Analysis of Human Gaze Interactions with Texture and Shape

Jiří Filip, Pavel Vácha, and Michal Haindl

Institute of Information Theory and Automation of the ASCR, Czech Republic

Abstract. Understanding of human perception of textured materials is one of the most difficult tasks of computer vision. In this paper we designed a strictly controlled psychophysical experiment with stimuli featuring different combinations of shape, illumination directions and surface texture. Appearance of five tested materials was represented by measured view and illumination dependent Bidirectional Texture Functions. Twelve subjects participated in visual search task - to find which of four identical three dimensional objects had its texture modified. We investigated the effect of shape and texture on subjects' attention. We are not looking at low level salience, as the task is to make a high level quality judgment. Our results revealed several interesting aspects of human perception of different textured materials and, surface shapes.

Keywords: texture, shape, human gaze, perception, psychophysics.

1 Introduction

Many research areas such as computer vision, computer graphics, image understanding, cognitive psychology, start to focus research effort also on human perception of real materials. This research has enormous positive impact on many practical application such as automatic object recognition, scene segmentation, data compression, efficiency of rendering algorithms etc. In this context the best practically available representation of real materials are their view and illumination direction dependent measurements. Such measurements are commonly thought as bidirectional texture functions (BTF) [1]. BTFs represent challenging data due to theirs huge size and thus high processing and rendering expenses [2]. In this paper we used this data to prepare controlled visual search experiment with stimuli preserving realistic appearance of textured materials. In this experiment we investigate the effect of object, surface and texture on overt attention in a demanding visual search task. We are not looking at low level salience, as the task is to make a high level quality judgment. We expect observers to make fixations to regions which they expect to get the most task relevant information from, rather than those which are the most salient.

2 Related Work

In the past visual psychophysics put great effort in understanding a way people perceive and assess objects of different reflectance properties under specific

E. Salerno, A.E. Çetin, and O. Salvetti (Eds.): MUSCLE 2011, LNCS 7252, pp. 160–171, 2012.

illumination conditions. Ho et al. [3] found that roughness perception is correlated with texture contrast. Lawson et al. [4] showed that human performance in matching 3D shapes is lower for varying view directions. Ostrovsky [5] pointed out that illumination inconsistency is hard to detect in geometrically irregular scenes. The importance of specular reflections of environment illuminations for recognition of 3D shape was analyzed in [7]. A psychophysical analysis of bidirectional reflectance distribution functions (BRDF) dependently on different shapes and illumination environments was performed in [8], [9],[10]. Basic perceptual interactions of texture and shape were investigated by Todd and Thaler [11] .

Also human gaze analysis has been widely used in many applications [12] such as visual search [13], web-sites usability [14], eye motion synthesis [15], predicting of fixation behavior in computer games [16], assessing visual realism of medical scenes [17] or quality of BTF resampling [18]. A gaze data analysis was employed in [19]. Authors analysed correlation of fixations density with of several statistical texture features. The best features were then used for perceptually driven BTF compression. Although, eye-tracking methods have been used to analyse the way people perceive or recognize simple 3D objects [20], to the best of our knowledge no one has analysed human gaze interaction with surface shape realistically textured by means of bidirectional texture function.

In this paper we build on our recent work [21], where the same experimental data were used but only subjects responses were analysed, i.e., not gaze fixations. We were comparing abilities of different statistical descriptors of texture to model subjects' responses. Here we are moving further on and thoroughly analyse human gaze attention to experimental stimuli.

In the context of multimedia understanding the effective processing of visual texture information is a must. One way to achieve such an effective processing is analysis of human visual perception and focusing processing algorithm preferably to visually salient part of the data.

Contribution of the Paper. The main aim of this paper is to explore the effect of object shape and texture on human gaze attention in a visually complex task using real textured materials measurements. As we are using one of the most accurate representation material texture appearance we believe that by taking into account texture statistics and their interaction with local geometry and illumination direction, we acquire information about our visual perception of real materials.

Organization of the Paper. The paper starts with description of psychophysical eye-tracking experiment in Section 3. Subjects' responses withing experiment are analysed in Section 4, while their gaze data analysis is subject of Section 5. Main conclusions and future work are outlined in Section 6.

3 Visual Search Experiment

We performed a visual search experiment in order to investigate effects of surface texture, shape and illumination direction and their interactions.

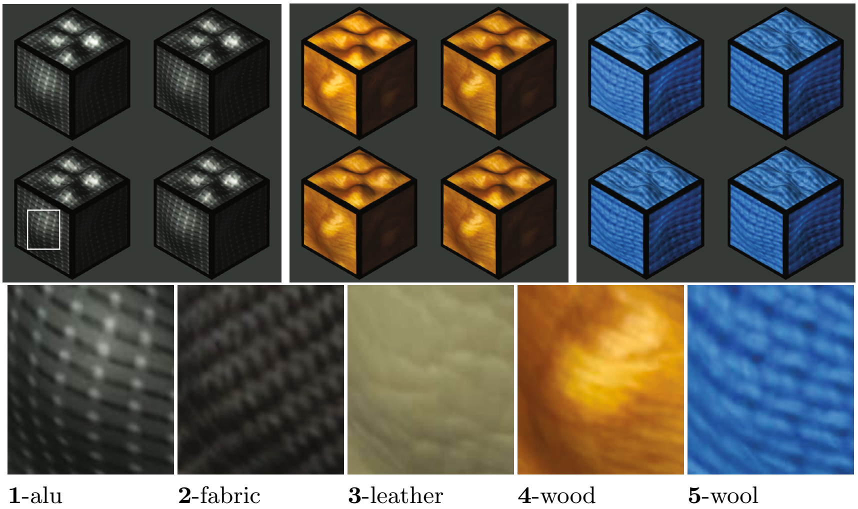

Fig. 1. Examples of the stimuli (first row) and five tested material samples (second row)

Experimental Stimuli. For experimental stimuli we have used static images of size 1000×1000 pixels, featuring four cubes in individual quadrants (first row of Fig. 1). We have used this stimuli layout to avoid the fixations central bias reported in [22], i.e. observers have a tendency to fixate the central area of the screen. The cube faces were modified in a way to feature different geometry on all three visible faces (top, left, right). We used different shape for each cube face: **I**-wide indent (1), **R**-random bumps (2), **B**-wide bump (3), **F**-flat face (4), **H**-horizontal waves (5), **V**-vertical waves (6) as it is show in Fig. 2. For illumination we used directional light from left and right directions parallel with the upper edge of the cubes. This configuration guarantee the same illumination of all cubes in stimuli and similar distribution of light across top and left/right faces in single cubes. Not all combinations of test cube orientations were used in the experiment as this would result in enormous number of stimuli. We used only eleven different orientations selected in a way to allow us to compare the most interesting combinations of faces geometry. Additionally, not all the orientations were illuminated from both directions as shown in Fig. 11. Figure shows orientation number (first row) and shapes of left, right, top faces (third row).

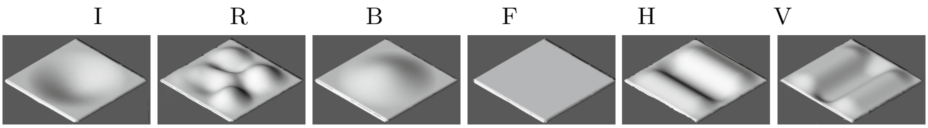

Fig. 2. Tested shapes modifying cube face

Finally, all cubes were rendered using textured materials. We used five different samples of view and illumination dependent textures represented by

Bidirectional Texture Functions (BTF) [23]. In each quadruple three cubes were showing the original data rendering and the remaining one was showing a slightly modified data. We have used three different filters for the original data modification:

A - illumination/view directions downsampling to 50%
B - spatial filtering (averaging by kernel 3×3)
C - spatial filtering (averaging by kernel 5×5).

The proposed filters introduce only very subtle differences (see Fig. 3) between the original and the modified data and force subjects to perform extensive visual search, which allows us to collect detailed gaze data. The edges of the cubes were set to black to mask potentially salient texture seams. Object edges and texture seams are interesting and important sources of visual information, but are deemed out-with the scope of this paper. The background and the remaining space on the screen was set to dark gray. Examples of stimuli and all tested textured materials are shown in Fig. 1.

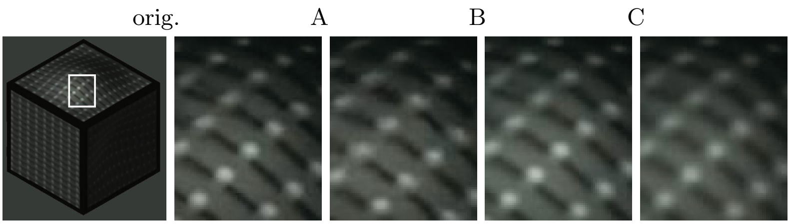

Fig. 3. Performance of the applied filters on sample 1-alu

First row of Fig. 11 shows the 13 conditions of cube orientation and illumination direction that were used. Together with five BTF texture samples, and three different filters, the total number of stimuli was 195 (13×5×3).

Participants. Twelve paid observers (three females, nine males) participated in the experiments. All were students or university employees working in different fields, were less than 35 years of age, and had normal or corrected to normal vision. All were naive with respect to the purpose and design of the experiment.

Experimental Procedure. The participants were shown the 195 stimuli in a random order and asked to identify which of the cubes has a modified/degraded surface texture, i.e., slightly different from the remaining three cubes. Stimulus was shown until one of four response keys, identifying the different cube, was pressed. There was a pause of one second between stimuli presentations, and participants took on average around 90 minutes to perform the whole experiment that was split in four sessions. All stimuli were presented on a

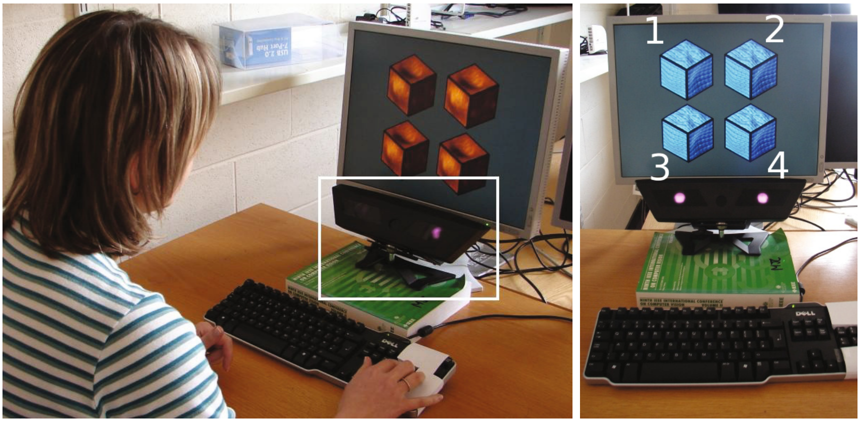

Fig. 4. Setup of the experiment with the eye-tracker highlighted

calibrated 20.1" NEC2090UXi LCD display (60Hz, resolution 1600×1200, color temperature 6500K, gamma 2.2, luminance 120 cd/m^2). The experiment was performed in a dark-room. Participants viewed the screen at a distance of 0.7m, so that each sphere in a pair subtended approximately 9^o of visual angle.

Subjects gaze data were recorded in a dark room using a Tobii x50 infrared-based binocular eye-tracking device as shown in Fig. 4-left. The device was calibrated for each subject individually and provided the locations and durations of fixations at a speed 50 samples/s. Maximum error specified by manufacturer is approximately $\pm 0.5^o$ of visual angle, which corresponds to ±30 pixels for our setup and stimuli resolution. The shortest fixation duration to be recorded was set to 100 ms.

4 Subjects Responses Analysis

First we analysed the subjects ability in finding the modified cube. In average the subjects were able to find the right cube in 67% stimuli, which was surprisingly high given the subtle changes introduced by applied filters (see Fig. 3) (chance level 25%). During an informal interview after experiment subjects mentioned that they were often certain in less than 50% of stimuli and for the rest they were only guessing the right answer. The obtained rates suggests that for difficult cases they often successfully relied on a low level visual perception. The responses accuracy of individual filters is shown in Fig. 5 and reveals that modifications introduced by angular resolution filter **A** are the hardest to spot while the spatial smoothing by filter **B** are the most apparent, as expected since smoothing effect is uniform and generally more apparent in comparison with slight illumination and view direction dependent change in reflectance caused by directions reduction (**A**). While success rates across textures were quite similar for smoothing filters **B** and **C**, their values for filter **A** varied much more. Average values for individual materials are shown in Fig.7-a.

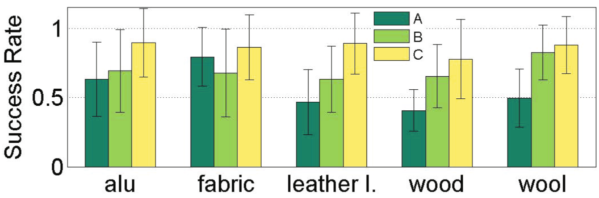

Fig. 5. Percentage of successful responses dependent on data filtration and texture sample

5 Gaze Data Analysis

This section has two parts. In the first, we analyse gaze fixation statistics such as their number, duration etc. across different tested materials. In the second, a spatial distribution of gaze fixation with regards to material and shape is analysed. In both parts a statistics predicting gaze data are offered. In total, twelve subjects performed 62916 fixations longer than 100 ms. Average fixation duration was 242 ms. Each stimulus image was in average observed 11 s and in average comprised 26 fixations.

5.1 Fixation Numbers and Duration Analysis

Fixation duration as a function of ordinal fixation number in Fig. 6-left shows that the average fixation duration was the lowest for the first three fixations and then increased almost linearly with trial duration. This behaviour is similar to results in [13] and suggests that subjects applied a coarse-to-fine approach during visual search. When the difference between cubes is more apparent the subjects notice it within the first few fixations, otherwise they spend more time by detail focused searching for a difference resulting in longer durations of fixations that increase proportionally with total length of the search. Fig. 6-right shows that the total number of fixations decreased almost exponentially and thus most responses to stimuli are given during the first forty fixations.

Figure Fig. 7 presents average success rate obtained from subject keyboard responses (a), image duration (b), number of fixations (c), and fixation duration

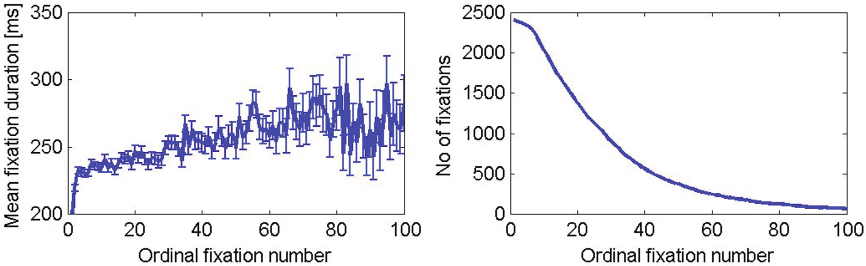

Fig. 6. Mean fixation duration (left) and number of fixations (right) as a function of ordinal fixation number, i.e., average length of first one hundred fixations

(d) for the five tested materials. Error-bars represent twice standard error across different cube orientations and illuminations.

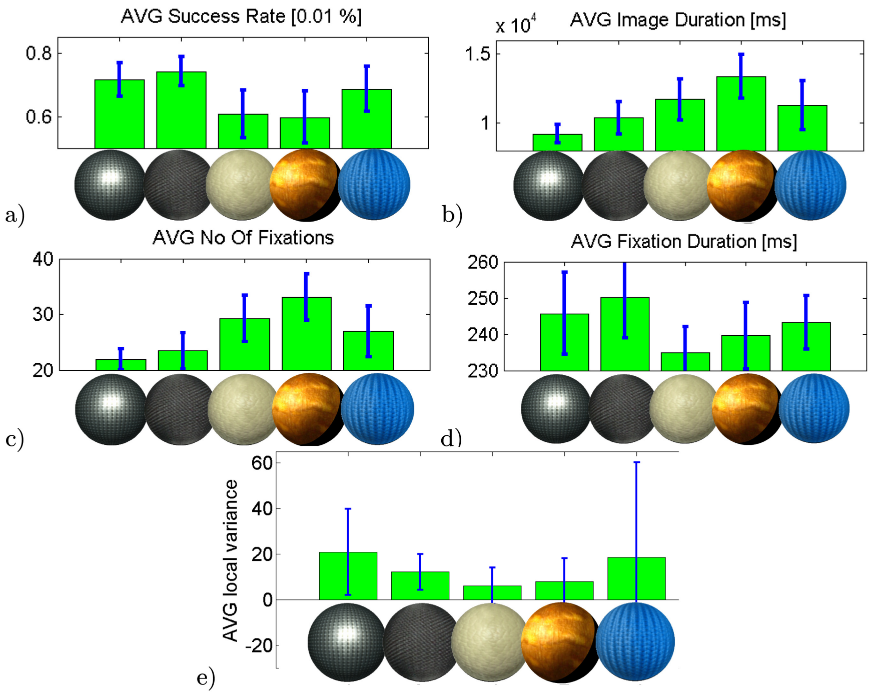

Fig. 7. Average subjects responses success rate (a), average stimulus duration [ms] (b), average number of fixations (c), and average fixation duration [ms] (d), average local variance (kernel 31×31) (e) for the tested material samples

There are apparent differences between materials. For non-regular / more smooth materials **3**-leather and **4**-wood the subjects were less successful in identification of the modified cube, they observed stimuli longer, made significantly more fixations which were much shorter in comparison with the other materials. We assume that when subjects detect regularity in the material, they focus on angular differences only while for non-regular samples they still perform search over wider spatial content of the material.

Motivated by research of Su et al. [24] who reduced texture variation to remove unwanted salient features and by results from [19], where a texture local variance was identified as predictor of fixation density, we computed average local variance of the sphere covered by the textures having the same resolution as the stimuli. We used square Gaussian weighted kernel of size 31×31 pixels. This setting roughly corresponds in our setup to $\pm 0.5^o$ of visual angle and thus the kernel should comprise approximately the same information as a focused eye. Values of average local variance filter for the tested material are shown in Fig. 7-e. Error-bars are twice the standard deviation of local variances across whole image. From the results it is apparent that the lower is average local variance of

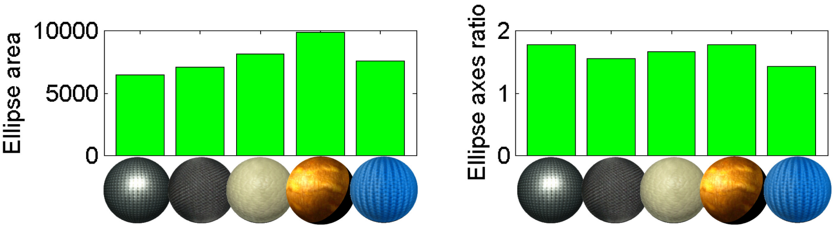

Fig. 8. Areas of fixation cloud fitted ellipses (left) and their axes ratio (right) for individual texture samples

the material, (1) the more time and fixations subjects need to carry on with a given task and (2) the subjects' success rate is lower as there more likely prevail only low frequency features, which is the case of smooth materials (**3**, **4**). For these materials it is more difficult for subjects to recognize the effect of applied filter. This conclusion was supported also by fitting PCA to fixations clouds for individual materials and representing the two principal components as axes of fixation ellipse. The area of such an ellipse for different materials is shown in Fig. 8-left. This suggests that non-regular/smoother textures, receive more but short fixations and the search area is wider. Fig. 8-right shows dependency of ratio of ellipse axes length on material sample (i.e., for 1 = circle). This ratio is higher for more glossy materials (**1**-alu, **3**-leather, **4**-wood) than for the other more diffuse materials. It might suggests that fixation cloud is shaped according specular reflection or other locally salient features of the sample and thus perhaps better follows places with important visual information.

5.2 Spatial Analysis of Fixations

Next a spatial analysis of gaze fixations was carried out. A typical search pattern is shown in Fig. 9-left. We can see that subject tended to do horizontal and vertical saccades (eye moves between fixations) in square pattern, while spending just one or two fixations on each cube. This fact is proved by histogram of a number of fixations on the same cube before jump to any other Fig. 9-right. As follows also from further analysis subjects very rarely made a diagonal saccade, which might be caused by the fact that in real life situations the majority of important visual information is organized either horizontally or vertically.

Interesting results revealed fixation analysis across different cube faces (shapes). Fig. 10-a shows such a distributions of saccades and fixations across different shapes Fig. 2, while Fig. 10-b shows avrage fixation duration. Surface with curvature (1-indent (I), 2-random bumps (R), 3-bump (B), 5,6-waves (V,H)) received more fixations while subjects were not interested in shape 4-flat surface. This fact can be reasoned by presence of less information on such a flat face in comparison with other shapes, which subjects tend to use more frequently.

We have found that distribution of attention across different shapes can be again relatively reliably predicted by the average local variance (Gaussian weighted kernel 31×31) as shown in Fig. 10-c. Interestingly, the number of

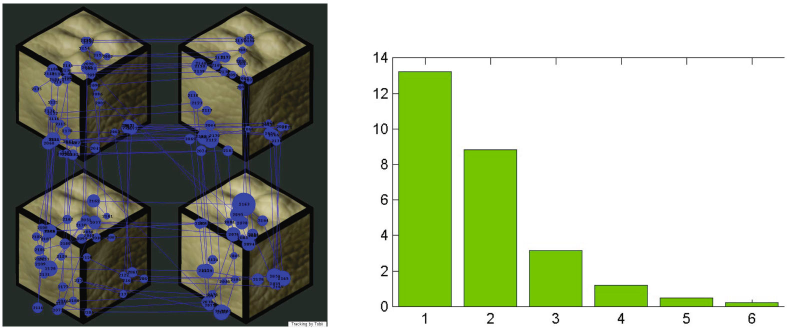

Fig. 9. Example of a typical search pattern (left), histogram of a number of fixations on the same cube before jump to any other (right)

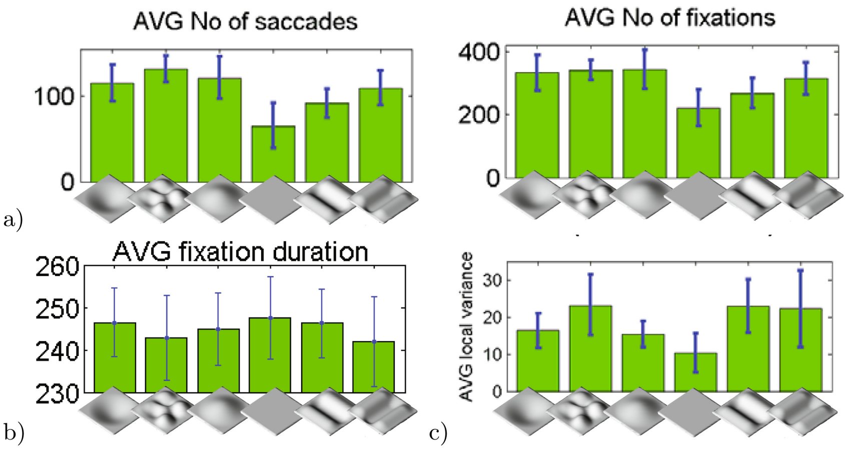

Fig. 10. Comparison of normalized numbers of saccades (left) and fixations (right) (a) and their average duration (b). Prediction based on average local variance (kernel 31×31) for individual cube faces (c).

fixations and saccades (Fig. 10-a) is inversely proportional their average duration (Fig. 10-b), i.e., the more fixations the shape receives the more information it conveys, and thus less time is required per each fixation.

Subjects attention to individual shapes illuminated from two different directions is shown in Fig. 11. The more thick is a dot/line in the second row of images the more fixations/saccades corresponding face or transitions between faces receives in average. Surprisingly, it is not always the case that the most fixations is focused on the two sides of a cube which are illuminated (Fig. 1). In the case when a flat face is illuminated it receives less fixations than shadowed face (i.e., poses 1-L, 2-R, 3-L). This suggests that subjects prefer to observe curved texture rather than uniform flat texture during their task even though such texture is dimly illuminated, i.e., has lower contrast. This conforms with findings of Mackworth and Morandi [25] showing that gaze selects informative details in the pictures.

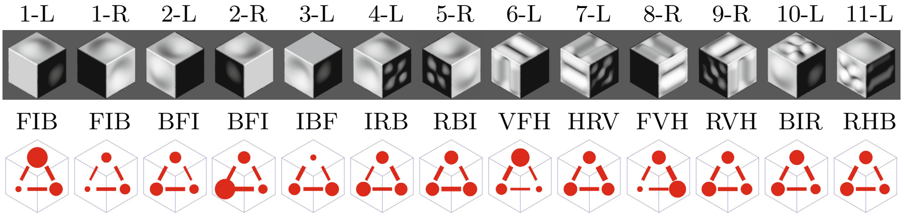

Fig. 11. Geometry significantly influences human gaze attention to texture. Intra-cube (second row) saccades and fixations for all tested cube poses (first row). Thickness of lines (inter-face saccades) and dots (intra-face fixations) represent a number of saccades/fixations.

Similar data across all cube orientations are visualized in the second row of Fig. 12, but there are additional plots in the first row, showing similarly average fixations/saccades between individual cubes in stimuli. It is apparent

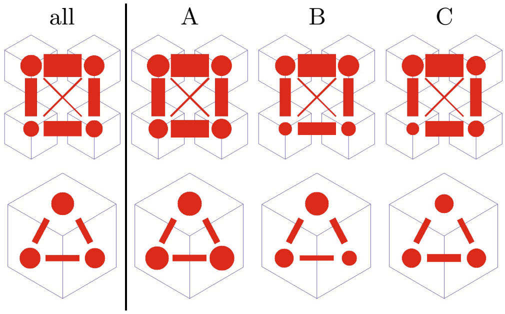

Fig. 12. Inter-cube (first row) and intra-cube (second row) saccades across all stimuli (left) and for individual filters. Thickness of lines (inter-cube/face saccades) and dots (intra-cube/face saccades) correlates with a number of saccades.

that subjects tend to avoid diagonal saccades between cubes and the upper cubes receive more of the saccades/fixations. Fig. 12 also compares gaze attention response to different data modification filters (**A**,**B**,**C**) and supports our previous conclusion that modification by filter **A** (angular downsampling) is the most difficult to spot as the number of fixations/saccades is similar regardless the position in the stimuli, and thus subjects are forced to perform extensive visual search.

6 Conclusions and Future Work

This paper provides an psychophysical analysis of human gaze attention to textured materials and its interaction with surface shape. It revealed several interesting facts. First, a shaped textured surface is definitely more attractive to look at and more informative for observer than a flat textured surface, that receives only half of the fixations in comparison with shaped surface. Second, average local variance of a curved surface texture can predict observers' gaze attention to both texture and its underlying geometry. In other words, the more higher frequencies and regularities are present in the material texture, the easier is identification of possible differences, which require lower number of shorter fixations. Third, angular degradation of view and illumination dependent data is material dependent and less apparent in comparison with a very subtle data spatial smoothing . Finally, upper parts of stimuli receive generally more attention mainly at the beginning of the trial.

Among applied contexts of this work are fast texture data search and retrieval algorithms predicting human attention by means of local variance. These algorithms can exploit the fact that observers tend to seek information in horizontal and vertical directions with more attention to information presented on non-planar shapes in the first half of shown visual frame. Another application is effective perceptually driven compression and rendering of texture data in virtual and mixed-reality applications.

As our future work we believe that these initial findings and their further investigation in multi-modal domain will help to improve performance of methods processing and visualizing accurate texture data.

Acknowledgements. We would like to thank Prof. P. Green, Dr. A. Clarke, Prof. M. Chantler for consultations and providing eye-tracking device, all subjects participating in the experiments, and the Bonn BTF database for providing their BTF measurements. This work has been supported by GAČR grants 103/11/0335 and 102/08/0593, EC Marie Curie ERG 239294, and CESNET grants 378 and 409.

References

1. Dana, K., van Ginneken, B., Nayar, S., Koenderink, J.: Reflectance and texture of real-world surfaces. ACM Transactions on Graphics 18(1), 1–34 (1999)
2. Filip, J., Haindl, M.: Bidirectional texture function modeling: A state of the art survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(11), 1921–1940 (2009)
3. Ho, Y., Landy, M., Maloney, L.: Conjoint measurement of gloss and surface texture. Psychological Science 19(2), 196–204 (2008)
4. Lawson, R., Bülthoff, H., Dumbell, S.: Interactions between view changes and shape changes in picture - picture matching. Perception 34(12), 1465–1498 (2003)
5. Ostrovsky, Y., Cavanagh, P., Sinha, P.: Perceiving illumination inconsistences in scenes. Perception 34, 1301–1314 (2005)

6. Fleming, R.W., Dror, R.O., Adelson, E.H.: Real-world illumination and the perception of surface reflectance properties. Jorunal of Vision (3), 347–368 (2003)
7. Fleming, R.W., Torralba, A., Adelson, E.H.: Specular reflections and the perception of shape. Journal of Vision 4(9) (2004)
8. Vangorp, P., Laurijssen, J., Dutre, P.: The influence of shape on the perception of material reflectance 26(3) (2007)
9. Ramanarayanan, G., Ferwerda, J., Walter, B., Bala, K.: Visual equivalence: Towards a new standard for image fidelity, 26(3) (2007)
10. Křivánek, J., Ferwerda, J.A., Bala, K.: Effects of global illumination approximations on material appearance. ACM Trans. Graph. 29, 112:1–112:10 (2010)
11. Todd, J., Thaler, L.: The perception of 3d shape from texture based on directional width gradients. Journal of Vision 10(5), 17 (2010)
12. Duchowski, A.T.: A breadth-first survey of eye-tracking applications. Behav. Res. Methods Instrum. Comput. 34(4), 455–470 (2002)
13. Over, E., Hooge, I., Vlaskamp, B., Erkelens, C.: Coarse-to-fine eye movement strategy in visual search. Vision Research 47(17), 2272–2280 (2007)
14. Nielsen, J., Pernice, K.: Eyetracking Web Usability. Voices That Matter. New Riders (2009)
15. Deng, Z., Lewis, J.P., Neumann, U.: Automated eye motion using texture synthesis. IEEE Computer Graphics and Applications 25(2), 24–30 (2005)
16. Sundstedt, V., Stavrakis, E., Wimmer, M., Reinhard, E.: A psychophysical study of fixation behavior in a computer game. In: APGV 2008, pp. 43–50. ACM (2008)
17. Elhelw, M.A., Nicolaou, M., Chung, J.A., Yang, G.Z., Atkins, M.S.: A gaze-based study for investigating the perception of visual realism in simulated scenes. ACM Transactions on Applied Perception 5(1) (2008)
18. Filip, J., Chantler, M., Haindl, M.: On uniform resampling and gaze analysis of bidirectional texture functions. ACM Transactions on Applied Perception 6(3), 15 (2009)
19. Filip, J., Haindl, M., Chantler, M.: Gaze-motivated compression of illumination and view dependent textures. In: Proceedings of the 20th International Conference on Pattern Recognition (ICPR), pp. 862–864 (August 2010)
20. Leek, E.C., Reppa, I., Rodriguez, E., Arguin, M.: Surface but not volumetric part structure mediates three-dimensional shape representation: Evidence from part-whole priming. The Quarterly Journal of Experimental Psychology 62(4), 814–830 (2009)
21. Filip, J., Vácha, P., Haindl, M., Green, P.R.: A Psychophysical Evaluation of Texture Degradation Descriptors. In: Hancock, E.R., Wilson, R.C., Windeatt, T., Ulusoy, I., Escolano, F. (eds.) SSPR & SPR 2010. LNCS, vol. 6218, pp. 423–433. Springer, Heidelberg (2010)
22. Tatler, B.W.: The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. Journal of Vision 7(14), 1–17 (2007)
23. Database BTF, Bonn (2003), `http://btf.cs.uni-bonn.de`
24. Su, S., Durand, F., Agrawala, M.: De-emphasis of distracting image regions using texture power maps. In: APGV 2005: 2nd Symposium on Applied Perception in Graphics and Visualization (2005)
25. Mackworth, N., Morandi, A.: The gaze selects informative details within pictures. Attention, Perception, & Psychophysics 2, 547–552 (1967) 10.3758/BF03210264