

LIGHT FIELD COMPRESSION BASED ON PREDICTION PROPAGATING AND WAVELET PACKET

Xu Dong, Dai Qionghai, Xu Wenli

Department of Automation, Tsinghua University, Beijing 100084, China

ABSTRACT

A novel data compression scheme of light field is presented. Different from the prior codecs, we perform image predicting and data coding in a set of subbands not directly in original images. In our approach, firstly the original images are decomposed into subbands using wavelet packet transform, and the corresponding wavelet packet bases are divided into two parts: the predictable bases and the unpredictable bases by some criterions. In coding, a propagating algorithm and a rhombuses structure is used, and the subbands corresponding to the basis in predictable and unpredictable bases are added sequently by relative energy until the reconstructing images meet the pre-established reconstruction quality. Experiments for two standard light fields verify the efficiency of our approach.

1. INTRODUCTION

Recently a new approach to rendering 3D object or scene has been proposed by Levoy et al. [1] and by Gortler et al. [2]: light field rendering. A light field is a collection of light rays flowing through space in all directions, and can be created from images captured at a set of view positions. And the set of images forms a 4D structure. A light field often contains lots of images and a large amount of data. The storage and transmission of the large amount of data is very difficult, hence, the compression of the data is one of the key challenges in light-field rendering.

Extensive researches have been carried out on light field compression for high compression ratio, random access, fast decoding and embedded representation. There are mainly three classes of compression techniques: vector quantization (VQ) [1], predictive coding [3,4,5,6] and progressive coding [7,8].

In this paper, we propose a novel light field coding algorithm, wavelet packet based prediction propagating coding (WPPPC), which decomposes images into subbands using wavelet packet (WP) transform and codes these subbands in two parts: predictable and unpredictable

part respectively. Different from Magnor's HDC, our approach predicts and codes image from the center of the image array, and propagates the predicting and coding along the diagonal directions with a certain interval until reach the border of the image array. And a rhombus structure is used to predict and code the other images. Besides high compression ratio our approach also provide scalarbility in some extend.

The paper is organized as follows. Section 2 introduces wavelet packet transform and the selecting of predictable and unpredictable basis. Section 3 discusses the coding performance of light field. Section 4 shows some experiments and results to verify the efficiency of our approach. Section 5 gives some conclusion of this paper.

2. WAVELET PACKET

Wavelet packet bases \mathbf{B} of $L^2(Z^2)$ whose elements are separable products of two wavelet packets with the same scale along x_1 and x_2 can be constructed by [9]:

$$\psi_j^{p,q}(x-2^j n) = \psi_j^p(x_1-2^j n_1) \psi_j^q(x_2-2^j n_2) \quad (1)$$

where $x = (x_1, x_2) \in Z^2$, $n = (n_1, n_2) \in Z^2$, $\psi_j^p(x_1-2^j n_1)$ and $\psi_j^q(x_2-2^j n_2)$ are the 1D wavelet packet along x_1 and x_2 respectively. These separable wavelet packet bases are associated to quad-trees, and divide the 2D Fourier plane (ω_1, ω_2) into square regions, called subbands, of varying size.

For a given image $f(x_1, x_2)$ of size $N_1 \times N_2$, the collection of all subband decompositions with the maximum number of j level forms a full, balanced quadtree of depth j . And at each leaf of this balanced quadtree, the original image $f(x_1, x_2)$ is represented by a component in the subband $W_j^{p,q}$ consisting of coefficients of $4^{-j} \times N_1 \times N_2$.

For still image compression, we can efficiently construct the original images with fewer important coefficients through searching for the best basis in a library of wavelet packet bases by some criterions [10]. But in light fields, if

the view positions are different, the best basis of images maybe different too. So we cannot efficiently reconstruct all the images in the light field using the same fewer coefficients, like still image compression.

Considering the inhomogeneous distribution of information in subbands and the similarity of images in light fields, we can discard those subbands, which have no important information by some criterions, so that the resulting approximation may differ minimally from the original images. Note, unlike still image compression, we use important subbands, not important coefficients, to reconstruct all of the original images. In order to select those subbands, we propose two criterions.

2.1. Relative energy

The energy of an image $f(x_1, x_2)$ given by

$$E(f(x_1, x_2)) = \sum_{x,y} |f(x_1, x_2)|^2 \quad (2)$$

usually has a very inhomogeneous distribution in subbands: only a few subbands have high relative energy, while others contain almost no energetically significant part of the image representation. Therefore we can use those sub-bands, which contain most energy, to reconstruct images.

For a light field LF , the relative energy $e_j^{p,q}$ of subbands corresponding to a given basis $B_j^{p,q}$ can be defined as the ratio of the sum of the energy of all subbands corresponding to the given basis $B_j^{p,q}$ and the sum of the energy of all subbands corresponding to the basis $B_L^{0,0}$:

$$e_j^{p,q} \stackrel{def}{=} \frac{\sum_{f(x,y) \in LF} E(\langle B_j^{p,q}, f(x_1, x_2) \rangle)}{\sum_{f(x,y) \in LF} E(\langle B_L^{0,0}, f(x_1, x_2) \rangle)} \quad (3)$$

Note if wavelet bases are orthogonal, the relative energies of all subbands sum to 1.

For a full balanced quadtree of depth j , we can construct energy-constrained bases $\mathcal{B}_j^e(LF, T_e)$ using a threshold T_e to discriminate those branches with less relative energy than T_e .

2.2. Correlations in light field

In light field, there are correlations between neighboring images. And the larger of the correlations, the more information can be predicted from neighboring images. But the correlations are different in subbands. Some subbands have large correlations, while others have little. In those subbands, which have large correlations, the coefficients can be efficiently predicted. And in those

subbands, which have little correlations, the coefficients can even not be predicted.

We define the correlation of the subband $W_j^{p,q}$ for an image $f(x_1, x_2)$ in light field as:

$$r_j^{p,q}(f(x_1, x_2)) \stackrel{def}{=} \frac{1}{4} \sum R(\langle B_j^{p,q}, f(x_1, x_2) \rangle, \langle B_j^{p,q}, f_i(x_1, x_2) \rangle) \quad (4)$$

where $f_i(x_1, x_2)$ is the images of the 4-neighbor of $f(x_1, x_2)$, and $R(\cdot)$ is a correlation function.

For the whole light field with size $N_{lf} \times M_{lf}$, the correlations of subbands corresponding to $B_j^{p,q}$ can be defined as:

$$r_j^{p,q} \stackrel{def}{=} \frac{1}{N_{lf} M_{lf}} \sum_{f(x,y) \in LF} r_j^{p,q}(f(x_1, x_2)) \quad (5)$$

In predicting coding, only the subbands, which have large correlations, can provide efficient prediction, other subbands mainly provide predicting residual error. For a full balanced quadtree of depth j , we can construct correlation-constrained bases $\mathcal{B}_j^c(LF, T_c)$ using a threshold T_c to detect those subbands with smaller correlation coefficients than T_c .

2.3. Predictable basis

For a light field, we can divide all the subbands into two parts: one contains the subbands, which have large relative energies and correlation coefficients, the other contains the subbands, which mainly provide isolated information of each image. We call the bases corresponding to the first part of subbands as predictable bases, and the bases corresponding to the second part of subbands as unpredictable bases.

Combination relative energy and correlation criterions, for a full balanced quadtree of depth j , we can define the predictable bases as:

$$B_j(LF, T_e, T_c) = B_j^e(LF, T_e) \cap B_j^c(LF, T_c) \quad (6)$$

3. LIGHT FIELD CODING

Considering a light field, which images are arranged in a regular grid of equal spacing, the four corners of the image array must be coded individually as I-images according to Magnor's HDC [3,4]. But there are some corresponding pixels in these images, and coding these images individually requires extra bits. And at the start steps, the predicting interval between images is too large to get precision results. Moreover, HDC has very high decoding complexity, which is not suitable for fast

rendering. Considering these disadvantages of HDC, we proposed a predicting propagating coder (PPC) based on wavelet packet transform. Fig.1 shows its coding order. In our approach, for a full balanced quadtree of depth j , images of size $N_1 \times N_2$ are decomposed into subbands of size $4^{-j} \times N_1 \times N_2$. In neighboring images the coefficients

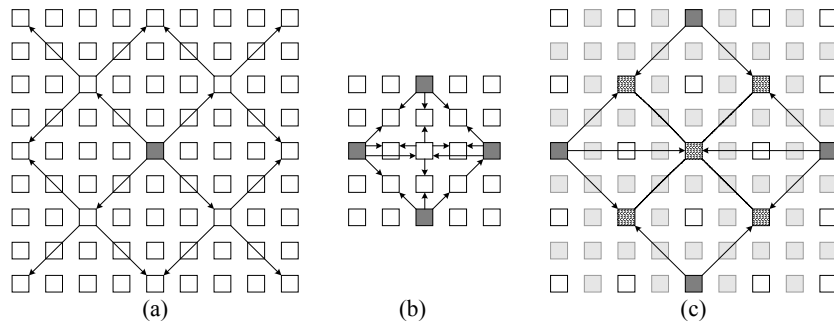


Fig.1 Coding order of our PPC

In coding, for the center image, the bases are sequentially selected from the predictable bases by relative energy, and coded using DCT followed by Huffman coding. If the reconstructed image meets the reconstruction quality q_{\min} , no more bases are added. Otherwise, the unpredictable bases are selected continuously.

And then, the four images in the diagonals with a certain image interval are predicted in the predictable subbands. If the predicted images meet the reconstruction quality, no more information is added. Otherwise, the predicting residual errors of each subband and the predictable subbands are coded sequentially by relative energy. Keeping the algorithm recursing until reach the sides of the image array, and subdividing the image array into several rhombuses.

In each rhombus, the center image is predicted with the images at the horizontal and vertical corners, and the other images are predicted with the symmetrical images at the closest corners or center, as shown in fig.1 (b). And the predicting errors are coded like above. When all of the rhombuses are processed, all images are coded.

If the size of light fields is too large, we can decrease it by sub-sampling. When light fields are up-sampled, the rhombuses can be subdivided into smaller sub-rhombuses by predicting the images at the center and the middle of the sides of the rhombuses, as shown in fig.1 (c).

Prior to rendering, a few of predictable subbands of the images at the nodes, including the center and the corners of each rhombus, are decoded and kept in local memory. Because the number of rhombus and subbands is much fewer, the memory requirements are modest.

In rendering, for the images at the nodes, the subbands in local memory are composed to create coarse images. If we require high quality images, we can decode more

of the subbands corresponding to the same basis are very similar. Hence, we can perform image predicting and data coding directly in each subband respectively. In order to simplify predicting, disparity maps can be estimated in original images with block-based or geometry-based estimating, and projected to subbands.

subbands. For other images, the predictable subbands are predicted from the pre-decoded subbands of the images at the nodes, and composed to create the required image besides the decoded residual errors and unpredictable subbands.

The coder has three advantages: first, only the center image is coded as I-Image that reduces the redundancy of pixels. Second, the intervals between predicted images are also deduced that proposed a more accuracy predicting. And third, if we only decode some of the subband images, we can get scalability features in some extend.

4. EXPERIMENTS AND RESULTS

We use two publically available light fields (Buddha and Dragon in fig. 2) to verify the proposed coding scheme. Both light fields contain $U \times V = 32 \times 32$ images of $S \times T = 256 \times 256$ 24-bit RGB pixels, totaling 192 MByte.

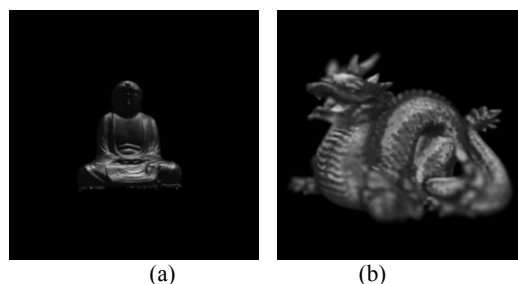


Fig.2 Images from light field Buddha and dragon

Each image is decomposed to varying levels, 4 and 5 levels, using wavelet packet transform with haar wavelet function, and the subband images have the size of 16×16 and 8×8 . Both light fields are coded for varying reconstruction quality parameter q_{\min} . The reconstructed

images are compared to the original light field images, and the distortion is measured (PSNR) and averaged over all images.

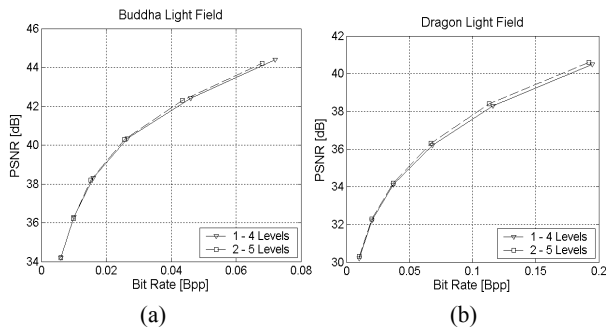


Fig. 3 Rate-distortion measurements of two light fields for different decomposition levels of wavelet packet transform

The rate-distortion curves for various decomposed levels are shown in fig. 3. Coding efficiency depends on reconstruction quality and light field scene characteristics. For the Dragon object, 543 kBytes are needed to code its light field at 36dB PSNR, when the decomposed level $j = 4$. For the Buddha statue, 218 kBytes are needed to code its light field at 40 dB PSNR, when the decomposed level $j = 4$.

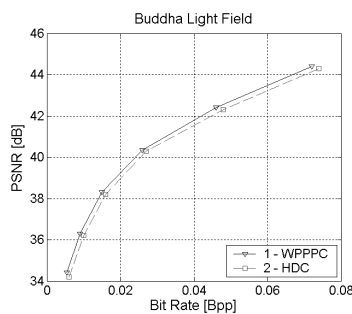


Fig. 4 The reconstructed images from our WPPPC and Magnor's HDC with different compression ratios

Fig. 4 gives the comparison of the rate-distortion curves of Buddha light field for our WPPPC and Magnor's HDC. Our WPPPC is performed at the level $j = 4$, and HDC is performed with block-based prediction of the size 16×16 . We can observe that our WPPPC can provide better reconstruction quality.

5. CONCLUSIONS

According to our approach, only a few of subbands, which have important information, are used to predict and reconstruct images. That reduces the decoding complexity and the local memory occupation. And the using of propagation algorithm and rhombus structure reduces the redundancy of pixels through only coding the center image as I-Image and reducing the predicting interval

between images. Furthermore, in our approach the subbands can be added to code sequently by relative energy until the reconstructed image meets the pre-established reconstruction criterion, this provides scalarability in some extend.

Although the introduction of wavelet packet transform reduces the disadvantages of decoding complexity and the local memory occupation, the complexity of wavelet packet transform limits the rendering speed. The fast reconstructing algorithm from some subbands is still an important problem.

5. ACKNOWLEDGEMENT

This work was performed in Broadband Networks & Digital Media Lab of Automation Dept, Tsinghua University, Beijing, China, and partly supported by China Postdoctoral Science Foundation (No. 2003034156) and National Nature Science Foundation of China (No. 60172026) and the Basic Research Foundation of Tsinghua University (No. JC2001028).

REFERENCES

- [1] Marc Levoy and Pat Hanrahan, "Light Field Rendering", in *Computer Graphics (Proceedings SIGGRAPH'96)*, pp 31-42, August 1996.
- [2] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The Lumigraph", *Computer Graphics (Proc. SIGGRAPH'96)*, pp. 43-54, August 1996.
- [3] Marcus Magnor and Bernd Girod, "Hierarchical Coding of Light Fields with Disparity Maps", *Proc. Int. Conf. Image Processing ICIP-99*, Kobe, Japan, pp. 334-338, Oct. 1999.
- [4] Marcus Magnor and Bernd Girod, "Data Compression for Light Field Rendering", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, no. 3, pp. 338-343, April 2000.
- [5] M. Magnor, P. Eisert, and B. Girod, "Model-aided Coding of Multi-viewpoint Image Data", *Proc. IEEE International Conference on Image Processing (ICIP-2000)*, Vancouver, Canada 2, pp. 919-922, Sept. 2000.
- [6] Marcus Magnor, Prashant Ramanathan and Bernd Girod, "Multi-view Coding for Image-based Rendering Using 3-D Scene Geometry", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 11, pp. 1092-1106, November 2003.
- [7] Marcus Magnor, Andreas Endmann, and Bernd Girod, "Progressive Compression and Rendering of Light Fields", in *Proceedings Vision, Modeling and Visualization 2000*, pp. 199-203, November 2000.
- [8] M. Magnor and B. Girod, "Model-based Coding of Multiviewpoint Imagery", in *Proceedings SPIE Visual Communications and Image Processing VCIP-2000*, Perth, Australia, vol. 1, pp. 14-22, June 2000.
- [9] S. Mallat, "A Wavelet Tour of signal processing", *Academic Press, A Harcourt Science and Technology Company, UK*, 1999.
- [10] Daechul Park, Moon Ho Lee, "Image Compression Based on Best Wavelet Packet Bases", *IEEE Trans. Consumer Electronics*, 40(3), pp. 527-536, 1994.