A NEW FRAMEWORK FOR MULTI-VIEW IMAGE CODING

Halûk Aydınoğlu, Faouzi Kossentini and Monson H. Hayes, III

School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332

ABSTRACT

In this paper, we propose a new framework for the compression of multi-view image sequences. We define three types of frames, and each type is coded with a different strategy. The first type of frame is independently coded and is called I-frame. The second is a B-frame and is coded using a bi-directional disparity estimator and a modified version of the Subspace Projection Technique (SPT), as proposed in [1]. The SPT algorithm compensates the photometric variations between the multi-view frames. Projection block size is chosen to be small so that coding of the residual image is not necessary. On the other hand, to decrease the overhead information both disparity vectors and projection coefficients are coded with a lossy scheme. Finally, the third type of frame is a P-frame and is coded by employing a unidirectional disparity estimator and dc level compensation.

1. INTRODUCTION

The human brain can process the subtle differences between the image on the left retina and the image on the right retina to perceive a three-dimensional outside world. This ability is called stereovision. A stereoscopic system is used to artificially stimulate the stereovision ability. A stereopair, a pair of images of the same scene taken from two slightly different angles, is presented to the observer so that the left image will be seen by the left eye and the right image will be seen by the right eye. Moreover, the sense of 3-D vision can be improved such that the observer will have look-around capability. In order to achieve this objective the display device should track the head movements of the observer so that the displayed frames will change accordingly to allow motion parallax in the horizontal direction.

Coding of multi-view sequences is composed of two parts: compression of the multi-view image data and view interpolation. We are interested in the data compression problem. In Section 2, a new framework for this problem is presented. The proposed framework can be extended to include view interpolation. B and P type predictors, that are employed in this framework, are explained in Section 3. In Section 4, the codec structures are explained. Finally, in Section 6, the experimental results are presented.

This work was supported in part by the Joint Services Electronic Program, Grant No. DAAH-04-93-G-0027.

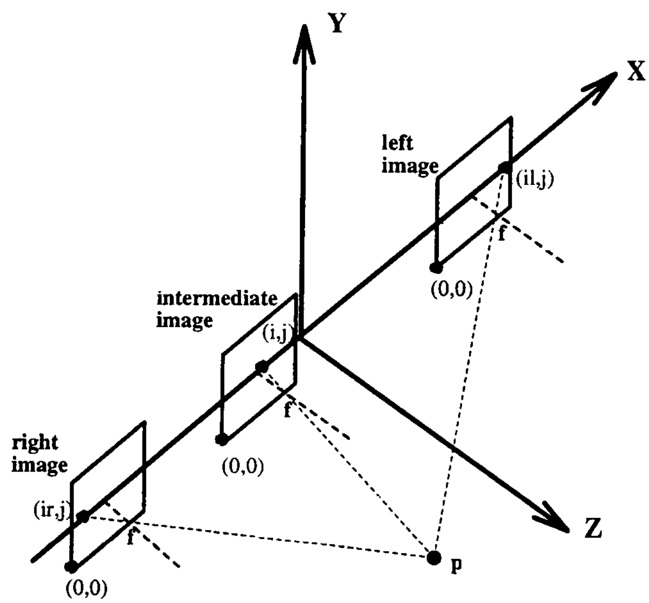

Figure 1: Multi-View Camera Geometry

2. THE FRAMEWORK

In the past, very few methods have been proposed for the coding of multi-view images. In [2], disparity compensation is proposed to reduce the inter-frame redundancy in a multi-view set. On the other hand, [3] investigated the view interpolation problem. Recently, the problem of compression of multi-view images has become popular [1] [4] [5] [6] [7].

Every object is displaced in a multi-view sequence. The displacement vectors are called disparity and they have only horizontal components if the camera geometry in Figure 1 is assumed. In this set-up, the cameras have parallel optical axes and coplanar optical centers. The displacements obtained from a multi-view sequence are generally larger than the displacements between successive frames in a video sequence. If a larger viewing angle is desired, independent coding of some of the frames is inevitable as in the MPEG coding standard. These frames are called I-frames. We employed a subband coding scheme for the I-frames as proposed in [8].

Occlusion is an important problem for multi-view images. The occlusion problem can be handled, if the inter-

0-7803-2431-5/95 $4.00 © 1995 IEEE

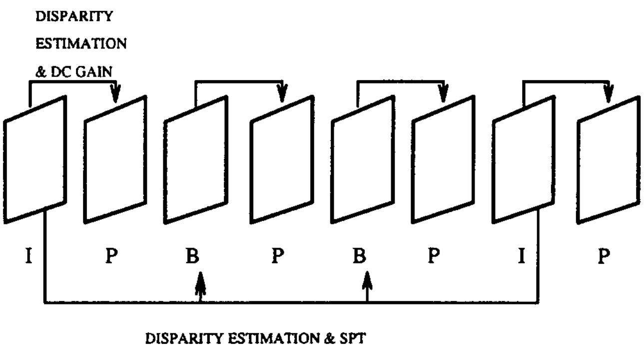

Figure 2: Frame Structure

mediate frames are coded based on both boundary I-frames by using a bi-directional disparity estimation.

The hypothesis of constant intensity along disparity trajectories is generally not valid for multi-view sequences. The photometric variations should be compensated either globally or locally. Empirical work has shown that local compensation methods are more satisfactory than the global compensation methods for coding applications, especially for space-varying characteristics of light reflection. We use a technique called subspace projection. This method compensates for dc-level and reflectivity variations. Our simulation results showed that after SPT is applied there is no need for either independent coding of some of the blocks or coding of disparity compensated residual frames.

B-frames are coded using both bi-directional disparity estimation and subspace projection technique. On the other hand, not all the intermediate frames are coded as B-frames. Some of the frames are coded with a uni-directional disparity estimation and dc-level compensation. These frames are called P-frames. The notion of P-frame is different from the P-frames in MPEG. Every other frame is a P-frame and each is coded based on either I-frames or B-frames. The structure of this coding scheme is illustrated in Figure 2.

Our approach is compatible to the special indexing characteristics of the multi-view sequences [1]. A multi-view sequence is indexed with respect to the position of the cameras. However, this indexing is meaningless at the display side. The frames are displayed depending on the observer's position. Contrary to some other techniques [4] [7], our algorithm emphasis this special indexing characteristics and, therefore, requires a smaller buffer at the display side.

In the next section we will explain the two types of predictors, B-frame and P-frame predictors.

3. B-FRAME AND P-FRAME PREDICTORS

Each block $\mathbf{b}_i$ from a B-Frame is estimated from the left and right I-frames. We run a disparity compensation algorithm, as in [9], to find the best matching blocks $\hat{\mathbf{b}}_{i_l}$ and $\hat{\mathbf{b}}_{i_r}$, from left and right I-frames respectively. An initial estimate $\hat{\mathbf{b}}_i$ is equal to either $\hat{\mathbf{b}}_{i_l}$ or $\hat{\mathbf{b}}_{i_r}$, and it is determined adaptively. The disparity vector corresponding to this estimate is sent through the channel. These disparity vectors can be near losslessly coded by employing an entropy-constrained residual scalar quantizer. This algorithm will be discussed in the next section.

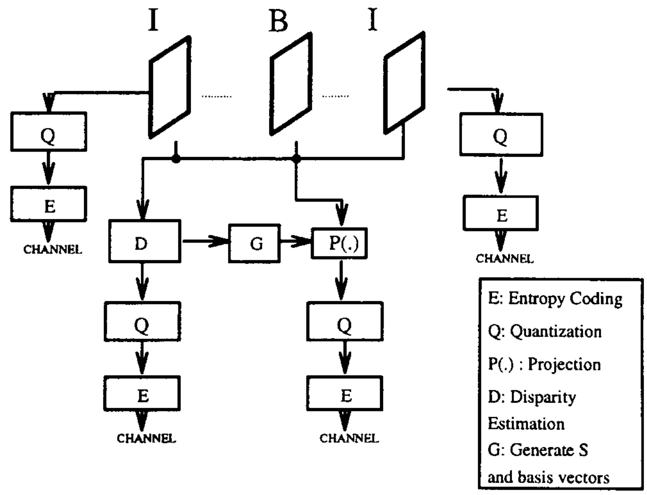

Figure 3: Block Diagram for B-Frame Encoder

Our aim is to modify the original estimate so that coding the disparity compensated difference images will not be necessary. We used an orthonormal basis approach for achieving this modification. The general idea of our approach is simple. We treated each block $\mathbf{b}_i$ as a vector. Our aim is find a subspace $\mathcal{S}$ of $\mathcal{R}^{n\times n}$ on which we can project these vectors. As with most compression techniques, we are concerned with reducing the dimensionality of the data to be processed. Thus, we wish to find a smaller dimensional subspace in which to accurately represent the vectors obtained from intermediate frame blocks. A varying subspace is more appropriate for this purpose. The estimates obtained from I-frames can be used in the span of this varying subspace. Also fixed vectors should be used to code the occlusion regions and compensate for illumination differences.

Our choice is to generate a vector subspace $\mathcal{S}$ for each vector $\mathbf{b}_i$ as the span of a constant dc term $\mathbf{C}$ and the best matching block $\mathbf{b}_i$ of the initial estimate. We generated an orthogonal basis and then projected each intermediate frame vector $\mathbf{b}_i$ onto the corresponding subspace $\mathcal{S}$ to obtain the modified vector $\hat{\mathbf{b}}_{im}$. We need to transmit the projection coefficients to generate the estimate at the decoder. The projection coefficients can be compressed for further bit rate reduction. This compression is also achieved by using entropy-constrained residual scalar quantization. However, the structure of the coder is slightly different than the one that is used for disparity vectors. In Figure 3 a block diagram of the encoding algorithm for B-frames is presented. Details of the SPT algorithm as well as other performance issues can be found in [1].

At the decoder, knowledge of the disparity will be enough to reconstruct each subspace. We code the disparity vectors near losslessly, and use the coded disparity values for constructing and reconstructing the subspaces. We also decode the projection coefficients and generate the coded vectors. We, then use these vectors to generate the B-frames.

The P-frames are coded either from I-frames or B-frames (Figure 2). A disparity compensation algorithm is employed to find the best matching blocks from the previous frame. The dc-level of each best matching block is corrected so that average intensity of the estimated block is equal to the average intensity of the original block. Both disparity vectors and dc shifts are also coded lossy.

4. ENTROPY-CONSTRAINED RESIDUAL SCALAR QUANTIZATION

Both the disparity vectors and the projection coefficient matrices are coded using a high order entropy-constrained residual scalar quantizer (HO-EC-RSQ), a special quantizer of the one addressed in [10]. The quantizer consists of a sequence of P stage scalar quantizers, each containing N_p ($1 \leq p \leq P$) levels. The output of each of the stage residual scalar quantizers (RSQ) is the input to one of many stage entropy coders as specified by a finite-state machine (FSM). The FSM derives its states based on the previously coded conditioning symbols as described in [10]. The entropy coder used in this work is a standard adaptive arithmetic coder, although other statistical-based entropy coders can be used.

Encoding using the HO-EC-RSQ is very simple. The RSQ codebooks are searched using a tree-based searching algorithm, which is usually the standard M-search algorithm. The state is then determined, and for each sequence of symbols emanating from each stage RSQ, a particular arithmetic encoder outputs a bit stream, which is then sent to the channel. Decoding involves using the corresponding arithmetic decoders to recover the symbols, which are then used to look-up the levels in the stage RSQ codebooks. Since the FSM employs information that is available to the decoder, both sides can use the same FSM and no side information is required.

The distortion measure used in the design HO-EC-RSQs for the projection coefficient matrices is the well-known squared error measure. However, we employ the absolute measure in the design of the disparity vector HO-EC-RSQs because each disparity vector is equally important (at least from the viewpoint of the quantizer) and there is no advantage to amplifying large differences. Moreover, the absolute measure is usually simpler to implement. The design procedure is the same for both measures, and is based on a Lagrangian minimization where encoder, decoder, and entropy coder are iteratively optimized in a complexity and entropy-constrained framework. The only exception is that the centroid used in the decoder optimization is different. It is the mean for the popular squared error measure, and it is the median for the absolute measure. Details of the design algorithm as well as the complexity issues can be found in [10].

The multistage residual structure allows relatively fast encoding/decoding, requires very low memory, and provides a framework where efficient and effective high order conditional entropy coding is performed. Moreover, it allows HO-EC-RSQ to operate over a very large range of bit rates, thereby making it suitable for near-lossless coding (as is usually required for disparity vectors) and high-distortion low-rate coding (typically allowed when coding the projection coefficient matrix), simultaneously. By operating over a wide range of bit rates, bit allocation between the disparity vectors and the projection coefficient matrices, performed here using the generalized BFOS algorithm, becomes very accurate.

Fr.No	dv.	c1	c2	total(bpp)	PSNR(dB)
1 (I)	-	-	-	0.64	37.07
2 (P)	0.034	0.023	-	0.057	32.17
3 (B)	0.066	0.021	0.10	0.187	30.53
4 (P)	0.038	0.031	-	0.069	28.92
5 (B)	0.07	0.022	0.094	0.186	30.01
6 (P)	0.043	0.018	-	0.061	29.94
7 (I)	-	-	-	0.64	37.13
8 (P)	0.036	0.025	-	0.061	28.06

Table 1: Experimental Results
dv=disparity vector, c1=coefficient 1, c2=coefficient2

5. EXPERIMENTAL RESULTS

The proposed framework was tested on the "Lab" sequence. The sequence is composed of eight 512 by 480 frames. Frame 1 and frame 7 are coded independently at 0.64 bpp. The PSNRs for these frames are 37.07 and 37.13 dB respectively. A block-based estimation algorithm is employed for obtaining disparities. The block size is 8 by 8 and mean square error is the similarity measure. The disparity range is 32 pixels for P-frames, and 120 pixels for the B-frames. However, disparity vectors may have both negative and positive values for B-frames as matching blocks are obtained adaptively from left and right I-frames. The block size for SPT algorithm is 4 by 4 and it is only applied to B-frames. The experimental results are presented in Table 1. Bit rates for disparity vectors, coefficient 1 (dc gain) and coefficient 2 (camera gain) and overall bit rates are given. In Figure 4 and 5 examples of coded P and B frames are presented.

6. CONCLUSION

We have tested a new framework for compression of multi-view sequences. Previous study showed that lossless compression of disparity vectors increases the bit-rate. For that reason, we employed a lossy scheme for disparity vectors. This scheme should be further improved such that distortion for blocks with vertical edges should be penalized. We are currently working on this topic. We also believe that, coding of coefficient 2 can be improved. If these two tasks are achieved, further reduction in bit rate and better reconstruction quality can be obtained.

7. REFERENCES

[1] H. Aydınoğlu and M. H. Hayes, "Compression of Multi-View Images," in *International Conference on Image Processing*, 1994.

[2] M. E. Lukacs, "Predictive Coding of Multi-Viewpoint Image Sets," in *ICASSP*, pp. 521–524, IEEE, 1986.

[3] J. Liu and R. Skerjanc, "Construction of Intermediate Pictures for a Multiview 3D System," in *Stereoscopic*

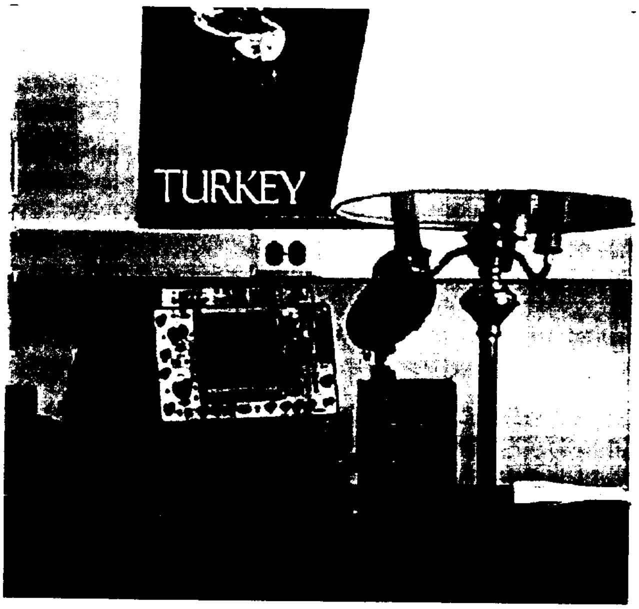

Figure 4: Coded P-frame: 32.17dB 0.057bpp

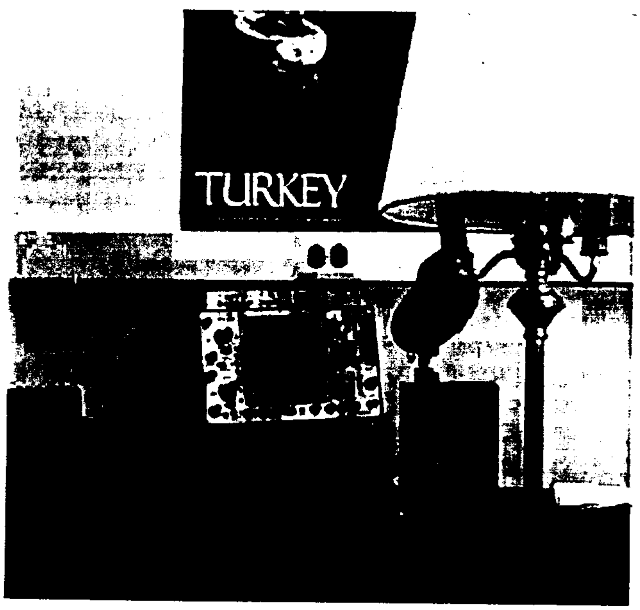

Figure 5: Coded B-frame: 30.53dB 0.187bpp

Displays and Applications III, vol. 1669, pp. 10–19, SPIE, 1992.

[4] T. Naemura and H. Harashima, "Self-Similarity Modeling for Interpolation and Extrapolation of Multi-Viewpoint Image Sets," in *ICASSP*, vol. V, pp. 369–372, 1994

[5] T. Fujii and H. Harashima, "3-D Image Coding Based on Affine Transform," in *ICASSP*, vol. V, pp. 557–580, 1994.

[6] T. Fujii and H. Harashima, "Data Compression of an Autostereoscopic 3-D Image," in *Stereoscopic Displays and Applications V, San Jose, CA*, vol. 2177A, SPIE, February 1994.

[7] R. Hsu, K. Kodama, and H. Harashima, "View Interpolation Using Epipolar Plane Images," in *ICIP*, vol. II, pp. 745–750, 1994.

[8] J. C. Darragh, *Subband and Transform Coding of Images*. PhD thesis, University of California, Los Angeles, Los Angeles, CA, 1989.

[9] H. Yamaguchi, Y.Tatehira, K. Akiyama, and Y. Kobayashi, "Stereoscopic Images Disparity for Predictive Coding," in *ICASSP*, pp. 1976–1979, IEEE, 1989.

[10] F. Kossentini, W. Chung, and M. Smith, "Image coding using high-order conditional entropy-constrained residual VQ," in *ICIP*, (Austin, Texas), nov 1994.