MODEL-AIDED CODING OF MULTI-VIEWPOINT IMAGE DATA

Marcus Magnor, Peter Eisert

Telecommunications Laboratory
University of Erlangen-Nuremberg
{magnor,eisert}@LNT.de

Bernd Girod

Information Systems Laboratory
Stanford University
girod@ee.stanford.edu

ABSTRACT

The paper presents a novel coding technique based on approximate geometry for images taken from arbitrary recording positions around a 3-D scene. Such data structures occur in image-based rendering applications, where many hundreds to thousands of images need to be stored and transmitted. After reconstructing and compressing the scene's approximate geometry, the 3-D model is used to infer disparity, occlusions and object silhouette, leading to improved prediction of images recorded from arbitrary angles. The images are hierarchically coded to ensure efficient exploitation of inter-image similarities. The coding scheme is capable of generating novel views by texturizing the available geometry model. The presented algorithm is validated using synthetic as well as natural image data sets, achieving up to 1000:1 compression at acceptable reconstruction quality.

1. INTRODUCTION

In image-based rendering (IBR), a set of images of a three-dimensional scene is used to create novel views of the depicted setting [1, 2]. To achieve photorealistic rendering results, however, many hundreds to thousands of images are required in order to sufficiently sample scene appearance. Several compression schemes have been proposed to store and transmit the large amount of image data as well as to fit all data into local memory during rendering. Vector quantization [1] and DCT coding [3] allow for swift decoding during rendering, but only low compression ratios are attained (below 30:1). Much higher compression ratios can be achieved if inter-image similarities are exploited [4]. By compensating for disparity between images, compression ratios on the order of 1000:1 have been obtained [5].

Until recently, camera recording positions were restricted to lie in a plane on a regular grid to allow interactive rendering rates [1, 2]. Lately, random recording positions within the plane [6] as well as spherical recording arrangements [7] have been shown to be applicable to IBR as well. This new freedom in choosing recording positions gives rise to the need for advanced coding schemes that are capable of efficiently compressing arbitrarily recorded image sets.

The proposed codec has been designed to code images of a 3-D scene taken from arbitrary, known viewpoints. From these views, the depicted scene's approximate geometry is derived. The geometry model allows estimating object silhouette, occlusions and disparity-compensated surface texture, leading to improved image predictions and enhanced coding performance. In addition, novel views can be efficiently generated by texturizing the model using one or several of the coded images.

This paper is organized as follows. In the next section, we briefly describe how approximate scene geometry is obtained and efficiently coded. We go on to explain how images are predicted using the 3-D model. The coding algorithm is described and experimental results are presented to validate the proposed coding scheme.

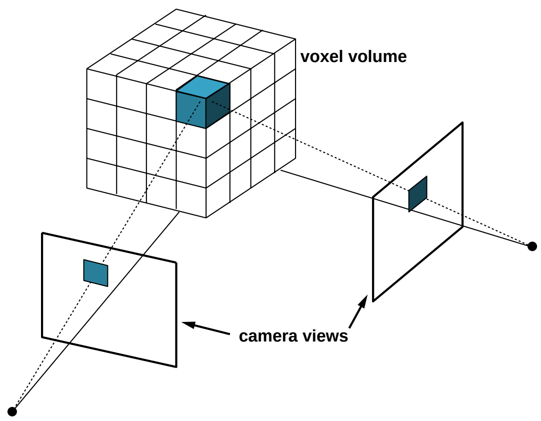
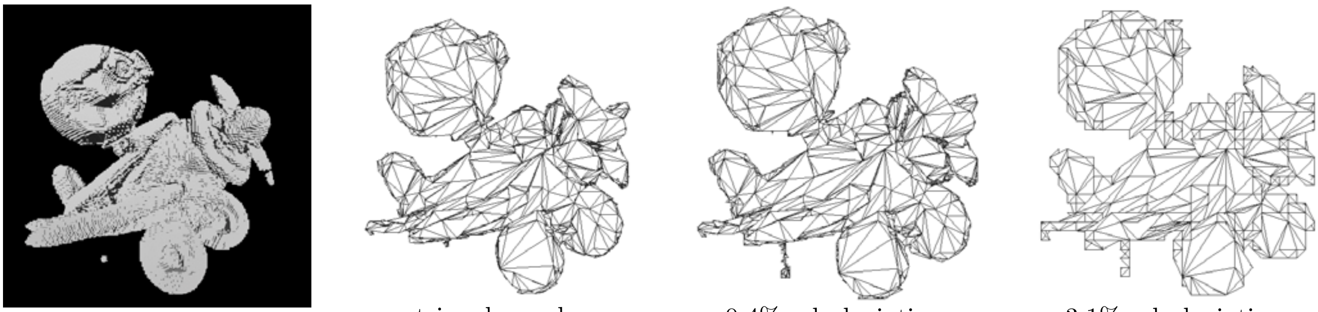


Figure 1: Volumetric geometry reconstruction [8]: the volume enclosing the scene is divided into voxels, each voxel is projected into all visible images, and a voxel is removed if no consistent color hypothesis can be derived.



reconstructed voxel model | triangle mesh 168104 bits | 0.4% rel. deviation 47061 bits | 3.1% rel. deviation 25156 bits

Figure 2: The reconstructed model consists of numerous small cubes. After triangulating the surface, the number of triangles is reduced while maintaining the initial model's resolution accuracy. The triangle mesh is progressively encoded, continuously adding detail to the model by refining vertex positions and increasing the number of triangles.

2. MODEL GENERATION AND GEOMETRY COMPRESSION

The scene's approximate 3-D geometry has to be inferred using all available images in conjunction with the known internal and external camera parameters. A robust scheme to reconstruct an approximate geometry model from multiple calibrated camera views has been described in [8]. The algorithm discretizes the volume enclosing the scene and projects voxels into all visible images, Fig. 1. For each voxel, color hypotheses are collected and checked for consistency. A voxel is removed if none of its color hypotheses is in accordance with its image projections. Iteration over all voxels ends when no further voxels can be removed. The voxel model is triangulated, and triangles not visible in any image are eliminated, Fig. 2. Triangle number is further reduced until geometry accuracy matches the volume model's surface accuracy which is determined by voxel size. The resulting triangle mesh is progressively encoded using the algorithm described in [9] which refines vertex positional information and increases the numbers of triangles, allowing nearly continuous adjustment of the model's level of detail. This way, bit-rate can be optimally allocated between geometry and residual error coding.

3. MODEL-AIDED IMAGE PREDICTION

Scene appearance from any viewpoint can be predicted using the scene's geometry model in conjunction with images recorded from nearby positions. The geometry model is projected into the image plane of the desired view. Each visible triangle is located in the surrounding images, and occluded areas are detected. Triangle texture is derived and averaged over the number of images. Image regions not visible in any of the surrounding images are estimated by local interpolation. An occlusion-compliant, disparity-compensated prediction of the considered image is obtained.

If model accuracy suffices, 3-D geometry can further provide silhouette information for each image which can be used to code only image regions of interest. If only coarse 3-D information is available, image content outside the silhouette must be additionally coded. As the model-aided prediction scheme can only deliver estimates for those image regions covered by the model's projection, accurate silhouettes reconstructed from the model are crucial to efficient model-aided coding.

4. HIERARCHICAL IMAGE CODING

To efficiently exploit inter-image similarities and allow fast access during decoding, the images are hierarchically ordered prior to coding. The hierarchical coding order presented in [5] for regular and planar recording patterns has been extended for arbitrary image record-

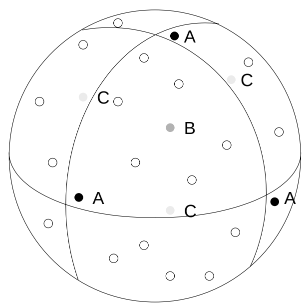


Figure 3: Image recording positions are projected onto a sphere surrounding the scene. The images closest to the sphere's poles and four images along the equator are intra-coded (A). The image closest to the center of each octant (B) is predicted from the corner images, and the nearest mid-side images (C) are predicted from the central and 2 corner images. Each octant is then subdivided until all images are coded.

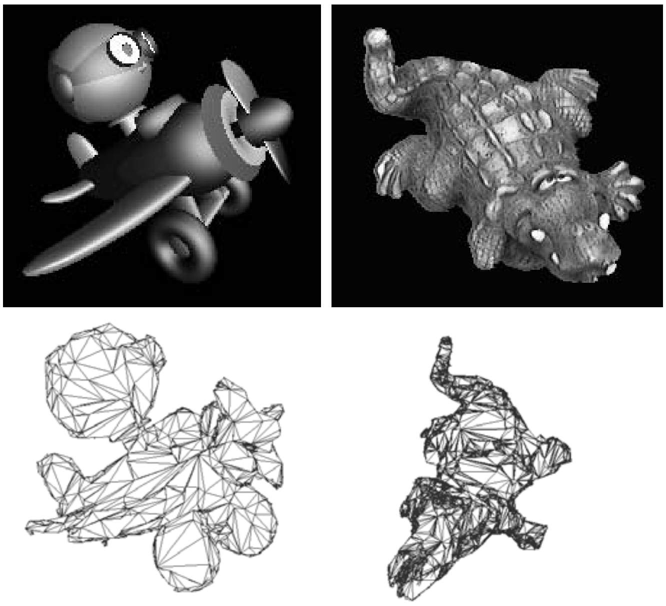


Figure 4: Images and reconstructed geometry models of the test data sets *Airplane* and *Crocodile*.

ing positions.

To obtain a two-dimensional parameterization, all image recording positions are projected onto a sphere around the scene, yielding the images' directional distribution in spherical coordinates. Two images closest to the sphere's poles and four images evenly distributed around the equator are intra-coded using a standard block-DCT scheme adopted from still-image compression (Images A in Fig. 3). For each image, the DCT quantization is individually chosen to ensure that the reconstructed image meets a preset minimum reconstruction quality. The 6 intra-coded images are arranged into eight groups, each consisting of one polar and two neighboring equatorial images, such that the sphere is divided into 8 octant regions. The image closest to an octant's center is estimated using the described model-aided prediction scheme from the 3 corner images (Image B). If the predicted image does not reach the required quality, the residual error is DCT-coded. Then, the three images closest to the edge midpoints are predicted from the center image and two corner images (Images C). As described in [5], the octant is divided into sub-regions with already coded corner images. Within the sub-regions, center- and side images are again predicted. Subdivision continues recursively until no more images need to be coded.

The evolving quadtree-structure allows fast access to different levels with increasing sampling density. The decoder can locally refine the image distribution by estimating scene appearance at intermediate positions using the 3-D model.

5. RESULTS

The proposed coding scheme is validated using two different image data sets, Fig. 4. The *Airplane* data set depicts a computer-rendered scene and consists of 256 images of 256×256 pixels. Recording positions lie in a plane and are arranged in a regular grid. The second test set of 257 images has been recorded with 384×288 pixels resolution from a real-world *Crocodile* clay figure in spherical configuration. Approximate geometry models are derived for both scenes using the described voxel-based method. The reconstructed airplane model is of higher accuracy than the crocodile's geometry because of calibration errors inherent to real-world recordings. Both geometry models are progressively coded and reconstructed at different levels of detail.

Fig. 5 depicts rate-distortion performance of the presented model-aided coding scheme for the *Airplane* image set. Geometry-coding bit-rate is taken into account. Reconstructed image quality is measured as the Peak-Signal-to-Noise Ratio (PSNR), averaged over all images' luminance component. For comparison, coding results from the disparity map-based codec described in [5] are shown. The progressive geometry compression scheme allows allocating optimal bit-rate to the 3-D model: full geometry information requires many bits which is not justified by improved image prediction. A coarser model needs fewer bits, but image prediction degrades to the point where more bits need to

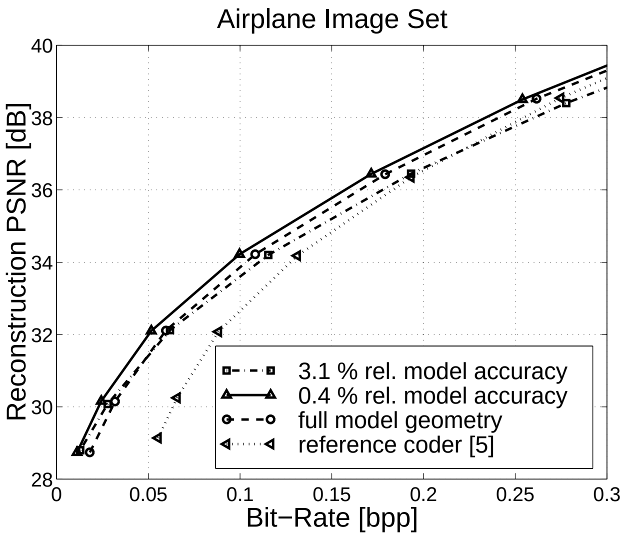


Figure 5: Rate-distortion performance for 256 *Airplane* images incl. geometry bit-rate; model accuracy is expressed as maximum deviation of vertex position relative to overall model size. The optimal amount of geometry detail does not depend on reconstruction quality. Coding results of the disparity map-based codec [5] are shown for comparison.

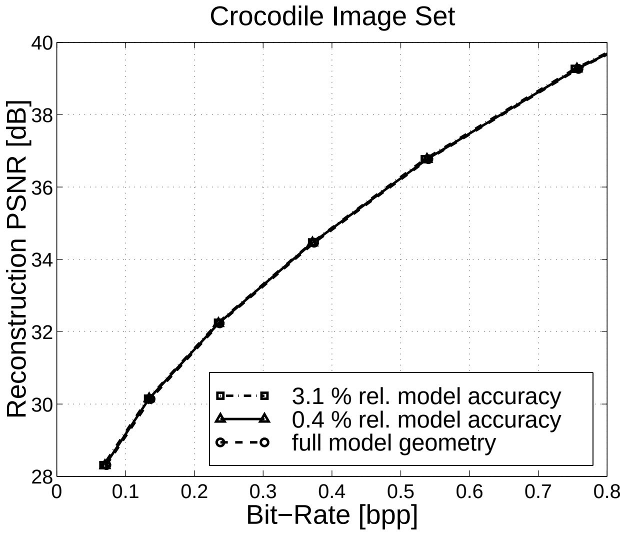


Figure 6: Rate-distortion performance for 257 *Crocodile* images incl. geometry bit-rate; coding performance is not effected by geometry detail, indicating that the reconstructed geometry model yields only modest image prediction.

be spent on residual error-coding than are saved on geometry compression. It is interesting to note that the optimal amount of geometry detail is the same over the entire range of considered bit-rates. Model-aided coding clearly excels at very low bit-rates, compressing the *Airplane* image data set by 1000:1 at 30 dB average PSNR. When compared to the disparity map-based codec at 0.07 bits per pixel, 3 dB better performance is observed, boosting reconstruction quality from 30 to 33 dB.

Model-aided prediction depends heavily on the quality of the scene's reconstructed geometry model. For the *Crocodile* image set, only a coarse geometry model was derived. The codec's rate-distortion characteristics using mediocre geometry is shown in Fig. 6. Coding performance varies only marginally with model accuracy, indicating that coding efficiency is limited due to the coarseness of the available geometry model.

6. CONCLUSIONS

A codec for multi-viewpoint imagery has been presented that makes use of approximate scene geometry in conjunction with an efficient model-aided image prediction scheme. Geometry is progressively coded to adjust geometry bit-rate depending on model accuracy. The image set is hierarchically coded, featuring progressive refinement of sampling density during decoding. If scene geometry is accurately reconstructed, the bit-rate overhead necessary to code model geometry is overcompensated by improved image predicton. At low bit-rates, superior coding performance is observed over existing compression techniques. Ongoing research focuses on improving geometry reconstruction to yield high coding efficiency for a broad range of image sets.

7. 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] M. Magnor and B. Girod, "Adaptive block-based light field coding", *Proc. International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging (IWSNHC3DI'99),* Santorini, Greece, pp. 140–143, Sept. 1999.

[5] M. Magnor and B. Girod, "Hierarchical coding of light fields with disparity maps", *Proc. International Conference on Image Processing (ICIP-99),* Kobe, Japan, pp. 334–338, Oct. 1999.

[6] 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.

[7] K. Hata and M. Etoh, "Epipolar geometry estimation and its application to image coding", *Proc. International Conference on Image Processing (ICIP-99),* Kobe, Japan, pp. 472–476, Oct. 1999.

[8] P. Eisert, E. Steinbach, and B. Girod, "Multi-hypothesis volumetric reconstruction of 3-D objects from multiple calibrated camera views", *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP'99)* Phoenix, USA, pp. 3509–3512, Mar. 1999.

[9] M. Magnor and B. Girod, "Fully embedded coding of triangle meshes", *Proc. Vision, Modeling, and Visualization (VMV'99),* Erlangen, Germany, pp. 253–259, Nov. 1999.