THE APPLICATION OF NONLINEAR FILTER BANKS TO EFFICIENT RENDERING AND PROGRESSIVE TRANSMISSION OF LIGHT FIELDS

King-To Ng, Shing-Chow Chan   and   *Heung-Yeung Shum*

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

Microsoft Research, Asia

hshum@microsoft.com

ABSTRACT

This paper studies the application of perfect reconstruction nonlinear filter banks (NFB) to the efficient rendering and progressive transmission of light fields. The reference pictures in the conventional disparity-compensated prediction encoder is decomposed using the NFB to reduce the amount of main memory needed to support fast rendering. The NFB has very low arithmetic complexity for reconstruction and small filter support which considerably simplifies the random access operations. It can also be applied to the predicted light field images to support progressive transmission. Different prediction and reconstruction strategies are also investigated to achieve different tradeoffs between memory requirement and decoding speed.

1. INTRODUCTION

Image based rendering (IBR) has recently emerged as a simple yet powerful photo-realistic representation of real world scenes [1-7]. Its basic principle is to render new views of a scene using rays that were previously captured in densely sampled pictures of the scene. It has plenty of applications including virtual walkthrough, electronic games, medical simulation, visualization, and many others, which require virtual camera motions. IBR provides superior image quality than 3D model building for very complicated real world scenes and requires much less computational power for rendering, regardless of the scene complexity. Central to IBR is the plenoptic function [2], which describes all the radiant energy that is perceived by the observer at any point in space and time. Due to its high dimensional nature, data reduction or compression of plenoptic function is essential to IBR systems. The dimension of the plenoptic function to be captured can be reduced by restricting the degree of viewing freedom in space. The light field [3] and lumigraph [4] reduce the dimensionality of the plenoptic function to four while the concentric mosaics [7] further reduce the dimensionality to three. In addition, it is possible to exploit the redundancy of the plenoptic function using waveform coding or other model-based techniques. Another important problem unique to compression/decompression of the high dimensional plenoptic function or IBR representations is to support efficient *random access* to the compressed data for real-time rendering. Unlike video coding, which supports random access at the *picture level*, higher dimensional IBR representations such as 4D light fields and lumigraphs require random access at the *pixel level*. As most existing compression algorithms employ entropy coding for better compression ratio, the symbols after compression will be of variable size. It is very time-consuming to retrieve and decode a pixel from the compressed data if there is no such provision for efficient random access. Also, the amount of raw data in the high dimensional plenoptic function becomes very large that it is very expensive to decode the complete bitstream in the main memory for rendering. The efficient compression, rendering and transmission of IBR representations are thus active subjects of

recent research [11-19]. While there are satisfactory solutions for the compression and rendering of lower dimensional IBR representations such as the concentric mosaics [8] and the panoramic video [13], similar issues in light fields and lumigraphs are far more complicated, due to their higher dimensional nature and lower level of random access requirement. An effective method for exploring the correlation in light field images is based on disparity-compensated prediction (DCP). Since the light field and lumigraph are obtained by sampling the plenoptic function in a 2D plane to form a 2D array of images of the scene, adjacent light field images appear to be shifted relative to each other, similar to linear motion in video coding. Such displacement of pixels is called disparity and is related to the object as well as the viewing geometries. Analogous to motion-compensated prediction, DCP can be employed to predict a given image from some reference images that are previously encoded. For fast rendering speed, most DCP based coders usually assume that the reference pictures are decoded and loaded into the memory [17] or in compressed form using vector quantization [16]. The number of reference pictures usually varies from 1/4 to 1/16 of the total light field images. This requires considerable amount of fast main memory, especially when the resolution gets higher and higher and there are several coexisting IBR representations in the scene. It is therefore desirable to devise simple and effective method for trading the memory requirement with decoding speed, as the processing power of workstations and desktops will continue to improve. One possible solution is to decompose each of these reference images into a series of subband signals with different resolutions using a perfect reconstruction filter bank (PR FB). The lowest resolution can either be decoded completely and stored in the main memory for fast rendering or further transcoded (not necessary if they are vector quantized) by Vector Quantization (VQ) to further reduce the amount of storage with a slight loss of performance. During rendering, the compressed data in higher resolutions are selectively decoded online before adding back to the lowest resolution to reconstruct the desired pixels in the reference images. In this paper, we study the application of the nonlinear filter bank proposed in [9] to the above problem. The main reason for employing a nonlinear perfect reconstruction filter bank, instead of the commonly used 9/7 wavelet filter bank, is its low arithmetic complexity and small filter support, which considerably simplifies the random access operations in rendering light field images, as we shall see later in Section 3. The same decomposition can be applied to the predicted light field images to support progressive transmission, at an expense of a slight decrease in coding efficiency. The paper is organized as follows: a brief introduction to disparity-compensated prediction (DCP) is given in Section 2. The application of the nonlinear filter bank to the rendering and progressive transmission of light fields is presented in Section 3. Experimental results are presented in Section 4. Finally, conclusions are given in Section 5.

2. DISPARITY-COMPENSATED PREDICTION

IEEE ICIP 2002

The light field and lumigraph are obtained by sampling the plenoptic function in a 2D plane such as the (u,v)-plane shown in Figure 1. A 2D array of images of the scene is then generated. The sampling points usually lie on a regular rectangular grid and an image of the object or scene is taken at each grid point, (u,v). We shall call the image plane as the (s,t)-plane. Using this *2D array of images*, it is possible to create different views of the object or scene at different viewing angles through a process called *image based rendering*. Adjacent light field images appear to be shifted relative to each other. Such displacement of pixels, called disparity, can therefore be employed to predict one image from the other analogous to motion-compensated prediction in video coding. This disparity-compensated prediction is illustrated in Figure 2, where the array of light field images is divided into I- and P-pictures. The P-pictures are predicted from the nearest encoded I-pictures. In the *V-coder* [17], conventional video compression scheme was proposed for light field compression. A number of images from the light field array, usually evenly distributed, are first encoded as intra or I-pictures. The remaining light field images are then predicted by disparity compensation using the encoded I-pictures as references. The P-pictures are divided into (16×16) blocks and are coded using video coding technique, hence the name *V-coder*. Prior to rendering, the I-pictures are decoded and kept in local memory to provide instantaneous access to a low-resolution version of the light field. Random access of light rays (pixel), however, was not provided which somewhat slow down the rendering speed. A similar approach called multiple reference frame (MRF) prediction was also proposed in [12,14]. A two-level index table is further incorporated into the bitstream for quick access to each individual picture and macroblock. The reference images, called the anchor frames (A-frames), are not decoded into the memory but a cache management is provided to speedup the rendering. Another hierarchical light field coder based on DCP and VQ (HDCP) was proposed in [16]. The 2D array of light field images is divided into layers, with the lowest layer vector quantized without any prediction. Images in higher layers are predicted, using disparity compensation, from images in the lower layers using single-frame or multiple-frame prediction. The prediction residuals are again vector quantized and different coding modes are incorporated to improve the coding efficiency. To facilitate random access, the residuals and the disparities are not entropy encoded. Moreover, the predicted coded images are divided into regions, and each is associated with a 4-byte offset to support random access. Next, we shall study methods for reducing the memory requirements for the reference images using the nonlinear filter banks, while minimizing the decoding complexity.

3. THE PROPOSED COMPRESSION SCHEME

3.1 The Nonlinear Filter Banks

We now describe the decomposition of the reference pictures using the nonlinear perfect reconstruction filter bank. The main reason of using a nonlinear perfect reconstruction filter bank, instead of the commonly used 9/7 wavelet filter bank is that it has a lower arithmetic complexity and lower filter support so that random access of the image pixels during rendering can be considerably simplified. Figure 3 shows a single level of decomposition of an I-picture using this nonlinear filter bank [9]. For simplicity, a notation similar to the conventional filter bank systems is adopted, namely LL, LH, HL and HH bands. The low-low (LL) band is simply obtained by decimating the original image by a factor of two both horizontally and vertically, and it

consists of pixels at positions $(2i, 2j)$. The HH band consists of prediction residuals, $e(2i+1, 2j+1)$, when pixels at positions $(2i+1, 2j+1)$ are predicted from its four quantized neighbours at positions $(2i, 2j)$, $(2i+2, 2j)$, $(2i, 2j+2)$, and $(2i+2, 2j+2)$. Similarly, HL and LH bands consist of prediction residuals when the pixels at $(2i+1, 2j)$ and $(2i, 2j+1)$ are predicted from their respective quantized neighbours. Different prediction methods can be used in this simple decomposition scheme. In this work, we adopted the simple but effective method in [9] and further consider several variations that simplify the decoding process. More precisely, the prediction method [9] can be summarized as follows: given the four input samples $\{p_1, p_2, p_3, p_4\}$, the minimum and maximum of $\{p_1, p_2, p_3, p_4\}$ are discarded. The predicted output is equal to the mean of the remaining two elements. This operation though very simple is able to enhance image edges from the subsampled image data. It can also be viewed as a post-processing operation for improving the visual quality of the reconstructed image. The arithmetic complexity is very low which requires only 4 comparisons and 2 additions per pixel. Similar decomposition can be applied to the LL band creating a multiresolution decomposition of the image, like an octave wavelet filter bank. Another important reason for using this nonlinear filter banks is that the reconstruction of a given pixel at a given resolution only requires its four adjacent neighbours, unlike their conventional linear counterparts. The subband coefficients can be encoded using VQ [16] or DCT-based coding. Since the lowest resolution subband will be decoded and preloaded to memory, random access mechanism such as the pointer structure [8] or indexing structure [16] is not necessary. On the other hand, such data structure has to be incorporated to provide fast access to the prediction residuals in higher resolution subbands of the I-pictures and the P-pictures. Figure 4 illustrates the proposed algorithm with two levels of decomposition.

3.2 Decoding Algorithm

Consider the simplest case where the I-picture is decomposed by one level. If the I-picture is encoded using say DCT-based transform coding, the pointer data structure similar to [8] can be used to provide efficient access to each compressed data block in the HL, LH and HH bands. Prior to rendering, the subband coefficients in the LL band are decoded and loaded into memory to provide fast random access. Since the LL band is a subset of the original image pixels, references to these pixels do not require any additional operations. To access a pixel associated with the HH band (position $(2i+1, 2j+1)$), its four nearest neighbours in the LL band have to be read. A predictor is then generated from these four pixels using the algorithm mentioned earlier. This requires 4 comparisons and 2 additions. The compressed data block in the HH band is then accessed using the pointer data structure or other equivalent structure. The compressed data block is decoded to obtain the prediction residual of the required pixel, which is added to the predictor calculated previously to reconstruct the value of the required pixel. For pixels associated with the HL (LH) band at positions $(2i+1, 2j)$ $((2i, 2j+1))$, additional prediction residuals from the HH band and the HL (LH) band has to be decoded to reconstruct the required pixels. We shall denote this scheme as intra-prediction scheme I (IPS-I).

To avoid this additional decoding of residuals in the HL or LH bands, a simple variation of the above prediction scheme can be used. The idea is to predict the pixels in the HL and LH bands entirely from the quantized values of the LL band. Instead of using the quantized values at position $(2i+1, 2j+1)$, the predictor

from the LL band is used. It is referred to as intra-prediction scheme II (IPS-II). During progressive rendering, the decoding of the residuals can further be skipped to provide faster interactive speed. This is called partial decoding (PD). The disadvantage of IPS-II is that larger prediction residuals will be generated in the HL and LH bands, degrading coding performance. If VQ is used to encode the HL, LH and HH bands, then the decoding of the residuals requires only simple table look up operations. Furthermore, if the LL band is encoded using VQ, further main memory saving can be obtained. However, for coding efficiency, it is preferable to encode the LL band by highly efficient schemes such as DCT- or wavelet transform-based methods. This method can be extended to more than one level by repeatedly decomposing the LL band. The reconstruction of a pixel in higher resolution bands will involve additional decoding of the residuals in lower resolution subbands. Due to the short filter support, the number of additional decoding of residuals is much smaller as compared with the conventional 9/7 wavelet filter banks. In addition, the arithmetic complexity associated with the synthesis filtering of the 9/7 wavelet filter banks is also much higher than the NFB. In practice, a few levels of decomposition is usually sufficient because the memory required to store an entire low-resolution I-picture is reduced by a factor of 4 whenever the number of level is increased by one. Also, the IPS-I and IPS-II can be used, respectively, in the lower and higher resolution subbands to simplify decoding. The P-pictures can be predicted from the I-pictures and its residuals can either be coded by VQ or DCT. Pointers or other equivalent structure should be provided for efficient access. To further reduce the decoding complexity of the P-pictures, the predictor in the I-pictures can be constructed from the IPS-II predictor which only involves quantized pixels in the LL band. This scheme is denoted by PPS-II. Conventional DCP predictor using the quantized pixels in the I-picture is referred to as PPS-I. The former scheme avoids the decoding of residuals in the HL, LH and HH bands, when forming the predictor for the P-pictures. The coding performance however will be degraded.

3.3 PROGRESSIVE TRANSMISSION

Using the subband decomposition proposed, there are two simple ways for progressive transmission. First, the P-pictures can be decomposed using the same structure of the I-pictures. The residuals are coded similar to that of the I-picture (using either VQ or DCT with random access mechanism). The LL bands of the P-pictures are encoded by DCP. This scheme, denoted by PT-I, allows the light field to be transmitted resolution by resolution. However, as the DCT is performed at the lowest resolution, coding performance degradation is to be expected. In the second progressive transmission scheme (PT-2), the I-pictures are transmitted first followed by the P-pictures.
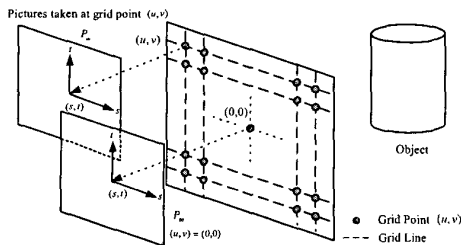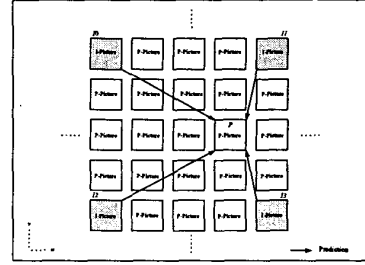

Fig. 1. Construction of the light field.

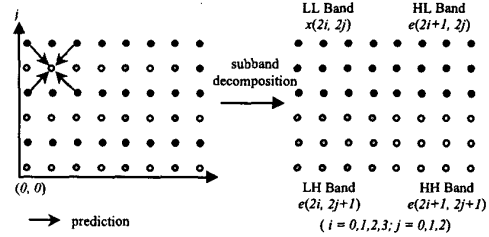Fig. 2. An array of light field images in DCP.

Fig. 3. One level of decomposition of the I-picture using the nonlinear filter bank. Samples marked with ● are used to predict the samples marked with ○, ◉ and ●. $e(\cdot, \cdot)$ are the prediction residuals.
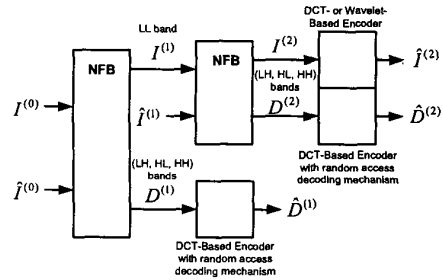

Fig. 4. Proposed encoder structure for the reference images. NFB: nonlinear filter banks. $\hat{I}$ ($\hat{D}$) means the quantized values of I (D).

4. EXPERIMENTAL RESULTS

The *Buddha* and *Dragon* light fields [3] are compressed using the proposed coding algorithm. Both data sets consist of 32×32 24-bit RGB images with 256×256 pixels. Similar to Figure 2, the reference I-pictures are evenly distributed among the whole image array and we have two P-pictures between neighbouring I-pictures. For simplicity, only one level of decomposition is used in our coder. The total memory size of the I-pictures is reduced from 18.75MB to 4.69MB. All the subbands in the reference pictures are encoded as I-pictures using a modified MPEG2 encoder. The P-pictures are encoded using DCP and transform coding, similar to the MPEG2 standard. Pointer structures, similar to [8], are used to provide random access. As all the images will be employed for rendering, a uniform quantizer is used in the modified MPEG2 encoder for all subbands. The following codec configurations are simulated: 1) MPEG2 like algorithm without NFB, 2) IPS-I with PPS-I, 3) IPS-II with PPS-I, and 4) progressive transmission using the LL band only, and the PT-I, IPS-I and PPS-I schemes. The coding results are plotted in Figures 5 and 6. It can be seen that the performances of the NFB with IPS-I and IPS-II are comparable to the MPEG2 like algorithm. The memory requirement for holding the decoded pixels in the I-pictures of the

former two schemes is reduced by a factor of four. There is however a significant performance degradation at high bit rate in supporting progressive transmission using PT-I. Figures 7 and 8 show the original and typical reconstructed images in the light fields *Buddha* and *Dragon*. They show good quality of reconstruction at bit rates between 0.05 and 0.1.

5. CONCLUSION

A new coding algorithm for light fields using perfect reconstruction nonlinear filter banks is presented. It is based on the conventional disparity-compensated prediction encoder with the reference pictures being decomposed using the NFB to reduce the amount of main memory needed to support fast rendering. The NFB has very low arithmetic complexity for reconstruction and small filter support which considerably simplifies the random access operations. Several prediction and reconstruction strategies are also presented to illustrate the flexibility of the proposed algorithm in achieving different tradeoffs between memory requirement and decoding speed. Generalization of the present method to include progressive transmission is also presented.

REFERENCES

[1] L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," in *Computer Graphics (SIGGRAPH'95)*, pp.39-46.

[2] 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.

[3] M. Levoy and P. Hanrahan, "Light field rendering," in *Computer Graphics (SIGGRAPH'96)*, pp. 31-42.

[4] S. J. Gortler, R. Grzeszczuk, R. Szeliski and M. F. Cohen, "The lumigraph," in *Computer Graphics (SIGGRAPH'96)*, pp. 43-54.

[5] S. E. Chen, "QuickTime VR –an image-based approach to virtual environment navigation," in *Computer Graphics (SIGGRAPH'95)*, pp.29-38.

[6] R. Szeliski and H. Y. Shum, "Creating full view panoramic image mosaics and texture-mapped models," in *Computer Graphics (SIGGRAPH'97)*, pp.251-258.

[7] H. Y. Shum and L. W. He, "Rendering with concentric mosaics," in *Computer Graphics (SIGGRAPH'97)*, pp. 299-306.

[8] 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 ICIP2000*, vol. 3, pp. 644-647, Sept. 2000.

[9] R. L. de Queiroz and D. A. F. Florencio, "A non-linear filter bank for image coding," in *Proc. of the 38th Midwest Symposium on Circuits and Systems*, vol. 2, pp. 1115 –1118, 1996.

[10] ITU-T Rec. H.262|ISO/IEC 13818-2, "Generic coding of moving pictures and associated audio information: video," Nov. 1994.

[11] K. T. Ng, S. C. Chan and H. Y. Shum, "Scalable coding and progressive transmission of concentric mosaic using nonlinear filter banks," in *Proc. of IEEE ICIP2001*, vol. 2, pp. 113-116, Oct. 2001.

[12] J. Li, H. Y. Shum and Y. Q. Zhang, "On the compression of image based rendering scene," in *Proc. of IEEE ICIP2000*, vol. 2, pp. 21-24, Sept. 2000.

[13] 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 ICIP2001*, vol. 2, pp. 105-108, Oct. 2001.

[14] 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 (DCC2000)*, pp. 253-262, Mar. 2000.

[15] C. Zhang and J. Li, "Compression and rendering of concentric mosaics with reference block codec (RBC)", in *Proc. SPIE VCIP'2000*, vol. 4067 (05), June 2000.

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

[17] 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, Apr. 2000.

[18] M. Magnor, A. Endmann and B. Girod, "Progressive compression and rendering of light fields," in *Proc. Vision, Modeling, and Visualization VMV-2000*, Saarbrucken, Germany, pp. 199-203, Nov. 2000.

[19] I. Peter and W. Strasser, "The wavelet stream – progressive transmission of compressed light field data," in *Proc. IEEE Visualization '99*, Oct. 1999.
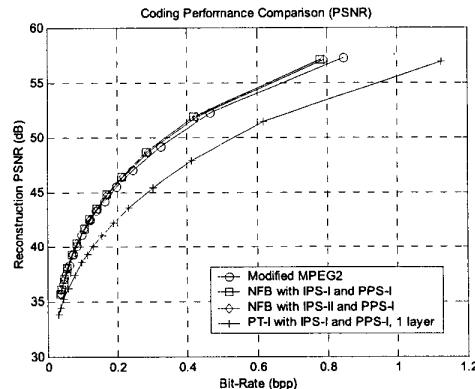
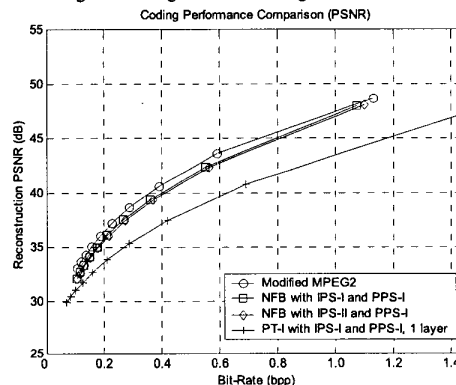Fig. 5. Coding results of the light field *Buddha*.

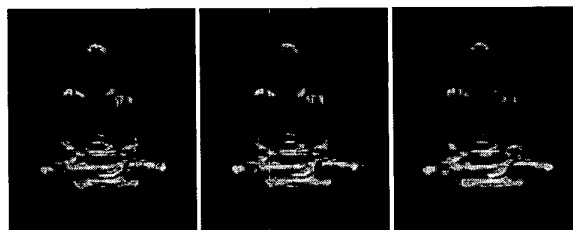Fig. 6. Coding results of the light field *Dragon*.

Fig. 7. Images from the light field *Buddha*. (Left) Original. (Middle) IPS-II with PPS-I. (0.05 bpp) (Right) PT-I with IPS-I and PPS-I. (0.049 bpp)
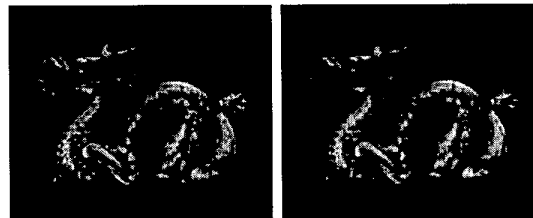

Fig. 8. Images from the light field *Dragon*. (Left) Original. (Right) IPS-II with PPS-I. (0.105 bpp)