SENSITIVITY OF IMAGE-BASED AND TEXTURE-BASED MULTI-VIEW CODING TO MODEL ACCURACY

Marcus Magnor and Bernd Girod

Computer Graphics Laboratory / Information Systems Laboratory
Stanford University

ABSTRACT

Multi-view image coding benefits from knowledge of the depicted scene's 3-D geometry. To exploit geometry information for compression, two different approaches can be distinguished. In texture-based coding, images are converted to texture maps prior to compression. In image-based predictive coding, geometry is used for disparity compensation and occlusion detection between images. Coding performance of both approaches depends on the accuracy of the available geometry model. In this paper, texture-based and image-based coding are compared with regard to the influence of geometry accuracy on coding efficiency. The results are theoretically explained. Experiments with natural as well as synthetic image sets show that texture-based coding is more sensitive to small geometry inaccuracies than image-based coding. For approximate geometry models, image-based coding performs best, while texture-based coding yields superior coding results if scene geometry is exactly known.

1. INTRODUCTION

Simultaneous views of a scene, recorded from many different viewpoints, constitute the basis of image-based rendering (IBR) techniques [1, 2, 3, 4]. Rendering quality thereby depends on the number of scene images available. Typically, many hundreds to thousands of images are necessary to obtain good rendering results. To store and transmit the large amount of multi-view image data, a number of compression schemes have been specifically developed for IBR [5, 6, 7]. Based on still-image or video coding techniques, these codecs vary in decoding complexity and achieve compression ratios ranging from 20:1 to 2000:1.

Coding efficiency, decoding speed, and rendering quality can be increased if 3-D scene geometry information is available [8, 9]. To exploit geometry information for multi-view coding, two different approaches can be distinguished. In texture-based coding, scene geometry is used to convert images to view-dependent texture maps prior to compression [10]. Because disparity-induced differences are eliminated, texture maps exhibit greater inter-map correlation than the original images. In image-based predictive coding, on the other hand, the images are successively estimated, and the prediction error is encoded [11]. With 3-D scene geometry available, disparity compensation and occlusion detection can be performed for prediction.

Coding efficiency of both texture-based and image-based predictive coding depends on model accuracy [8]. This paper investigates the influence of scene geometry accuracy on both approaches' coding performance. In the following, texture-based and image-based coding are outlined, and coding results for real-world image sets are presented. Experiments with synthetic image sets show that while texture-based coding yields superior results for exact geometry, image-based predictive coding is less susceptible to model inaccuracies. The results are theoretically explained by examining the approximation-induced disparity compensation error for both coding approaches.

2. MULTI-VIEW COMPRESSION

Fig. 1 depicts the hemispherical arrangement of image recording positions around the scene used throughout this paper. Each multi-view image set consists of 257 calibrated RGB-images. Real-world objects are recorded using a camera on a computer-controlled lever arm and a turntable. A volumetric reconstruction algorithm is used to obtain a 3-D model of the scene [12]. For encoding, the reconstructed voxel model is converted to a triangle-mesh surface representation [10].

2.1. Image-based Predictive Coding

The predictive coding scheme outlined in the following is adapted from the *model-aided coder* described in more detail in [11]. First, the image closest to the pole of the hemisphere in Fig. 1 and four images spaced evenly around the equator are intra-encoded, dividing the hemisphere into 4 quadrants. The center image of each quadrant is predicted by warping the corner images of the respective quadrant. To do so, the geometry model is first rendered for the central image position. Each image pixel is assigned its corresponding geometry triangle and relative position within the triangle. The geometry model is then rendered for the

0-7803-6725-1/01/$10.00 ©2001 IEEE

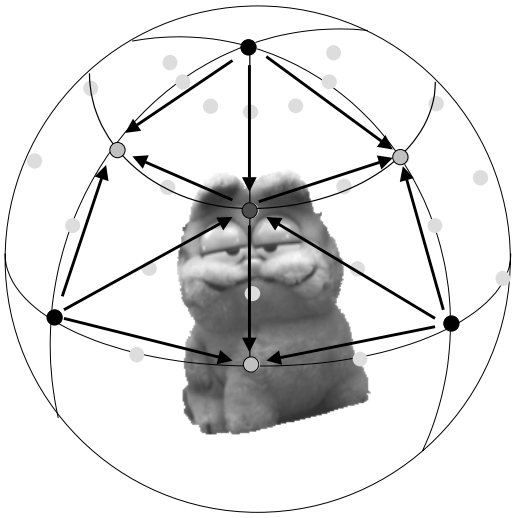

Fig. 1. Multi-view image recording: Image positions are arranged on a hemisphere around the scene.

corner images. For each pixel in the center image, the corresponding pixels in the reference images are identified from triangle index and relative position. Pixels that are not visible in a reference image are automatically detected. A partially occluded image region is predicted only from those reference images that show the respective region. Because several reference images are used for prediction, the amount of completely occluded regions is minimal, while multiple pixel predictions are averaged. Completely occluded areas are filled by interpolation using a resolution pyramid of the predicted image. After prediction, the remaining error is DCT-encoded. Next, the images in the middle of the quadrant's sides are predicted likewise using the just-encoded center image and the two closest corner images as reference. Each quadrant is then subdivided into four sub-quadrants whose center and mid-side images are predicted from the already-encoded sub-quadrant's corner images. Quadrant subdivision continues until all images are encoded.

As all images are predicted from previously encoded images, a multi-level hierarchy is established among the image set. The hierarchical coding order provides short prediction distances while keeping the entire image set quickly accessible.

2.2. Texture-based Coding

The second approach to exploit 3-D scene geometry for multi-view coding consists of converting the image set to multiple view-dependent texture maps. The texture-based coder implementation investigated in this paper is identical to the *model-based coder* described more thoroughly in [10]. Intrinsically, texture-based coding can encode only image regions within the projected model silhouette. To convert all images to texture maps, a suitable mapping must

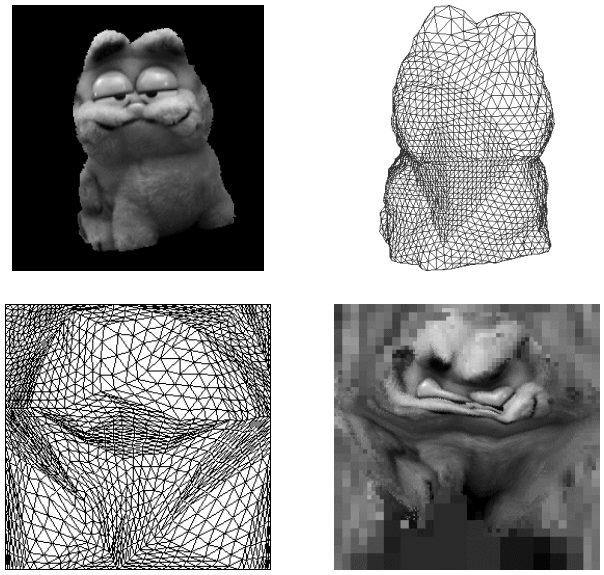

Fig. 2. Texture map generation: Model surface must be suitably parameterized to be mappable onto a planar rectangle.

be found that projects model surface onto a texture plane (Fig. 2). For objects with sphere-like topology, a contiguous model-to-plane mapping can be derived based on octahedral geometry [10]. To convert the images into texture maps, the geometry model is rendered, and each pixel inside the projected model silhouette is assigned its corresponding geometry triangle and the relative coordinates within the triangle. Triangle number and coordinates determine the texel to which the color value of the image pixel is copied. By mapping image pixels onto the texture plane, the texture map is only sparsely filled and must be suitably interpolated [7]. The interpolated texture-map array is then decomposed into 4-D subbands, and the wavelet coefficients are progressively encoded [10].

3. EXPERIMENTAL EVALUATION

To evaluate the performance of image-based and texture-based coding, three real-world image sets depicting stuffed toy animals are recorded. Object geometry must be re-

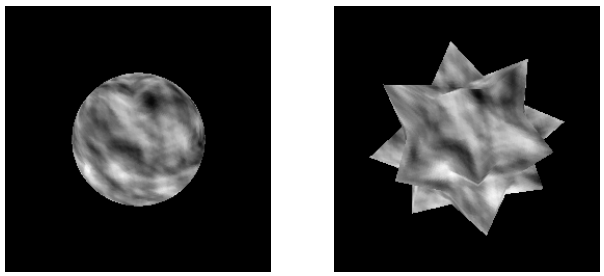

Fig. 3. Synthetic image sets *Sphere* and *Star*.

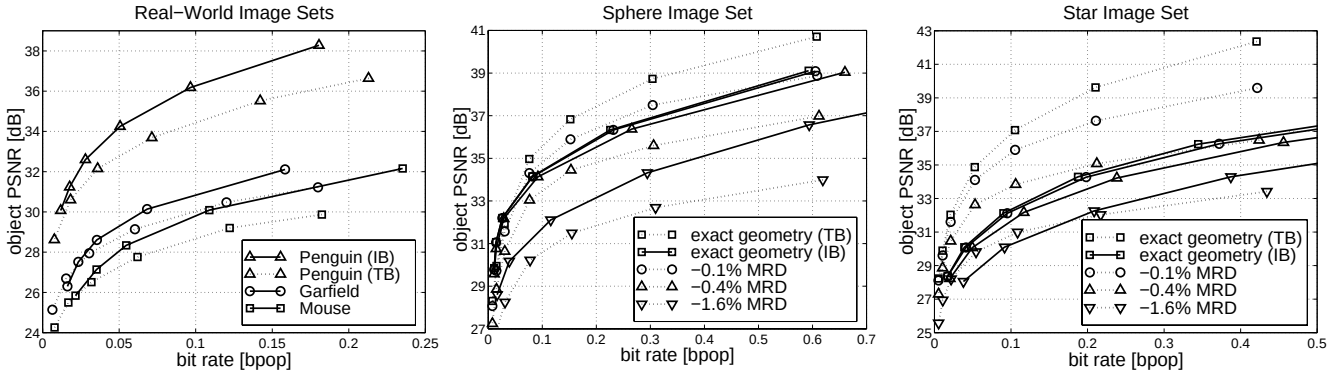

Fig. 4. Rate-distortion curves of image-based (IB) and texture-based coding (TB) for three real-world image sets, encoded with reconstructed approximate geometry (Fig. a), and two synthetic image sets using exact geometry and deliberately approximated models (Figs. b,c).

constructed directly from the images, so the models exhibit only finite precision. In addition, two synthetic image sets are rendered using geometry models of a sphere and a star (Fig. 3). For encoding the synthetic image sets, exact 3-D geometry is available. Fig. 4 depicts rate-distortion performance of image-based and texture-based coding for all image sets. Reconstruction quality is measured as the mean PSNR of all object pixels, while bit rate is expressed in bits per object pixel (bpop). For the real-world objects (Fig. 4a), the predictive coder yields up to 40 % better compression than the texture-based coder. The synthetic image sets (Figs. 4b,4c), however, are much more efficiently encoded using the texture-based coder.

To examine if the coders' different performance for synthetic and real-world images originates from inaccurate geometry, the synthetic image sets are encoded again after the vertices of the exact geometry models have been randomly displaced in radial direction. Figs. 4b,4c illustrate coding performance for different mean radial deviations (MRD) of the vertices, expressed in relation to distance from the center. Texture-based coding performance decreases rapidly with increasing model inaccuracy. Image-based predictive coding, however, is much less susceptible to approximate scene geometry. Minute alterations of the exact geometry are already sufficient for texture-based coding performance to fall below the corresponding RD-curves of the predictive coder. The experimental results indicate that the real-world image sets' only approximately reconstructable geometry is responsible for their inferior texture-based coding performance.

4. DISPARITY ANALYSIS

The observed experimental results can be theoretically understood by regarding the amount of texture mismatch and prediction error caused by approximate scene geometry. For the sake of simplicity, we consider a planar object on the floor of the hemispherical dome of image recording positions, as depicted in Fig. 5a. Let the geometry model for the object be a parallel plane, displaced by a distance Δz. Assuming orthographic projection, Fig. 5b illustrates that the mismatch on the surface of the geometry model Δd_{TB} depends on the angle α between viewing direction and the normal of the plane:

$$\Delta d_{\mathrm{TB}} \quad = \quad \Delta z \, \tan \alpha = \Delta z \, \frac{\sin \alpha}{\cos \alpha}. \qquad (1)$$

In image-based predictive coding, reference images are disparity-compensated to yield an estimate of the prediction image. The displacement-induced disparity compensation error Δd_{IB} in Fig. 5b depends also on the angular distance β between reference and prediction image:

$$\Delta d_{\mathrm{IB}} \quad = \quad \Delta z \, \frac{\sin \beta}{\cos \alpha}. \qquad (2)$$

Because of the hierarchical image coding order employed, the prediction distance β is smaller or at most equal to the viewing angle α. (2) therefore yields always an equal or smaller error than (1).

Fig. 5c depicts mean texture and disparity mismatch, averaged over all 257 image recording positions. Plane displacement Δz is expressed in units of texel extend ds (Fig. 5b). Image pixel size is set equal to texel size, matching image and texture resolution. With increasing plane displacement, texture mismatch increases much more rapidly than disparity compensation error. Displacing the model plane by one pixel results in a mean disparity compensation error of 0.5 pixels, whereas the generated texture maps are misaligned by more than 1.6 texels, on average. Greater texture mismatch causes a decrease in correlation between texture maps, deteriorating texture-based coding performance for approximate geometry models.

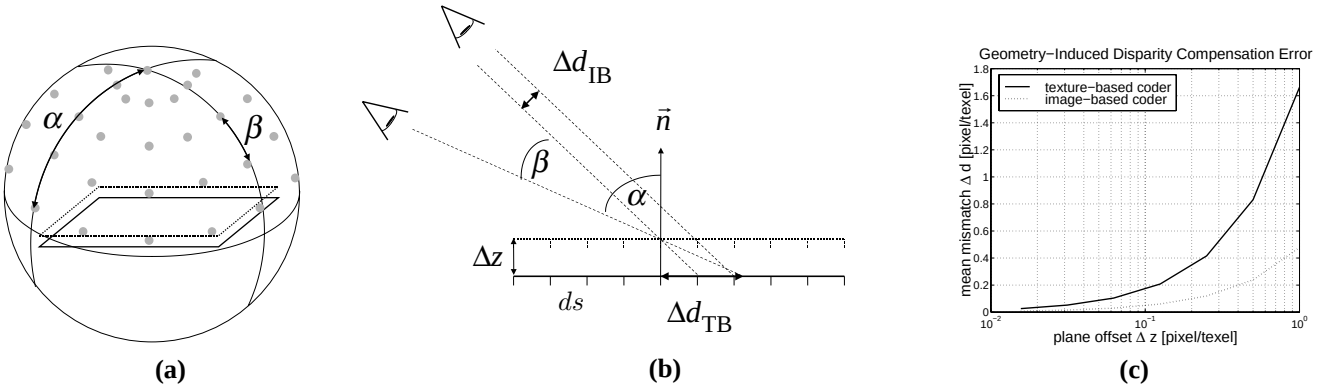

Fig. 5. Multi-view coding of a planar object (Fig. a): Displacing the plane by Δz (Fig. b) results in a texture mismatch Δd_{TB} that depends on viewing angle α. The disparity compensation error Δd_{IB} in image-based prediction is additionally dependent on the angular distance between reference and prediction image β. Averaged over all recording positions, texture mismatch increases considerably faster with increasing plane offset than the disparity-compensation error (Fig. c).

5. SUMMARY

In this paper, the influence of model accuracy on geometry-based coding performance is examined. Two different coding approaches are experimentally and theoretically evaluated. Coding experiments with real-world and synthetic image sets show that for approximate geometry models, image-based predictive coding is more efficient, while texture-based coding is superior if exact geometry is available. The results are theoretically explained by considering approximation-induced disparity compensation error and texture mismatch. Geometry inaccuracies are shown to have a large impact on texture generation, whereas hierarchical image-based prediction is rather insensitive to approximate geometry.

6. ACKNOWLEDGMENTS

The first author gratefully acknowledges a Feodor Lynen postdoctoral research fellowship from the Alexander von Humboldt Foundation, Germany, which allowed him to continue his work at Stanford University, as well as conference funding from the Stanford Immersive Television Project and Stanford Information Systems Laboratory.

7. REFERENCES

[1] M. Levoy and P. Hanrahan, "Light field rendering," *Proc. ACM Conference on Computer Graphics (SIGGRAPH'96)*, New Orleans, USA, pp. 31–42, Aug. 1996.

[2] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The lumigraph," *Proc. ACM Conference on Computer Graphics (SIGGRAPH'96)*, New Orleans, USA, pp. 43–54, Aug. 1996.

[3] J. Lengyel, "The convergence of graphics and vision," *Computer*, vol. 31, no. 7, pp. 46–53, July 1998.

[4] D. Wood, D. Azuma, K. Aldinger, B. Curless, T. Duchamp, D. Salesin, and W. Stuetzle, "Surface light fields for 3D photography," *Proc. ACM Conference on Computer Graphics (SIGGRAPH-2000)*, New Orleans, USA, pp. 287–296, July 2000.

[5] P. Lalonde and A. Fournier, "Interactive rendering of wavelet projected light fields," *Proc. Graphics Interface'99*, Kingston, Canada, pp. 107–114, June 1999.

[6] X. Tong and R.M. Gray, "Coding of multi-view images for immersive viewing," *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP-2000)*, Istanbul, Turkey, vol. 4, pp. 1879–1882, June 2000.

[7] M. Magnor, *Geometry-Adaptive Multi-View Coding Techniques for Image-based Rendering*, Shaker Verlag, Aachen, Germany, ISBN 3-8265-8315-9, 2001, Ph.D. Thesis, University Erlangen-Nuremberg, Germany, Nov. 2000.

[8] B. Girod and M. Magnor, "Two approaches to incorporate approximate geometry into multi-view image coding," *Proc. IEEE International Conference on Image Processing (ICIP-2000)*, Vancouver, Canada, vol. 2, pp. 5–8, Sept. 2000.

[9] H. Schirmacher, W. Heidrich, and H.-P. Seidel, "High-quality interactive Lumigraph rendering through warping," *Proc. Graphics Interface 2000*, Montreal, Canada, pp. 87–94, May 2000.

[10] M. Magnor and B. Girod, "Model-based coding of multi-viewpoint imagery," *Proc. SPIE Visual Communication and Image Processing (VCIP-2000)*, Perth, Australia, vol. 1, pp. 14–22, June 2000.

[11] M. Magnor, P. Eisert, and B. Girod, "Model-aided coding of multi-viewpoint image data," *Proc. IEEE International Conference on Image Processing (ICIP-2000)*, Vancouver, Canada, vol. 2, pp. 919–922, Sept. 2000.

[12] 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.