IEEE CONFERENCE PROCEEDINGS: INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP2000)
VANCOUVER, CANADA, VOL. 2, PP. 919-922, SEPTEMBER 2000

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 new coding technique for images taken from arbitrary recording positions around a static scene, based on reconstructed object geometry. Such data structures occur in image-based rendering applications where many hundreds to thousands of images need to be stored and transmitted. Approximate scene geometry enables disparity compensation as well as occlusion detection, leading to improved image prediction. Images are coded in hierarchical order to ensure efficient exploitation of inter-image similarities. The 3-D geometry model allows rendering new views by warping recorded images. The presented algorithm is validated using real-world image data sets, achieving better than 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 generate new views of the depicted setting. To achieve photo-realistic rendering results, however, many hundreds to thousands of images are required. Different 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. Direct vector quantization features swift decoding, but only low compression ratios are attained (below 30:1) [1]. Much higher compression ratios can be achieved if inter-image disparities are compensated. For light fields, compression ratios of up to 1000:1 have been reported [2].

Image recording positions in light field rendering are restricted to lie in a plane on a regular grid [1]. Recently, random recording positions within a plane [3] as well as spherical recording arrangements [4] have been successfully applied to IBR, stimulating the need for efficient coding schemes for more general multi-view imagery.

The *Model-aided Coder* presented in this paper has been specifically designed to code multiple calibrated images of a static 3-D scene recorded from arbitrary viewpoints. Prior to coding, approximate scene geometry is reconstructed from the imagery. The geometry model allows disparity compensation and occlusion detection, yielding accurate image predictions and efficient coding performance. Scene geometry further enables generating new views by warping recorded images.

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 geometry model. Images are coded in hierarchical order. Experimental results obtained for real-world image data sets are presented to validate the proposed coding scheme.

2. MODEL RECONSTRUCTION AND GEOMETRY COMPRESSION

Images of three stuffed toy animals on a turntable are recorded from multiple viewpoints using a camera on a lever arm (Figs. 1,2). The calibrated images are used to reconstruct approximate 3-D geometry. A robust scheme to deduce approximate scene geometry from multiple unsegmented camera views is described in [5]. The algorithm discretizes the volume enclosing

Figure 1: 3 stuffed animals serve as test objects.

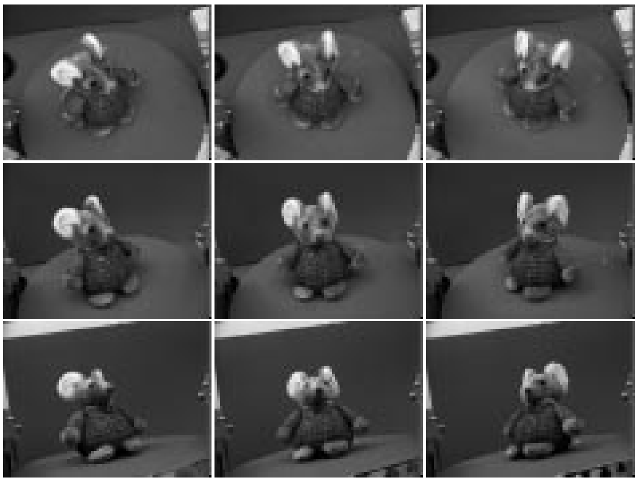


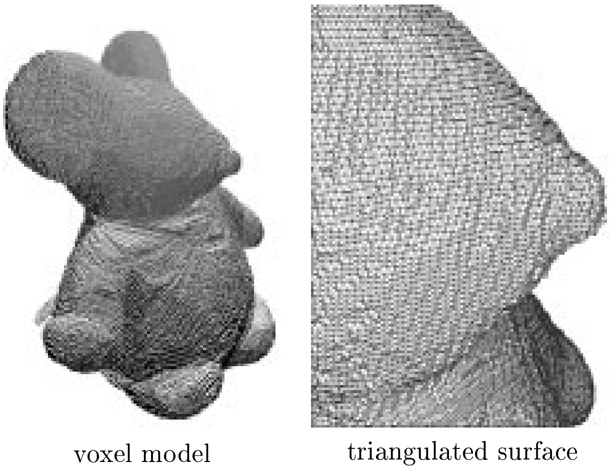
Figure 2: Natural object appearance is captured from multiple viewpoints.

the scene into voxels and projects each voxel into all visible images. 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 all image projections. Iteration over all voxels ends when no further voxels can be removed (Fig. 3). The voxel model surface is triangulated, yielding several hundred thousand triangles.

Because model surface accuracy is determined by voxel size, the number of triangles can be significantly reduced using a progressive mesh-decimation scheme [6] (Fig. 4). The resulting triangle mesh is progressively encoded applying *Embedded Mesh Coding (EMC)* [7]. The EMC algorithm progressively refines vertex positional information and triangle-mesh connectivity, providing multiple geometry approximations. Reconstructed model accuracy is expressed as a vertex's maximum deviation from its original position relative to the model extension (Fig. 4). Bit-rate allocation between geometry coding and predicton error coding can be varied to find optimal model accuracy.

3. MODEL-AIDED IMAGE PREDICTION

Scene appearance from arbitrary viewpoints can be predicted using 3-D geometry in conjunction with images recorded from nearby positions. The geometry model is projected onto the image plane of the desired view and the reference images. Each triangle visible in the target image is located in the reference images, detecting any occluded image regions at the same time. Multiple reference images are used to minimize occlusions. Visible triangle texture is copied from reference images while averaging multiple pixel predictions, yielding a disparity-compensated estimate of the target image. The object model can only predict the image region inside the projected model's silhouette. To compensate for silhouette inaccuracies and to account for image background, the prediction error is coded for the entire image.

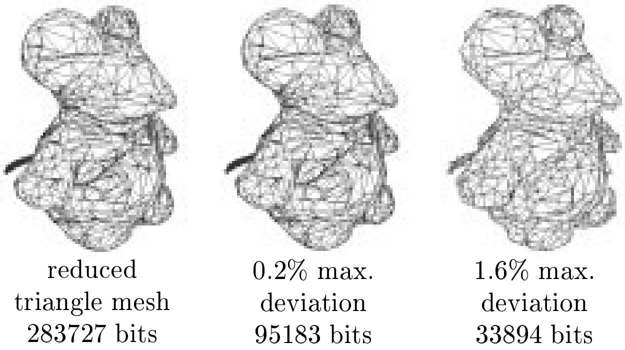


voxel model triangulated surface

Figure 3: From multiple images, a solid 3-D model is reconstructed; the triangulated surface consists of several hundred thousand triangles.

4. MULTI-VIEW IMAGE CODING

Prior to coding, the images are hierarchically ordered to efficiently exploit inter-image similarities. Image recording positions are expressed in spherical coordinates with the origin at the scene's center. As the object typically stands on a solid surface, the recording positions are distributed on a half-sphere around the

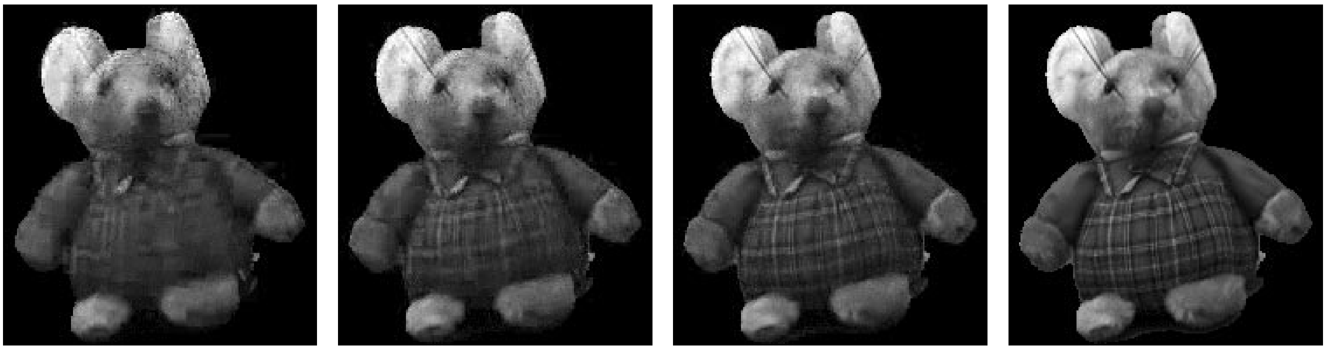


reduced triangle mesh 283727 bits

0.2% max. deviation 95183 bits

1.6% max. deviation 33894 bits

Figure 4: Triangle mesh accuracy is matched to voxel resolution by progressively decimating the number of triangles; the reduced triangle mesh is *EMC*-coded [7], and approximate geometry models are derived by decoding only a fraction of the bitstream; model accuracy is expressed as the maximum deviation of vertex position relative to model size.



0.009 bpp 0.039 bpp 0.136 bpp original image

Figure 5: Reconstructed images for different bit-rates (in bits-per-pixel).

scene (Fig. 6).

The image closest to the sphere's pole and four images roughly evenly spaced around the equator are intra-coded using the block-DCT scheme familiar from still-image compression (Images A in Fig. 6). For each image, the DCT quantization parameter is individually adjusted such that the coded image meets a preset minimum reconstruction quality. The five intra-coded images are arranged into four groups, each consisting of one polar and two neighboring equatorial images, subdividing the half-sphere into four quadrants. In each quadrant, the image closest to the center position (image B in Fig. 6) is sought. The center image is predicted using the three corner images A as reference. The residual error is DCT-coded if needed, again adjusting the quantization parameter. The three images closest to the quadrant sides' mid-positions (images C in Fig. 6) are predicted and coded likewise from the center and two corner images. After all quadrants have been considered, each quadrant is divided into four sub-quadrants, i.e., the center image B, a corner image A and one or two mid-side images C form a sub-quadrant. The sub-quadrants' corner images are already coded, and center and side images are again predictively coded. Subdivision continues until all images are coded.

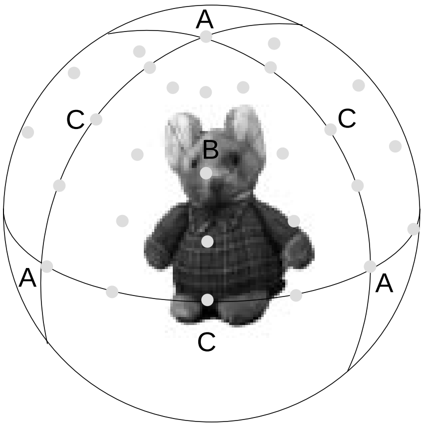


Figure 6: Image recording positions are projected onto a sphere surrounding the scene. The images closest to the sphere's pole and four images along the equator are intra-coded (A). The image closest to the center of each quadrant (B) is predicted from the corner images, and mid-side images (C) are predicted from the central and 2 corner images. Each quadrant is then subdivided and coded likewise until all images are coded.

During decoding, views from different directions can be quickly accessed. To enhance rendering quality, intermediate images between recorded views can be warped using the 3-D model.

5. RESULTS

The Model-aided Codec is validated using the *Mouse*, *Penguin* and *Garfield*^{TM} image data sets. Each data set consists of 257 images of 384 × 288 24-bit RGB pixels. Prior to coding, the images are transformed to YUV color space, and the chrominance components are downsampled by a factor 2. Image recording positions are arranged on a half-sphere around the object. Angular step size is $\sim 11.25°$ in horizontal and vertical direction, yielding 32 × 8 images plus one image from the zenith. For each image set, a voxel model is reconstructed, model surface is triangulated, the triangle mesh is reduced, EMC-coded, and the progressive EMC bitstream is decoded up to different bit-rates providing approximate geometry models of varying accuracy. The image sets are coded using the different model approximations.

Bit-rate includes geometry coding bits and is expressed in *bits-per-pixel (bpp).* Image reconstruction quality is expressed as the *Peak-Signal-to-Noise-Ratio (PSNR)* of the images' luminance component, averaged

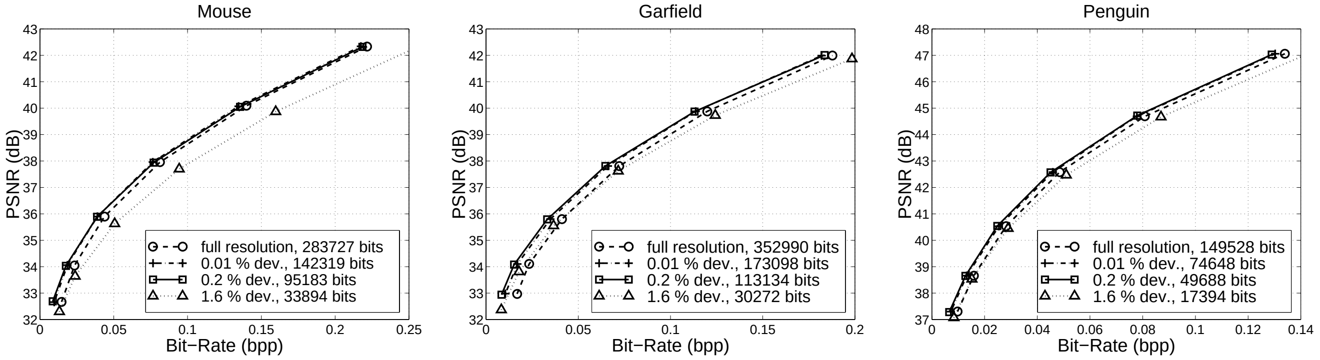


Figure 7: Rate-distortion performance of the Model-aided Coder for the *Mouse*, *Garfield*^{TM} and *Penguin*; different model approximations are tested to optimally allocate bit-rate between geometry and prediction error coding.

over all 257 images.

Fig. 7 illustrates coding performance for the *Mouse*, *Penguin* and *Garfield*^{TM} data sets. Model-aided Coding attains 1000 : 1 compression and higher at reasonable reconstruction quality (Fig. 5). Overall coding efficiency depends on the depicted object's characteristics. The *Penguin* is more efficiently coded than the *Mouse* or *Garfield*^{TM} because it exhibits much less detailed texture (Fig. 1). Coding performance depends on model accuracy, expressed as maximum deviation of a vertex from its original position, relative to model size. Coarse geometry approximations yield inferior coding results regardless of reconstruction quality (1.6% deviation). Coding performance improves with enhanced model accuracy until the increase in geometry bits is balanced by the decrease in prediction-error coding bits (0.2% and 0.01% deviation). Over the entire range of considered bit-rates, one geometry approximation can be used to yield the best coding results. Because model-aided prediction is limited by the accuracy of the reconstructed voxel model, overly accurate model geometry does not increase coding performance. Diminishing coding efficiency for high model accuracy is noticable only at low bit-rates as geometry coding bit-rate is negligible at medium and high reconstruction quality (full model resolution in Fig. 7).

6. CONCLUSIONS

A codec for multi-viewpoint imagery has been presented that applies approximate scene geometry for prediction. Geometry is progressively coded which allows allocating optimal bit-rate to geometry coding. The image set is coded in hierarchical order, featuring progressive refinement of sampling density during decoding. If scene geometry is accurately reconstructed, the bit-rate overhead necessary to code model geometry is well worth it since it improves image prediction. Attained compression factors of 1000 : 1 illustrate the usefulness of model-aided coding for multi-view image data compression.

7. REFERENCES

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

[2] M. Magnor and B. Girod, "Data compression for light field rendering", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, no. 3, pp. 338 – 343, Apr. 2000.

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

[4] 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.

[5] 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.

[6] S. Campagna, L. Kobbelt, and H.-P. Seidel, "Efficient decimation of complex triangle meshes", Tech. Rep., IMMD9, University Erlangen-Nuremberg, Mar. 1998.

[7] 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.