LIGHT FIELD COMPRESSION USING HOMOGRAPHY AND 2D WARPING

Shinjini Kundu

School of Medicine, University of Pittsburgh
Department of ECE, Carnegie Mellon University

ABSTRACT

This paper describes a new method for light field compression that exploits inter-view correlation. The proposed method uses homography and 2D warping to predict views and does not require additional camera parameters or a 3D geometry model. The method utilizes angular shift between views, which is neglected in conventional motion compensation methods. Results indicate improved coding efficiency of the proposed method over traditional motion compensation schemes. A full light field coder based on video-compression demonstrates 1-1.5 dB additional improvement in PSNR, or equivalently a 4-12% additional reduction in bitrate when the new method is introduced as a prediction mode.

Index Terms— Light field, Compression, Homography, Inter-view prediction, 2D warping

1. INTRODUCTION

A light field [1][2] is a data set used to render 3D interactive photorealistic graphics. A light field image is captured by a camera whose position in 3D space can be described by five variables: three for the planar viewing positions and two for the viewing angles [3]. Sampling of the 3D space yields a two-dimensional array of 2D light field views. Novel views are constructed by appropriately combing image pixels from existing views. Dense sampling leads to a large volume of data; for example, the light field data set of the statue David by Michelangelo has an uncompressed size of 36 gigabytes [3]. Thus, compression is imperative in any light field coding scheme.

Previous approaches in light field compression have focused primarily on disparity-compensation [5] and model-based coding [5]. Disparity compensation is analogous to motion compensation for video, and exploits translational interdependence between image blocks. It assumes solely translational motion in the camera and neglects changes in camera angle. While temporal movement in a video sequence can be modeled using translational shift, the displacement between multiple views of a 3D scene cannot.

The inadequacy of simple translational modeling has been recognized in multi-view video compression. Researchers have explored methods that utilize warp [7][8] that take into account camera angles and zooms to exploit correlations between views. In this paper, an analogous warping technique is presented for light field compression that does not require depth maps or actual camera parameters.

The proposed method for light field compression uses 2D feature-based warping and homography. Subsequent views of light field images are computed using homography from corresponding feature points, followed by interpolation and linear extrapolation. In this paper, we compare the coding performance of the proposed method with motion compensation methods over a varied range of block sizes. Finally, we investigate utility of the proposed method in a full light field coder based on video-compression when projection is incorporated as a prediction mode [5].

The remainder of the paper is organized as follows: in section 2, we provide an overview of the proposed method. In section 3, we review the practical light field coding scheme based on video-compression used in this paper. Experimental results and analysis are presented in section 4.

2. COMPRESSION METHOD

The basic system architecture of the proposed method is illustrated in Figure 1.

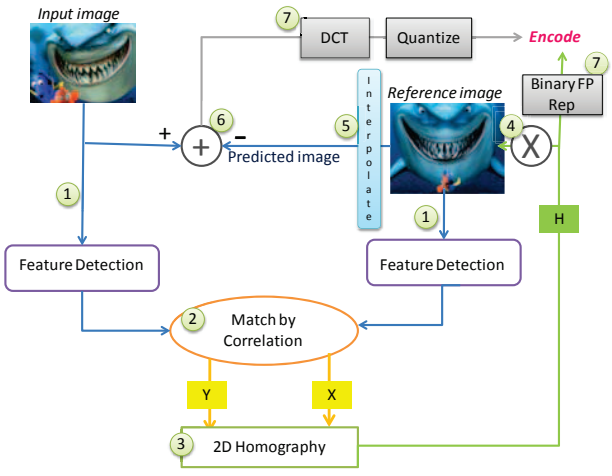

Fig. 1 System architecture for proposed method

2.1 Feature Detection

Points of interest are detected on the reference image and input image independently using Harris corner and edge detection algorithm [9]. The images are first filtered by a smoothing Gaussian filter. The threshold to the Harris detector was set by experimentation at 50. A mathematical derivation of the Harris algorithm can be found in [9].

2.2 Match by Correlation

The feature points are matched pairwise between the reference image and input image using a correlation criterion. A window is created around every feature point in each image, and the correlation is computed between each possible pair of windows. Two feature points are considered pairs if their correlation is the highest amongst the possible pairs and meets a threshold condition. In this work, an experimental window of radius 63 pixels was used.

2.3 Homography

A linear mapping between sets of corresponding points is determined by 2D homography, producing projection matrix H. The projection matrix is used to warp the reference image pixel by pixel to predict the input image. Suppose that x is a pixel in the reference image and x' is its matching pixel in the input image.

$$x = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \qquad x' = \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} \tag{1}$$

A 3x3 matrix H that relates the two pixels is computed such that

$$x' = Hx \tag{2}$$

In this work, the 2D homography matrix H is estimated using Random Sample Consensus (RANSAC). Mathematical details of the RANSAC algorithm can be found in [10].

2.4 Interpolation and Extrapolation

Linear extrapolation and interpolation using the Laplacian operator is used to fill in any missing information in the predicted image.

2.5 Encoding of the Projection Matrix and Residual

The residual image is computed by finding the difference between the actual and predicted input image. The projection matrix and residual are encoded. The residual is intra-coded using discrete cosine transform (DCT), followed by coefficient quantization; the elements of the projection matrix are encoded using a truncated binary representation of a floating-point number. Projection matrix entries are rounded to 3-decimal accuracy, as determined experimentally that does not compromise the predicted image quality. The mantissa and exponent were encoded separately, as well as the sign bits.

3. FULL LIGHT FIELD CODER

The performance of a full light field coder is explored when projection is introduced as a prediction mode. The light field coder in this paper is based on a video-compression scheme adapted for light fields, as described by Magnor *et al* [5]. The authors' terminology is adopted in this paper, and the coder will henceforth be referred to as *V-coder*.

Figure 2 is an illustration of the encoding scheme. The interested reader is referred to this paper [5] for description of other practical light field encoders.

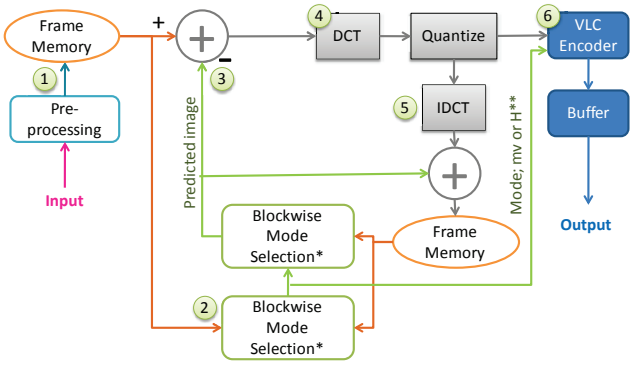

Fig. 2. System architecture for practical coder

The light field images are classified into two coding hierarchies: *I-frames,* and *P-frames.* The I-frames are selected to uniformly subsample the light field and are coded using blockwise DCT followed by coefficient quantization. The P-frames are images in the set that are predicted from the I-frames [5].

3.1 Prediction Modes

The P-frames are divided into 16x16 blocks and each block is encoded independently using mode selection among 5 prediction modes. The selected mode is the one that minimizes the Lagrangian cost function (see [5]).

- CLOSEST/NO DISP: The block on the P-frame is directly copied from the corresponding block on the closest I-frame which has the smallest difference
- PROJECTION: The projection matrices and projected images are computed in turn using each of the I-frames as reference
- MOTION-COMPENSATED: All the possible I-frames are used to find motion vectors and residuals for the block using half-pel motion compensation
- AVERAGE: The corresponding blocks from each I-frame are all averaged together
- INTRA: The block is encoded using DCT in intra-mode, with no reference to any I-frame.

4. EXPERIMENTS

Two data sets were used in the experiments: *Lego Men* and *Crystal* (data for crystal not shown in this paper)[1]. (See Figure 6). The images were transformed into YUV color space, then downsampled by a factor of 4 and interpolated using bicubic interpolation. For this work, only the luminance component is encoded. All of the simulation code was implemented using MATLAB and tested on a 4x4 array of 256x256 8-bit grayscale images that had a total uncompressed size of 1 MB.

The light field images were read in zig-zag order. The results are presented as rate-PSNR curves, for 8 quantization steps, where $\Delta \in \{2^i\}$ for i ranging from 1 to 8.

We note our assumption that the distortion, or mean-squared error, between adjacent views is uncorrelated. We therefore approximate the total distortion for the entire light field as the summation of the reconstruction distortion in each of the reference views [5].

4.1 Projection vs. Sub-pel Motion Compensation for Fixed Block Size

The performance of our proposed projection method is compared with sub-pel motion compensation to half-pixel accuracy. Figure 3 shows the rate-PSNR curve.

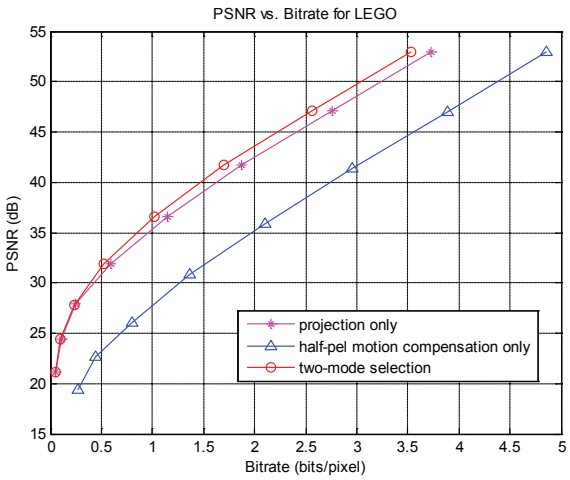

Fig. 3 Performance of projection warping method compared to to half-pel motion compensation using 16x16 blocks. Data set: *Lego Men*

At the same bit rate, projection warping performs about 7 dB better than half-pel motion compensation in terms of the PSNR, or, reconstruction quality. Equivalently, we achieve about 33% reduction in bitrate for the same reconstruction quality.

Furthermore, when bi-mode selection is implemented, there is only 1 dB gain as compared to the projection method alone. The results suggest that the proposed method outperforms sub-pel motion compensation on the data sets tested.

The projection method introduces less overhead per block. A 3x3 projection matrix takes 0.002 bits per pixel to encode, while an average motion vector takes 0.004 bits per pixel to encode. Even when the residuals are comparable, the projection warping method has a bitrate advantage.

4.2 Bi-mode Selection for Variable Block Sizes

Figure 4 suggests that as block size is increased, the gain magnitude of the projection method decreases over sub-pel motion compensation. The gain drops from 11 dB for 8x8 pixel block size to 8 dB for 16x16 pixels to 5 dB for 32x32 pixels using bi-mode selection.

As block size is increased, the overhead per pixel for motion compensation is decreased. The projection method does not depend on block size, and requires a constant bitrate of 0.002 to encode the projection matrix. For 8x8 blocks, the average bitrate per pixel needed is 0.015 to encode the motion vectors. For 16x16 blocks, the average bitrate needed is 0.004 bits per pixel, and for 32x32 blocks, the overhead becomes 0.001 bit per pixel.

The reduced overhead is not the sole explanation for why the proposed method provides better gains over motion compensation. The bitrate to encode 32x32 blocks is lower than that for the projection method, yet the proejction method is still favored. We may infer that the proposed method produces better behaved residuals, or is a better model for inter-view prediction than motion compensation alone, at least for the data sets in this paper.

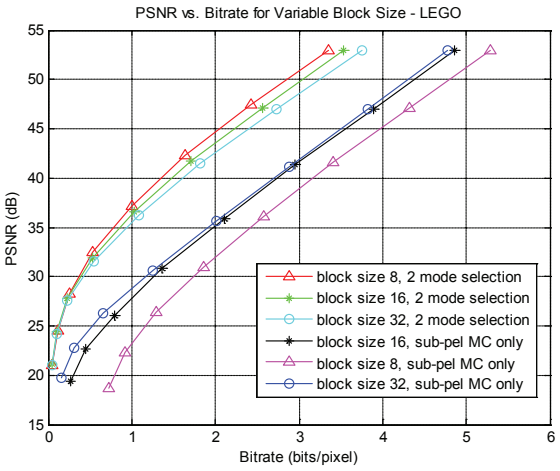

Fig. 4 Performance of proposed method with two mode selection, compared to sub-pel motion compensation for variable block size. Data set: *Lego men.*

4.3 Performance of Projection Method in Full V-Coder

As Figure 5 indicates, incorporating *projection* and *motion-compensation* predictive modes produces the most improvement in gains for the light field images. We achieve

[1] Credit: Andrew Adams, from http://lightfield.stanford.edu

about 1-2 dB gain in reconstruction quality at the same bitrate, or equivalently, 4-12% reduction in bitrate for the same reconstruction quality, when motion compensation is introduced as a mode. When projection is introduced as a mode, we achieve an additional 1-1.5 dB gain in reconstruction quality at the same bitrate, or equivalently, 4-12% reduction in bitrate for the same reconstruction quality.

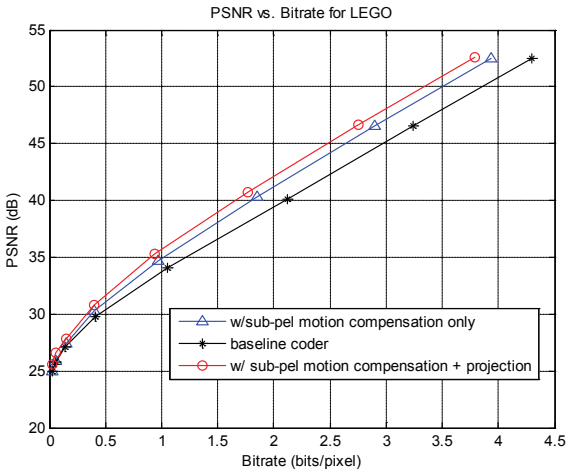

Fig. 5 Performance of proposed method with five mode selection in practical light field coder. Block size of 16x16 pixels were used. Data set: *Lego men.*

6. CONCLUSION

A new compression scheme for light fields is presented that uses 2D feature detection and homography for inter-view prediction. Results demonstrate improved compression efficiency of the proposed method compared to conventional motion-compensation methods. A gain of more than 7 dB in overall reconstruction quality at the same bitrate is observed with the proposed scheme over half-pel motion compensation. When bi-mode selection is implemented with sub-pel motion compensation, an overall gain of almost 8 dB is achieved. A 1.5 dB improvement in reconstruction quality at the same bitrate is achieved when the projection mode is introduced into a full light field coder. The results indicate that the proposed method complements motion-compensation and analogous disparity compensation in inter-view prediction.

7. ACKNOWLEDGEMENT

The author would like to thank Andrew Adams for access to *Lego men* and *Crystal* data sets used in this work. Special thanks to Professor Bernd Girod for introduction to the field. Thanks to Huizhong (Frank) Chen and Derek Pang for discussions, and to Professor Kovesi at the University of Western Australia for his open-source computer vision code.

8. REFERENCES

[1] M. Levoy and P. Hanrahan, "Light field rendering," *Computer Graphics (Proceedings SIGGRAPH 96),* pp. 31-42, 1996.

[2] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M. E Cohen,"The lumigraph," *Computer Graphics (Proceedings SIGGRAPH 96)*, pp. 43-54, 1996.

[3] T. Kanade, P. Rander, and P.J. Narayanan, "Virtualized reality: constructing virtual worlds from real scenes", *IEEE MultiMedia,* vol.4, no.1, pp.34-47, Jan-Mar 1997.

[4] M. Levoy, K. Pulli. *et al*, "The Digital Michelangelo project: 3D scanning of large statues," *Computer Graphics (Proceedings SIGGRAPH),* pp. 131-144, 2000.

[5] 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, 2000.

[6] M. Magnor, P. Eisert, and B. Girod, "Model-aided coding of multi-viewpoint image data," *Proc. ICIP-2000,* vol. 2, pp. 919-922, 2000.

[7] K. N. Iyer, K. Maiti, B. Navathe, H. Kannan, A. Sharma, "Multiview video coding using depth based 3D warping," *IEEE ICME,* pp.1108-1113, 2010.

[8] M. Zamarin, P. Zanuttigh, S. Milani, G. M. Cortelazzo, and S. Forchhammer, "A Joint Multi-View Plus Depth Image Coding Scheme Based on 3D-Warping," in *Proc. of 3DVP 2010*, pp. 7–12, 2010.

[9] C. Harris and M. Stephens, "A combined corner and edge detector," *Proc. of the 4th Alvey Vision Conference*, pp.147 – 151, 1998.

[10] M. A. Fischler and R.C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography." *Comm. Of the ACM,* vol. 24, issue 6, pp. 381-395, 1981.

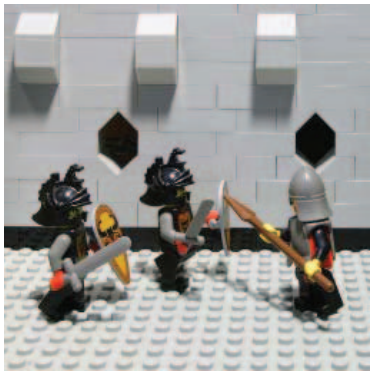

Fig. 6 *Test images* 4x4 arrays of 256x256-pixel data sets were used. Sample image from *Lego men*