MULTI-HYPOTHESIS PREDICTION FOR DISPARITY-COMPENSATED LIGHT FIELD COMPRESSION

Prashant Ramanathan, Markus Flierl[*]*, Bernd Girod*

Information Systems Laboratory
Stanford University, Stanford CA 94305
{pramanat, mflierl, bgirod}@stanford.edu

ABSTRACT

In this paper, we describe a multi-hypothesis prediction technique for disparity-compensated light field compression. Multi-hypothesis prediction has been used extensively in video compression. In this work, we apply multi-hypothesis prediction to the problem of light field compression. Most current techniques for light field compression utilize some form of disparity compensation with one hypothesis. We demonstrate that a multi-hypothesis approach, where two hypotheses are used, improves the overall efficiency of a light field coder. Our experimental results show an image quality gain of up to 1 dB in PSNR on our test data sets.

1. INTRODUCTION

Image-based rendering has emerged as an important new alternative to traditional image synthesis techniques in computer graphics. With image-based rendering, scenes can be rendered by sampling previously acquired image data, instead of synthesizing them from light and surface shading models and scene geometry. Light field rendering [1, 2] is one such image-based technique that is particularly useful for interactive applications.

A light field is a 4-D data set which can be parameterized as a 2-D array of 2-D light field images. For photo-realistic quality, a large number of high-resolution light field images is required, resulting in extremely large data sets. For example, a light field of an interesting object, such as Michelangelo's statue of *Night*, contains tens of thousands of images and requires over 90 Gigabytes of storage for raw data [3]. Compression is therefore essential for light fields.

Currently, the most efficient techniques for light field compression use *disparity compensation*, analogous to motion compensation in video compression. In disparity compensation, images are predicted from previously encoded reference images. A disparity or depth value is specified for a block of pixels. In our work we use depth values, as it simplifies our implementation. In this paper, we will use the terms disparity and depth interchangeably, with the understanding that they are equivalent to one another.

In previous work with disparity compensation [4, 5, 6], a single disparity value is specified for a block of pixels. In this paper, we describe a multi-hypothesis scheme for disparity compensation, where two depth values are specified, along with their corresponding reference images.

[*]Markus Flierl is a visiting student from the Telecommunications Laboratory, University of Erlangen-Nuremburg, Germany

Using multi-hypothesis prediction improves the prediction signal, and reduces the bit-rate required to encode the residual error. The cost is a larger bit-rate for the compensation. Multi-hypothesis prediction is employed in a practical light field coder as a mode that is selected using a rate-distortion criterion. We demonstrate in this paper that by using multi-hypothesis prediction, we improve the efficiency of a light field coder.

The remainder of the paper is organized as follows. In Section 2, we describe some of the basic concepts of disparity-compensated light field compression and multi-hypothesis prediction as used in video compression. We then describe the light field coder and the multi-hypothesis light field coding scheme in Section 3. In Section 4, we show the experimental results, and evaluate the efficiency of multi-hypothesis prediction for light field compression.

2. BACKGROUND

2.1. Disparity-compensated light field compression

Disparity-compensated prediction is originally proposed for stereo vision compression in [7, 8, 9]. Disparity compensation is used in most current light field compression algorithms [4, 5, 6]. The underlying idea of disparity-compensated prediction is that a pixel in a light field image can be predicted from pixels in one or more other light field images.

This prediction is made possible by specifying a depth value for a given pixel. Since we assume that what appears at a pixel corresponds to some surface in the light field, we are, in fact, specifying the position of this surface element. In a light field, the recording geometry is known, which means that it is possible to find a pixel in another view that corresponds to this surface element. This corresponding pixel is used to predict the original pixel as shown in Figure 1.

For such a prediction scheme to work, this light field surface must appear similar in both views. This is true if we assume that that the surface is Lambertian and is not occluded in these views.

Appropriate depth values must be used for prediction. In [5], depth values are calculated from an explicit geometry model. In [4, 6], depth values are estimated for a block of pixels from the image data using block-matching. When selecting depth values in our light field coder, we consider both the rate to encode the depth values, as well as the prediction error.

In a practical light field coder, the reference images that are used to predict a particular image must be defined. We follow the hierarchical structure described in [4]. Here, each image is predicted from a unique set of three or four reference images. The

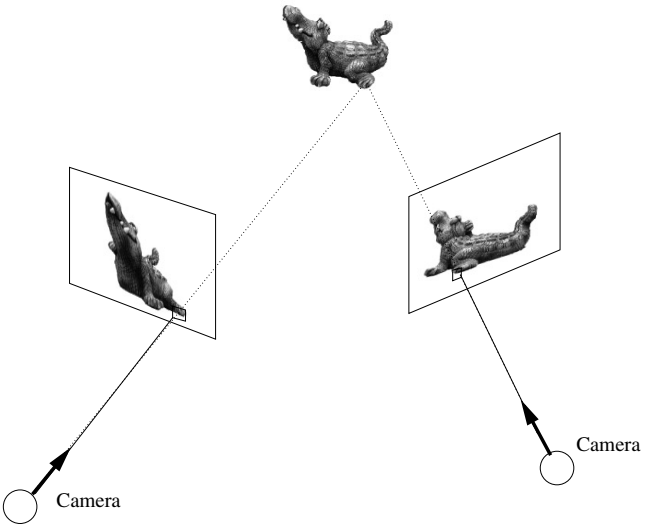

Fig. 1. A surface in the scene appears in two different views of the scene. Therefore, a pixel correspondence between the two views may be established.

order in which images are encoded is also defined, so that images are predicted from images that are already encoded.

2.2. Multi-hypothesis Prediction in Video Compression

Multi-hypothesis prediction has been extensively used in video compression. In multi-hypothesis schemes, several prediction signals are combined together to give one overall prediction signal. One example of multi-hypothesis prediction is B-frames in H.263 [10] and MPEG. In a B-frame, there is a multi-hypothesis prediction mode where blocks can be predicted from both the previous and future frames. In this mode, two independent motion vectors are sent, and the resulting prediction signals are averaged together. Multi-hypothesis prediction is also used in long-term memory motion-compensated prediction [11]. A theoretical analysis of multi-hypothesis prediction [12] shows that significant gains are possible over simply increasing pixel accuracy during motion estimation.

Typically, more bits are required for compensation in a multi-hypothesis prediction scheme, but fewer bits are required for residual error coding due to an improved prediction signal. In a practical coder, the multi-hypothesis scheme is usually implemented as a mode that can be chosen on a per-block basis. Mode selection is based on a Lagrangian cost function that incorporates the reconstruction quality and the overall bit-rate for both compensation and residual encoding.

3. MULTI-HYPOTHESIS PREDICTION FOR LIGHT FIELD COMPRESSION

3.1. Multi-Hypothesis Prediction Mode

The multi-hypothesis prediction mode H2 is one of four modes in our light field coder that can be selected to encode a block of pixels. In this sub-section, we describe the H2 mode in detail. The other three modes are described in the following subsection.

In the H2 mode, we specify two depth values for a block of pixels. For each depth value, we also specify the index of the corresponding reference image that is used for prediction. The depth value is used to find a corresponding block of pixels in the specified reference image. This corresponding block of pixels is a prediction signal for the original block. We find two such prediction signals, since we use two hypotheses, and average them together to obtain the overall prediction block. The error between the original block and the predicted block is encoded by a DCT coder.

The depth values and corresponding reference image indices that are used for prediction must be selected appropriately. A codebook with variable-length codes is used to encode the depth values and the reference indices. The representative depth values in this codebook are selected such that they correspond to approximately integer-pixel disparity accuracy in the images. The reference image can be one of three or four possible reference images.

It is important to jointly select the combination of two depth values and reference image indices for prediction. In order to do this, we perform a full search over all possible combinations of two depth values and reference images.

To evaluate a particular combination of depth values and reference images, we need to take into account both the bits R_e required to represents this combination, as well as the resulting overall distortion D_e. We define distortion to be the sum of absolute differences between the pixels in the original and predicted block. The Lagrangian cost function

$$J_e = D_e + \lambda_e R_e,$$

takes both bit-rate and distortion into account. The Lagrange multiplier λ_e indicates the trade-off between rate and distortion. The quantity λ_e is ultimately related to the quantization level Q in the DCT residual encoder by the equation

$$\lambda_e = \sqrt{0.85Q^2}$$

as described in [13].

3.2. Other Modes

In addition to the multi-hypothesis mode, three other modes, INTRA, H1 and SKIP, are used in our light field coder. Here, we give a brief description of each mode.

The INTRA mode is conceptually the simplest. In this mode, no prediction is used, and a DCT coder is used directly on the original image block.

The H1 mode is the single-hypothesis counterpart to the multi-hypothesis H2 mode. In the H1 mode, one depth value is specified for a block of pixels. This depth value, along with the corresponding reference image to be used for prediction, is specified using a variable-length codebook. As in the H2 mode, the residual error is DCT-encoded. Full search using a Lagrangian cost function is employed to select the depth value and reference image index.

The final mode SKIP predicts the depth value for a block using the depth values from its reference images. This is based on the disparity-map warping approach described in [4]. The depth values already encoded in the reference images can be warped, using the known light field recording geometry, to give depth values in the current image. These depth values are used to obtain prediction blocks from all reference images. These prediction signals are

averaged together to give the overall prediction block. No residual error coding is used in this mode. Bits are required only to specify the mode. The mode is specified using a variable-length codebook.

3.3. Light Field Coder Control

Each block of pixels can be coded in one of the four modes. The mode is selected based on the Lagrangian cost function

$$J_m = D_m + \lambda_m R_m.$$

which considers both rate and distortion. The mode with the smallest Lagrangian cost is selected.

The distortion D_m is calculated between the original block and the reconstructed block as the sum of squared errors between corresponding pixels. The reconstructed block is the result after compensation and residual error coding. The bit rate R_m includes the bits used for compensation, residual error coding, and mode specification. The Lagrange multiplier λ_m defines the trade-off between rate and distortion. This trade-off is dependent on the required image quality which is defined by the DCT quantization parameter Q. The relationship between Q and λ_m has been determined empirically [13] to be

$$\lambda_m = 0.85Q^2.$$

4. EXPERIMENTAL RESULTS

We now investigate the benefits of using multi-hypothesis prediction in a light field coder. Two data sets, *Garfield* and *Crocodile*, are used in the experiments. Representative images from these data sets are shown in Figure 2. The *Garfield* data set consists of 288 images at a resolution of 192×144 pixels. The *Crocodile* data set consists of 257 images at a resolution of 192×144 pixels. Both light fields were recorded using a hemispherical light field recording geometry, as shown in Figure 3. For the results in this paper, the images are converted to YUV format, and only the intensity component Y is used.

Garfield *Crocodile*

Fig. 2. Representative images from the two light fields.

There are two configurations of the light field coder that are tested. In the first configuration, only the INTRA, H1 and SKIP modes are used. In the second configuration, the H2 mode is also used. The operational rate-distortion performance is obtained for both of these configurations by the varying the DCT quantization parameter Q. Recall that this parameter is directly related to the Lagrange multipliers used in disparity estimation and mode selection, and determines, therefore, the trade-off between rate and distortion.

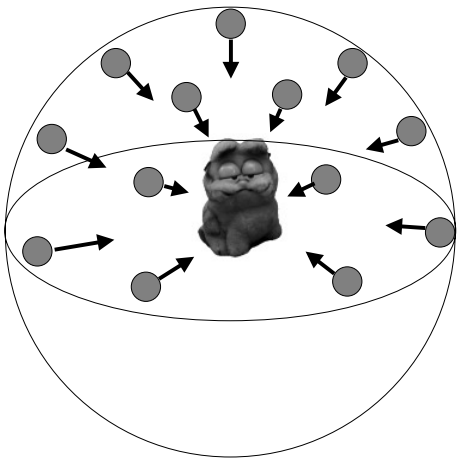

Fig. 3. Hemispherical Light Field Arrangement. The small spheres represent the positions of the light field cameras, and the arrows indicate their orientations. The object is placed at the center of the hemisphere.

A block size of 8×8 is used for all experiments. Figures 4 and 5 show the rate-PSNR performance for the two configurations. The rate is given in bits per pixel (bpp), which is the total number of bits required to code the light field divided by the total number of pixels in the light field. PSNR, *Peak Signal-to-Noise Ratio*, a measure of image quality, is given in decibels and is defined by the equation

$$PSNR = 10 \log_{10}(\frac{255^2}{D})$$

where D is the mean squared error between original and reconstructed images, averaged over all images.

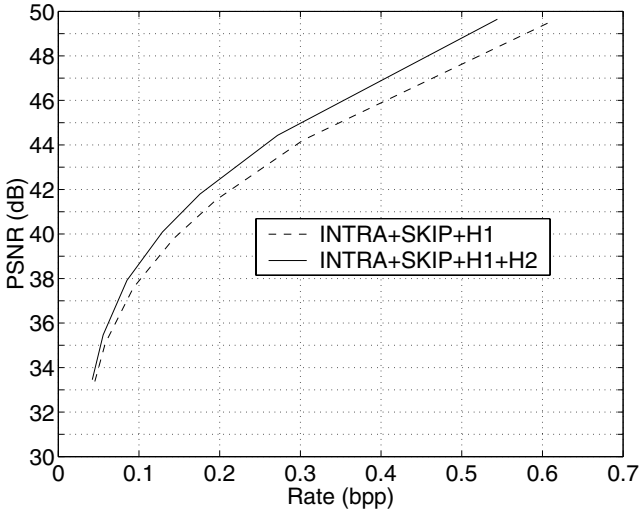

Fig. 4. PSNR vs. Rate for the Garfield data set. A decrease in bit-rate of over 10% is observed at high bit-rates.

We see that multi-hypothesis prediction is particularly effective for high bit-rates. For the *Garfield* data set, we require 10% fewer bits to code the light field, and for the *Crocodile* data set,

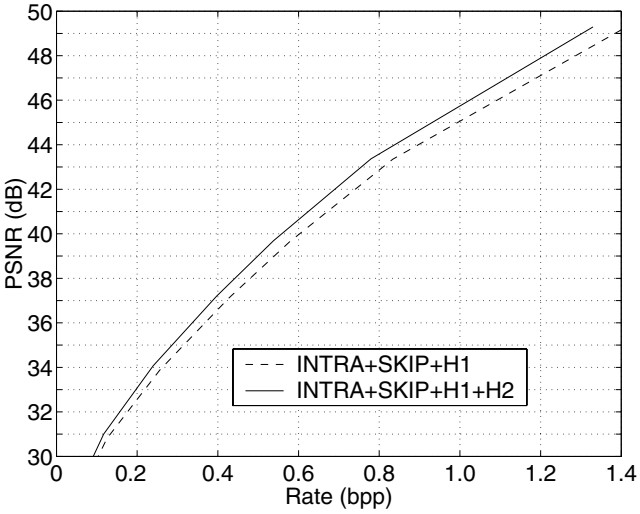

Fig. 5. PSNR vs. Rate for the Crocodile data set. A decrease in bit-rate of over 5% is observed at high bit-rates.

we require 5% fewer bits. This translates to a gain of approximately 0.5 to 1 dB in PSNR. At lower bit-rates, the gains from using multi-hypothesis prediction are smaller.

5. CONCLUSIONS

We have described a multi-hypothesis prediction scheme for disparity-compensated light field compression. Using multiple hypotheses improves the disparity-compensated prediction error, and reduces the bit rate required by the residual error coder. Multi-hypothesis prediction was used as one of several modes in our light field coder. We applied a rate-distortion cost function to select the best coding mode for each block in the image. When employed in this manner, multi-hypothesis prediction improved the overall efficiency of the light field coder. Our experimental results showed an improvement in image quality of up to 1 dB in PSNR for our test data sets.

6. REFERENCES

[1] Marc Levoy and Pat Hanrahan, "Light field rendering," in *Computer Graphics (Proceedings SIGGRAPH96)*, August 1996, pp. 31–42.

[2] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen, "The lumigraph," in *Computer Graphics (Proceedings SIGGRAPH96)*, August 1996, pp. 43–54.

[3] Marc Levoy, Kari Pulli, et al., "The Digital Michelangelo project: 3d scanning of large statues," in *Computer Graphics (Proceedings SIGGRAPH00)*, August 2000, pp. 131–144.

[4] Marcus Magnor and Bernd Girod, "Data compression for light field rendering," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 3, pp. 338–343, April 2000.

[5] Marcus Magnor, Peter Eisert, and Bernd Girod, "Model-aided coding of multi-viewpoint image data," in *Proceedings of the IEEE International Conference on Image Processing ICIP-2000*, Vancouver, Canada, September 2000, vol. 2, pp. 919–922.

[6] Xin Tong and Robert M. Gray, "Coding of multi-view images for immersive viewing," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing ICASSP 2000*, Instanbul, Turkey, June 2000, vol. 4, pp. 1879–1882.

[7] M. Lukacs, "Predictive coding of multi-viewpoint image sets," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing ICASSP 1986*, 1986, pp. 521–524.

[8] I. Dinstein, G. Guy, and J. Rabany, "On the compression of stereo images: Preliminary results," *Signal Processing*, vol. 17, pp. 373–382, 1989.

[9] Michael G. Perkins, "Data compression of stereopairs," *IEEE Transactions on Communications*, vol. 40, no. 4, pp. 684–696, April 1992.

[10] ITU-T, "Video coding for low bitrate communication: Recommendation H.263, Version 2," 1998.

[11] Markus Flierl, Thomas Wiegand, and Bernd Girod, "Rate-constrained multi-hypothesis motion-compensated prediction for video coding," in *Proceedings of the IEEE International Conference on Image Processing ICIP-2000*, Vancouver, BC, Canada, September 2000.

[12] Bernd Girod, "Efficiency analysis of multihypothesis motion-compensated prediction for video coding," *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 173–183, February 2000.

[13] Gary J. Sullivan and Thomas Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 74–90, November 1998.