Gaze-Motivated Compression of Illumination and View Dependent Textures

Jiří Filip   Michal Haindl
Institute of Information Theory and Automation of the ASCR
Prague, Czech Republic
Email: {filipj,haindl}@utia.cas.cz

Michael J. Chantler
School of Mathematical and Computer Sciences
Heriot-Watt University
Edinburgh, Scotland
Email: M.J.Chantler@hw.ac.uk

Abstract—Illumination and view dependent texture provide ample information on the appearance of real materials at the cost of enormous data storage requirements. Hence, past research focused mainly on compression and modelling of these data, however, few papers have explicitly addressed the way in which humans perceive these compressed data. We analyzed human gaze information to determine appropriate texture statistics. These statistics were then exploited in a pilot illumination and view direction dependent data compression algorithm. Our results showed that taking into account local texture variance can increase compression of current methods more than twofold, while preserving original realistic appearance and allowing fast data reconstruction.

Keywords-texture; BTF; compression; eye-tracking; visual psychophysics;

I. INTRODUCTION

Advanced applications of texture recognition, classification, and visualization, among others, use view and illumination dependent texture to achieve accurate computational and visual results. A data structure often representing these type of textures is called a *bidirectional texture function* [1]

$$BTF(x, y, \theta_i, \varphi_i, \theta_v, \varphi_v) \ , \tag{1}$$

i.e., a function describing the dependence of each pixel (x, y) on elevation θ and azimuthal φ angles specifying illumination and view direction (i, v). These data provide ample information about material appearance in the real world and thus are essential for accurate applications in computer graphics and computer vision. However, for practical usage the data has to be compressed to avoid storage of gigabytes per material sample. A number of approaches to compression of these data have been proposed in the past (see a survey [2]). They were based on clustering, linear decomposition methods, pixel-wise reflectance models or on Markov random fields models. In recent years researchers have started to investigate human perception of BTF data and proposed methods using standard [3] or experimentally derived [4], [5] rules for enhancement of data compression while still maintaining rapid access to the data.

We build on this research and exploit gaze data analysis to provide a better understanding of a way humans perceive different real-world materials, and consequently apply this knowledge for improvement of BTF data compression. Approaches exist, e.g. [6], that use visual masking for control of texture compression, e.g., more intensive compression of texture regions is employed where the imperfections are masked by contrast caused by a higher local texture contrast or object curvature. In our case, we are looking for feature statistics that correlate well with captured spatial gaze statistics and which are fast to evaluate on thousands of BTF images. In contrary to the approach varying compression across a textured plane, we focused on features controlling compression across different illumination and view directions, i.e., providing a single value evaluating the perceptual significance of a texture image.

II. GAZE DATA ANALYSIS

To obtain gaze data for analysis we designed a psychophysical experiment presenting stimuli on a calibrated LCD screen to a group of eleven naive subjects. Stimuli comprised two spheres illuminated from left (Fig. 2-a). The spheres were covered by view and illumination dependent textures represented by BTFs from the University Bonn database [7]. We used eight samples shown in Fig. 1-a. The BTF on one of the spheres is slightly downsampled across illumination and view angles. The task of the subjects was to find out whether the textures on both spheres are identical or different. Subjects had unlimited time for their decision while their gaze data were recorded using a Tobii x50 infrared-based binocular eye-tracking device. The device was calibrated for each subject individually and provided the locations and durations of fixations at a speed 50 samples/s. Maximum error is approximately $\pm 0.5^o$ of visual angle, which corresponds to ± 32 pixels for our setup. The shortest fixation duration to be recorded was set to 100 ms. Average

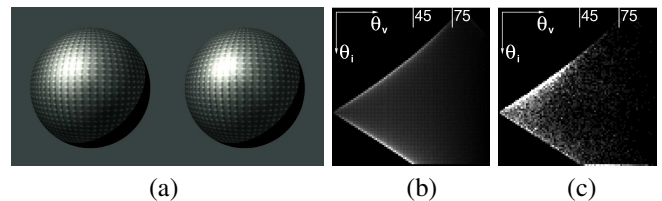(a)   (b)   (c)

Figure 2. (a) Stimulus example. An average distribution of elevation angles θ_i, θ_v (b) all over stimuli, (c) in the fixated locations only.

spatial distributions of gaze fixations of all subjects for front-

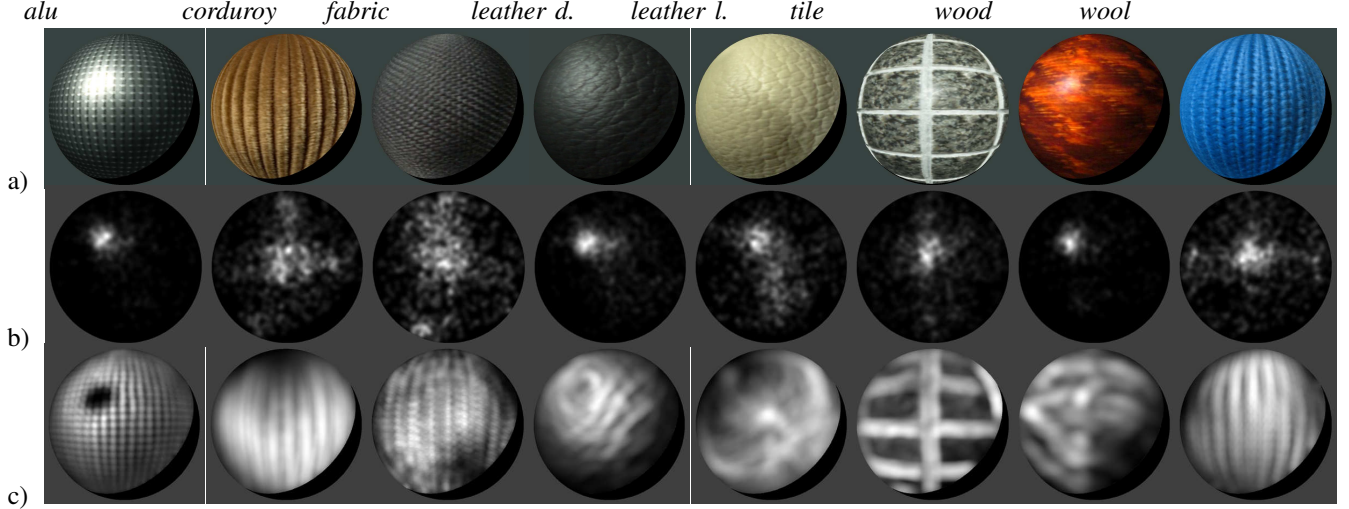| | alu | corduroy | fabric | leather d. | leather l. | tile | wood | wool |

a)

b)

c)

Figure 1. (a) Tested material samples, (b) corresponding normalized distributions of gaze fixations, and (c) an example of image statistic based on a local variance V.

top illumination are shown in Fig. 1-b. There are obvious differences across individual samples and the results suggest that subjects gaze was attracted either by specular highlights or by object's center having less perspective distorted surface texture. Our hypothesis is that the shape of the distributions depends on the underlying stimuli image, hence our motivation was to find image statistics that resemble most closely the captured distributions. We tested seven different statistics computed over the stimuli. The first two were local mean (M) and local variance (V) computed in contextual neighborhood. The size of the neighborhood was estimated from the highest response of Gabor features to stimuli over different frequencies. Next we computed local energy (E) over the image I as an average magnitude of a gradient in the local neighborhood using

$$e(x,y) = \sqrt{(\partial I/\partial x)^2 + (\partial I/\partial y)^2} \ . \qquad (2)$$

Next, the two statistics luminance (L) and chromatic (C) contrast were estimated from pixel-wise cone responses to stimuli images ($L - M$ and $S - (L + M)$ cone channels respectively) according to [8] and [9]. The last two statistics are based on image salience [10] and describe luminance (SL) and texture (ST) contrast based on per-pixel spatial gradients over stimuli luminance represented by a Gaussian-Laplacian pyramid.

For each statistic we computed the Pearson correlation coefficient with gaze fixation distribution using

$$R_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \ , \qquad (3)$$

where X, Y are compared pixels vectors, μ and σ are their means and variances. The feature images (e.g. Fig. 1-c) were computed from gamma compensated halves of stimuli images, and together with the smoothed spatial gaze

distribution images were downsampled to resolution 30×30 pixels prior to correlation, to preserve locality information. Resulting correlation coefficients for all statistics and samples are shown in Tab. I. Mean correlation coefficient values

Table I
CORRELATION OF THE TESTED STATISTICS WITH GAZE FIXATION
DISTRIBUTIONS.

sample	M	V	E	L	C	SL	SC
alu	0.86	0.52	0.35	0.90	0.93	0.46	0.38
cord.	0.48	0.50	0.16	–0.57	0.55	0.31	0.22
fabric	0.53	0.71	0.58	0.39	0.45	0.63	0.57
leath.d.	0.77	0.80	0.77	0.80	0.80	0.89	0.85
leath.l.	0.46	0.52	0.46	0.41	0.53	0.15	-0.04
tile	0.58	0.58	0.58	0.66	0.63	0.32	0.22
wood	0.77	0.61	0.77	0.78	0.87	0.51	0.51
wool	0.48	0.71	0.48	0.52	-0.52	0.58	0.41
AVG	**0.62**	**0.62**	0.35	0.29	0.53	0.48	0.39

for individual samples are shown in Fig. 3. Errorbars indicate twice the standard deviation across samples. The highest correlation value and the lowest standard deviation across samples was obtained for local variance (V) statistics showed in Fig. 1-c.
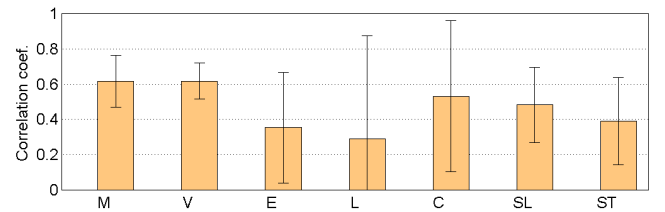

Figure 3. Average correlations of gaze fixations with the tested statistics.

III. Texture Compression Application

Based on the performance of tested features and their computational demands we used only the statistics M, V, E, L, and C in the following validation test. For this purpose we designed a exemplary BTF data compression method controlled by the precomputed statistics. The method is based on data downsampling in the spatial domain. First, each statistic was computed over local window of each measured illumination and view dependent texture in BTF data. Then the histogram of the statistics values was divided into three parts, each having the same area. The texture images having feature values within the first area were downsampled three times, while those falling into middle area were downsampled two times, and those falling into the last area were left untouched, i.e., the textures having the highest values were downsampled the least. This scheme assumes that a higher valued statistic represents a higher perceptual importance of underlying texture image, which should be downsampled less than images having lower values. Additionally, as oblique views θ_v are apparently less important for human observers, as was shown in Fig. 2-bc, we forced all BTF images having view direction elevation $\theta_v > 70^o$, to the highest downsampling level regardless to the feature value. This compression scheme allows us to store approximately $\sim 38\%$ of the original data. This value can change slightly dependently on actual balance of feature values in oblique views of the data.

Performance evaluation of individual features would require running another psychophysical experiment visually comparing original and compressed data. Instead of this we used *visual difference predictor* (VDP) [11] that simulates low level human perception for known viewing conditions (in our case: display size 37×30 cm, resolution 1280×1024 pixels, observer's distance 0.7 m) and thus is sufficient for our task of perceptually plausible detection of subtle downsampling artifacts. Fig. 4 shows renderings of original and compressed data on a region of a more complex object shape together with the responses of the VDP. Table II shows the percentage of different object pixels detected with probability higher than 75% (green and red pixels in Fig. 4).

Table II
PERCENTAGE OF DIFFERENT PIXELS OF ALL TEXTURED PIXELS.

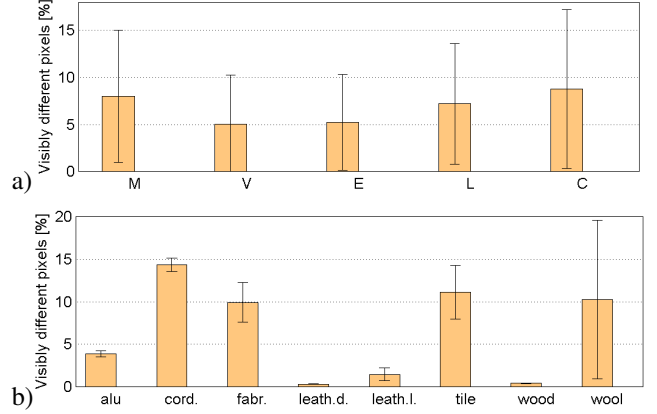| sample	M	V	E	L	C
alu	3.67	3.55	4.23	8.40	3.61
corduroy	13.42	14.44	15.12	7.45	13.71
fabric	12.69	8.52	8.07	3.27	12.30
leath.d.	0.34	0.34	0.28	0.79	0.34
leath.l.	2.03	0.17	1.58	0.45	1.97
tile	14.55	9.25	8.57	11.73	14.50
wood	0.45	0.40	0.40	5.58	0.40
wool	16.92	3.84	3.61	19.91	23.41
AVG	8.01	**5.06**	**5.23**	7.20	8.78

Figure 5.   Average percentage of different pixels across (a) samples, (b) statistics.

The visual performance differs significantly depending on both the statistics and the material samples as is shown in Fig. 5. However, both the lowest distortion and standard deviations across samples were obtained for local variance (V) and local energy (E) statistics. This is in accordance with the gaze correlation results in Section II. As for the limitations Fig. 5-b suggests that the method is more suitable for material samples having a less rough surface structure causing less occlusions and masking.

Although the compression is relatively small compared with a state of the art [2], it shows a promising way of making current compression methods more effective by taking into account observer's perceptual sensitivity to illumination and view dependent textures. Moreover, as the compression was performed in the spatial domain it can be effectively combined with another compression of individual downsamling levels, or with compression in space of illuminations and views.

IV. Conclusions

In this paper we have proposed an illumination and view dependent texture compression enhancement scheme motivated by a human gaze data analysis. We have found a local variance computed in a smallest texture element as the most appropriate predictor of texture perceptual significance. This was validated in the exemplary compression experiment. The proposed technique is able to extend many existing compression methods, has negligible computational overhead and can be easily implemented on graphics hardware for purpose of interactive visualization of materials.
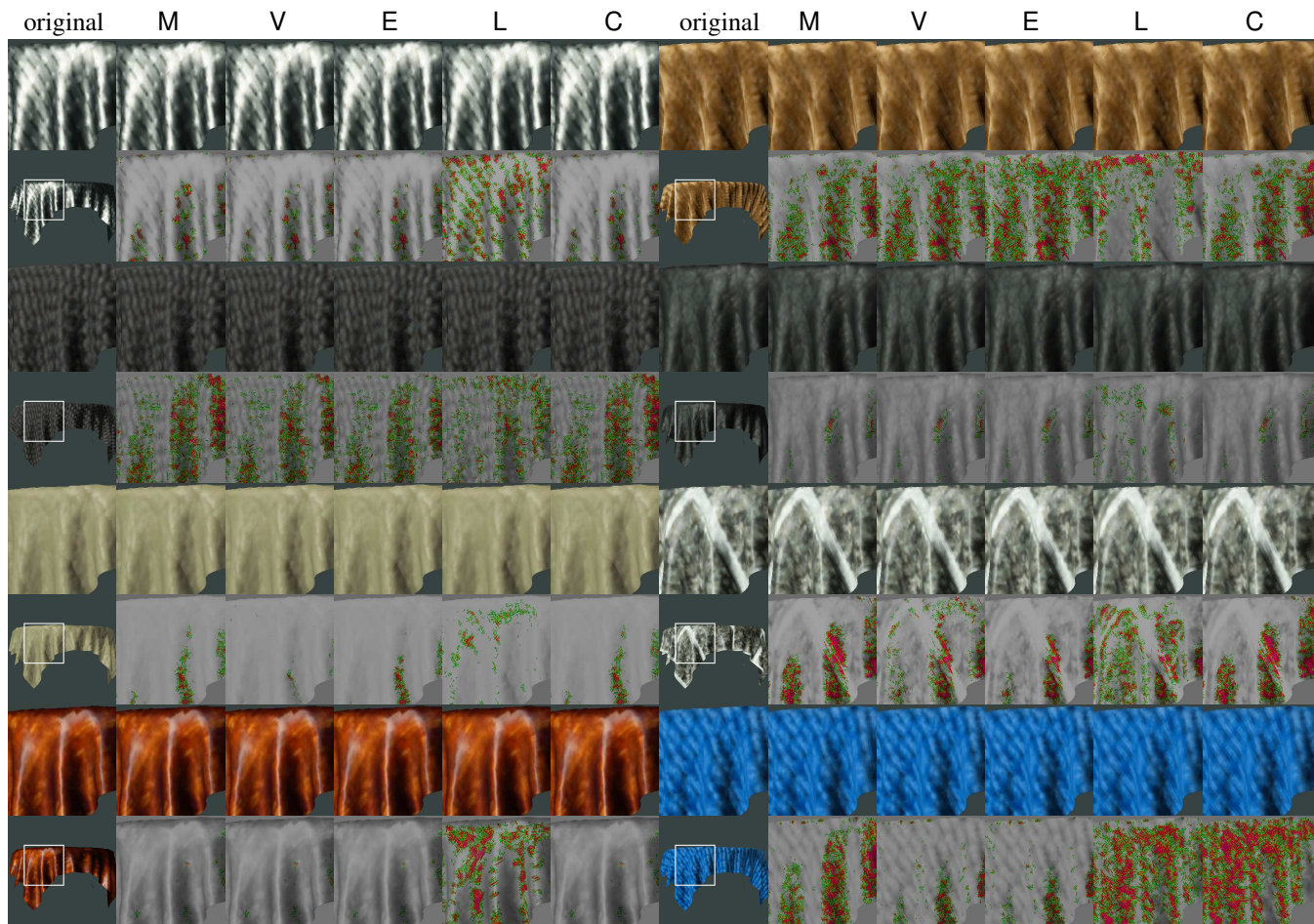
Acknowledgments

Figure 4. Compression performance of the statistics (odd rows) evaluated by VDP (even rows).

REFERENCES

[1] K. Dana, B. van Ginneken, S. Nayar, and J. Koenderink, "Reflectance and texture of real-world surfaces," *ACM Transactions on Graphics*, vol. 18, no. 1, pp. 1–34, 1999.

[2] J. Filip and M. Haindl, "Bidirectional texture function modeling: A state of the art survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1921–1940, October 2009.

[3] M. Guthe, G. Müller, M. Schneider, and R. Klein, "BTF-CIELab: A perceptual difference measure for quality assessment and compression of BTFs," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 101–113, 2009.

[4] J. Filip, M. Chantler, P. Green, and M. Haindl, "A psychophysically validated metric for bidirectional texture data reduction," *ACM Transactions on Graphics*, vol. 27, no. 5, p. 138, December 2008. [Online]. Available: http://staff.utia.cas.cz/filip/projects/pertex

[5] J. Filip, M. Chantler, and M. Haindl, "On uniform resampling and gaze analysis of bidirectional texture functions," *ACM Transactions on Applied Perception*, vol. 6, no. 3, p. 15, August 2009.

[6] J. Ferwerda, P. Shirley, S. Pattanaik, and D. Greenberg, "A model of visual masking for computer graphics," in *Computer graphics and interactive techniques*, 1997, pp. 143–152.

[7] BTF Database Bonn, "http://btf.cs.uni-bonn.de," 2003.

[8] D. Rudeman, T. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: implications for visual coding," *J. Opt. Soc. Am. A*, vol. 15, no. 8, pp. 2036–2045, 1998.

[9] M. Webster and J. Mollon, "Adaptation and the color statistics of natural images," *Vision Research*, vol. 37, no. 23, pp. 3283–3298, 1997.

[10] D. Parkhurst and E. Niebur, "Texture contrast attracts overt visual attention in natural scenes," *European Journal of Neuroscience*, vol. 19, pp. 783–789, 2004.

[11] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Visible difference predictor for high dynamic range images," in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, 2004, pp. 2763–2769.