-74-438, July 1974.

[8] A. N. Netravali, J. D Robbins, "Motion compensated television coding -part 1",Bell System Technical Journal, 58(3), pp.629-668,1979.

[9] Q. Wang, R.J. Clarke, "Motion estimation and compensation for image sequence coding", Signal Processing; Image Communication, Vol. 4. 1992.