A Perception-Based Threshold for Bidirectional Texture Functions

Banafsheh Azari (banafsheh.azari@uni-weimar.de)
CogVis/MMC, Faculty of Media, Bauhaus-Universität Weimar, Bauhausstraße 11, 99423 Weimar, Germany

Sven Bertel (sven.bertel@uni-weimar.de)
Usability Research Group, Faculty of Media, Bauhaus-Universität Weimar, Bauhausstraße 11, 99423 Weimar, Germany

Charles A. Wuethrich (charles.wuethrich@uni-weimar.de)
CogVis/MMC, Faculty of Media, Bauhaus-Universität Weimar, Bauhausstraße 11, 99423 Weimar, Germany

Abstract

For creating photorealistic images, Computer Graphics researchers introduced Bidirectional Texture Functions (BTFs), which use view- and illumination-dependent textures for rendering. BTFs require massive storage, and several proposals were made on how to compress them, but very few take into account human perception. We present and discuss an experimental study on how decreasing the texture resolution influences perceived quality of the rendered images. In a visual comparison task, observer quality judgments and gaze data were collected and analysed to determine the optimal downsampling of BTF data without significant loss of their perceived visual quality.

Keywords: Perceived image quality; realistic rendering; threshold in image perception; eye tracking.

Introduction

One of the main aims of Computer Graphics is the simulation of the complex reflection behaviour of real world materials. Among different types of materials, particular importance is given to fabrics. Graphical simulations of fabrics are used not only in interior design and architecture, but also increasingly in the clothing, car, film, and computer gaming industries.

Fabrics possess highly complex reflection behaviour, as reflection of the incoming light changes dramatically from material to material, depending, among other factors, on meso- and micro-structures of the thread, an example of which is shown in Figure 1, on the type of weaving, which influences the position of the thread in the fabric, on the interreflections between the components of the fabric, and on surface and subsurface scattering of light. Fabrics exhibit not only simple reflection characteristics, such as diffuse and specular reflection, but are characterised also by thread-dependent highlights and self shadowing, as well as anisotropic reflection.

To exactly simulate the correct reflection behaviour of fabrics, the paths of light within and on the surface of the material would have to be computed. Given the number of individual components of a thread, such task is computationally not feasible. Bidirectional Texture Functions (BTFs; Dana et al. (1999)) represent an alternative solution to this problem. A BTF contains reflectance information of points on a surface under fixed lighting and viewing conditions. It involves a function which depends on coordinates in texture space (x_p, y_p) on the surface of a simulated object, as well as the spherical angles of a vector from the viewer to the surface (θ_o, ϕ_o) and a vector to the illuminant (θ_i, ϕ_i).

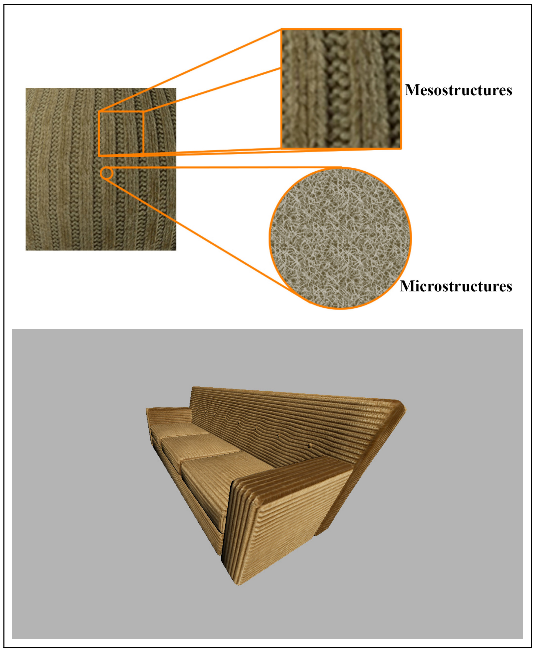

Figure 1: *Above:* Meso- and micro-structures of a woollen fabric. *Below:* An example of the stimuli used in our study.

In practical use, BTFs rely on large collections of digitally acquired pictures of a material (e.g., of a woollen fabric) that were taken for ranges of discretely varied illumination and viewing angles. When a simulation of the material needs to be computed to texturise a virtual object's surface, viewer and illuminant vectors are used to pick the matching pictures from this collection. The texture of an object then results from an interpolation of these (Nicodemus et al. (1977)).

A severe disadvantage of BTFs lies in the sheer size of picture collections needed, as they contain one photograph for each combination of viewing and illumination angles. The disadvantage is particular acute for the purposes of real time rendering, as the entire collection of pictures needs to be kept in the computer memory. Various past projects have therefore focussed on efficient compression methods for BTFs (including reflectance models based on linear factorization and pixelwise bidirectional reflection distribution functions, in short BRDFs, which are the general reflection model from which BTFs are derived, Filip and Haindl (2009)).

While the existing approaches are often technically well motivated, we believe that, before starting to choose how and how strongly to compress BTF data, it makes sense to first take a step back and see how the human observer perceives and judges compressed and non-compressed BTF textures in comparison tasks. Specifically, we look at BTF-based synthetic renderings of three–dimensional objects and ask: when does using high-resolution textures make sense because the high resolution leads to a perceived increase in texture quality? And when can one do just as well with lower resolution textures without perceived loss in quality? In this contribution, we present and discuss an investigation aimed at locating a robust threshold for downsampling BTF images without loosing perceptual quality. Information about the location of such a threshold is not only of importance to a better understanding of visual perception of textures, especially in object comparison tasks, but also of importance for developing novel data compression methods in synthetic rendering.

In the next section, we will review relevant related work. We will then describe our method of experimentation, which involves quality comparison tasks with pairs of texturized objects of varying BTF quality levels and varying exposure times. Gaze data was collected to aid visual comparison strategy detection. The presentation of study results is then followed by a discussion, conclusions, and an outlook.

Previous work

Compression methods for BTF data have been studied for many years in order to accelerate rendering and to compress data. However, only rarely the focus was on the perceived quality of the results of compression. Fleming et al. (2003) studied how humans perceive reflections on surfaces, while Lawson et al. (2003) demonstrated the importance of view changes in synthetic picture matching tasks. te Pas and Pont (2005a) showed that differences in the microstructure of a material are hard to distinguish from differences in the illumination, and that light source direction estimation depends on the material's bidirectional reflection distribution functions or BDRFs (te Pas and Pont (2005b)).

Work by Pellacini et al. (2000) introduced a new light reflection model for image synthesis based on experimental studies of surface gloss perception. Two experiments were conducted to explore the relationships between the physical parameters used to describe the reflectance properties of glossy surfaces and perceptual dimensions of glossy appearance. Psychophysical tests by Mcmillan et al. (2003) showed consistent transitions in perceived properties between interpolate and extrapolate BRDFs in the space of acquisition.

The accurate reproduction of material structures that can be achieved by using measured BTFs was investigated in Meseth et al. (2006), while Filip et al. (2008) performed a psychophysical study to optimize sparse sampling of BTFs data. In a further study, Filip et al. (2009) assessed different uniform reduced samplings of BTF data based on azimuthal angles of view and illumination as well as on elevation angles.

Few of the existing approaches include an investigation of viewers' gaze behavior while viewing the rendered images. A notable exception is the work by Filip et al. (2009) in which location, duration, and frequency of fixations were recorded. Fixation data was used to analyze strategies of the subjects over the course of the experiment (e.g., did locations and durations of fixations change as the study progressed? Both were found to be the case.).

In short, previous work focussed on the influence of light, viewing, material reflectance, shape, and angular sampling density of BTF data. In this contribution, we investigate the influence on perceived image quality that the *size of the individual BTF texture pictures* has, based on which a synthetic object's texture is interpolated. This variable has not been addressed previously. Our aim is to find a threshold for downsampling BTF resolution — that is, for reducing the image size of the individual BTF textures — without any perceived degradation in the quality of the rendered image. Similar to the procedure by Filip et al., we will collect gaze data to aid the detection of visual strategy and its change.

Method

In a pilot study using different self shadowing fabrics, like corduroy and wool, available in the BTF database of the University Bonn[1] we established that there are no differences in gaze behavior or perceived quality jugments between fabrics. We therefore decided to here focus on the corduroy dataset, which we will refer to as *Cord-256*, as its texture pictures are 256x256 pixels.

We then generated two new datasets by downscaling the *Cord-256* set through bilinear interpolation to respective resolutions of 128x128 pixels (*Cord-128*) and 64x64 pixels (*Cord-64*). For each of the three texture data sets, a three-dimensional textured model of a sofa was rendered through the standard BTF rendering method at a screen resolution of 1920x1084 pixels (compare Figure 1, below). The sofa model was oriented for display to the viewer to present textured parts across a large range of picture depth.

We chose a sofa object model for three main reasons: first, to present an everyday object that viewers are familiar with and instantly recognize. Second, to have an object with a structured surface and composition (e.g., individual buttons, cushions, etc.). This is important in order to ensure that a large set of fitting BTF pictures will be selected as basis for the object's texture, with widely varying illumination and viewing angles. And, third, a sofa is a type of object for which a cord texture would be commonly found.

Stimulus Pairs

Pairs of the rendered images displayed in full screen, native resolution mode were used as experimental stimuli. Each pair consisted of a sequentially presented rendering using two of the three texture resolutions as shown in Table 1. A total of 72 image pairs were shown to test subjects.

---

[1] http://btf.cs.uni-bonn.de/.

Table 1: Image pairs as experimental stimuli.

First Image	Second Image
Cord-256	*Cord-64*
Cord-256	*Cord-128*
Cord-128	*Cord-256*
Cord-128	*Cord-64*
Cord-64	*Cord-256*
Cord-64	*Cord-128*

Table 2: Answer posibilities.

1	First image has higher quality.
2	Second image has the higher quality.
=	The images have the same quality.
None	Subject is not sure.

The experiment was performed in three blocks of 24 image pairs each, between which image exposure time was varied. Exposure time per image was either 1000, 2000, or 3000 milliseconds (ms), respectively labeled as *short*, *medium* and *long* test conditions. Presentation order of the three blocks was balanced across subjects based on a Latin square design. Our rationale behind introducing variation of image exposure time was to test for effects it may have on comparative perceived image quality. It seems possible that, for pairs of different images, longer exposures could lead to higher frequencies of detecting that a difference exists.

Presentation of images in each pair was separated by 200 ms. After the presentation of the second image in a pair, subjects had 3000 ms to make a decision about the comparative image quality within the pair: was the first or second image of better visual quality? Or were the two images of the same visual quality? Responses were given on a three-key keyboard and were possible at any time after the start of the presentation of the second image. Subjects were also instructed that they could choose not to press any button if they felt unsure about the comparison. Please see Table 2 for an overview. When looking at the six image pairs in Table 1, it becomes clear that all pairs are different and that, consequently, any judgment that a pair shows that same image quality will be incorrect. However, subjects were not previously instructed that no same-quality pairs would be shown. After the decision time of 3000 ms had lapsed, the next image pair was automatically presented.

Experimental Setup

The images were presented on a 24-inch monitor with a resolution of 1920x1080 pixels at a distance of 70 cm from the viewer. The screen measured 22.35x15.80 inches and subtended approximately 33 degrees of visual angle. Due to the texture pattern, the minimal texture detail (i.e., for the parts of the sofa at the greatest depth in the image) had a cycle of 4 pixels, which means a subtended angle for a viewer of about 6 cycles per minute of a degree of arc.

An SMI RED250 remote eye tracking system was used in binocular mode with 250 Hz fixation detection, in order to record subjects' fixation behavior. *SMI BeGaze 2.4* software was used for subsequent analysis of gaze data.

Subjects. A total number of 20 subjects, 12 males and 8 females, participated in the experiment. Subjects were undergraduate or graduate students or department members in Computer Science or Civil Engineering, and they were not informed about the purpose of the experiment prior to conducting it. The age of the test subjects ranged from 22 to 39 years ($mean = 30.5$). Subjects had normal or corrected-to-normal visual acuity.

Procedure. Test subjects were seated in front of the monitor and eye tracker, introduced to the setup and to the experimental procedure, including the answer options. Before the start of the experiment, subjects were asked to read and sign a declaration of informed consent. Subjects could abort the experiment at any time and were guaranteed anonymous treatment of all collected data. They were familiarized with the used sofa images through a preliminary test round with eight image pair comparisons, the results of which were discarded for the subsequent analysis. Then, the subjects were calibrated on the eye tracker and the first of the three test blocks was presented. Calibration was repeated before each subsequent block. Each subject needed about 30 minutes to complete all three blocks.

Results

The section consists of two parts: an analysis of *subject performance* (i.e., the subjects' ability to judge image quality differences for the six pairs of Table 1) and an analysis of *gaze data* (locations, frequencies, and durations of fixations).

Subject Performance Analysis

The first three columns of Table 3 illustrate the numbers of correct and incorrect answers given for each of the six image pairs. Incorrect answers are provided as incorrect *equal* answers and as other incorrect answers. Looking at the numbers suggests that differences exist between the six pair conditions for numbers of correct answers. A Friedman ANOVA confirms the existence of significant differences ($\chi^2(2) = 41.989, p < 0.001, r = 0.952$). Two groups of pair comparisons exist, irrespective of presentation order: as a first group, *Cord-256* and *Cord-128* with lower performance, as a second group *Cord-256* and *Cord-64* as well as *Cord-128* and *Cord-64*, with higher performance. The same groups can be formed for the number of incorrect *equal* answers ($\chi^2(2) = 73.935, p < 0.001, r = 0.920$). The first group has many more incorrect *equal* answers than the second. A breakdown of performance and incorrect *equal* counts for the three exposure duration conditions (*short*, *medium*, *long*) revealed no significant differences.

In order to check for training effects, we compared numbers of correct answers for the three blocks (first: 268, sec-

Table 3: Frequencies of correct answers, incorrect equal-quality answers, and other incorrect answers (accumulated over all 20 subjects; sum of answers per pair: 240); average fixation durations and average fixation frequencies per image pair presentation.

	# correct	# equal (incorrect)	# other (incorrect)	av. fix. dur. [ms]	av. fix. freq.
Cord-256 / Cord-128	24	167	49	386.48	2.32
Cord-128 / Cord-256	39	156	45	398.51	2.25
Cord-64 / Cord-256	195	25	20	404.14	2.21
Cord-256 / Cord-64	193	24	23	412.01	2.20
Cord-64 / Cord-128	193	25	22	418.08	2.14
Cord-128 / Cord-64	196	19	25	407.59	2.20

ond: 274, third: 297). For each block, a total of 480 answers were collected across all 20 participants. A Friedman test revealed significant differences between the blocks ($\chi^2(2) = 6.195, p < 0.05, r = 0.952$). A comparisons of means shows a positive training effect.

Gaze Fixation Analysis

We next analyzed subjects' gaze fixation distributions across the sofa image in order to assess whether differences exist for different exposure durations and for different image pair comparisons. Fixation counts for cells in an overlaid 16x16 grid are shown in Figure 2 (upper part) for nine conditions. Fixation count patterns between any pair of these nine conditions are significantly correlated with all $r > 0.850$ and $p < 0.001$.

Table 4 shows average fixation duration (AFD, in ms) and fixation frequencies (FF). For the three BTF resolution conditions, a Friedman ANOVA shows significant differences in FF ($\chi^2(2) = 6.495, p = 0.039, r = 0.697$) and AFD ($\chi^2(2) = 7.777, p < 0.03, r = 0.649$). AFDs decrease and FFs increase from lower to higher resolution textures. For first and second images, a Wilcoxon test shows a significantly lower FF on the second image ($Z = 3.062, p < 0.003, r = 0.684$) and a longer AFD on the second ($Z = 2.420, p = 0.025, r = 0.541$). For the first, second, and third blocks we find an increase in AFDs ($\chi^2(2) = 8.527, p = 0.045, r = 0.623$) and a decrease in FFs ($\chi^2(2) = 8.954, p = 0.011, r = 0.608$).

In order to check whether the subjects' fixation location patterns were driven by visually perceivable differences between images in our BTF image pairs, we employed the Visible Difference Predictor (VDP) (Mantiuk, Daly, Myszkowski, and Seidel (2005)). VDP simulates low level human perception for known viewing conditions (in our case: a resolution of 1920x1080 pixels at an observer's distance of 0.7m). The last row of Figure 2 shows the visually perceivable differences per image pair (irrespective of presentation order) as predicted by VDP. Correlations between VDP results and respective fixation location patterns can be seen in Table 5 (as averaged over exposure durations; displayed in the columns above each VDP result in Figure 2). The results confirm the two groups of image pairs found in the subject performance analysis: (1) a weak correlation for *Cord-256* and *Cord-128* pairs and (2) strong correlations for the pairs within the group of *Cord-256* and *Cord-64* as well as *Cord-128* and *Cord-64*. Lastly, existence of the two groups is further supported by average fixation durations and fixation frequencies for the individual image pairs as seen in the right-hand part of Table 3. AFDs in the first group are significantly lower than in the second ($\chi^2(2) = 73.935, p < 0.001, r = 0.920$), while FFs are significantly higher ($\chi^2(2) = 41.989, p < 0.001, r = 0.952$).

Table 4: Average Fixation Duration[ms] (AFD) and Fixation Frequency (FF) for different image quality levels, first and second images, and for blocks.

	AFD[ms]	FF
Cord-256	429.78	2.24
Cord-128	436.45	2.23
Cord-64	444.90	2.17
First Image	422.43	2.38
Second Image	493.09	2.05
First Block	375.08	2.33
Second Block	405.25	2.30
Third Block	448.71	2.02

Table 5: Correlations between VDP results and fixations independently of exposure durations and presentation order.

	r	p
Cord-256 _ Cord- 64	0.808	0.0001
Cord-128 _ Cord- 64	0.753	0.0001
Cord-256 _ Cord-128	0.015	0.0175

Discussion

The results show that two groups of image comparisons exist in our study. The first group consists of comparisons between *Cord-256* and *Cord-128*. For this group, subjects are largely unable to perceive existing differences between the images. Instead, they frequently judge the pair to consist of the same image. The higher average FFs and lower AFDs in this group suggest more visual search for existing differences. The VDP model predicts few visually perceivable differences for image pairs in this group.

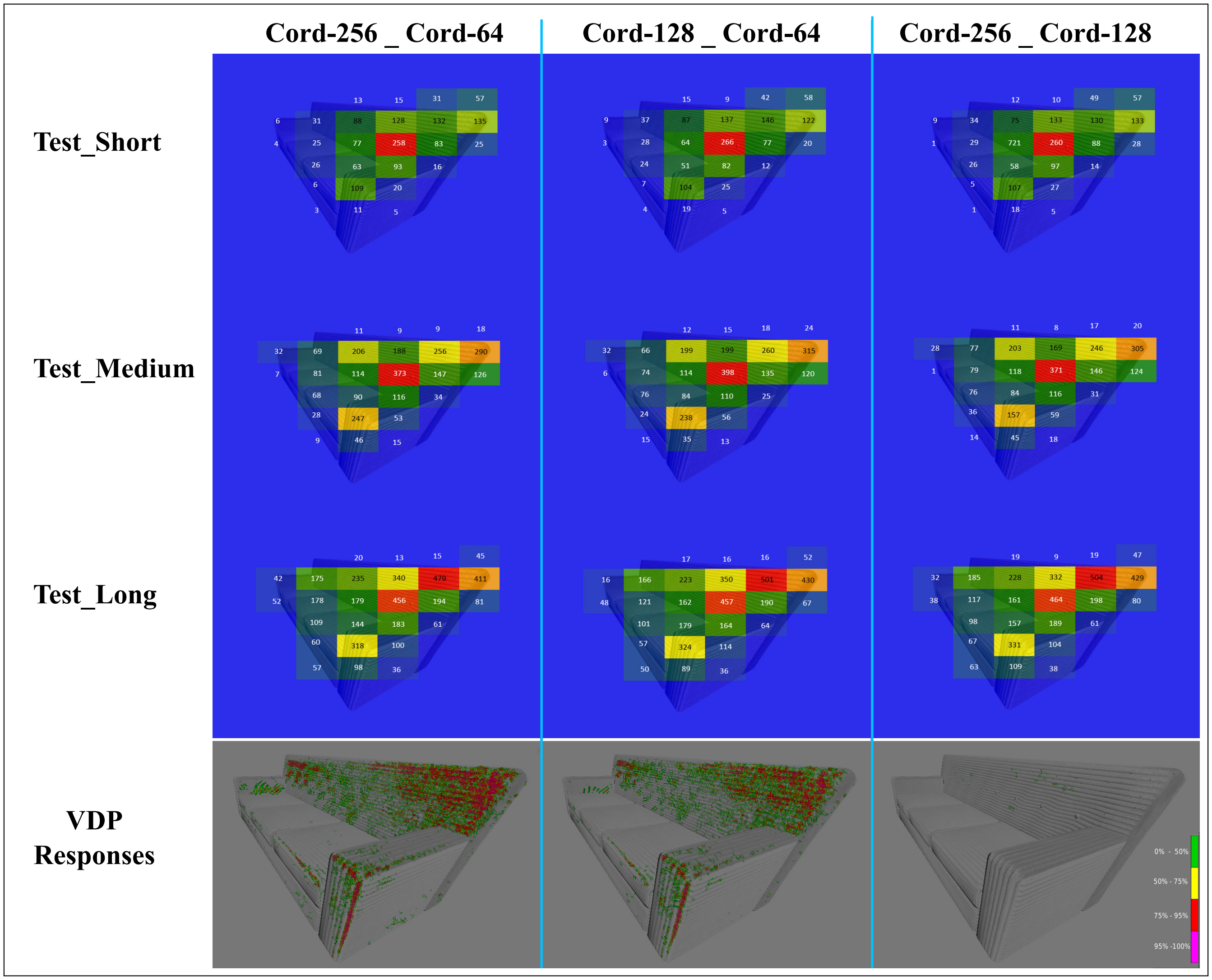

Figure 2: Fixation count in different test duration and responses of visual difference predictor for tested image pairs.

The second group consists of comparisons between *Cord-256* and *Cord-64* as well as between *Cord-128* and *Cord-64*. For this group, subjects are largely able to see the differences among the pairs. Occurrences of incorrectly labeling pairs as *equal* are few. The lower FF counts and higher AFDs suggest that subjects are better able to concentrate on informative locations (i.e., on locations at which the images within a pair differ). The VDP model predicts a higher number of differences which are also detectable with higher probability.

A comparison between the fixation location patterns between the first and the second group reveals that, irrespective of group, subjects seem to fixate on similar locations, and do so with similar frequencies. One conclusion is that they employ similar strategies while inspecting image pairs of any of the six types. VDP predictions differ markedly between the groups. We observed strong correlations between locations of predicted visually perceivable differences and observed fixation patterns only for the second group of comparisons. We interpret this as evidence for subjects' ability to pick up on differences in the second group and use information about the location of these differences for image comparisons. A significant, albeit very weak, correlation exists for the first group.

When comparing AFDs and FFs between the three BTF image qualities, it seems likely that low image quality leads to less visual search, suggesting that subjects are fast at discerning features that hint at low quality.

AFD was lowest for the first block and then increased over the course of the experiment, while the average FF decreased. This pattern is in line with the one presented in (Over et al. (2007)) and suggests that subjects may have applied a coarse-to-fine approach during visual search. Within the first comparisons, subjects may notice locations at which differences between images of different visual quality are located, leading to more fixations at them. This may differ for behavioral patterns in the beginning, when subjects spend more time carefully searching for differences among image pairs, resulting in shorter and a larger number of fixations.

Longer AFDs in the second image in a pair compared to the first indicate that by the time subjects look at the second image they already have formed hypotheses about where to look for differences.

There were no differences in performance and gaze fixation for different exposure durations.

The main purpose of this study was to locate a threshold for robust, effective BTF compression based on a downsampling of BTF pictures. Above the threshold, differences between pictures are not visually perceivable by a human observer. Our results clearly indicate that differences between *Cord-256* and *Cord-128* lie above such threshold, while differences between *Cord-256* and *Cord-64* as well as between *Cord-128* and *Cord-64* lie below it.

The results are likely to apply to all self shadowing fabrics.

Conclusion

The results of our study narrowed the bracket in which the threshold is located that separates visually perceivable differences in BTF renderings from those that are not. Consequently, we can now suggest a perception-based criterion for downscaling BTFs. A result for image synthesis is that, above the threshold, the lowest texture resolution available can be used without visually perceivable degradation of image quality. This allows to significantly reduce computer memory usage in BTF rendering.

A logical next step would be to conduct a localized search within this established bracket, that is, between *Cord-128* on one side and *Cord-64* on the other, since our study showed that observers cannot distinguish between *Cord-256* and *Cord-128*.

In the future, we also plan to look for ability- and/or skill-dependent differences in the ability to distinguish BTFs at different quality levels. We have already conducted pilot studies with groups of engineers and artists.

In general, there are few studies on perceptual measures of rendering algorithms. This study is a first step in this direction.

Also, this study could open up new research insights for the field of perception of textures of real objects, especially in object comparison tasks. For example, future questions that can be addressed could relate to the categorization of textures in object perception, either general or with regard to group-dependent or individual differences, to effects of attention in object texture perception, or to effects of expertise which may be acquired through completing series of object texture comparisons similar to the ones employed in this study.

Acknowledgments

We would like to thank our test subjects for participating, Jakob Gomoll for helping in implementing and conducting the pilot study and Stefanie Wetzel for the constructive discussions.

References

Dana, K. J., Van Ginneken, B., Nayar, S. K., & Koenderink, J. J. (1999). Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics (TOG)*, *18*(1), 1–34.

Filip, J., Chantler, M. J., & Haindl, M. (2008). On optimal resampling of view and illumination dependent textures. In *Proceedings of the 5th symposium on applied perception in graphics and visualization* (pp. 131–134).

Filip, J., Chantler, M. J., & Haindl, M. (2009). On uniform resampling and gaze analysis of bidirectional texture functions. *ACM Transactions on Applied Perception (TAP)*, *6*(3), 18.

Filip, J., & Haindl, M. (2009). Bidirectional texture function modeling: A state of the art survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *31*(11), 1921–1940.

Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, *3*(5), 3.

Lawson, R., Bulthoff, H. H., & Dumbell, S. (2003). Interactions between view changes and shape changes in picture-picture matching. *PERCEPTION-LONDON-*, *32*(12), 1465–1498.

Mantiuk, R., Daly, S. J., Myszkowski, K., & Seidel, H.-P. (2005). Predicting visible differences in high dynamic range images: model and its calibration. In *Electronic imaging 2005* (pp. 204–214).

Mcmillan, L., Smith, A. C., Matusik, W., & Matusik, W. (2003). A data-driven reflectance model. In *in proc. of siggraph.*

Meseth, J., Müller, G., Klein, R., Röder, F., & Arnold, M. (2006). Verification of rendering quality from measured btfs. In *Proceedings of the 3rd symposium on applied perception in graphics and visualization* (pp. 127–134).

Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsberg, I. W., & Limperis, T. (1977). *Geometrical considerations and nomenclature for reflectance* (Vol. 160). US Department of Commerce, National Bureau of Standards Washington, DC, USA.

Over, E., Hooge, I., Vlaskamp, B., & Erkelens, C. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Research*, *47*(17), 2272–2280.

Pellacini, F., Ferwerda, J. A., & Greenberg, D. P. (2000). Toward a psychophysically-based light reflection model for image synthesis. In *Proceedings of the 27th annual conference on computer graphics and interactive techniques* (pp. 55–64).

te Pas, S. F., & Pont, S. C. (2005a). A comparison of material and illumination discrimination performance for real rough, real smooth and computer generated smooth spheres. In *Proceedings of the 2nd symposium on applied perception in graphics and visualization* (pp. 75–81).

te Pas, S. F., & Pont, S. C. (2005b). Estimations of light-source direction depend critically on material brdfs. *Perception ECVP abstract*, *34*, 0–0.