Rapid Sensing of Material Affordances

Martin Giesel and Qasim Zaidi
Graduate Center for Vision Research, SUNY College of Optometry,
New York, NY10036, USA

***Abstract*— People can make rapid visual judgments of the affordances of materials. The inferential problems created by the interactions of illumination-geometry with 3D material-structure and object-shape can be simplified by heuristics based on rapidly extracted image information. We demonstrate that material properties, such as roughness, thickness and flexibility, are characterized by specific scales of luminance variations, and that percepts of these qualities are a function of the relative energy in corresponding spatial-frequency bands. Cortical mechanisms could estimate material properties by combining the parallel outputs of frequency-selective neurons. Similarly, a hardware implementation using oriented multi-scale filters in parallel could also do rapid remote sensing of material properties.**

Keywords: material affordances, material perception, spatial frequency filters, parallel processing

I. Introduction

Estimating material properties is often at least as important as recognizing object classes [1-5]. When running on a path it is necessary to make rapid judgments to avoid areas that may be slippery, muddy, or flooded. When reaching for sandpapers while working on wood, we usually make rapid visual judgments of the roughness. Visual inferences are especially important when judgments have to be made rapidly for materials at greater distance than arms' length. In all of the examples above, the intended use of the material specifies the relevant affordances [6], i.e., properties that allow particular uses. Affordance judgments are based on the retinal images projected from interactions between illumination-geometry, material-structure and object-shape. Without external constraints, the physics of these interactions are too involved [7,8] for the visual system to estimate material structure by inverse optics, yet perceiving material roughness or slipperiness shows that humans can, to a certain degree, rapidly infer the underlying physical structure from the retinal image.

An attractive possibility is that we use heuristics based on rapidly extracted image features to infer affordances. In human perception, such image features have been proposed for various material properties, for example, X-junctions [9] and contrast relations for transparency [10], or image motion [11] and skewness of luminance statistics [12] for gloss. Heuristics work well when they are reliably related to 3D configurations, e.g. a correlation between the perceived roughness of textures and their fractal dimension works, because the fractal dimension of 3D scenes and the fractal dimension of their images are identical [13].

We summarize and build on the low-level perceptual mechanisms demonstrated by [4] for the identification of the material properties flexibility, thickness, and roughness. This mechanism estimates 3D surface properties from the 2D frequency representation of images.

Our results suggest, that a parallel processing scheme of spatial frequency based filters might enable rapid identification of many material properties. Such a scheme would utilize the repetitive 3-D structure of fabric-like materials that is mainly conveyed by 3-D shape from the scale of luminance variations and could thus be generalized to real-time robotic vision.

II. Frequency-Based Heuristics for Material Properties

To investigate the perception of material properties, we chose images of fabrics, since they are familiar materials that come in a wide variety with diverse affordances. Fabric properties are a function of the nature of the fiber and the structure of the knit or weave. Since these structures vary within a restricted spatial scale, an additional advantage is that most fabrics are examined within a narrow range of distances, similar to what we used in our experiments.

Nine paid observers rated 256 color images of fabrics on four opponent affordance dimensions: soft – rough, flexible – stiff, warm – cool, and water-absorbent – water-repellent, using five-point scales. To guide observers, we gave them questions aimed at potential uses: "If you felt this material on your skin, would it feel soft or rough?", "If you folded or draped this material, would it be stiff or flexible?", "Would clothes made of this material keep you warm or cool?", "Would you use this material to repel water or would you use it to absorb water?".

Fig. 1A shows examples of the classifications for each affordance dimension, illustrating that even images placed in the same affordance category vary on multiple perceptual dimensions. In addition, some properties were strongly associated with others (Fig. 1B, 1C). We tabulated the contingency table of frequencies with which an image had been rated in the two highest scale values for each of the eight properties, and used Correspondence Analysis (SVD applied to $\chi 2$ association statistics [17]) to reduce the dimensionality of the problem. The first dimension (CA1), which explained 57.1% of the total variance in the data matrix, largely coincided with the properties soft and flexible on the positive side, and their opponents, rough and stiff, on the negative side. The second dimension (CA2), which explained 24.8% of the total variance, was closest to the material properties cool on one end and warm on the other. As would be expected, the perceived warmth of a fabric is often a function of its perceived thickness. Absorbent sensibly is closer to warm and soft, while repellent is closer to stiff and rough. Based on the dimensionality reduction, we focused the subsequent image analyses on

Support: NEI grants EY07556 & EY13312 (QZ), DFG grant GI 806/1-1 (MG)

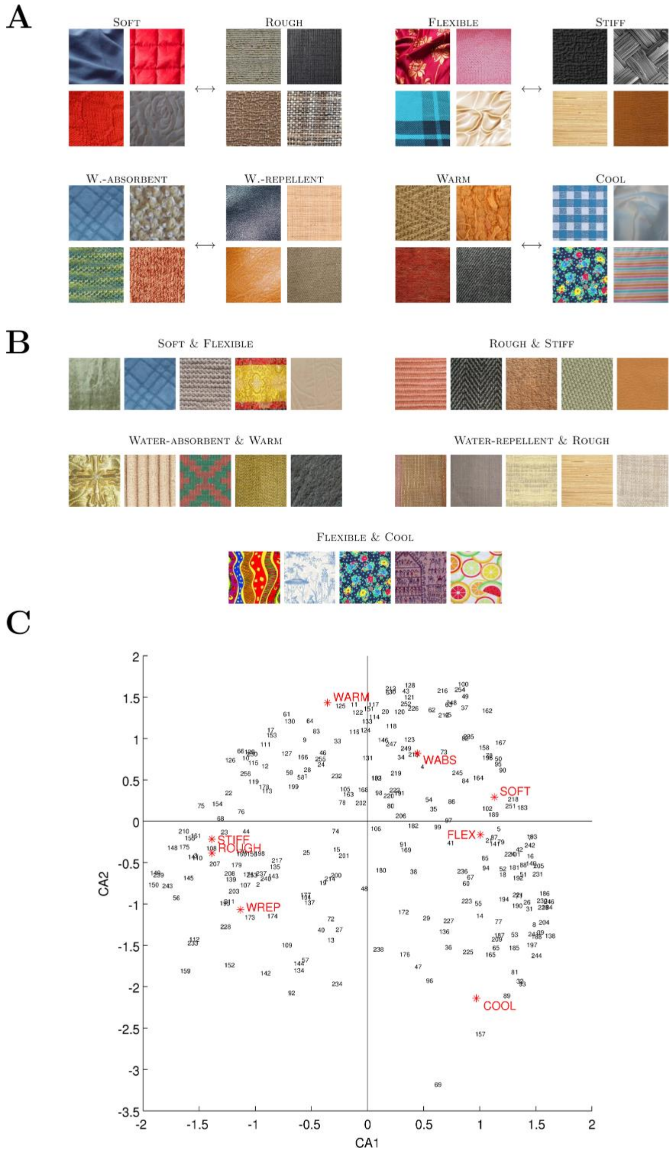

Figure 1. Results of the rating experiment. (A) Examples of materials with highest observer consensus for four opponent material property pairs. (B) Strongest associations across material properties. (C) Results of the Correspondence Analysis. The two axes are the two orthogonal dimension determined by the Correspondence Analysis. The locations of the properties on these axes are shown in red (FLEX = flexible, WABS = water-absorbent, WREP = water-repellent).

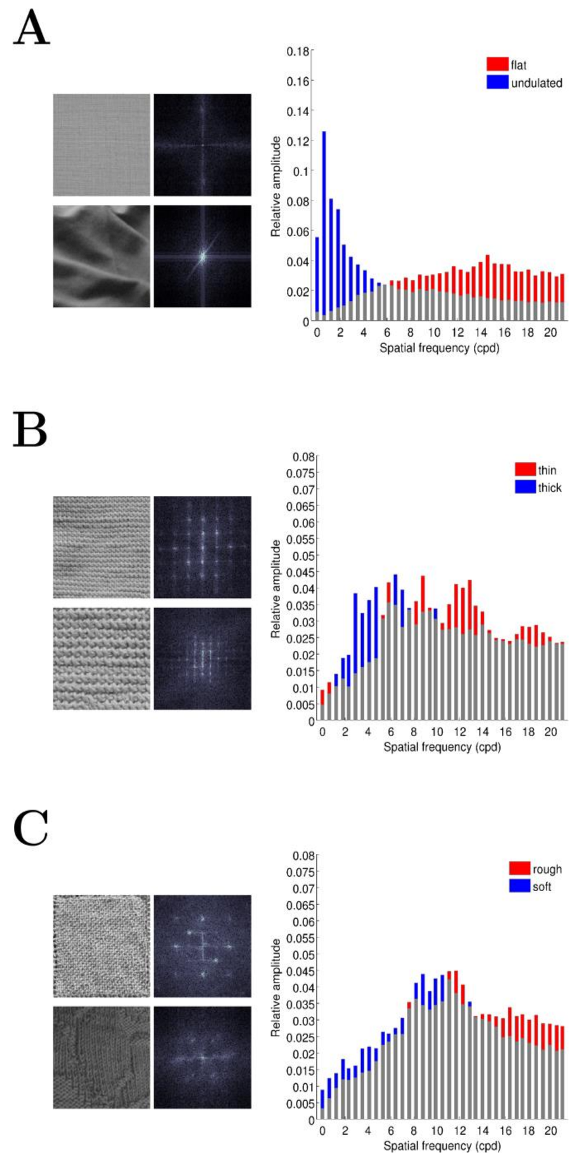

Figure 2. Comparisons of amplitude spectra for opponent material properties summarized by spatial frequency histograms. (Left column) Fabric images with their amplitude spectra. (Right column) Histograms of amplitude distributions across spatial frequencies. The colored parts of the bars indicate the amount by which one image exceeds the other. (A) undulation. (B) thickness. (C) roughness.

materials classified as soft or rough and on thin and thick appearing materials. A visual inspection of the soft and rough images in Fig. 1A suggests that the size of the dominant structure or pattern is a distinguishing cue between them. Rough materials have a fine-grained structure with sharp transitions, whereas soft materials have larger structures with smooth transitions. Within the soft materials there seems to be a further subdivision into a group of thicker looking fabrics, and a group with thinner fabrics that contain broad undulations, probably due to the suppleness of the fabrics.

To determine the dominant scale of a material's structure, we analyzed the images' amplitude spectra. Since we found no obvious effect of color in the classification experiment, we used gray-scale versions of the images for the frequency analysis. Fig. 2 shows amplitude spectra of pairs of fabrics chosen to be exemplary of the opposite ends of the undulation (Fig. 2A), thickness (Fig. 2B), and roughness (Fig. 2C) properties. The histograms in the second column of Fig. 2 show the relative energy in various bands of spatial frequencies. The colored parts of the bars indicate the amount of energy by which one of the fabrics exceeds the other one in a given frequency band. The spectra of undulated fabrics contained more energy at low frequencies as compared to spectra of flat fabrics. Spectra of thick and thin fabrics differed in a frequency band slightly higher than the first band, and spectra of rough fabrics contained more energy at middle-frequencies than spectra of soft fabrics.

III. IMAGE MANIPULATIONS

To determine whether the amount of energy at certain spatial scales systematically influences the perception of the material properties undulation, thickness, and roughness, we chose three bands of spatial frequencies based on the image analysis: A low-frequency band corresponding to undulations in fabrics covering 2–8 cycles per image (cpi) or 0.57–2.29 cycles per degree (cpd), a frequency band corresponding to the thickness of the weave or knit (8–15 cpi & 2.29–4.28 cpd), and a middle-frequency band corresponding to the roughness of the fabric (23–53 cpi & 6.57–15.14 cpd).

To verify that these bands are related to their corresponding material properties, we assessed the appearance of images as a function of relative energy in the three bands. To increase or decrease the energy in a frequency band, the frequency band was multiplicatively scaled. To keep the sum of the energy across the amplitude spectrum constant, the remainder of the amplitude spectrum was scaled accordingly. The manipulated images had the same mean as the original images.

Fig. 3 shows the results of multiplicatively scaling the energy in each of the three bands while keeping the overall energy constant. The icons in the left-most column of Fig. 3 indicate the spatial frequency bands. Increasing the energy in the low-frequency band (Fig. 3A, Movie 1), inflates the quilt, whereas decreasing the energy deflates it. The three dimensional appearance of the quilt is largely due to shading variations that are generally gradual, so the energy is concentrated at low spatial frequencies. Increasing the energy in the second frequency band (Fig. 3B, Movie 2), increases the thickness of the weave, whereas decreasing the energy results in a flatter, thinner appearance. Increasing the middle to high-frequency energy (Fig. 3C, Movie 3), leads to a coarser or rougher texture, while decreasing the energy in this frequency range results in a smoother texture. Varying the relative energy of a frequency band, influences how much structures at a certain spatial scale contribute to the overall appearance of the material. It does not alter existing structures or creates new structure. If a material's original spectrum has no structure in a band, multiplying the energy in this band will not lead to the desired appearance changes, manipulating, e.g., the amplitude spectrum of white noise will in general not result in a appearance change that is related to a change in material property. However, the frequency-band analysis suggests a method to transfer qualities across materials, e.g. the soft and flexible appearance conveyed by folds can be transferred to a material originally rated as rough and stiff [4].

IV. RETINAL VERSUS MATERIAL SPATIAL FREQUENCY

We have expressed frequency bands in retinal spatial frequencies, but because all measurements were done at one distance, they could equivalently have been expressed in material spatial frequency. A change in distance alters retinal spatial frequency (cpd), but leaves unchanged the material spatial frequency (cpm), a concept analogous to object spatial frequencies [15]. To determine whether the perceived material properties are determined by material spatial frequencies or retinal spatial frequencies, [4] conducted a control experiment in which we presented images of materials at three different

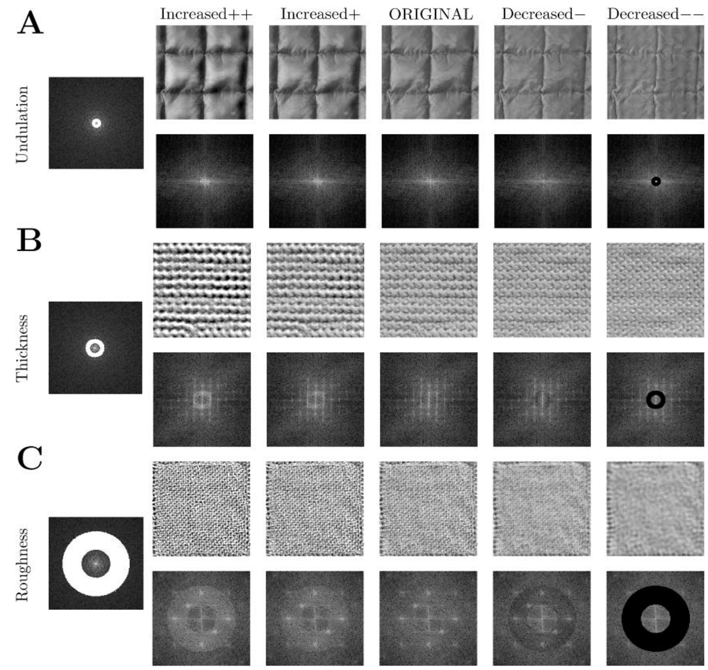

Figure 3. Original and manipulated images and their amplitude spectra. The middle column shows the original images, the 1st and 2nd column show images with increased energy in the frequency bands, and the 4th and 5th column show images with decreased energy. (A) undulation band, (B) thickness band, (C) roughness band (see also Movies 1–3).

distances, and showed that at best there was a weak overall effect of distance.

Since retinal spatial frequencies increase with distance, the band-pass nature of human visual sensitivity to spatial frequencies would be expected to play a role. For example, the two images used for roughness were affected in opposite ways by the increase in the viewing distance. The first image, which was rated as soft, had the critical variations at low to middle frequencies. These were shifted to higher spatial frequencies with increasing distance, and resulted in a rougher appearance. The dominant variations in the second image were already in a high frequency region, so increasing the distance moved them to a low sensitivity domain. Overall, material judgments remained stable over a range of distances from which an observer would commonly examine materials. This suggests that visual inferences of material properties are more likely to be based on estimated material spatial frequencies than on retinal frequencies. This finding is in accordance with our everyday experience where material properties do not change massively with viewing distance.

V. SPATIAL FILTER BASED IMPLEMENTATION

As a start towards a rapid hardware implementation of the frequency based heuristic for judging material properties, we have examined how well a limited number of spatial frequency and orientation tuned filters can separate the three opponent properties discussed above. We have used the pyramid scheme of [16]. We first ascertained that the filters in the set “sp5filters” (6 orientations times 4 spatial frequencies) gave

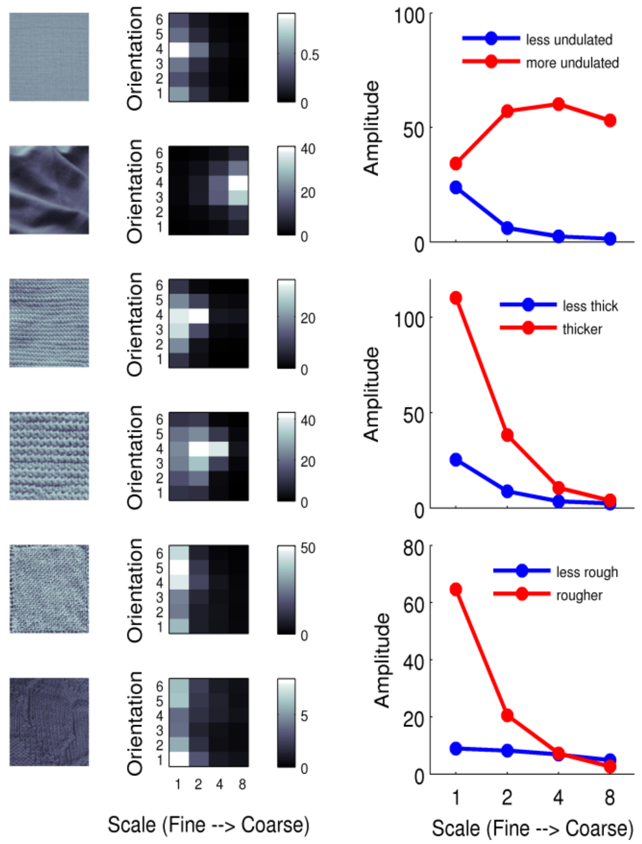

Figure 4. Filter responses for fabrics chosen to be exemplary of the opposite ends of the properties (A, same as in Figure2). (B) Gray-level heat map of filter responses. (C) Filter responses summed over orientation for each spatial-frequency class for materials in the top three and bottom three categories of each material ranking.

separated responses to white noise filtered in the bands of Fig. 3. Then we applied the filters to all of the 161 non-patterned images of fabrics in the original analysis. Fig. 4B shows a gray-level heat map of filter responses to the same images as Fig. 2. It is clear that the filter responses can be used to differentiate between the soft and rough, undulated and flat, and thick and thin fabrics. To strengthen the generality of the frequency-based analysis of material properties, [4] conducted a ranking experiment to derive the frequency bands from experimental data. In this experiment they used printouts of the 161 non-patterned images. The observers' task was to sort the images independently in nine point classes from "least" to "most" undulated, thick, and rough. The sorting was carried out on a large table so that images could be seen simultaneously and compared directly. No specific instructions regarding the viewing distance and the properties were given to the observers. However, observers were shown samples of real fabrics: A fabric lying flat on the table, and the same fabric with folds to illustrate the undulation property, a thinner and a thicker fabric to illustrate thickness, and a softer and rougher fabric to illustrate roughness. This experiment was less well controlled than the monitor based experiments, but had the advantage that observers made relative judgments of the material properties without being required to explicitly label a material as belonging to a certain property category. Fig. 4C shows the responses summed over orientation for each spatial-frequency class, for materials in the top three and bottom three categories of each ranking. The results are promising in how well the filters perform.

Since primate striate cortex is a massively parallel configuration of scale and orientation tuned filters, these results suggest that the frequency-based heuristic can function on rapidly extracted image features. One possible implementation could be to design a Bayesian categorization procedure that uses conditional priors as the embodiment of the heuristic, i.e. a prior probability of the presence of a property or its opposite as a function of spatial frequency, and then calculates the posterior for all six opponent categories. Decisions can then be made based on the ratio of posteriors for each opponent pair. This scheme allows for categorization based on both high and low responses of filters, and where materials have multiple attributes, e.g. rough and undulated. Since the main procedure is multi-scale spatial filtering, just like the cortex, it can be implemented in parallel using GPU programming

VI. DISCUSSION

The main results of this paper are that material properties, such as roughness, thickness and undulations, are characterized by specific scales of luminance variations. The 2D luminance variations arise from the 3D textures of the materials, and human judgments of 3D roughness, thickness, and undulations vary continuously as a function of relative contrast in corresponding 2D frequency-bands. The appearance changes that result from the manipulations of the amplitude distributions in the three frequency bands are all caused by changes of the shading components at different spatial scales. The perceived material properties are thus functions of the 3D structures of the materials, and are mainly conveyed by shape-from-shading cues.

We started our investigation with an experiment asking observers to classify images of fabrics with respect to four material property dimensions using affordance-related questions (soft vs. rough, flexible vs. stiff, warm vs. cool, water-absorbent vs. water-repellent). Using Correspondence Analysis, we inferred basic material property dimensions (undulation, thickness and roughness) potentially underlying the affordance dimensions. In another experiment, we asked observers to rate a subset of the images with respect to the inferred material properties. To better understand the relationship between the affordance-based classifications and the material property ratings, we computed the correlations between the number of observers who classified an image as belonging to a certain affordance category and the median ratings of the same image obtained from the rating experiment. The results are shown in Table 1. Overall, the classifications and ratings exhibit medium to weak correlations. The three material properties clearly do not capture all the variation in the classification data and do not allow differentiating between all the affordance-related classifications (e.g. soft vs. rough and flexible vs. stiff). Further research has to show whether it is possible to describe affordances-related categories as combinations of more general material properties and derive them directly from images.

TABLE I. CORRELATIONS BETWEEN AFFORDANCE-BASED CLASSIFICATIONS AND MATERIAL PROPERTY RATINGS

Affordance-based classifications	**Material property ratings**		
	Undulation	***Thickness***	***Roughness***
Soft vs	0.49	0.07	-0.56
Rough	-0.29	0.19	0.69
Flexible vs	0.51	0.00	-0.41
Stiff	-0.23	0.26	0.60
Warm vs	0.18	0.52	0.15
Cool	0.10	-0.41	-0.27
Absorbent vs	0.35	0.30	-0.17
Repellent	-0.19	0.01	0.42

Since the earliest stage of cortical visual processing consists of neurons that filter the visual scene in terms of spatial frequencies and orientations [17], it is not surprising that spatial frequencies play an important role in pre-attentive texture discrimination [18], and texture matching [19]. More recently, direct scene categorization schemes proposed correlations between specific configurations of power spectra and perceptual scene dimensions such as naturalness and openness [20].

While the visual perception of roughness, thickness or undulation has not been investigated extensively, the estimation of the roughness of surfaces or terrains from images has long been an important topic of research in machine vision. A wide array of methods has been employed to this end, including spatial frequency analysis. In this context, it has been found that spatial frequency analysis was often inferior to other methods, e.g., statistics derived from gray-tone co-occurrence probabilities [21,22]. However, the focus of this line of research was on the reliable identification of physical structures as, for example, required in remote sensing. Here, we were primarily concerned with material appearance and not with the veridical recovery of surface properties. Surface properties and illumination geometry are conflated in the spatial frequency information. The amplitude distribution changes systematically with changes in pose, scale, and illumination, and that seems correlated with the resulting changes in material appearance. An interesting case is presented by slanted surfaces. When 2D textures are slanted, the spatial frequencies are increased in the image, orientation flows are created [23], and the brightness is reduced for Lambertian surfaces. However, the case is more complicated when 3D structures are slanted as the structure determines the change in brightness [24] and spatial frequency [25]. To analyze the interaction of pose, scale, and illumination on material perception, [4] analyzed images from the KTH-TIPS database [26]. In general, slanting materials increased the spatial frequencies in the image at short distances, and the effect is small or even absent for larger distances. Illumination from the side emphasized the finer structure of the fabrics, thus causing a shift to higher image frequencies. Interestingly, the energy peaks for the fabrics were generally located in one of the previously identified frequency bands, and the perceived qualities followed the bands, e.g., when the spatial frequency peak moved to frequencies higher than the roughness band, the material appeared increasingly flat and smooth.

It is obvious that we have looked at a very limited set of material affordances and properties. We intend to use our own high-resolution images of a much larger set of materials to test the usefulness of the frequency-based heuristic. We also intend to use filters at more spatial scales to make more fine-grained frequency distinctions. In the analysis above it made only a little difference whether we summed filter outputs over orientations or used a winner-take-all rule. Whether this is true in general remains to be tested. Finally, patterns on fabrics or painted images on other materials will need to be separated from the luminance variations germane to judging properties of 3-D textures. Since pattern distortions in perspective images provide orientation and frequency information that gives 3-D shape information [27], solving this problem will provide insights on how parallel outputs of orientation and frequency tuned filters are used simultaneously to solve shape from shading and 2-D texture/pattern. Another thorny problem that we still have to tackle is to calculate material versus retinal spatial frequency. This will require judging size of objects across distance variations, and connecting to work on the perceived structure of space.

The neural substrate for the proposed frequency heuristic remains to be investigated. fMRI studies suggest that cortical areas V4 and PIT are important stages for constructing information about material properties [28]. An understanding of how neurons in these areas use the spatial frequency information from earlier areas to make decisions about material affordances, especially across patterns and distance, would be extremely insightful about how to accomplish efficient machine implementations for robotic vision.

REFERENCES

[1] E. H. Adelson, “Seeing stuff: The perception of materials by humans and machines,” in Proceedings of the SPIE: Human Vision and Electronic Imaging VI, 4299, 1—12, 2001.

[2] Q. Zaidi, “Visual inferences of material changes: Color as clue and distraction,” Wiley Interdisciplinary Reviews: Cognitive Science, 2: 686–700, 2011.

[3] B. L. Anderson, “Visual perception of materials and surfaces,” Current Biology, 21: R978—R983, 2011.

[4] M. Giesel and Q. Zaidi, “Frequency-based heuristics for material perception,” Journal of Vision, 13(14): 1—19, 2013.

[5] R. W. Fleming, “Visual perception of materials and their properties,” Vision research, 94: 62—75, 2014.

[6] J. J. Gibson, The Ecological Approach to Visual Perception, New York: Psychology Press, 1986.

[7] J. J. Koenderink and A. J. Doorn, “Illuminance texture due to surface mesostructure,” Journal of the Optical Society of America A, 13: 452—463, 1996.

[8] J. J. Koenderink, A. J. Doorn, K. J. Dana, and S. Nayar, “Bidirectional Reflection Distribution Function of Thoroughly Pitted Surfaces,” International Journal of Computer Vision, 31: 129—144, 1999.

[9] F. Metelli, “Stimulation and perception of transparency,” Psychological Research, 7: 185—202, 1985.

[10] R. Robilotto, B.-G. Khang, and Q. Zaidi, ”Sensory and physical determinants of perceived achromatic transparency,” Journal of Vision, 2(5): 338—403, 2002.

[11] K. Doerschner, R. W. Fleming, O. Yilmaz, P. R. Schrater, B. Hartung, and D. Kersten, "Visual motion and the perception of surface material," Current Biology, 21: 2010—2016, 2011.

[12] L. Sharan, Y. Li, I. Motoyoshi, S. Nishida, and E. H. Adelson, "Image statistics for surface reflectance perception," Journal of the Optical Society of America A, 25: 846—865, 2008.

[13] A. P. Pentland, "Fractal-based description of natural scenes," IEEE Transactions on Pattern Analysis and Machine Intelligence, 6: 661—675, 1984.

[14] H. O. Hirschfeld, A connection between correlation and contingency," Proceedings of the Cambridge Philosophical Society, 31: 520—524, 1935.

[15] D. H. Parish and G. Sperling, "Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination," Vision Research, 31(7): 1399—1415, 1991.

[16] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," International Conference on Image Processing, Vol. 3, IEEE Computer Society, 1995.

[17] M. J. Hawken and A. J. Parker, "Spatial properties of neurons in the monkey striate cortex," Proceedings of the Royal Society of London Series B: Biological Sciences, 231(1263): 251—288, 1987.

[18] B. Julesz, "Visual pattern discrimination," IRE Transactions on Information, 8(2): 84—92, 1962.

[19] W. Richards and A. Polit, "Texture matching," Kybernetik, 16: 155–162, 1974.

[20] A. Oliva and A. Torralba, "Modeling the Shape of the Scene: A holistic representation of the spatial envelope," International Journal of Computer Vision, 42: 145—175, 2001.

[21] R. M. Haralick, "Statistical and structural approaches to texture," in Proceedings of the IEEE, 67, 5, 786—804, 1979.

[22] M. Tuceryan and A. K. Jain, "Texture analysis," in The Handbook of Pattern Recognition and Computer Vision, C. H. Chen, L. F. Pau, and P. S. P. Wang, Eds, Singapore: World Scientific, 1998, pp. 207—248.

[23] A. Li and Q. Zaidi, "Three-dimensional shape from non-homogeneous textures: carved and stretched surfaces," Journal of Vision, 4(10): 860—878, 2004.

[24] S. K. Nayar and M. Oren, "Visual appearance of matte surfaces," Science 267: 1153—1156, 1995.

[25] K. J. Dana, B. Van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real-world surfaces," ACM Transactions on Graphics, 18: 1—34, 1999.

[26] M. Fritz, E. Hayman, B. Caputo, and J.-O Eklundh, "The KTH-TIPS database," available at www.nada.kth.se/cvap/databases/kth-tips, 2004.

[27] Q. Zaidi, A. Li, C. Wong, E. Cohen, and X. Meng, "Hardwired and plastic mechanisms in 3-D shape perception," in Shape Perception in Human and Computer Vision: An interdisciplinary Perspective, Dickinson and Pizlo, Eds, Springer, 2013.

[28] N. Goda, A. Tachibana, G. Okazawa, and H. Komatsu, "Representation of the material properties of objects in the visual cortex of non-human primates," Journal of Neuroscience, 34(7): 2660—2673, 2014.