IEEE CONFERENCE PROCEEDINGS: INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP-99)
KOBE, JAPAN, VOL. 3, PP. 334-338, OCTOBER 1999

HIERARCHICAL CODING OF LIGHT FIELDS WITH DISPARITY MAPS

Marcus Magnor and Bernd Girod

Telecommunications Laboratory, University of Erlangen-Nuremberg
Cauerstraße 7, 91058 Erlangen, Germany
email: {magnor,girod}@nt.e-technik.uni-erlangen.de

ABSTRACT

A coder for light fields is presented. Due to the large amount of data needed to represent a complete light field, a hierarchical decomposition is employed. The full light field is built up by recursively predicting intermediate images from a small subset of light-field images. Intermediate images are predicted by disparity-compensating multiple surrounding images. The predicted images are refined using DCT coding. Rate-distortion measurements for two standard light fields verify the efficiency of the proposed scheme. Compression ratios of 1000:1 are achieved at acceptable quality of the rendered views.

1. INTRODUCTION

Common computer graphics rendering schemes are based on descriptions of object geometry, texture properties and illumination conditions. Photorealistic rendering results require very detailed geometry models and the computationally demanding simulation of global illumination effects. Acquisition of geometry information from natural objects is difficult and can even prove impossible if objects exhibit ill-defined surfaces (e.g., fur, hair, smoke, etc.).

In 1996, a new technique called *Light Field Rendering* (LFR) [1, 2] was proposed that promises to overcome the limitations of traditional geometry-based rendering. Instead of 3-D geometry descriptions, LFR relies on a set of ordinary 2-D images of a static scene, enabling easy data acquisition from virtual as well as real-world scenes. If the light field is sufficiently densely sampled, LFR can provide photorealistic rendered images at interactive frame rates, regardless of scene complexity.

Unfortunately, huge amounts of data are required to attain photorealistic rendering results: tens to hundreds of thousands of full-sized color images need to be compressed efficiently while maintaining fast access to arbitrary data segments. In [1], vector quantization of the light field yields compression ratios of 27:1. A block-based DCT coding scheme is used in [3] to compress light-field data by a factor of 24:1. For fast data transmission and rendering on standard hardware, much higher compression ratios are desirable.

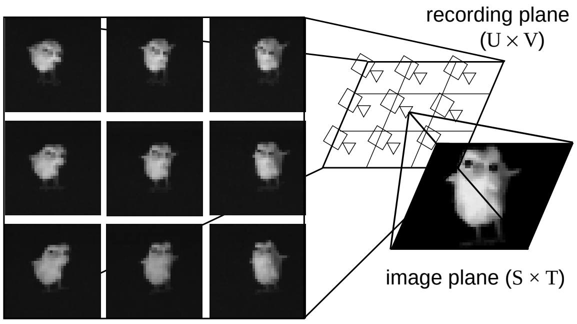
We propose a light-field coder that exploits inter-image similarities by compensating disparity between light-field images. All images are coded in hierarchical order for fast access to multiple resolution levels. Images that are not contained in the recorded light-field image set can be estimated by disparity compensation to facilitate rendering of photorealistic views from sub-sampled light fields.

This paper is organized as follows. In the next section, we describe the light-field data structure and its sub-sampled representation. We go on to explain how disparity maps are derived for all images and how light-field images are predicted. The light-field coding algorithm is described and experimental results are presented to validate the proposed coding scheme.

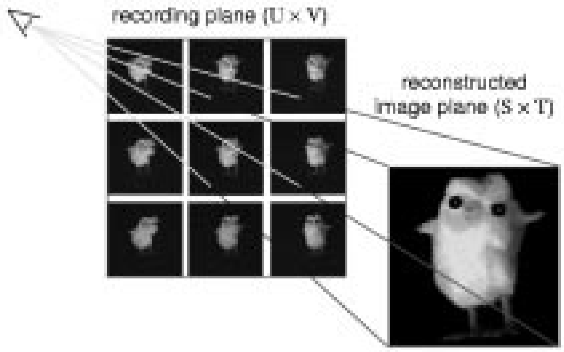
2. THE LIGHT FIELD

An illuminated, static object fills the surrounding space with light reflected off its surfaces, establishing the object's *light field* [4]. The light field can be captured by recording images from multiple view points around the object (Fig. 1). The light-field images' recording positions are arranged on planes facing the object [1, 2]. On each plane, the recording positions are arranged in a regular grid, so the light-field structure resembles a two-dimensional array of images.

Optimal rendering quality is achieved if disparity between adjacent images does not exceed one pixel. Respecting this constraint, the number of images needed to capture a complete light field is proportional to image resolution. For example, to record a light field at 256^2 pixels resolution requires ≈ 262000 images, which



Recording a light field.



Light Field Rendering.

Figure 1: An object's light field is recorded by taking a number of conventional images from regularly spaced positions in a plane. During Light Field Rendering, the viewing rays' intersection points with the recording plane and image plane are calculated, and the image array is resampled.

corresponds to 48 GBytes, given that each pixel is represented by 24 bits. Acquisition of such a large number of images is practically impossible. Therefore, light fields are almost always a sub-sampled representation of the complete light-field information.

During rendering, the image array is resampled by tracing rays from the view point along the viewing direction. The rays' intersection points with recording and image plane determine which light-field segments to render. Because intersection coordinates do not, in general, coincide with image recording positions, in-between image segments must be estimated based on the available sub-sampled light-field data. To avoid aliasing, quadralinear interpolation is applied in [1, 2], resulting in low-pass filtered (blurred) rendered images. To avoid unsharp rendering results, the available array of light-field images has to be refined by disparity-compensated intermediate images. The presented coder allows to estimate disparity-compensated intermediate images, enhancing rendering quality from sub-sampled light fields.

3. DISPARITY MAP ESTIMATION

When comparing two adjacent light-field images, a point on the surface of the depicted object appears at different positions in the images. This displacement is the point's parallax or *disparity*.

From known image recording positions, the disparity *direction* between two images can be inferred. The disparity *magnitude* is a scalar value associated with each pixel that describes the amount of shift along the disparity direction. An image's *disparity map* is an array of scalar values that lists the amount of shift for each pixel when compared to neighboring images.

As object geometry is not recorded during light field acquisition, disparity information must be retrieved from the light-field images. Disparity maps are estimated on image blocks (Fig. 2). Because the light-field images are arranged in a regular grid of equal spacing, an image block's true disparity magnitude is the same when comparing the block to all four neighboring images along the respective disparity direction. To find the magnitude of disparity for each block, a number of disparity values within a predefined search range are considered. For each disparity value, the corresponding blocks from all four neighboring images are extracted. The extracted blocks are averaged and compared to the original block. The disparity value resulting in the smallest prediction error is chosen as the block's disparity magnitude. Full-pel disparity accuracy was found to be sufficient, in accordance with considerations in [5] for multi-hypothesis compensation. A fixed Huffman-code tabe is used to code the disparity values.

In contrast to geometry reconstruction algorithms, the described method converges regardless of light-field scene characteristics. The estimated low-resolution disparity maps allow optimal disparity compensation of light-field images from surrounding images, and disparity-compensated prediction of intermediate (missing) images becomes possible. Coding efficiency can be optimized by varying the disparity maps' block size.

If the light-field object's full 3-D geometry can be estimated, high resolution disparity maps are available. Instead of Huffman-coding each disparity map entry, arbitrarily many maps can be generated from the object model at constant geometry coding bit-rate. Results for geometry-aided light-field coding are presented in [6] in these proceedings.

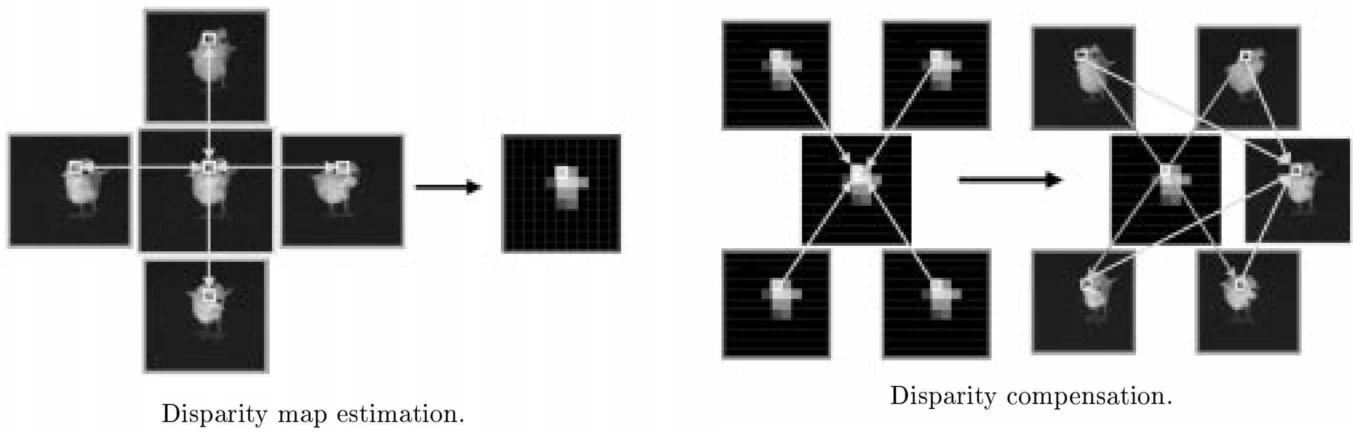


Figure 2: Disparity maps are estimated for all light-field images by comparing image blocks to adjacent images. Image prediction is performed by first estimating the target image's disparity map which is used for disparity-compensating from all reference images.

4. DISPARITY COMPENSATION

The disparity maps are estimated with one-pixel accuracy in respect to neighboring images. If reference and target images are farther apart, full-pixel accurate compensation is possible only at lower image resolution. Therefore, prediction over greater distances is performed after the images have been downsampled by an appropriate factor. The disparity-compensated image is upsampled again to yield the target image's prediction. Given multiple reference images and their corresponding disparity maps, it was experimentally found that a target image and its disparity map can best be predicted in the following manner (Fig. 2):

1. All reference images' disparity maps are low-pass filtered and forward-compensated to the target image's position, yielding an estimate for the target image's disparity map.
2. The estimated disparity map is used to backward-compensate the target image from all reference images.

The disparity maps are low-pass filtered to yield smooth disparity fields. If reference and target positions are farther apart, the smoothed disparity maps are downsampled appropriately. The disparity maps are then forward-compensated to the target image's position. The resulting disparity maps are averaged and undefined patches, if any, are filled by interpolation. By upsampling to original image size, an estimate for the target image's disparity map is derived.

The target image itself is predicted by using the estimated disparity map and backward-compensating from all reference images. If reference and target images are not adjacent, the reference images are downsampled. All reference images are disparity-compensated, and the prediction results are averaged. Finally, the estimated image is upsampled to yield the target image.

5. THE CODER

Prior to coding a light field containing $U \times V$ images, a minimum reconstruction quality parameter q_{min} is set. All reconstructed light-field images must meet this quality criterion. Image quality is measured as the peak-signal-to-noise ratio (PSNR) of the image's luminance signal.

First, the image array's four corner images $(1,1)$, $(U,1)$, (U,V), $(1,V)$ are intra-coded using a standard block-DCT scheme adopted from still-image compression (positions A in Fig. 3). For each image, the DCT quantization parameter is individually chosen to ensure that the reconstructed image meets the reconstruction value q_{min}. The corner images' disparity maps are Huffman-coded applying a fixed table.

From the four reconstructed corner images, the center image $(U/2, V/2)$ and its disparity map (position B in Fig. 3) is predicted as described. If the image estimate meets the reconstruction criterion q_{min}, no further information regarding the center image is coded. Otherwise, the center image is again predicted using its original disparity map which is then Huffman-coded. If image quality still doesn't suffice, the prediction error is additionally DCT-coded. The DCT quantizer level is thereby adjusted to yield minimum bit-rate for the required image quality. Then, the four middle images on the array sides $(U/2,1)$, $(U,V/2)$, $(U/2,V)$, $(1,V/2)$ (positions C in Fig. 3) are predicted from the reconstructed center image and the two closest corner images. If necessary, the residual error and disparity

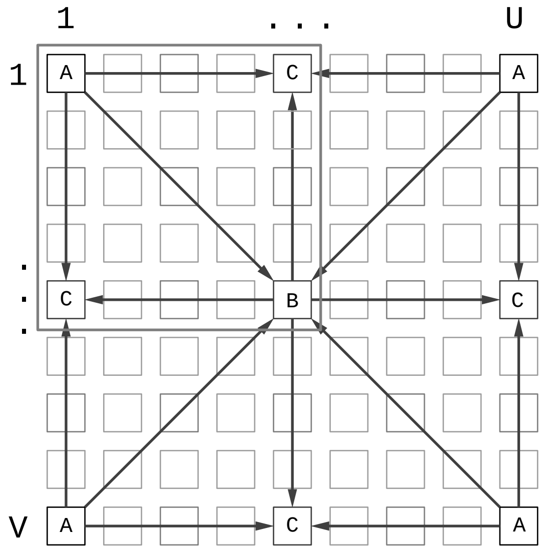


Figure 3: Coding order of the light field image array: from the corner images (A), the center image (B) is predicted. The images at the middle of the sides (C) are predicted from the center image and the two closest corner images. The array is subdivided into quadrants and each quadrant is coded likewise. The algorithm keeps recursing until all images are coded.

maps are coded.

At this point, 9 light field images spanning the full recording plane are available. The image array is now divided into four quadrants. The four corner images of each quadrant are already coded and, as before, the center and side images in each quadrant are predicted. The algorithm keeps recursing through the quadtree structure until all images are coded.

The proposed decoder allows access to different levels of increasing sampling density of the $U \times V$ image array. Further, the decoder can locally refine the light field by estimating disparity-compensated intermediate light-field images that were not originally recorded. These interpolated images can greatly enhance rendering performance [7].

6. RESULTS

The proposed coding scheme is validated using two publically available light fields (*Buddha, Dragon*, Fig. 4)[1]. Both light fields contain $U \times V = 32 \times 32$ images of $S \times T = 256 \times 256$ 24-bit RGB pixels, totalling 192 MBytes. In both data sets, disparity between adjacent images ranges from -3 to 3 pixels.

Disparity maps with different block sizes are estimated for all light-field images. 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.

[1] www-graphics.stanford.edu/software/lightpack/lifs.html

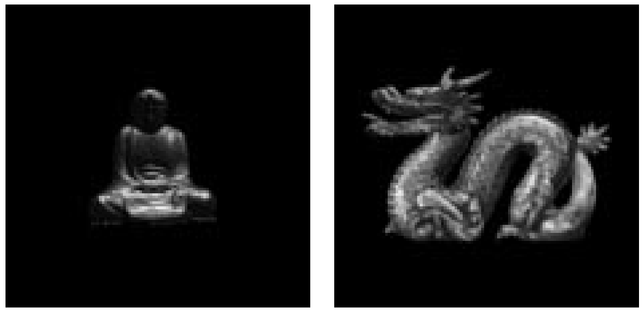


Figure 4: Images from the light fields *Buddha* and *Dragon*.

The rate-distortion curves for various disparity map resolutions are shown in Fig. 5. For both light fields, maps of block size 8×8 and 16×16 yield the best allocation of bit-rate between disparity information and residual error for all reconstruction distortion values considered.

Coding efficiency depends on reconstruction quality and light-field scene characteristics. The *Dragon* object exhibits detailed texture and complex geometry. 569 kBytes are needed to code its light field at 36 dB PSNR. The *Buddha* statue appears smaller in the images and has a more compact geometry, enabling the light field to be compressed to 222 kBytes at 40 dB PSNR.

Fig. 6 depicts the number of light-field images that need encoding of the residual error. At lower image reconstruction quality, many light-field images are pre-

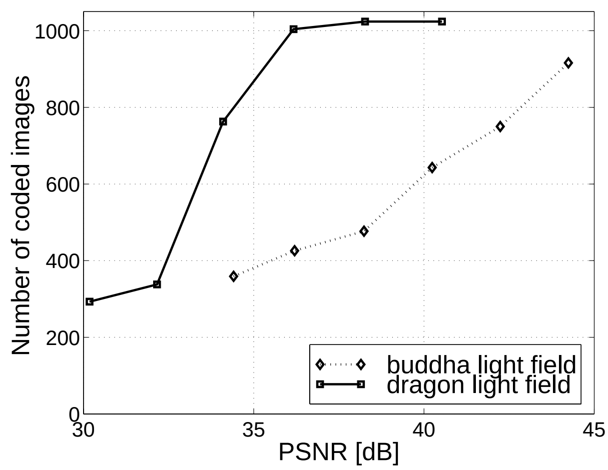


Figure 6: Number of residual-error coded light-field images (disparity map block size 16×16 pixels).

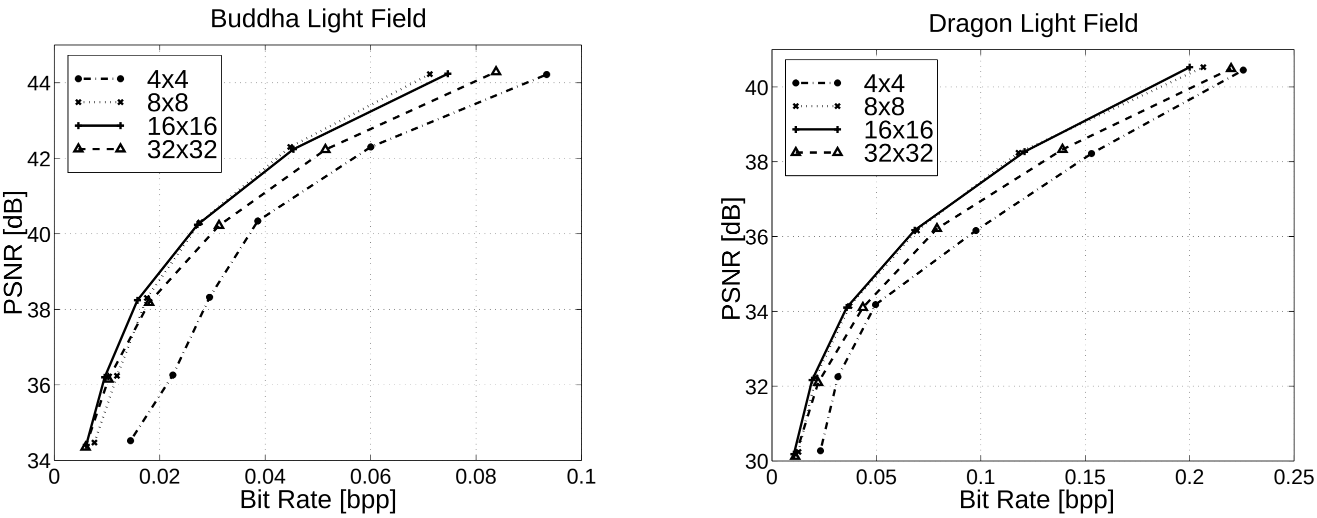


Figure 5: Rate-distortion measurements for 1024 images in each light field at different disparity map block sizes.

dicted well enough by disparity compensation to neglect the residual error. This way, the coder decimates the number of light-field images and only codes relevant images. The discarded images are reconstructed by merely disparity-compensating surrounding images.

7. CONCLUSIONS

A coding scheme has been presented that facilitates hierarchical decoding of the light-field data structure. The presented algorithm allows disparity-compensated estimation of missing light-field information, yielding enhanced rendering results. At reasonable reconstruction distortions, light fields can be compressed by factors of 1000:1 and higher if compared to the amount of recorded light-field data available as input. With regard to the complete light-field information necessary for optimal rendering performance, compression ratios exceeding 10,000:1 are achieved. These bit-rates allow convenient storage of all light-field data in local memory for rendering on standard hardware. Also, for possible future 3D-television or Internet distribution applications, capacity requirements for transmitting light-field data is greatly reduced.

8. REFERENCES

[1] Marc Levoy and Pat Hanrahan, "Light field rendering", *Computer Graphics (SIGGRAPH '96 Proceedings)*, pp. 31–42, Aug. 1996.

[2] Steven J. Gortler, Radek Grzeszczuk, Richard Szelinski, and Michael F. Cohen, "The Lumigraph", *Computer Graphics (SIGGRAPH '96 Proceedings)*, pp. 43–54, Aug. 1996.

[3] Gavin Miller, Steven Rubin, and Dulce Ponceleon, "Lazy decompression of surface light fields for precomputed global illumination", *Rendering Techniques '98 (Eurographics Proceedings)*, pp. 281–292, June 1998.

[4] A. Gershun, "The light field", *J. Math Phys.*, vol. 18, pp. 51–151, 1939.

[5] B. Girod, "Efficiency analysis of multi-hypothesis motion-compensated prediction for video coding", *IEEE Trans. Image Processing*, in print.

[6] B. Girod, P. Eisert, M. Magnor, E. Steinbach, and T. Wiegand, "3-d image models and compression - synthetic hybrid or natural fit ?", *Proc. International Conference on Image Processing ICIP-99,* Kobe, Japan, Oct. 1999.

[7] W. Heidrich, H. Schirmacher, H. Kück, and H.-P. Seidel, "A warping-based refinement of lumigraphs", *Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization*, 1998.