RATE-DISTORTION OPTIMIZED STREAMING OF COMPRESSED LIGHT FIELDS

Prashant Ramanathan, Mark Kalman, Bernd Girod

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

ABSTRACT

We propose a framework for the streaming of light fields over a lossy error-prone packet network. This system is optimized for the end-user according to a rate-distortion criterion. We build upon recent rate-distortion optimized streaming work for audio and video data. We extend this work to light field image data by introducing view-dependent distortion, multiple playout deadlines, and state-based distortion. In our experimental results, the rate-distortion optimized framework has a bit-rate reduction of up to 75% over a heuristic rule-based system.

1. INTRODUCTION

Interactive photorealistic 3-D graphics has increasingly started to appear on the Internet, and has now entered the domain of the typical home user. Many interesting graphics data sets [1], however, are very large. It is therefore attractive to *stream* such large data sets to the user while they interact with it, instead of requiring the user to download the data in its entirety before viewing it. In this paper, we propose a system for the streaming of compressed light fields that is optimized for the user according to a rate-distortion criterion.

A *light field* [2, 3] 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. 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. Since image-based rendering involves only re-sampling the acquired image data, it is particularly attractive for interactive applications.

Since light fields can be extremely large, compression is an important problem and has attracted much research interest. In this paper, we use a particular compression method based on disparity compensation [4, 5]. In this approach, images are either predicted from previously encoded reference images, using a geometry model, or encoded independently. A DCT transform is used for image and residual error coding.

The light field coder uses a hierarchical prediction structure. This is designed in such a way as to provide some amount of view-point scalability. That is, with only the images from the lower levels of the prediction hierarchy, a rendered image of reasonable quality can be generated. This property is important for our streaming system.

Relatively little work has been reported on the streaming of 3-D graphics data. In [6], a streaming framework for extremely large geometry models is presented. A system for streaming high-resolution panoramic images, compressed using JPEG, is described in [7]. In [8], the authors propose a system to stream concentric mosaics, an image-based rendering data set.

In previous work such as [6, 7, 8], heuristics are used to control the transmission of packets across a network. Streaming over an error-prone network, such as the Internet, however, requires mechanisms to control against errors that are encountered. These range from simple schemes such as re-transmitting packets to incorporating Forward Error Correction (FEC) or scheduling more important packets earlier than other packets, in order to increase the chance of correct delivery by retransmission.

There has been much recent work in the area of a rate-distortion optimized streaming framework for audio and video data [9, 10, 11]. In this paper, we propose a rate-distortion optimized streaming framework for light field data sets. We build upon the prior rate-distortion optimized streaming work, and extend it to the interactive remote viewing of light fields.

The remainder of this paper is organized as follows. In Section 2, we summarize prior work on rate-distortion optimized streaming. We then describe our approach to the streaming of light fields in Section 3. We show experimental results and comparisons with a heuristic algorithm in Section 4.

2. BACKGROUND

In [10, 11], a rate-distortion optimized framework for the streaming of media over a lossy packetized network is presented. This framework assumes a compressed media representation that has been assembled into packets or data units.

Associated with each data unit is the data unit size, for instance, in Bytes, and the deadline by which the data unit must arrive in order for it to be useful for playout.

There is also a notion of distortion that the user experiences when only a subset of data units is available for decoding and playout. In order to estimate the distortion in a computationally efficient manner, the authors assume that each data unit contributes to reducing the distortion that the user experiences. A distortion reduction value is associated with each data unit, which represents the contribution of a particular data unit, given that it arrives in time, and all ancestor data units that it depends upon are also available. These distortion reductions are assumed to be additive, and the overall distortion is computed by using an acyclic directed graph that describes the inter-dependencies between the data units in the media presentation.

A transmission policy π_l is associated with each data unit l that, for instance, describes whether or not there is a transmission at each transmission interval. For each given transmission policy

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.

π_l, there is an associated cost for each byte of the data unit, $\rho(\pi_l)$, and error probability $\epsilon(\pi_l)$, the probability that the data unit does not arrive by the playout deadline.

The goal is to determine the optimal transmission policies for all data units, $\boldsymbol{\pi} = [\pi_1 \pi_2 \cdots \pi_N]$, given the per-byte costs, deadlines and distortion reductions, along with the knowledge of network packet loss and delay characteristics, acknowledgments from the receiver, and transmission history. The policy that minimizes the overall rate-distortion Lagrangian cost,

$$J(\boldsymbol{\pi}) = D(\boldsymbol{\pi}) + \lambda R(\boldsymbol{\pi}), \tag{1}$$

where λ controls the trade-off between rate and distortion, is selected as the optimal policy.

The expected rate $R(\boldsymbol{\pi})$ depends upon the data unit sizes B_l, and the per-byte costs $\rho(\pi_l)$. The expected distortion $D(\boldsymbol{\pi})$ depends upon the error probabilities $\epsilon(\pi_l)$, and uses the distortion reduction values and the inter-dependency graph for the data units.

With a large number of data units, it is difficult to exactly solve the minimization problem in (1). A reasonable approximate solution is to find the optimal policy for each data unit, while keeping the policies for all other data units fixed, and iterate until the overall solution converges. The cost function for data unit l is given by the equation

$$J_l(\pi_l) = \epsilon(\pi_l) + \lambda' \rho(\pi_l) \tag{2}$$

where $\epsilon(\pi_l)$ is the error probability and $\rho(\pi_l)$ is the per-byte cost as before. It can be shown that $\lambda' = \frac{\lambda B_l}{S_l}$, incorporating the rate-distortion trade-off operator λ from (1), the data unit size B_l, and S_l, the sensitivity of the overall distortion to not having received data unit l by its deadline. The sensitivity term represents the relative importance of a particular data unit.

3. LIGHT FIELD STREAMING

In the following sub-sections, we describe how to extend the rate-distortion optimized streaming framework to light field streaming, and we present our complete system.

3.1. View-dependent Distortion

For light field streaming, we must consider the view desired by the end-user when calculating distortion values. Hence, distortion is parameterized by the view variable v in our system. In theory, we would need to compute and store the distortions for every possible view. Instead, it appears more practical to interpolate these values from the distortion values at a small set of views. In our experiments, we assume knowledge of the view trajectory. Therefore, we can use the distortions at the views in the trajectory, and do not have to interpolate.

3.2. Multiple Deadlines

In conventional media like audio or video, a data unit is usually associated with one particular instance in time, its playout deadline. Light fields, on the other hand, typically require a particular data unit to render several different views along the user view trajectory, at different time instances.

Hence, we associate multiple playout deadlines with a data unit. This results in several changes in our overall distortion, as well as our minimization algorithm. We first redefine the distortion that the end-user experiences as the sum of the distortions at each time instance over a viewing window of the user's view trajectory.

The result is that we can no longer discuss a single error probability for a data unit. Instead, we have an entire set of error probabilities, $\{\epsilon(\pi_l, t_i)\}$, for each of the playout deadlines associated with that data unit. As a result, we also need to modify our minimization procedure.

We still minimize independently for each data unit, in an iterative fashion. Our cost function for each data unit now becomes

$$J_l(\pi_l) = \rho(\pi_l) + \sum_{i \in \mathcal{T}} \nu_{t_i} \epsilon(\pi_l, t_i) \tag{3}$$

where we consider all time instances t_i in the viewing window, indexed by $i \in \mathcal{T}$. The quantity $\epsilon(\pi_l, t_i)$ is the probability that data unit l does not arrive by time t_i; the per-byte cost $\rho(\pi_l)$ does not depend on time. The quantity ν_{t_i} is given by $\nu_{t_i} = \frac{S_{l,t_i}}{\lambda B_l}$, analogous to the reciprocal of λ' in (2). Note that the sensitivity term S_{l,t_i} is also indexed by time; it is the sensitivity of the overall distortion to the non-arrival of data unit l by time instance t_i.

The cost (3) is computed for each policy, and the one with the lowest cost is selected as the optimal policy. Since the policy length tends to determine the complexity of the algorithm, using multiple deadlines does not greatly impact the computational complexity of the algorithm.

3.3. State-dependent Distortion

In this subsection, we present our calculation of the distortion values. Suppose that for a particular view $v(t_i)$ at time instance t_i, we require the set of L data units, $\mathcal{L}_{v(t_i)}$, to render that view. There are 2^L possible combinations of arrivals and non-arrivals for this set of data units. We consider each such combination as a state s that belongs to the set of 2^L states $\mathcal{P}_{v(t_i)}$.

For every $s \in \mathcal{P}_{v(t_i)}$, we can compute the corresponding distortion that end-user would experience, so that we have a set of 2^L distortion values, one corresponding to each state. These distortion values are used to compute the expected distortion for a transmission policy.

This state-based approach can quickly become intractable for even a modest number of data units. For instance, the data sets that we use in our experiments typically require 20 data units to render a view, which would result in over $1,000,000$ states. We can, however, exploit the hierarchical structure of our light field data set to reduce the state space.

According to the hierarchy, we can organize the data units into several different levels, where the data units at a particular level depend upon some of the data units at next lower level. To simplify the state-space, we assume that a data unit requires *all* the data units at the next lower level; this appears to be a reasonable assumption for our data sets.

We consider all combinations of arrivals and non-arrivals of data units at a particular level, only if all data units at all lower levels have arrived. The resulting total number of states is simply the sum of the number of states for the data units for each level. If the data units are evenly distributed across levels, then this results in a significant reduction in the size of the state space. We denote this reduced state space as $\mathcal{P}'_{v(t_i)}$, and we associate a unique distortion value with each $s \in \mathcal{P}'_{v(t_i)}$.

In order to use these state-based distortion values in our framework, we need to compute the expected distortion, $D(\boldsymbol{\pi})$, for a

given policy. Given a transmission policy, error probabilities can be calculated for each data unit. Given these error probabilities, we can compute the probability of any state. We compute the expected distortion for this policy using the state distortion values and the probability of each state occurring. The next section details the calculation of the expected distortion, and the sensitivity values, incorporating view-dependence, multiple deadlines, and a state-based approach.

3.4. Overall System

In our system, for a view trajectory $\mathbf{v}$ and policy $\boldsymbol{\pi}$, the overall distortion is

$$D(\boldsymbol{\pi};\mathbf{v}) = \sum_{i\in\mathcal{T}} \left[\sum_{s\in\mathcal{P}'_{v(t_i)}} D(s, v(t_i)) \Pr\{s\} \right] \tag{4}$$

where the probability of a state $\Pr\{s\}$ is given by

$$\Pr\{s\} = \prod_{\substack{l\in\mathcal{L}_{v(t_i)}\\ l:s_l=1}} (1-\epsilon(\pi_l,t_i)) \prod_{\substack{l\in\mathcal{L}_{v(t_i)}\\ l:s_l=0}} (\epsilon(\pi_l,t_i)). \tag{5}$$

The binary-valued variable s_l indicates whether data unit l has arrived or not in state s. The rate, given by

$$R(\boldsymbol{\pi}) = \sum_l B_l \rho(\pi_l), \tag{6}$$

does not depend on the view trajectory directly. Note, however, that it is coupled indirectly through the choice of the transmission policy $\boldsymbol{\pi}$ appropriate for that trajectory.

In order to minimize the overall Lagrangian cost (1), as before, we minimize the Lagrangian cost (3) for each data unit. For our overall system, we must take into account the new distortion calculation in (4). This involves deriving a new sensitivity term

$$S_{l,t_i} = \begin{cases} \sum_{\substack{s\in\mathcal{P}'_{v(t_i)}\\ s_l=0}} S'_{s,l,t_i} - \sum_{\substack{s\in\mathcal{P}'_{v(t_i)}\\ s_l=1}} S'_{s,l,t_i} & \text{if } l\in\mathcal{L}_{v(t_i)}, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where

$$S'_{s,l,t_i} = D(s,v(t_i)) \prod_{\substack{l'\in\mathcal{L}_{v(t_i)}\\ l':s_{l'}=1\\ l'\neq l}} (1-\epsilon(\pi_{l'},t_i)) \prod_{\substack{l'\in\mathcal{L}_{v(t_i)}\\ l':s_{l'}=0\\ l'\neq l}} (\epsilon(\pi_{l'},t_i)). \tag{8}$$

4. EXPERIMENTAL RESULTS

We compare our rate-distortion optimized streaming framework with a heuristic scheme that we describe in Subsection 4.1. In both cases, we simulate the network conditions, compute transmission sequences for both algorithms, and derive the set of available data units at the receiver for each time instance in the simulation.

We use the *Garfield* light field data set for comparison. The *Garfield* data set consists of 288 views, each of pixel resolution 192×144. There are 7 levels in the hierarchy for *Garfield*. We encode this data set at the highest possible quality, using the hierarchical disparity-compensated prediction scheme in [4, 5]. We render from the light field data set using a geometry model. The geometry model, along with the camera parameters of the light field, are assumed to be sent initially, and are not counted in the transmission bit-rate.

In this work, we assume perfect advance knowledge of the user's view trajectory. We consider 2 different trajectories of the *Garfield* data set, both involving rotating about the object in a particular direction. For each view in a trajectory, the state-based distortion values are computed and stored for use by our algorithm. The distortion for a state is measured by comparing the rendered view from the original uncompressed light field to the rendered view using only the available data units, as defined by the state, of the decoded light field.

Our model of the network assumes identical delay and loss probability distributions for each data unit. Successive packet delays and losses are assumed to be independent. We assume a loss probability on the forward channel from the sender to the receiver of 20%; data units that are not lost are subject to delay according to a shifted gamma distribution with a shift of 10ms, a mean of 50ms, and a standard deviation of 23ms. Acknowledgments are sent from the receiver to the sender on the back channel with no loss, but with an identical gamma delay distribution as the forward channel. We allow for a 400ms delay between the first transmission and the rendering of the first view.

In order to determine which views in the trajectory should be considered for transmission, we use the windowing function in [11] to generate the view window. In our Lagrangian cost minimization, we use a policy length of 6. Increasing the length beyond this value increases computational complexity without much benefit.

4.1. Heuristic Packet Scheduling

In order to provide a baseline for our rate-distortion optimized streaming system, we implement a packet-scheduling algorithm based on some heuristic rules. This heuristic scheduler has knowledge of the view trajectory, the viewing window, the deadlines and level of each data unit. It also has knowledge of the network conditions, and can calculate the probability of arrival for each data unit given its transmission history.

Using this information, the scheduler can sort the data units to be sent according to importance, and send them in the order of importance until the bit-rate constraint is met. The importance of a data unit is determined by whether or not it has already been sent, its level, and its playout deadline, approximately in that order. Additionally, if the conditional probability of a data unit not arriving, given that it has not been acknowledged, is over the arbitrary threshold of 99%, we consider this to be a negative acknowledgment. Such a data unit is considered to be as important as a data unit that has not been sent.

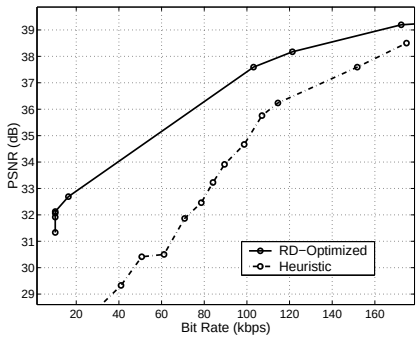
4.2. Comparison

In our comparison, we simulate network conditions, and record the transmissions computed by both our rate-distortion optimized and heuristic scheduling algorithms. This determines a set of data units that are available for rendering at each time instance. We calculate the image quality of these renderings using PSNR, with the rendered view from the original uncompressed light field as reference.

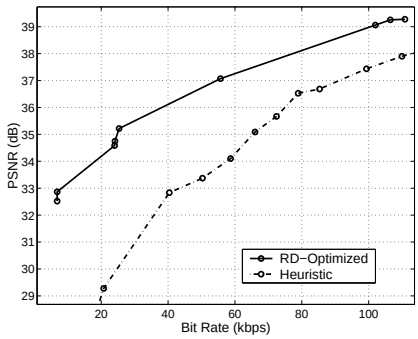
For the rate-distortion optimized system, we vary the Lagrangian multiplier λ in (1) to vary the bit-rate. For the heuristic

scheme, we simply vary the target average bit-rate that the scheme strives to meet.

For our two trajectories, the comparison rate-PSNR curves are shown in Figure 1. With a policy length of 6, the rate-distortion optimized scheme outperforms the heuristic scheme by several dB at low bit-rates, and at least 1dB at higher bit-rates. At low bit-rates, the rate-distortion optimized scheme only transmits the few data units that are crucial to the view. The heuristic-based scheme does not know which data units are important within a given level. The rate-distortion optimized scheme, however, tends to send data units in bursts, whereas with the heuristic scheme, the bit-rate varies less over time.



(a)



(b)

Fig. 1. Rate-PSNR curves for the *Garfield* data set, for trajectories of (a) 101 views and (b) 201 views, showing the streaming performance of a rate-distortion optimized scheme and a heuristic scheme.

5. CONCLUSIONS

We have presented a rate-distortion optimized framework for the streaming of light fields. This framework takes into account the network conditions, history of transmissions, and acknowledgments, and the importance of each data unit. Our main contribution is in extending the prior rate-distortion optimized framework for audio and video to light fields. We introduce the notion of a view-dependent image distortion; we associate with each data unit multiple playout deadlines, corresponding to the views when this data unit is required for rendering; and we consider state-based distortion values to give a more accurate estimate of the overall image distortion as experienced by the end-user. We show in our results with the *Garfield* light field that our rate-distortion optimized streaming system out-performs a heuristic scheme. We report gains of over 1dB in image quality at higher bit-rates, and of up to several dB at lower bit-rates.

6. REFERENCES

[1] 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.

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

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

[4] 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.

[5] Prashant Ramanathan, Eckehard Steinbach, Peter Eisert, and Bernd Girod, "Geometry refinement for light field compression," in *Proceedings of the IEEE International Conference on Image Processing ICIP-2002*, Rochester, NY, USA, September 2002, vol. 2, pp. 225–228.

[6] Szymon Rusinkiewicz and Marc Levoy, "Streaming QSplat: A viewer for networked visualization of large,dense models," in *Proc. 2001 Symposium on Interactive 3D Graphics*, 2001.

[7] Carsten Grünheit, Aljoscha Smolić, and Thomas Wiegand, "Efficient representation and interactive streaming of high-resolution panaromic views," in *Proceedings of the IEEE International Conference on Image Processing ICIP-2002*, Rochester, NY, USA, September 2002, vol. 3, pp. 209–212.

[8] Cha Zhang and Jin Li, "Interactive browsing of 3D environment over the Internet," in *Proceedings SPIE Visual Communications and Image Processing VCIP-2001*, 2001.

[9] Matthew Podolsky, Steven McCanne, and Martin Vetterli, "Soft ARQ for layered streaming media," Tech. Rep. UCB/CSD-98-1024, University of California, Computer Science Department, Berkeley, CA, November 1998.

[10] Philip A. Chou and Anshul Seghal, "Rate-distortion optimized receiver-driven streaming over best-effort networks," in *Packet Video Workshop*, Pittsburgh, PA, USA, April 2002.

[11] Philip A. Chou and Zhourong Miao, "Rate-distortion optimized streaming of packetized media," Tech. Rep. MSR-TR-2001-35, Microsoft Research, February 2001, (also submitted to IEEE Transactions on Multimedia).