THE COMPRESSION OF SIMPLIFIED DYNAMIC LIGHT FIELDS

Shing-Chow Chan, King-To Ng, Zhi-Feng Gan, Kin-Lok Chan, *and* *Heung-Yeung Shum**

Department of Electrical and Electronic Engineering
The University of Hong Kong
{ scchan, ktng, zfgan, kinlok }@eee.hku.hk

* Microsoft Research, Asia
hshum@microsoft.com

ABSTRACT

This paper studies the compression of a dynamic IBR representation called the simplified dynamic light fields (SDLF). It is obtained by constraining the viewpoints in a dynamic environment along a line instead of a 2D plane. The simplified dynamic light fields have a dimensionality of four, which considerably simplifies their capturing and data compression. A new coding algorithm for this simplified dynamic light fields using a modified MPEG2 algorithm is proposed. It employs both temporal and spatial predictions from the reference video streams to better explore the redundancy among the light field images. Experimental results, using a synthetic SDLF, show that the proposed compression scheme offers a 2 dB improvement in PSNR over a similar coding scheme using only temporal prediction.

1. INTRODUCTION

Image-based rendering (IBR) is an emerging and promising technology for rendering photo-realistic views of scenes and objects from a collection of densely sampled images. Central to IBR is the plenoptic function, which is a versatile generalization of traditional images and videos and new framework for developing revolutionary virtual reality and visualization systems such as interactive or new immersive television systems. Another important advantage of IBR is the superior image quality that it offers over 3D model building, especially for very complicated real world scenes. Furthermore, it requires much less computational power for rendering, regardless of the scene complexity. Thus, it is also very useful in re-rendering complicated synthetic environment, which can be very time consuming. Unfortunately, IBR representations usually consist of hundreds or thousands of images, which involve a large amount of data. To simplify the capturing and storage, various IBR representations with lower dimensions have been advocated [2-7].

One of the important problems in IBR research is the compression of dynamic IBR representations of real world scene [16] or synthetic data. Most IBR representations reported so far deal with static scenes. This is largely attributed to the logistical difficulties in capturing and transmitting dynamic representations, which involve a huge amount of data. The latter has simulated considerable research effort into efficient compression methods for various IBR representations such as the light field, lumigraph, and concentric mosaics [10-14]. In this paper, we study the compression of a four-dimensional dynamic IBR representation, called the simplified dynamic light field. More precisely, we focus on the light field of dynamic scene with viewpoints being constrained along a line instead of a 2D plane to obtain a simplified plenoptic function with four dimensions. This can greatly reduce the complexity of the dynamic IBR system. At the same time, the user can still observe significant parallax and lighting changes along the horizontal direction. Furthermore, the given number of cameras can be used to maximize the sampling rate along the horizontal direction and thus reduce the risk of insufficient sampling in a 2D configuration with the same number of cameras and horizontal panning range. The proposed compressed algorithm is a generalization of the MPEG2 algorithm for coding multiple adjacent video streams in the simplified dynamic light fields. It employs both temporal and spatial predictions to better explore the redundancy in dynamic light fields. Spatial prediction or disparity compensated prediction has been used in coding of static light fields [10-12] and stereo images [19]. The coding algorithm considered here can be viewed as their generalization to the dynamic situation. Experimental results were performed to evaluate the efficiency of this coding scheme. Using a synthetic dynamic light field, it was found that spatial prediction significantly improves the coding efficiency. Together with temporal prediction, most of the macroblocks can be predicted satisfactorily, yielding about 2 dB improvements in performance.

The paper is organized as follows: a brief introduction to the plenoptic function and the simplified dynamic light field (SDLF) is given in Section 2. The proposed compression algorithm for SDLFs is described in Section 3. Experimental results are presented in Section 4. Finally, conclusions are given in Section 5.

2. SIMPLIFIED DYNAMIC LIGHT FIELD (SDLF)

Central to IBR is the *plenoptic function*, which was first used by Adelson and Bergen [1] to describe all the radiant energy that can be perceived by the observer at any point (V_x, V_y, V_z) in space and time τ. At each point in space, we can select any of the viewable rays by choosing an azimuth and elevation angle (θ, ϕ) as well as a band of wavelengths, λ. For dynamic scenes, there is an additional time variable τ. Therefore, the plenoptic function p has seven dimensions. Due to its high dimensional nature, data reduction or compression of the plenoptic function is essential to IBR systems. The simplest two-dimensional plenoptic function is the panorama [2,6]. A panorama can be constructed by first taking a set of images at different angles along a given axis. The images are then re-projected onto a cylinder to create the panoramic image. During rendering, part of the panoramic image is re-projected onto the screen to produce the view at a given angle. QuickTime VR was one of the first virtual reality systems using this concept [2]. Chen and Williams' view interpolation [8] and McMillan and Bishop's plenoptic modeling [5] also used a similar technique. Other important four-dimensional IBR representations are the lumigraph of Gortler et al [3] and the light field of Levoy and Hanrahan [4]. They demonstrated that new views of a scene could be rendered from images taken at points in a 2D plane, as illustrated in Figure 1. Using this *2D array of images*, it is possible to render different views of the object or scene at different viewing angles. By scarifying to different extents the degree of freedom in space, lumigraph and light field reduce the dimensionality of the plenoptic function to four, after ignoring wavelength and time. The *lumigraph* differs from the light field in that an *approximate geometric model* of the object, in addition to the 2D array of images, can be used to improve the quality of the reconstruction at *lower sampling densities*. Another simpler IBR representation is the 3D concentric mosaics [7], which is obtained by constraining the viewpoints on a horizontal plane. An interesting 3D dynamic IBR representation is the panoramic video or time-varying environment map [2]. Panoramic video is a sequence of panoramas created at different locations along a path in space, which can be used to capture dynamic scenes at a stationary location or in general along a path with 360 degree of viewing freedom [18].

One of the important problems in IBR research is the *capturing* and *compression of dynamic IBR representations*. By dynamic IBR, we meant higher dimensional plenoptic function with time variable, other than traditional videos. For example, if multiple video cameras are employed in capturing the light field or lumigraph in Figure 1, we obtain a 5D plenoptic function or dynamic light field. This will allow us to render time-varying or dynamic scenes at any position on the 2D plane and develop photo-realistic and interactive virtual reality and visualization systems. Unfortunately, from the sampling analysis in [15], the sampling rate of a static scene will depend on the depth and its variation of the scene. A large number of cameras in a 2D arrangement, say 64×64 might be needed. This creates hundreds of videos, which have to be compressed and stored in real-time. To avoid such a high dimensionality and the excessive hardware cost, we shall focus on light field with viewpoints being constrained along a line instead of a 2D plane. This simplified dynamic light field (SDLF) has a dimensionality of four, i.e. (u,v,s,τ). Apart from the simplicity of the overall system, there are several reasons for such a choice. First of all, the user can still observe significant parallax and lighting changes along the horizontal direction. Secondly, the given number of cameras can be used to maximize the sampling rate along the horizontal direction and thus reduce the risk of insufficient sampling in a 2D configuration with the same number of cameras and horizontal panning range. In this paper, we shall limit our scope to the compression of this simplified dynamic light field for synthetic scenes. The capturing problem will be reported in future publications. The main objective is to reduce the re-render time for time-varying computer graphics, as we shall demonstrate later in Section 4. The proposed algorithm, however, also applies to real-world scenes.

3. COMPRESSION OF SDLF

Since dynamic IBR representations usually have large data sizes, an efficient compression algorithm is always desirable to reduce the amount of storage and bandwidth for transmission. In simplified dynamic light fields (SDLF), video streams are obtained by sampling the plenoptic function in a 1D array. The adjacent video streams will appear to be shifted relative to each other. In order to explore the correlation in the SDLF, we divide the video streams into groups and compress them together. The proposed method for coding the SDLF is shown in Figure 3. Only three videos are shown for simplicity, and we call it a group of field (GOF). To provide *random access to individual pictures*, we have adopted a modified MPEG-2 video compression algorithm [9] to encode the image frames. There are two types of video streams in the proposed dynamic light field: *main* and *secondary* video streams. Main video streams are encoded using the MPEG-2 algorithm, which can be decoded without reference to other video streams. The image frames in a main stream are divided into I-, P-, and B-pictures, where I-pictures are coded using intra-frame DCT-based transform coding, while P-pictures are coded by hybrid motion compensated/transform coding using previous I- or P-pictures as references. B-pictures are coded by a similar method except that forward and backward motion compensations are performed by using nearby I- or P-pictures as references, these are indicated by the block arrows in Figure 3. The light field images captured at the same time instant as the I-pictures in a main stream constitute an *I-field*. Similarly, we define the *P-* and *B-fields* as the light field images containing respectively the P- and B-pictures of the main video stream. Pictures from the secondary stream in the I-field are encoded using "*spatial prediction*" from the reference I-picture in the I-field. It is because adjacent light field images appear to be shifted relative to each other, similar to the effect of linear motion in video coding. Such displacement of pixels is called *disparity* and is related to the objects and viewing geometries. This kind of prediction, which is called disparity compensated prediction, has been used in coding of static light fields [10-12]. The coding algorithm considered here can be viewed as their generalization to the dynamic situation. Pictures from the secondary stream in a P-field are predicted using spatial prediction from adjacent P-picture in the main stream, and the forward motion compensation from the reference I- or P- fields in the same secondary stream. Pictures from the secondary stream in B-field are predicted using spatial prediction from adjacent B-picture in the main stream, and the forward/backward motion compensation from nearby reference I- and/or P- fields in the same secondary stream.

For simplicity, we have only included one main stream in each GOF. More sophisticated disparity compensation schemes such as bi-directional prediction with multiple main streams can be incorporated in a single GOF or successive GOFs. Our scheme can also be generalized to 2D GOFs in the compression of 5D dynamic light fields, with main streams distributed on certain points in the 2D array, instead of a 1D array considered here. In order to maintain a uniform reconstruction quality among the SDLF, we allocate a higher bit rate to the main stream than the secondary streams because the I-pictures in the main stream usually require considerably more bits than P- and B-pictures. Furthermore, the rate control algorithm of the MPEG2 Test Model 5 is used to prevent buffer overflow and underflow problems, although other more sophisticated rate control algorithms can also be applied. Efficient accessing mechanism is a major requirement in IBR compression. For static and dynamic light fields and lumigraphs, random access at the *pixel level* is required. As most existing compression algorithms employ entropy coding, such as Huffman or arithmetic coding, the symbols after compression will be variable sizes. It is, therefore, very time-consuming and memory intensive to retrieve and decode a single line or pixel from the compressed data if there is no such provision for random access. To address the random access problem, *pointers* are embedded into the compressed data stream as in [11,13,17]. During rendering, the required macroblocks will be selectively decoded from the compressed data stream. However, these pointers add the overheads to the compressed data stream.

4. EXPERIMENTAL RESULTS

The SDLF used in our experiments is synthetic and it was rendered by using the 3D Studio Max software. Figure 2 shows part of the SDLF used in this study. The data sets consist of 16×1 24-bit RGB videos with 320×240 pixels and 24 frames per second. This SDLF is compressed using the proposed coding algorithm. Coding results for different number of video streams in a group of field (GOF) are investigated, and they are plotted in Figure 4. For SP3, we have three video streams in the GOF as illustrated previously in Figure 3. For SP5 and SP7, we have five and seven video streams, respectively. As a comparison, we also compressed all the video streams of the synthetic SDLF by MPEG2 algorithm independently. It can be seen that the performances of the proposed algorithm using both temporal and disparity compensations have significant improvement, about 2 dB, over the independent coding scheme. This shows that there is significant amount of spatial redundancy among the video sequences. When the number of video streams in the GOF, and hence the number of secondary streams, is increased, the PSNR improves because less I-pictures are coded and better disparity prediction is obtained in the SDLF. However, the difference between SP5 and SP7 is small because disparity compensation will be less effective when video streams are far apart. Figure 5 shows two typical reconstructed images in the SDLF using SP7. They show good quality of reconstruction at a bit rate of 194k bits/s per video stream (compression ratio=228). In order to study the performance of the spatial prediction, we calculated the number of macroblocks used in different prediction types and they are summarized in Table 1. At a bit rate of 1.78M bits/s per stream, secondary video streams which are next to the main video stream (distance *d*=1) have 35.2% of their macroblocks predicted by disparity compensation prediction. When the distance (*d*) increases, there are fewer macroblocks predicted spatially. This drops to 29.4 % when the distance is increased to 3. The reason is that the prediction will become more difficult when the distance from the main video stream increases. This might be improved by using bi-directional disparity compensation prediction. Furthermore, it is noted that this percentage depends on the target bit rate. For example, when we decrease the bit rate, more macroblocks (up to 50%) will employ spatial predication.

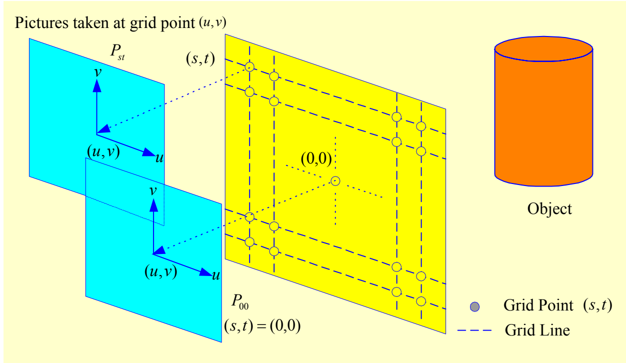

Fig. 1. Light field and lumigraph capturing.

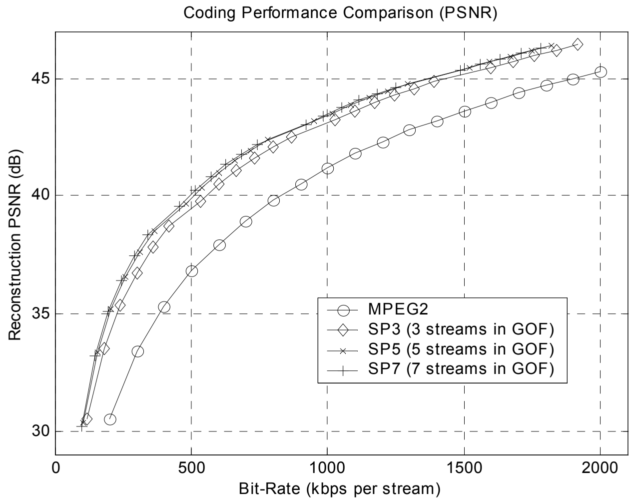

Fig. 4. Coding results of the simplified dynamic light field.

	Main Stream	Secondary Stream		
		d = 1	d = 2	d = 3
97Kbps				
Intra MB	9.3%	0.1%	0.2%	0.3%
Temporary MB	90.7%	44.5%	49.0%	50.9%
Spatial MB	0.0%	55.4%	50.8%	48.8%
743Kbps				
Intra MB	9.0%	0.1%	0.2%	0.2%
Temporary MB	91.0%	62.1%	66.8%	67.4%
Spatial MB	0.0%	37.8%	33.0%	32.4%
1.78Mbps				
Intra MB	9.0%	0.1%	0.2%	0.2%
Temporary MB	91.0%	64.7%	69.4%	70.4%
Spatial MB	0.0%	35.2%	30.4%	29.4%

Table 1. The number of macroblocks used for different types.

5. CONCLUSION

A new compression algorithm for coding a dynamic IBR representation called the simplified dynamic light fields (SDLF) is presented. The SDLF is obtained by constraining the users' viewpoints in a dynamic environment along a line, instead of a 2D plane in the static light field. The SDLF have a dimensionality of four, which considerably simplifies its capturing and data compression. The proposed coding algorithm is based on a modified MPEG2 algorithm. It employs both temporal and spatial predictions from the reference video streams to better explore the redundancy among the light field images. Experimental results, using a synthetic SDLF, show that the proposed compression scheme offers a 2 dB improvement in PSNR over a similar coding scheme using only temporal prediction.

REFERENCES

[1] E. H. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, pp. 3-20, MIT Press, Cambridge, MA, 1991.

[2] S. E. Chen, "QuickTime VR – an image-based approach to virtual environment navigation," in *Proc. of the annual conference on Computer Graphics (SIGGRAPH'95)*, pp. 29-38, Aug. 1995.

[3] S. J. Gortler, R. Grzeszczuk, R. Szeliski and M. F. Cohen, "The lumigraph," in *Proc. of the annual conference* on *Computer Graphics (SIGGRAPH'96)*, pp. 43-54, Aug. 1996.

[4] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. of the annual conference on Computer Graphics (SIGGRAPH'96)*, pp. 31-42, Aug. 1996.

[5] L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," in *Proc. of the annual conference on Computer Graphics (SIGGRAPH'95)*, pp. 39-46, Aug. 1995.

[6] R. Szeliski and H. Y. Shum, "Creating full view panoramic image mosaics and environment maps," in *Proc. of the annual conference on Computer Graphics (SIGGRAPH'97)*, pp. 251-258, Aug. 1997.

[7] H. Y. Shum and L. W. He, "Rendering with concentric mosaics," in *Proc. of the annual conference on Computer Graphics (SIGGRAPH'99)*, pp. 299 - 306, Aug. 1999.

[8] S. E. Chen and L. Williams, "View interpolation for image synthesis," in *Proc. of the annual conference on Computer Graphics (SIGGRAPH'93)*, pp. 279-288, Aug. 1993.

[9] ITU-T Recommendation H.262|ISO/IEC 13818-2, "Information Technology-Generic Coding of Moving Pictures and Associated Audio Information: Video," Nov. 1994.

[10] M. Magnor and B. Girod, "Data compression for light-field rendering," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, no. 3, pp. 338-343, April 2000.

[11] X. Tong and R. M. Gray, "Coding of multi-view images for immersive viewing," in *Proc. of IEEE Intl. Conf .on Acoustics, Speech, and Signal Processing,* vol. 4, pp. 1879-1882, June 2000.

[12] C. Zhang and J. Li, "Compression of Lumigraph with multiple reference frame (MRF) prediction and just-in-time rendering", in *Proc. of IEEE Data Compression Conference*, pp. 253-262, Mar. 2000.

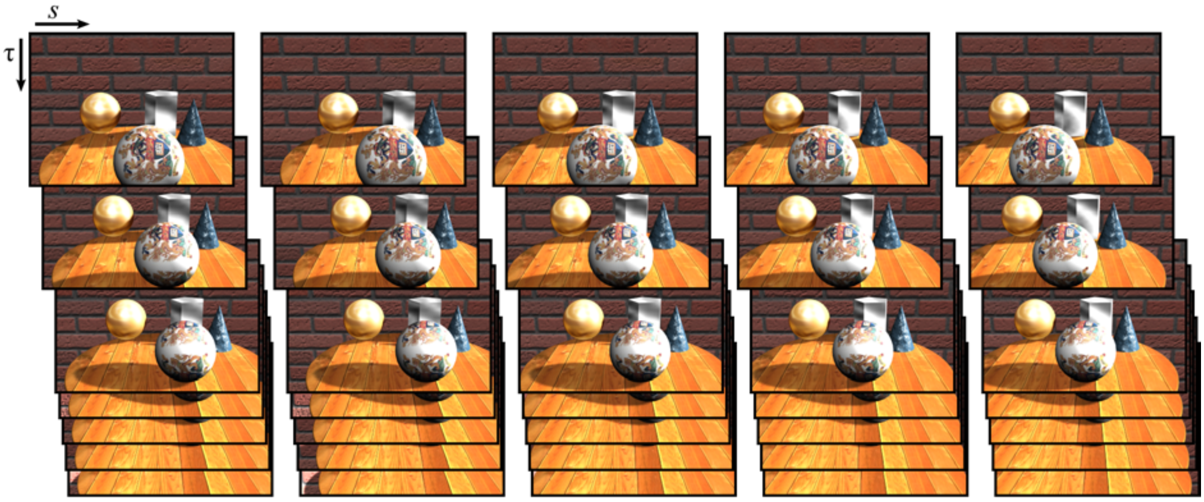

Fig. 2. Simplified dynamic light field images.

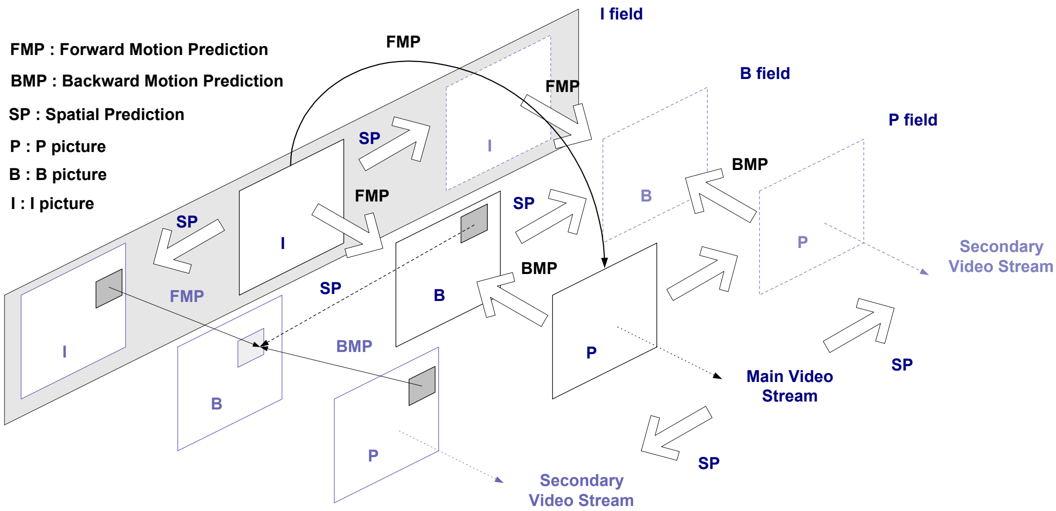

Fig. 3. Compression of 4D simplified dynamic light field.

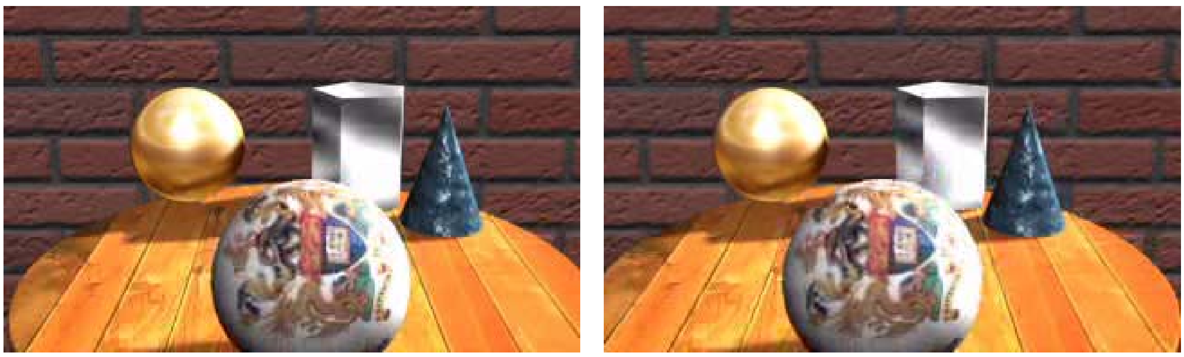

Fig. 5. Frame 488 of reconstructed images in the SDLF, from main (left) and secondary (right) video streams (194kbps per stream).

[13] H. Y. Shum, K. T. Ng and S. C. Chan, "Virtual reality using the concentric mosaic: construction, rendering and data compression," in *Proc. of IEEE Intl. Conf. on Image Processing*, vol. 3, pp. 644-647, Sept. 2000.

[14] C. Zhang and J. Li, "Compression and rendering of concentric mosaics with reference block codec (RBC)", in *Proc. of SPIE Visual Communication and Image Processing*, vol. 4067(05), June 2000.

[15] J. X. Chai, X. Tong, S. C. Chan and H. Y. Shum, "Plenoptic sampling," in *Proc. of the annual conference on Computer Graphics (SIGGRAPH'00)*, pp. 307-318, July 2000.

[16] B. Wilburn, M. Smulski, K. Lee, and M. Horowitz, "The Light Field Video Camera," *Proc. Media Processors 2002, SPIE Electronic Imaging 2002*.

[17] H. Y. Shum, K. T. Ng and S. C. Chan, "A virtual reality system using the concentric mosaic: construction, rendering and data compression," accepted for publication in *IEEE Trans. Multimedia*, 2002.

[18] K. T. Ng, S. C. Chan, H. Y. Shum, and S. B. Kong, "On the data compression and transmission aspects of panoramic video," in *Proc. of IEEE Intl. Conf. on Image Processing,* vol. 3, pp. 644-647, Oct. 2001.

[19] W. Woo and A. Orteta, "Stereo image compression based on the disparity compensation using the MRF model", in *Proc. of SPIE Conf. on Visual Comm. and Image Processing*, vol.2727, pp.28-41, 1996.