RATE-DISTORTION ANALYSIS OF RANDOM ACCESS FOR COMPRESSED LIGHT FIELDS

Prashant Ramanathan and Bernd Girod

Information Systems Laboratory, Department of Electrical Engineering
Stanford University
{pramanat,bgirod}@Stanford.EDU

ABSTRACT

Image-based rendering data sets, such as light fields, require efficient compression due to their large data size, but also easy random access when rendering from the data set. Efficient compression usually depends upon prediction between images, which creates dependencies between them, conflicting with the requirement of having easy random access. Existing light field coders concentrate either on compression efficiency, or use ad hoc methods to design prediction that balances random access and compression efficiency requirements. In this paper, we study this joint problem of compression efficiency and random access. We propose a model for light field image generation, light field image coding, and rendering novel views from these light field images. We present a view-dependent rate-distortion measure that allows us to consider random access and compression efficiency simultaneously. We compare the theoretical results from the model with the experimental results from our DCT-based coder, and show that they qualitatively give similar results. Finally, we suggest how, with this model, we can better optimize the prediction dependency structure in our coder for random access and compression efficiency performance.

1. INTRODUCTION

A *light field* [1, 2] is an image-based rendering data set, that represents the outgoing radiance from a particular scene or object, at all points in 3-D space and in all directions. A light field is a 4-D data set which is often parameterized as a 2-D array of images. In this paper, we use a 2-D hemispherical arrangement of cameras surrounding the object of interest in the light field.

Efficient representation is typically a concern with light fields, due to the large amount of data involved. The most efficient compression techniques use disparity compensation, which utilizes geometry information to predict one image from one or more other images. Examples of this include DPCM-like coders [3, 4, 5] and recent wavelet-based scalable approaches such as [6, 7].

One of the main uses for image-based rendering, however, is in interactive applications, since image-based rendering involves only re-sampling the acquired image data. This is much faster than traditional approaches such as ray-tracing that synthesize scenes from light and surface shading models and scene geometry. In order to allow for re-sampling, there must be random access into the light field at the image level, and often at pixel level.

This work was supported, in part, by a gift from Intel Corporation and, in part, by Grant No. ECS-0225315 of the National Science Foundation.

While disparity-compensated prediction typically improves performance by exploiting the correlation between views, it introduces dependencies between images, which restricts random access to the data. Several approaches to balance the requirements of random access and compression efficiency have been suggested.

The vector quantization (VQ) based compression scheme in the original light field work [1] allows for random access to 2-D or 4-D blocks of pixels, but is fairly inefficient as it does not use disparity compensation. Tong and Gray [8] propose a VQ approach which does incorporate disparity compensation, and uses an indexing scheme to allow for random access. For the prediction structure, they empirically choose a three-level hierarchical decomposition of the images.

Zhang and Li [4] use disparity-compensated prediction with a two-level hierarchical decomposition, along with an indexing scheme, to give better random access. In [6, 7], varying levels of inter-view wavelet decomposition have been investigated. In [5], a full hierarchical decomposition has been used. However, in [6, 7, 5], the focus is on overall compression efficiency and not on random access.

Our motivation in this paper is to analyze this joint problem. We use a view-dependent rate-distortion metric, and propose a theoretical model for the light field coding and rendering system. Previous work on a statistical model for analyzing rate-distortion compression efficiency of light fields is presented in [9], which examines the role of the accuracy of geometry information used for disparity compensation on compression performance. There is related work on motion compensation accuracy and video compression efficiency [10].

The remainder of the paper is organized as follows. In Section 2, we start by summarizing our work in [9]. We then incorporate rendering into the model, and show how we can measure view-dependent rate and distortion. We also show how to design prediction for optimal random access and compression efficiency. In Section 3, we compare data from experiments using our DCT-based light field coder, and from our theoretical model.

2. LIGHT FIELD SIGNAL MODEL FOR CODING AND RENDERING

2.1. Background

In this section we summarize the signal model presented in [9, 5]. We model the light field object as a planar surface. We assume that this surface has a 2-D texture signal $v(x, y)$, and is viewed by N cameras from directions $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N$.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

The geometry error is modelled as an offset Δz of the planar surface from its true position. When the image is back-projected from view i onto this inaccurate geometry, this results in a texture c_i that is a shifted version of the original texture signal v. We also include image noise and non-Lambertian view-dependent effects, both modelled as additive noise, to this shifted texture image giving us the equation $c_i(x, y) = v(x - \Delta_{xi}, y - \Delta_{yi}) + n_i(x, y)$ where n_i is the additive noise component. The shift depends only upon the camera's viewing direction $\mathbf{r}_i = [r_{ix}\ r_{iy}\ r_{iz}]^T$ and the geometry error Δz.

The image vector $\mathbf{c} = [c_1\ c_2\ \cdots\ c_N]^T$ represents the set of light field images or texture maps that have already been compensated or corrected with our (inaccurate) knowledge of the geometry. A light field coder does not encode these geometry-compensated images directly, but rather tries to exploit the correlation between them. Note that perfect knowledge of the geometry would mean that the geometry-compensated images are perfectly aligned.

We represent the prediction scheme with a linear transform T, that takes as input the geometry-compensated light field images $\mathbf{c}$, and produces the set of "error" images $\{e_i\}$ that are independently encoded. In a closed-loop scheme, we predict from reconstructed images instead of the original images implied by this transform model. At high rates, however, the quantization error of the reconstructed images can be accounted for by the additive noise terms. Figure 1 shows a block diagram of this model of the light field coder, which we use in our statistical analysis.

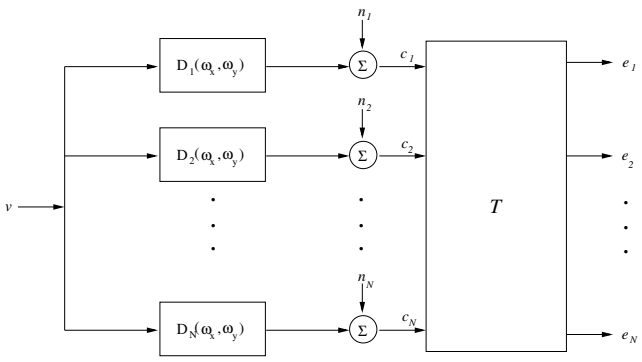


Fig. 1. The signals c_i are produced by shifting the original texture signal v by $(\Delta_{xi}, \Delta_{yi})$ (described by the transfer function $D_i(\omega_x, \omega_y)$) and adding signal-independent noise n_i. The signals $\{c_i\}$ are then encoded by first linearly transforming them with matrix T to give the signals $\{e_i\}$. These signals are then independently encoded.

We model our texture signal v as a wide-sense stationary random process. The noise signals $\{n_i\}$ are also assumed to be wide-sense stationary and independent of each other, as well as the texture signal. The shifts, represented in the frequency domain, are denoted by D_i. Using the block diagram, we can show that the error signals e_i are also wide-sense stationary, and we can calculate their statistics in terms of their power spectral densities (PSD).

2.2. Incorporating Rendering and View-Dependent Rate-Distortion Performance

Knowing the PSD, assuming Gaussian statistics and using a mean squared error distortion measure, we can compute the rate R_i and distortion D_i for each error image e_i, in terms of the rate-distortion trade-off parameter ϕ_i [11].

These individual rates and distortions can be used to calculate the rate and distortion for encoding the entire light field, but we would instead like to find the rate-distortion performance of rendering one or more novel views. We start by considering rendering a single view, and extend to a set of views. We model a rendered view v_S as a linear combination of the light field images $\{c_i\}$, given by the equation $v_S = \sum_{i=1}^{N} K_{S,i} c_i$ where $\{K_{S,i}\}$ are the rendering weights. Using the reconstructed light field images $\{\hat{c}_i\}$, we obtain the distorted rendered view $\hat{v_S} = \sum_{i=1}^{N} K_{S,i} \hat{c}_i$. The constants $K_{S,i}$ depend on the rendering algorithm. In our rendering algorithm, for instance, we use a maximum of four images to render a novel view.

For the distortion calculation, we first note that for a closed-loop predictive coder, the quantization error for e_i is identical to that of the signal c_i, i.e., $e_i - \hat{e_i} = c_i - \hat{c_i}$. The error due to coding in the rendered novel view is $v_S - \hat{v_S} = \sum_{i=1}^{N} K_{S,i}(e_i - \hat{e_i})$.

We assume that for high rates, the quantization errors $e_i - \hat{e_i}$ become uncorrelated with each other, allowing us to ignore the cross-terms in the distortion expression. The distortion due in the rendered novel view due to coding is $D_S = \sum_{i=1}^{N} K_{S,i}^2 D_i$ where D_i is the individual distortion calculated from the PSD of e_i.

Calculating the rate to decode and render a novel view requires examining the prediction dependency structure. An image required for rendering can only be decoded if other images upon which it is predicted are present. For the rate, we count all the images that need to be decoded. We denote the set of all images needed to decode the images to render a view S as $\mathcal{C}_S$. The total rate is given by $R_S = \sum_{i \in \mathcal{C}_S} R_i$ where R_i is the individual rate from calculated from the PSD of e_i.

This can now be easily extended from rendering a single view S, to rendering a set of views $\{S_t\}$, where $t = 1, \cdots, T$ and T is the number of rendered views. The distortion simply is the average of the distortion to render each view: $D = \frac{1}{T} \sum_{t=1}^{T} D_{S_t}$. The set of images required to render the set of views is now $\mathcal{C} = \bigcup_{t=1}^{T} \mathcal{C}_{S_t}$, and the overall rate is given by $R = \sum_{i \in \mathcal{C}} R_i$.

3. EXPERIMENTS

3.1. Model Parameters

In order to faithfully model the coding and rendering of a light field, we must set several parameters related to the statistics of the light field data, the light field coder, and the rendering algorithm. In this section, we describe these parameters and list the values that we use in our simulations.

There are three main parameters that we choose related to the statistics of the light field data. First, we must choose the PSD of the texture image v. We assume an isotropic exponential autocorrelation function, as in [10], with spatial correlation ρ. The quantity ρ can be estimated from the light field image.

The geometry inaccuracy Δz is modelled as a Gaussian random variable with zero mean, and variance σ_G^2. The geometry accuracy can be estimated either from *a priori* knowledge of the true geometry model, or from prediction error images.

The third parameter of importance is the independent noise component that is added to each light field. We assume white noise with variance σ_N^2. It is possible to estimate this by comparing the variance of prediction error images to the variance of the original images. It may be difficult, however, to separate error due to view-dependent and non-Lambertian effects and those due to inaccurate geometry.

For the light field coder, the main parameter of interest is the prediction dependency structure. This is encoded in the transform matrix T which is discussed in further detail in [9]. We can examine many different prediction structures simply by using a new transform matrix T. The prediction dependency structure also changes the decoding dependency between images, and must be considered when computing the rate.

The final component of our system is the renderer. Our actual rendering system follows the principles of unstructured lumigraph rendering [12]. We combine at most 4 images to create a novel view. While in the real renderer, the weighting of warped image pixels from original light field images varies spatially across the novel view, we approximate it in our model as a constant. We obtain these weights from the renderer, which calculates them based on angular distance between the original images and the novel view to be rendered.

3.2. Experimental and Model Results

We use the *Garfield* light field in our experiments. It consists of 288 images, each of resolution 192×144. From the image data, we estimate the spatial correlation to be $\rho = 0.91$. The geometry used for the experiments is estimated from the image information and is not exact. We set $\sigma_G^2 = 0.01$, which correspond to approximately half-pixel geometry-compensation accuracy in the image domain. The independent noise component is estimated from the image data to be approximately $\sigma_N^2 = 0.02$.

A DCT-based light field coder, similar to the one in [5], is used to encode the light field. We try several different prediction dependency structures with this coder, but for sake of presentation, we show results for four representative structures. The prediction structures are: INTRA coding, where each image is independently encoded with no prediction; PAIRS, where images are grouped into neighbouring pairs, with one image encoded without prediction, and the other predicted from the first; QUADS, similar to PAIRS, except that we group 4 neighbouring images together, encode one image independently and predict the others using this image; and HIERARCHICAL, as described in [5], where there are many layers of prediction, with the first layer of images independently encoded, and then the subsequent layers encoded using the layers above, until all images are encoded.

In our experiments, three different viewing patterns were investigated, reflecting three different types of random access to the light field. The first, called *SingleRandom*, consisting of a single random view from the light field, represents the situation where a user wants only a snapshot of the light field. The second, called *DenseViews*, consisting of 100 views densely clustered around some random position on the hemisphere of viewing directions, represents a sitation where the user wants to examine a specific part of the light field. The third, called *UniformViews*, consists of 100 views randomly scattered around the hemisphere of viewing directions. Here, the user wants to see all of the light field.

We run the coder on the *Garfield* data set, for each of the prediction dependency structures, and render one set of views for each of the three viewing patterns. We calculate the rate for a set of views by counting the bits for all images required for decoding, using our rendering algorithm and the specified prediction structure, and dividing by the total number of rendered pixels in that set of views. The distortion is computed by comparing rendered views from the original and reconstructed light field images, using mean squared error. For the theoretical simulations, we set the parameters to match the statistics of the light field. The distortion is calculated according to the previous section, and the bit-rate is measured in terms of bits per rendered pixel.

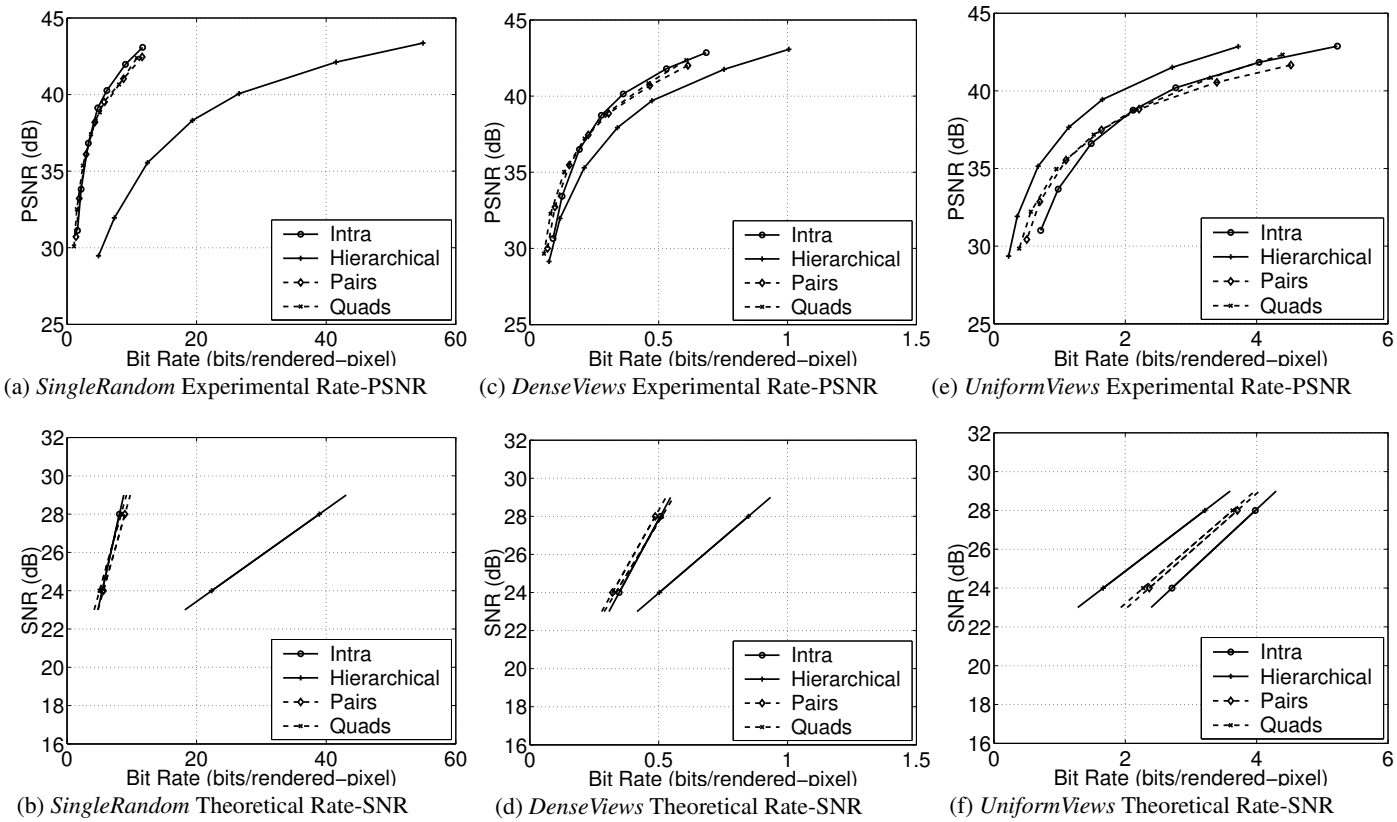
Figures 2(a)(b) shows the Rate-PSNR curve from the actual coder, and the Rate-SNR curve from the theoretical simulations, for the view set *SingleRandom*. The theoretical results agree with the experiments on two key observations about the view-dependent rate-distortion performance. The first is that the performance of the schemes INTRA, PAIRS and QUADS are all fairly similar. Here, the prediction gain in PAIRS and QUADS is balanced by having to decode more images for rendering. The second observation is that the performance of these three prediction dependency structure is far superior to that of the HIERARCHICAL prediction structure. Here, it is obvious that the cost of having to decode the necessary images in the hierarchy overshadows the prediction gain benefit.

Figures 2(c)(d) shows experimental and theoretical results for the *DenseViews* view set. Again, the theoretical results agree with the experimental in the fact that the prediction schemes INTRA, PAIRS and QUADS all perform similarly, and perform superior to the HIERARCHICAL prediction structure. There is some discrepancy between the theoretical and experimental in the magnitude of difference between the prediction schemes. The experimental results show a difference of 2 dB between INTRA and HIERARCHICAL at 0.5 bits per rendered pixel, whereas the theory suggest a difference of approximately 4 dB at the same rate.

Figures 2(e)(f) shows theoretical and experimental results for the *UniformViews* view set. The experimental results confirm those from the theory that the HIERARCHICAL prediction scheme has the superior performance in this case. At a rate of 3 bits per rendered pixel, both theory and experiments agree that the difference between the HIERARCHICAL prediction scheme and INTRA prediction is approximately 2 dB. The experimental results show that the performance of the INTRA, PAIRS and QUADS prediction schemes is within 0.25 dB of each other at 3 bits per rendered pixel, whereas the the theory has a spread of 0.5 dB between the three scheme at the same rate.

While these results are not exhaustive, they do suggest that our proposed theoretical model can describe the general trends of the performance of an actual light field coding and rendering system. The correspondence between theory and experiment is not exact. We can attribute this to numerous assumptions and simplifications used to make this model tractable, such as stationarity and Gaussian statistics, and a planar geometry model.

What we can hope to learn from this model is understanding into how the performance of light field coding and random access is affected by the various image and coder parameters. One immediate practical use is to optimize over various prediction dependency structures and find ones that are suitable to different light field rendering and viewing scenarios. Preliminary results suggest that for the *Garfield* data set, INTRA coding can give comparable or superior performance relative to several other tested prediction structures, when the entire light field is not viewed. This is hinted at in Figures 2(a-d).



(a) *SingleRandom* Experimental Rate-PSNR (c) *DenseViews* Experimental Rate-PSNR (e) *UniformViews* Experimental Rate-PSNR

(b) *SingleRandom* Theoretical Rate-SNR (d) *DenseViews* Theoretical Rate-SNR (f) *UniformViews* Theoretical Rate-SNR

Fig. 2. Experimental and theoretical results for the *Garfield* light field

4. CONCLUSIONS

In this paper, we analyze the inter-related problems of efficient coding and easy random access for light fields. We study this problem using a rate-distortion measure that depends on the novel views that a user accesses from a light field. We model the coding and rendering of light fields, and theoretically derive the view-dependent rate-distortion performance. We show that with this model, calculating the rate-distortion performance is tractable, and gives results that qualitatively match those of an actual light field coding and rendering system. Using experimental and theoretical results, we show that for some viewing scenarios, using little or no prediction can actually be optimal. Additionally, we suggest how this model can be used to optimize the prediction dependency structure given a specific light field and viewing pattern.

5. REFERENCES

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

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

[3] Marcus Magnor and Bernd 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.

[4] Cha Zhang and Jin Li, "Compression of lumigraph with multiple reference frame (MRF) prediction and just-in-time rendering," in *Proc. of the Data Compression Conference 2000*, 2000, pp. 253–262.

[5] Marcus Magnor, Prashant Ramanathan, and Bernd Girod, "Multi-view coding for image-based rendering using 3-D scene geometry," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 11, pp. 1092–1106, November 2003.

[6] Bernd Girod, Chuo-Ling Chang, Prashant Ramanathan, and Xiaoqing Zhu, "Light field compression using disparity-compensated lifting," in *Proc. of the IEEE Intl. Conf. on Acoustics, Speech and Signal Processing 2003*, Hong Kong, China, April 2003, vol. IV, pp. 761–764.

[7] Chuo-Ling Chang, Xiaoqing Zhu, Prashant Ramanathan, and Bernd Girod, "Inter-view wavelet compression of light fields with disparity-compensated lifting," in *Proc. SPIE Visual Comm. and Image Processing VCIP-1999*, Lugano, Switzerland, July 2003.

[8] Xin Tong and Robert M. Gray, "Interactive rendering from compressed light fields," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 11, pp. 1080–1091, November 2003.

[9] Prashant Ramanathan and Bernd Girod, "Theoretical analysis of geometry inaccuracy for light field compression," in *Proc. IEEE Intl. Conf. on Image Processing ICIP-2002*, Rochester, NY, USA, September 2002, vol. 2, pp. 229–232.

[10] 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.

[11] Robert Gallager, *Information Theory and Reliable Communication*, Wiley, 1968.

[12] Chris Buehler, Michael Bosse, Leonard McMillan, Steven Gortler, and Michael Cohen, "Unstructured lumigraph rendering," in *Computer Graphics (Proc. SIGGRAPH01)*, 2001.