Invited Paper

The Rough Side of Texture: Texture Analysis Through the Lens of HVEI

Thrasyvoulos N. Pappas

EECS Department, Northwestern University, Address, Evanston, IL 60208, USA

ABSTRACT

We take a look at texture analysis research over the past 25 years, from the persective of the Human Vision and Electronic Imaging conference. We consider advances in the understanding of human perception of textures and the development of texture analysis algorithms for practical applications. We cover perceptual models and algorithms for image halftoning, texture discrimination, texture segmentation, texture analysis/synthesis, perceptually and structurally lossless compression, content-based retrieval, and sense substitution.

Keywords: Filter-rectify-filter (FRF) models, structural texture similarity metrics STSIM)

1. INTRODUCTION

Texture is an important visual attribute that has attracted the attention of both visual scientists and engineers developing electronic imaging systems. Indeed, there are strong links in the developments in the two disciplines that have been brought together, over the past 25 years, by the Human Vision and Electronic Imaging (HVEI) conference. For example, understanding the principles and underlying models for texture segregation have been closely linked to techniques for texture analysis/synthesis and texture segmentation, the role of texture in surface material perception has found applications in graphics, and recently, a better understanding of texture perception has lead to the development of objective metrics for texture similarity for content-based retrieval and texture-based compression.

Texture can take different forms, from film grain and halftoning noise, to material surface, to patterns on cloth, tapestry, or rug, to tree foliage, to forest, to a crown viewed from a distance, to terrain in a satellite image. Figure 1 shows several examples, including synthetic textures, for example, a pair of random-dot stereograms, a pattern of crossed lines of different orientations, as well as a portrait of Bela Julesz with a variety of textures (wall pattern, hair, skin, shirt and jacket patterns). Given the variety of visual textures, random or regular, rough or smooth, directional or nondirectional, glossy or matte, natural or synthetic, with different scales and color compositions, it is difficult to come up with a precise definition. However, a number of authors (e.g., Portilla and Simoncelli[1]) agree that more or less *texture images are spatially homogeneous and typically contain repeated structures, often with some random variation (e.g., random positions, size, orientations or colors).* A more difficult task, as we will see in the following sections, is to define perceptual or mathematical features for the characterization of texture. For a concise review on texture perception, we refer the reader to Ref. 2, and for more detailed reviews to Bergen[3] and Landy and Graham.[4]

Texture provides important information to the human visual system (HVS). It enables it to identify the surface geometry and material properties of the object; it allows it to localize the boundaries between objects; and it provides information about object shape. Each of these uses of texture have been studied by vision scientists and engineers and have been addressed by papers at the HVEI conference. Thus, there were papers on surface material perception,[5–7] texture segregation,[8,9] and shape from texture.[10–12]

Texture is also important for electronic imaging systems, whether they try to preserve it (e.g., in image rendering and compression), leverage it for image display (e.g., in image halftoning), or utilize it for image analysis and understanding, as is done by the HVS. Our journey through the past 25 years of the HVEI conference starts with image halftoning, which received a lot of attention over the first decade of the conference (Section 2).

Further author information: (T.N.P.: E-mail: pappas@eecs.northwestern.edu

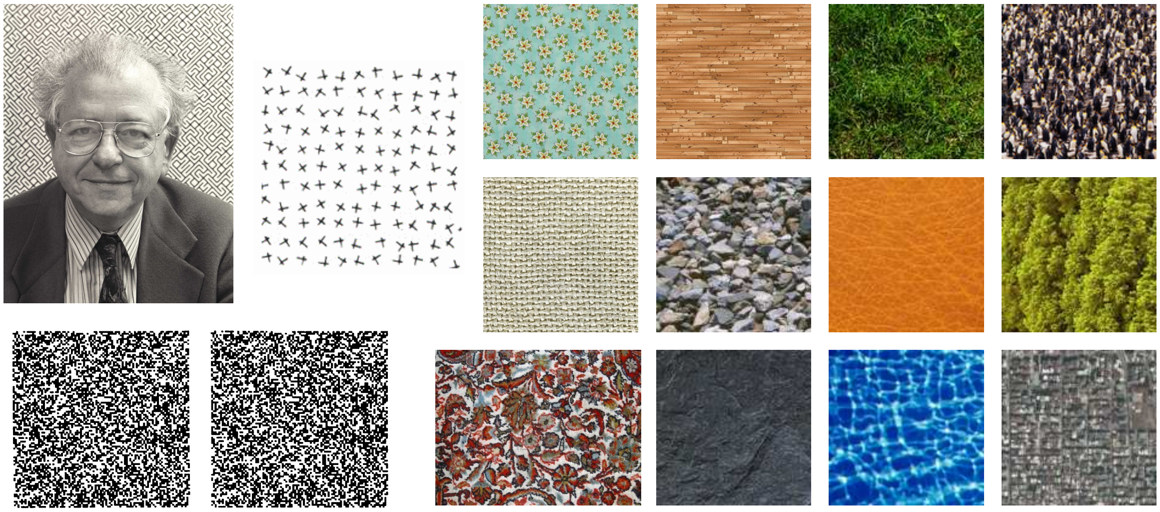


Figure 1. Examples of textures

Texture segregation was one of the topics that received the most attention over the 25 years of the conference. In Section 3, we consider the problem of preattentive texture discrimination, from Bela Julesz's conjecture that textures with identical second-order statistics are indistinguishable, to his texton theory, to the multistage FRF (filter, rectify, filter) models. In the process, we discuss the importance of spatial-frequency-channel decompositions, which leads to the discussion of models and algorithms for texture analysis/synthesis, in Section 4. In addition to their utility for synthesizing realistic textures, such techniques provide a framework for investigating fundamental problems in texture discrimination. Then, in Section 5, we consider the development of objective texture similarity metrics, followed by a brief discussion of texture segmentation techniques in Section 6. In Section 7, we look at applications of texture similarity, including image compression, content-based retrieval, and sense substitution. Based on the application requirements and the capabilities of human perception, we then discuss domains for testing similarity metrics (Section 8). Finally, in Section 9, we discuss high-level features for texture classification, and close with the outlook for the future in Section 10.

2. MAKING TEXTURE INVISIBLE: DIGITAL HALFTONING

In early applications of imaging systems, texture played an essential role, e.g., manifesting itself as film grain noise or digital halftoning patterns, which are necessary for image rendering, but must be invisible to the eye (ignoring artistic effects). In digital halftoning, which was a dominant topic in the first decade of HVEI, the rendering of gray or color tones depends on generating patterns of pixels with a limited set of colors (two for black and white printers and eight for CMYK printers) that, when integrated by the eye, create the illusion of a continuous-tone image.

Figure 2 shows different image halftoning techniques at 100 dots per inch (dpi), including screening[13] (classical, Bayer, blue-noise[14]), Floyd-Steinberg (FS) error diffusion,[13] FS error diffusion with perturbations,[13] Jarvis-Judice-Ninke (JJN) error diffusion,[13] and JJN model-based error diffusion with ink-spreading simulation.[15,16] Note that at 100 dpi there are virtually no printer distortions, such as ink-spreading, except in Figure 2 (h), where ink-spreading has been simulated. However, at 300 dpi or higher, printer distortions are significant and must be taken into account when designing a halftoning technique.[15,16]

The visibility of the halftoning patterns (artifacts) can be predicted to a large extend by a simple eye model that consists of a linear filter of Gaussian shape with standard deviation determined by the viewing distance.[17–19] This model was based on estimates of the spatial frequency sensitivity of the eye by Mannos and Sakrison,[20] but flattened at low spatial frequencies, as the drop in in sensitivity at low frequencies is does not enter in the design of a halftoning algorithm. The frequency response of this simple eye filter is also very close to the MTF proposed by Daly.[21]

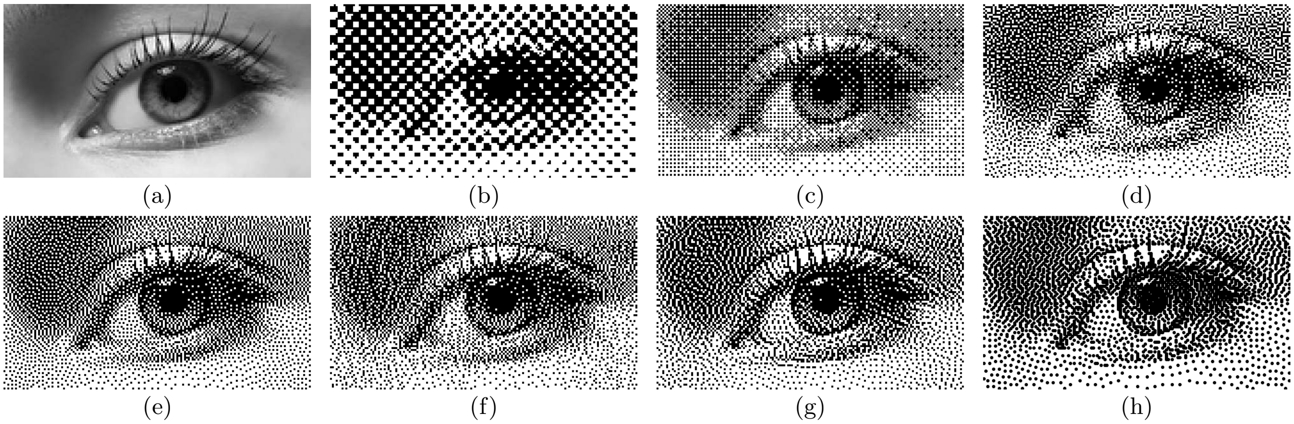


Figure 2. Image halftoning techniques at 100 dpi: (a) Original, (b) classical screening, (c) Bayer screening, (d) blue-noise screening, (e) FS error diffusion, (f) FS error diffusion with perturbations, (g) JJN error diffusion, (h) JJN model-based error diffusion with ink-spreading simulation

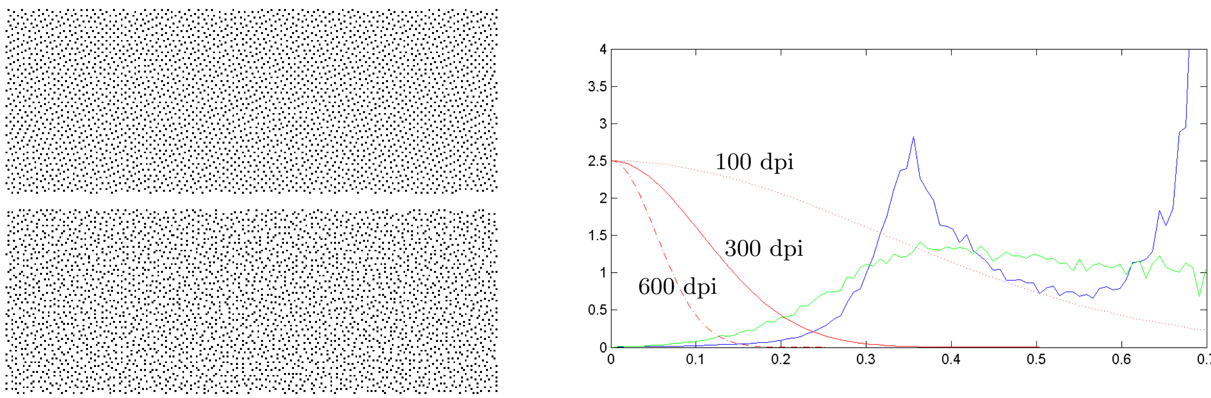


Figure 3. Visibility of halftone patterns at viewing distance of 30 inches and different printer resolutions

A nice illustration of the usefulness of such eye models is in determining the visibility of halftoning artifacts at different viewing distances and printer resolutions. Figure 3 shows two halftoning patterns designed to approximate a constant gray level, 0.125, in a scale of 0 (white) to 1 (black). The pattern on the left is generated by FS error diffusion, and the pattern on the right by FS error diffusion with perturbations.[13] The plot shows the radial frequency spectrum of the two techniques, (FS in blue and FS with perturbations in green). The eye filter for different printer resolutions at a fixed viewing distance (30 inches) is also shown in red. Note that that at relatively high resolutions (300 to 600 dpi), the FS pattern is less visible, while at low resolutions (100 dpi), the FS pattern is more visible, due to a peak at about 0.35 dots/degree.

3. FROM TEXTURE DISCRIMINATION TO THE FRF MODELS

In his 1989 HVEI paper,[22] Bela Julesz presented a brief history and critical discussion of his contributions to preattentive (effortless) texture discrimination. In contrast to the instant success of the random-dot stereograms,[23] which demonstrated that depth perception is possible without any monocular cues, Julesz explained the difficulties of developing a theory of texture discrimination. His work in the 60s and 70s focused on the statistical properties of texture that affect preattentive discrimination.[24–26] To eliminate familiarity cues, and to force the subjects to rely on primitive mechanisms, Julesz conducted systematic tests with synthetic textures.[24] His initial hypothesis was that textures with identical second-order statistics are preattentively indistinguishable.[24,25] This is known as the initial *Julesz conjecture*. However, Julesz himself and his colleagues showed that the conjecture was wrong.[26,27] Setting aside the hypothesis that discrimination can be based on Nth-order statistics, Julesz emphasized the importance of quasi-local features, which he called “textons.”[28] While Julesz's research relied primarily on synthetic textures, Voorhees and Poggio were interested in applying the theory to natural images, and in Ref. 29 presented an algorithm for detecting and comparing textons in natural images.

While Julesz argued that no linear filter model can account for preattentive texture discrimination,[22] Bergen and Adelson[30] showed that adding a nonlinearity after a linear filter, and a second stage of linear filtering after the nonlinearity, leads to models of texture discrimination that are quite effective. A number of other researchers proposed similar models, for example, Malik and Perona.[31] Such models are collectively known as the *LNL* (linear, nonlinear, linear) or *FRF* (filter, rectify, filter) model.[4] An extended discussion of the FRF model can be found in Landy and Graham.[4]

The first stage of the FRF/LNL model consists of a multi-scale/multi-orientation frequency decomposition, like the Gabor[32] and steerable[33] filters. Such decompositions into spatial frequency channels have been widely used to model early visual processing.[34] Similar approaches have been adopted by the signal processing community, which relied on subband decompositions (e.g., QMF[35,36] and GQMF[37] filterbanks and DCTs) for signal analysis, fidelity, and compression. With their perceptually-tuned subband image coder, Safranek and Johnston[38,39] introduced the idea of *perceptually lossless* image compression, and associated perceptual image fidelity models. Similar models were developed by Watson and colleagues for DCTs[40,41] and wavelets,[42] while Lubin[43,44] and Daly[45] proposed fidelity metrics that were based on the more elaborate steerable filter decomposition. Note, however, that with a few notable exceptions,[46] the image fidelity and compression work of the 80s and 90s did not use any models of texture, nor did it make any explicit references to texture other than masking.

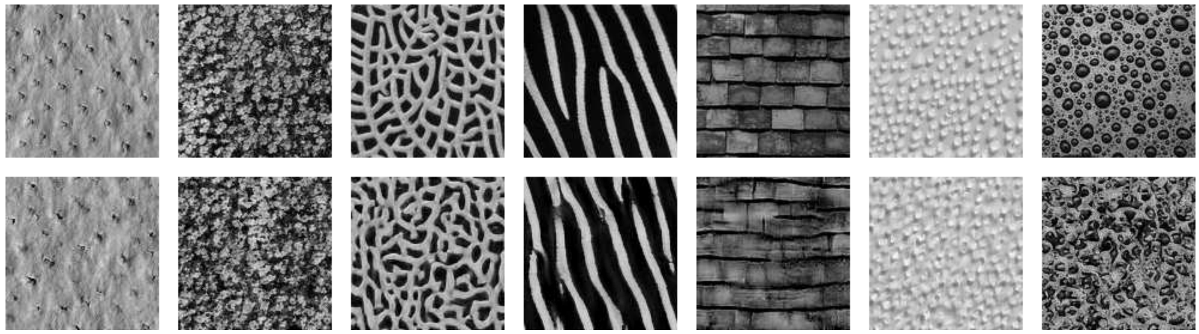
In the next two sections, we discuss the application of spatial frequency channel decompositions and the FRF model to texture analysis/synthesis and texture similarity metrics. As we will see, both of these developments had important implications for image compression.

4. TEXTURE ANALYSIS/SYNTHESIS

As we mentioned in the introduction, texture analysis/synthesis provides a good example of the tight linking between understanding human perception and developing image processing algorithms. Our interest is on texture analysis/synthesis algorithms that are based on spatial frequency channel decompositions.[1,32,46–51] There are a number of other approaches for texture analysis/synthesis that are based on completely different principles. For example, Efros and Freeman[52] proposed a technique for generating a texture based on a small sample. This is done by stitching together patches from the sample. The key to their approach is ensuring continuity between block boundaries. An extended discussion of such techniques is beyond the scope of this paper.

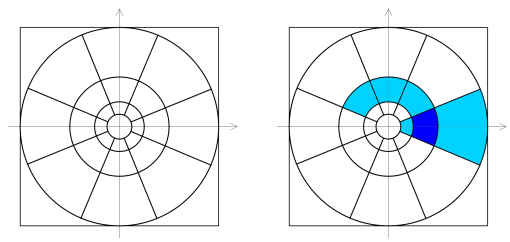
Safranek and Johnston[38,39] demonstrated that when simple, well-established, but limited techniques are applied in the appropriate domain, they can lead to significant advances. Their perceptually-tuned subband image coder[38,39] applied standard quantization techniques (PCM, DPCM[53]) in the subband domain, which allows the incorporation of a perceptual model (subband sensitivities, contrast masking, etc.). In similar fashion, Heeger and Bergen[48] applied simple histogram matching in the subband (steerable filter) domain, and demonstrated impressive texture synthesis results. However, their analysis/synthesis technique is limited to stochastic homogeneous textures. Portilla and Simoncelli[1] relied on the steerable filter decomposition to develop a more elaborate statistical model for synthesizing a much broader set of textures. Their model incorporates a wide variety of subband statistics, including the range, mean, variance, skewness, and kurtosis of each subband, as well as subband auto-correlations and cross-correlations between subbands. This results in a fairly large set of over 800 parameters. Examples are shown in Fig. 4. Note that the texture reconstructions are quite similar to the original texture, even for regular textures like the one in Fig. 4 (e). The only problems arise when the repeating elements become more complicated, as in Fig. 4 (f) and (g). Of course, as the authors demonstrate in Ref. 1, the technique breaks down when the images do not meet the definition of texture.

More importantly, Portilla and Simoncelli[1] used the analysis/synthesis framework to investigate the fundamental underlying principles for texture discrimination. Their goal was to obtain a universal statistical model that parametrizes the space of visual textures. They thus formulated a more elaborate version of the Julesz conjecture, which stipulates that it is possible to find a set of statistics that provide necessary and sufficient conditions for the perceptual equivalence of the textures that obey these statistics. For example, the original and synthesized textures must obey the same statistics.



(a) (b) (c) (d) (e) (f) (g)

Figure 4. Texture analysis/synthesis of Portilla and Simoncelli:[1] Original above, synthesized below



(a) (b)

Figure 5. (a) Steerable filter decomposition; (b) Crossband correlations of blue band with each of the cyan bands

Among the papers that paved the way for the work of Portilla and Simoncelli, we should mention the HVEI papers by Navarro and Portilla[54] and by Zhu, Wu, and Mumford.[50] The Portilla-Simoncelli work, which was highlighted in Simoncelli's 2002 keynote address at HVEI, had a major impact on subsequent texture perception research. Balas[55] further explored the importance of each of the Portilla-Simoncelli statistics for natural textures, while Balas and Rosenholtz[56–58] and Freeman and Simoncelli[59] have used the model to explore the importance of texture perception in peripheral vision. In addition, the structural similarity metrics we discuss in the next section have been inspired by this work.

5. STRUCTURAL TEXTURE SIMILARITY METRICS

Traditional image fidelity metrics that compare two images on a point-by-point basis, like the peak signal-to-noise ratio (PSNR) and some of the metrics that incorporate a perceptual model in the subband/wavelet or DCT domain,[38,39,41,42,44,45] are not well-suited for the objective evaluation of texture similarity. This is because the repetitive, stochastic nature of typically encountered textures allows substantial point-by-point variations that do not have a significant effect on overall texture perception. Thus, it is necessary to adopt a statistical approach for the development of texture similarity metrics.

The development of structural texture similarity metrics (STSIMs)[60–62] was inspired by structural similarity metrics (SSIMs)[63,64] and texture analysis/synthesis.[1] SSIMs introduced the idea of comparing region statistics instead of point-by-point comparisons. Wang *et al.* proposed metrics in both the space domain[63] and the wavelet domain.[64] However, the cross-correlations in the "structure" term are point-by-point comparisons, which result in low similarity scores for textures that are perceptually similar. On the other hand, STSIMs depend entirely on region statistics. The key elements of a STSIM are the following:

A subband decomposition: This is typically a steerable filter decomposition, as in Ref. 1, with 3 scales and 4 orientations, plus a lowpass and a highpass subband, for a total of 14 subbands, as shown in Fig. 5 (a).

A set of statistics: Each of the statistics corresponds to one image, one subband or pair of subbands, and one window in that subband. The window can be local or global (the entire subband). The statistics are typically the mean, variance, horizontal and vertical autocorrelation. Crossband correlations between subbands at adjacent scales for a given orientation and between all orientations for a given scale are also included; these are illustrated

(a) (b) (c) (a) (b) (c)

Figure 6. Texture Segmentation: (a) Original, (b) Chen *et al.*,[67] (c) He[69]

in Fig. 5 (b). The statistics are computed on the complex values or the magnitudes of the coefficients, depending on the particular statistic, and the particular metric implementation.[62]

Formulas for comparing statistics: These take different forms depending on the values that the particular statistics take and the particular metric implementation. The result is a non-negative value that represents the similarity or dissimilarity (depending on the particular metric implementation) score for the particular statistic.

A pooling strategy for combining similarity or dissimilarity scores across statistics, subbands, and window positions in order to obtain an overall STSIM value.

Zujovic *et al.* proposed two different STSIM implementations, STSIM-2 and STSIM-M.[62] The STSIMs are consistent with the FRF model. The three FRF stages correspond to the subband decomposition (linear), the statistic comparisons (nonlinear), and the pooling (linear).

The STSIMs we described above, are applied to grayscale images or to the grayscale component of a color image. For color images, Zujovic *et al.* combine the grayscale metric value with a metric that compares the color compositions of two image regions.[61,65,66] They thus separate the color composition of a texture from the spatial structure, which to a large extend is reflected in the grayscale component. Note that this does not account for any pure chrominance structure, which is not typical for most natural textures.[65]

6. TEXTURE SEGMENTATION

Our interest is in image segmentation techniques that can partition an image into perceptually uniform regions based on low-level features, not relying on any semantics. In this respect, they correspond to preattentive texture discrimination (and the resulting segregation) discussed in Section 3. In principle, the texture similarity metrics we discussed in the previous section could be used for texture segmentation. However, in natural scenes, perceptually uniform regions have spatially varying statistical characteristics. Thus, segmentation must be based on simple texture models and similarity metrics that can adapt to local variations.

We consider two color-texture segmentation algorithms that rely on simple features inspired by human perception. Chen *et al.*[67] used spatially adaptive dominant colors and the associated percentages to represent the color composition of the texture in the vicinity of each pixel, and dominant orientations (based on a steerable filter decomposition) to represent the texture structure in the vicinity of each pixel. The color composition comparisons were based on the optimal color composition distance (OCCD).[68] This algorithm provides good segmentation performance, albeit at a high computational cost.

He[69] proposed a more efficient approach that relies on the fact that perceptually uniform natural textures are typically characterized by one or two dominant colors to obtain a much simpler texture model. This considerably simplifies the OCCD metric. He then used a *feature-aligned* clustering approach to segment the image into perceptually uniform regions. The algorithm is computationally efficient without any performance sacrifices relative to Ref. 67. Examples of the two algorithms are shown in Fig. 6.

7. APPLICATIONS OF TEXTURE SIMILARITY

7.1 Structurally Lossless Image Compression

The goal of traditional, mathematically lossy, image compression techniques is to exploit image redundancy without significant sacrifices in perceived image quality. To accomplish this, they rely on transform-based techniques (DCTs, subband/wavelet decompositions) and point-by-point image fidelity metrics, both of which are not well-suited for exploiting redundancies in texture areas. Transform-based techniques are quite inefficient in textured areas, while, as we saw, traditional fidelity metrics are not well-suited for predicting perceptual texture similarity. The gold standard is *perceptually lossless* compression (see also Section 3), whereby the original and compressed image are indistinguishable in a side-by-side comparison. Instead of transform-based compression, the key to making gains in texture compression is exploiting texture self-similarity. This requires STSIM metrics that allow substantial point-by-point differences in textured regions that are perceptually similar. However, this necessitates replacing the goal of perceptually lossless compression with *structurally lossless* compression, whereby there may be visible differences between the original and the compressed image, even though neither of the images appears distorted, and both could be considered as original images.[70] Similar ideas have been explored in graphics; for example, Ferwerda *et al.*[71] introduced the notion of *visual equivalence,* whereby two "images convey the same impressions of scene appearance, even if they are visibly different."

Based on the above ideas, Jin *et al.* proposed a new structurally lossless compression technique, which they call matched-texture coding (MTC).[72] The main idea is that MTC encodes textured image patches by pointing to previously encoded perceptually similar patches, while the rest of the image (non-textured regions) is encoded with a baseline method, such as JPEG. A number of other approaches have been proposed for exploiting texture in image compression. The most common approach is to replace textured image regions with synthesized texture. One of the earliest such techniques was proposed by Popat and Picard.[46] Another technique, utilizing the Heeger and Bergen model for texture synthesis,[48] was presented a few years later at HVEI by Yoon and Adelson,[73] while Haenselmann and Effelsberg, also at HVEI,[74] presented a texture resynthesis approach based on principle component analysis. At the 2008 HVEI conference, Kramm presented a texture analysis/synthesis based compression technique that exploits redundancies in groups of images containing textures.[75] Recently, the topic of compression-by-synthesis has seen a lot of activity, e.g., in Refs. 76–80.

7.2 Content-Based Retrieval

In his 2001 keynote presentation at HVEI, Adelson emphasized the importance of texture in material perception, and contrasted it to the focus of the computer vision community on object extraction, as opposed to material perception.[5] When we combine this with Koenderink's suggestion, in his 2011 keynote presentation, that the human visual system searches for "clues" for content extraction,[81] we realize that an important clue is provided by texture. Another way to think about this is that the human visual system can recognize an object or scene from its visible parts and their spatial relations;[82,83] each part and spatial relation provides a clue. Now, in order to implement this idea, we need to segment the objects in a scene into regions of perceptually uniform textures, using one of the algorithms we reviewed in Section 6. The material clues can then be obtained by semantic analysis of the texture segments, or by direct comparison with reference textures using a texture similarity metric. There are two problems of interest.

The first is the retrieval of *identical* textures,[62] whereby the goal is finding samples from the same, perceptually uniform, texture. An example of identical textures is shown in Fig. 7(f). This is important for retrieving textures from a database of fabrics, rugs, marble tiles, as well as any other patterns of interest. It is also important when searching for images that correspond to a particular scene (for which the component textures are known) or images of scenes that contain characteristic textures (a brick wall, granite tiles, concrete floor, etc.). To get a database for this problem, all that is necessary is to cut patches from perceptually uniform textures. Then, all the patches that come from the same perceptually uniform texture form a class of identical textures.[62] Thus, the ground truth follows from database construction. This obviates the need for subjective tests and allows testing on a large number of images. All that is needed is to make sure that we start with a set of perceptually uniform textures, and to make sure that they are distinct from each other.

The second problem is the retrieval of *similar* textures.[65,66] This is a difficult problem, because texture similarity can be defined on the basis of several attributes: directionality, scale, regularity, color, shape of generating elements, etc. Thus, when two texture pairs are dissimilar, for example the pairs in Fig. 7 (a) and (b), it is difficult for subjects to make consistent judgments about their relative similarity. A similar problem arises when textures are similar on the basis of one attribute and different on the basis of another. For example, the textures in Fig. 7(c) have similar structure (scale, shape of structuring element) but different colors, while the textures in Fig. 7(d) have similar colors but different structure. In such cases, one subject may consider

(a) (b) (c) (d) (e) (f)

Figure 7. Examples of texture pairs (a),(b) dissimilar textures; (c) similar structure, different color; (d) similar color, different structure; (e) similar in all perceptual dimensions; (f) "identical"

the pair similar and another dissimilar, depending on the attribute they focus on. It is only when textures are similar along all perceptual dimensions, as in the textures in Fig. 7(e), that subjects consistently classify them as similar. As we will see in Section 8, a reasonable question to ask is whether two textures are similar or dissimilar. If they are dissimilar, it is difficult, if not impossible to further quantify similarity.

7.3 Visual to tactile and auditory mapping

The limited spatial resolution of the touch necessitates a segmentation-based approach for transforming visual to tactile information. The idea is illustrated in Figure 8. In order to transform an array of image intensities into a pattern of bumps to be displayed on a tactile device, we first segment the image into perceptually uniform segments, and then we map each segment into a distinct tactile pattern.[84] Each black dot in Fig. 8 (c) represents a bump, typically with a fixed radius and height. The mapping of the visual textures of the segments into tactile textures can be arbitrary, or can be based on some intuitive mapping between the two modalities (e.g., on the basis of the roughness, directionality, or regularity of the patterns). In many cases the visual information is already available in segmented form, for example, in maps, pie charts, and other graphics; in such cases, the segmentation step is not needed. Finally, if the system is to work, the brain must be able to integrate all the information into a mental picture, which necessitates an understanding of haptic space perception.

8. DOMAINS FOR TESTING TEXTURE SIMILARITY METRICS

In Section 7.2 we saw that it does not make sense to quantify similarity when textures are dissimilar. Indeed, Zujovic *et al.*[65,66] found that, apart from a relatively narrow range at the top of the similarity scale, where images exhibit similarity in every texture attribute, subjects do not make consistent judgments about the relative similarity of texture pairs, except for distinguishing similar from dissimilar pairs. Since our goal is to design metrics that mimic human perception, we should not expect metrics to do what humans cannot do.

Now, we look at the application requirements. In image compression, it is desirable to have a monotonic relationship between measured and perceived similarity. However, this is important only when the images are fairly similar; when they are dissimilar, it should be sufficient that the metric value be low. Another important requirement is that the metric provide consistent values for different image content, so that a uniform quality

(a) (b) (c)

Figure 8. Visual to tactile mapping. (a) Original color image (b) Segmentation (c) Tactile display

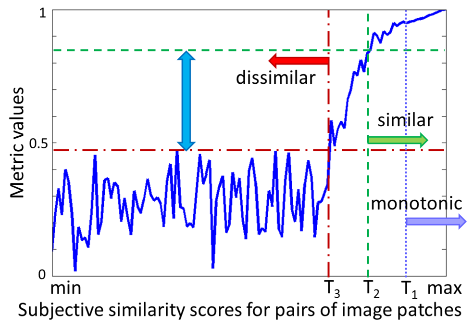


Figure 9. Desirable metric behavior (metric values vs. subjective similarity scores)

criterion can be established. In CBR, as we discussed in Section 7.2, it is important to distinguish between similar and dissimilar pairs. Ordering images according their similarity to a query image may also be desired, but again, this is important only at the high end of the scale, i.e., for the most similar images. For the remainder of the images, it is sufficient to reject them as dissimilar. As in the compression case, there may also be a need for thresholds that establish whether two textures are sufficiently similar. Thus, in both applications, similarity needs to be quantified only over a small range at the top of the similarity scale, while for the rest of the range, it is sufficient to declare that texture pairs are dissimilar. Finally, the determination whether two textures are identical is important for both applications.

Thus, both human perception and application requirements agree that a monotonic relationship between metric values and subjective ratings is desired only in the region of very similar textures, and that otherwise, the metric should be able to distinguish between similar and dissimilar texture pairs. Figure 9 provides a schematic illustration of the desirable metric behavior.[65,66] It plots subjective similarity ratings versus objective metric values. Observe that the monotonic relationship is limited to the region of very similar textures – to the right of T_1, which includes identical textures. There is also a similar range – to the right of T_2 – where subjects agree that textures are similar but do not assign consistent similarity scores. Metric values should be high in this range, but monotonic behavior is not expected. Finally, there is a region of *dissimilar* textures – to the left of T_3) – where the subjects agree that the textures are dissimilar but cannot assign consistent similarity scores. Metric values should be low in this range.

In summary, there are three operating domains where a similarity metric can be tested:

1. The metric ability to differentiate perceptually similar and dissimilar textures.
2. The existence of a monotonic region at the top of the scale.
3. The metric ability to retrieve identical textures.

Each of these domains requires different metric testing procedures. The first domain is explored in Ref. 65, the second in Ref. 85, and the third in Ref. 62.

9. HIGH-LEVEL TEXTURE FEATURES

In Sections 7.2 and 8, we discussed several attributes (scale, regularity, color) that can be used for the perceptual characterization of texture. The identification of such attributes, most of which are high-level in the sense that they relate to semantics, has been an important topic of research. Several authors have proposed high-level attributes (or features) for the perception of visual texture. Tamura *et al.*,[86] identified six basic features: coarseness, contrast, directionality, line-likeness, regularity, and roughness. On the basis of subjective experiments, Rao and Lohse[87] identified three important perceptual dimensions for texture perception: repetitiveness versus irregularity, directional versus nondirectional, and structurally complex versus simple. Mojsilovic *et al.*[88] conducted subjective experiments with a special class of color patterns (fabrics and carpets), and identified five perceptual dimensions: overall color, directionality and orientation, regularity and placement rules, color purity, and pattern complexity and heaviness. There have also been several attempts, by these and other authors, to link such perceptual features to low-level image parameters.[86,88,89] However, the linking of low-level features to high-level semantics is a difficult problem that remains an active topic of research.

10. CONCLUSIONS

The past 25 years have seen a lot of advances in our understanding of texture perception and the development of image processing algorithms for exploiting this understanding. However, a lot of problems remain open. In addition, the development of new electronic imaging technologies (displays, mobile devices, communication networks), has opened up new promising avenues for research. One of the most promising research directions is the study of individual texture attributes, such as surface reflectance (gloss)[90–92] and roughness.[93] Other attributes, such as directionality, regularity/periodicity, are also interesting to explore. One of the lessons we learned from the past 25 years is the importance of image statistics in tackling these problems. Of course, the most important lesson is that a better understanding of texture is key to further advances in image analysis, compression, and semantic extraction, as well as electronic imaging in general.

REFERENCES

[1] Portilla, J. and Simoncelli, E. P., "A parametric texture model based on joint statictics of complex wavelet coefficients," *Int. J. Computer Vision* **40**, 49–71 (Oct. 2000).

[2] Landy, M. S., "Texture perception," in [*Encyclopedia of Neuroscience*], Adelman, G. and Smith, B. H., eds., Elsevier, Amsterdam, 3rd, on the www ed. (2004).

[3] Bergen, J. R., "Theories of visual texture perception," in [*Spatial Vision*], Regan, D., ed., *Vision and Visual Dysfunction* **10**, 114–134, CRC Press, Cambridge, MA (1991).

[4] Landy, M. S. and Graham, N., "Visual perception of texture," in [*The Visual Neurosciences*], Chalupa, L. M. and Werner, J. S., eds., 1106–1118, MIT Press, Cambridge, MA (2004).

[5] Adelson, E. H., "On seeing stuff: The perception of materials by humans and machines," in [*Human Vision and Electronic Imaging VI*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **4299**, 1–12 (Jan. 2001).

[6] Adelson, E. H., "Image statistics and surface perception," in [*Human Vision and Electronic Imaging XIII*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **6806**, 680602–1–9 (Jan. 28–31 2008).

[7] Rushmeier, H., "The perception of simulated materials," in [*Human Vision and Electronic Imaging XIII*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **6806**, 680603–1–12 (Jan. 28–31 2008).

[8] Landy, M. S., "Combining multiple cues for texture edge localization," in [*Human Vision, visual processing and digital display IV*], Rogowitz, B. E. and Allebach, J. R., eds., *Proc. SPIE* **1913**, 506–517 (1993).

[9] Rosenholtz, R., "General-purpose localization of textured image regions," in [*Human Vision and Electronic Imaging IV*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **3644**, 454–460 (Jan. 1999).

[10] Rushmeier, H., Rogowitz, B. E., and Piatko, C., "Perceptual issues in substituting texture for geometry," in [*Human Vision and Electronic Imaging V*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **3959**, 372–383 (Jan. 24–27 2000).

[11] Interrante, V. and Kim, S., "Investigating the effect of texture orientation on the perception of 3D shape," in [*Human Vision and Electronic Imaging VI*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **4299**, 330–339 (Jan. 2001).

[12] Müller, W. and Hildebrand, A., "Shape from texture – an evaluation of visual cues," in [*Human Vision, Visual Proc., and Digital Display V*], Rogowitz, B. E. and Allebach, J. P., eds., *Proc. SPIE* **2179**, 220–234 (Feb. 1994).

[13] Ulichney, R., [*Digital Halftoning*], The MIT Press, Cambridge, MA (1987).

[14] Ulichney, R., "The void-and-cluster method for dither array generation," in [*Human Vision, Visual Proc., Dig. Display IV*], Allebach, J.P. and Rogowitz, B.E., eds., **Proc. SPIE, Vol. 1913**, 332–343 (Feb. 1993).

[15] Pappas, T. N. and Neuhoff, D. L., "Model-based halftoning," in [*Human Vision, Visual Proc., Dig. Display II*], Rogowitz, B.E., Brill, M.H., Allebach, J.P., eds., **Proc. SPIE, Vol. 1453**, 244–255 (Feb. 1991).

[16] Pappas, T. N. and Neuhoff, D. L., "Printer models and error diffusion," *IEEE Trans. Image Processing* **4**, 66–80 (Jan. 1995).

[17] Pappas, T. N. and Neuhoff, D. L., "Least-squares model-based halftoning," in [*Human Vision, Visual Proc., and Digital Display III*], Rogowitz, B. E., ed., **Proc. SPIE, Vol. 1666**, 165–176 (Feb. 1992).

[18] Pappas, T. N. and Neuhoff, D. L., "Least-squares model-based halftoning," *IEEE Trans. Image Processing* **8**, 1102–1116 (Aug. 1999).

[19] Pappas, T. N., Allebach, J. P., and Neuhoff, D. L., "Model-based digital halftoning," *IEEE Signal Processing Mag.* **20**, 14–27 (July 2003).

[20] Mannos, J. L. and Sakrison, D. J., "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory* **IT-20**, 525–536 (July 1974).

[21] Daly, S., "Subroutine for the generation of a two dimensional human visual contrast sensitivity function," Technical Report 233203Y, Eastman Kodak, Rochester, NY (1987).

[22] Julesz, B., "AI and early vision – Part II," in [*Human Vision, Visual Proc., and Digital Display*], Rogowitz, B. E., ed., *Proc. SPIE* **1077**, 246–268 (Jan. 18–20 1989).

[23] Julesz, B., "Binocular depth perception of computer-generated patterns," *Bell System Technical Journal* **39**, 1125–1162 (1960).

[24] Julesz, B., "Visual pattern discrimination," *IRE Trans. Information Theory* **8**, 84–92 (Feb. 1962).

[25] Julesz, B., Gilbert, E. N., Shepp, L. A., and Frisch, H. L., "Inability of humans to discriminate between visual textures that agree in second-order statistics – revisited," *Perception* **2**(4), 391–405 (1973).

[26] Julesz, B., Gilbert, E. N., and Victor, J. D., "Visual discrimination of textures with identical third-order statistics," *Biological Cybernetics* **31**(3), 137–140 (1978).

[27] Caelli, T. and Julesz, B., "On perceptual analyzers underlying visual texture discrimination: part I," *Biological Cybernetics* **28**(3), 167–175 (1978).

[28] Julesz, B., "Textons, the elements of texture perception and their interactions," *Nature* **290**, 91–97 (1981).

[29] Voorhees, H. and Poggio, T., "Computing texture boundaries," *Nature* **333**, 364–367 (May 1988).

[30] Bergen, J. R. and Adelson, E. H., "Early vision and texture perception," *Nature* **333**, 363–364 (May 1988).

[31] Malik, J. and Perona, P., "Preattentive texture discrimination with early vision mechanisms," *Journal of the Optical Society of America A* **7**, 923–932 (1990).

[32] Porat, M. and Zeevi, Y. Y., "Localized texture processing in vision: Analysis and synthesis in Gaborian space," *IEEE Trans. Biomed. Eng.* **36**(1), 115–129 (1989).

[33] Simoncelli, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J., "Shiftable multi-scale transforms," *IEEE Trans. Inform. Theory* **38**, 587–607 (Mar. 1992).

[34] Valois, R. L. D. and Valois, K. K. D., [*Spatial Vision*], Oxford University Press, New York (1990).

[35] Esteban, D. and Galand, C., "Application of quadrature mirror filters to split band voice coding schemes," in [*Proceedings ICASSP*], 191–195 (1978).

[36] Vaidyanathan, P. P., "Quadrature mirror filter banks, M-band extensions and perfect-reconstruction techniques," *IEEE ASSP Mag.* , 4–20 (July 1987).

[37] Cox, R. V., "The design of uniformly and nonuniformly spaced pseudoquadrature mirror filters," *IEEE Trans. Acoust., Speech, Signal Processing* **34**, 1090–1096 (Oct. 1986).

[38] Safranek, R. J. and Johnston, J. D., "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in [*Proc. ICASSP-89*], **3**, 1945–1948 (May 1989).

[39] Safranek, R. J., Johnston, J. D., and Rosenholtz, R. E., "A perceptually tuned sub-band image coder," in [*Proc. SPIE, Human Vision and Electronic Imaging: Models, Methods, and Applications*], Rogowitz, B. E. and Allebach, J. P., eds., **Proc. SPIE, Vol. 1249**, 284–293 (Feb. 1990).

[40] Peterson, H. A., Ahumada, Jr., A. J., and Watson, A. B., "An improved detection model for DCT coefficient quantization," in [*Human Vision, Visual Proc., and Digital Display IV*], Allebach, J. P. and Rogowitz, B. E., eds., **Proc. SPIE, Vol. 1913**, 191–201 (Feb. 1993).

[41] Watson, A. B., "DCT quantization matrices visually optimized for individual images," in [*Human Vision, Visual Proc., and Digital Display IV*], Allebach, J. P. and Rogowitz, B. E., eds., **Proc. SPIE, Vol. 1913**, 202–216 (Feb. 1993).

[42] Watson, A. B., Yang, G. Y., Solomon, J. A., and Villasenor, J., "Visual thresholds for wavelet quantization error," in [*Proc. SPIE, vol. 2657, Human Vision and Electronic Imaging*], 382–392 (Jan. 29–Feb. 1, 1996).

[43] Lubin, J. and Bergen, J. R., "Pattern discrimination in the fovea and periphery," *Investigative Ophthalmology and Visual Science Supplement* **32**, 1024 (1991).

[44] Lubin, J., "The use of psychophysical data and models in the analysis of display system performance," in [*Digital Images and Human Vision*], Watson, A. B., ed., 163–178, The MIT Press, Cambridge, MA (1993).

[45] Daly, S., "The visible differences predictor: an algorithm for the assessment of image fidelity," in [*Digital Images and Human Vision*], Watson, A. B., ed., 179–206, The MIT Press, Cambridge, MA (1993).

[46] Popat, K. and Picard, R. W., "Novel cluster-based probability model for texture synthesis, classification, and compression," in [*Proc. SPIE Visual Communications*], (1993).

[47] Cano, D. and Minh, T. H., "Texture synthesis using hierarchical linear transforms," *Signal Processing* **15**, 131–148 (1988).

[48] Heeger, D. J. and Bergen, J. R., "Pyramid-based texture analysis/synthesis," in [*Proc. Int. Conf. Image Processing (ICIP), vol. III*], 648–651 (Oct. 1995).

[49] Portilla, J., Navarro, R., Nestares, O., and Tabernero, A., "Texture synthesis-by-analysis based on a multi-scale early-vision model," *Optical Engineering* **35**(8), 2403–2417 (1996).

[50] Zhu, S. C., Wu, Y. N., and Mumford, D., "Modeling images and textures by minimax entropy," in [*Human Vision and Electronic Imaging II*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **3016**, 387–400 (Feb. 10–13 1997).

[51] De Bonet, J. S. and Viola, P. A., "A non-parametric multi-scale statistical model for natural images," *Adv. in Neural Info. Processing Systems* **9** (1997).

[52] Efros, A. A. and Freeman, W. T., "Image quilting for texture synthesis and transfer," in [*Proc. 28th Intl. Conf. Computer Graphics and Interactive Techniques (SIGGRAPH-01)*], 341–346 (Aug. 2001).

[53] Sayood, K., [*Introduction to Data Compression*], Morgan Kaufmann (2006).

[54] Navarro, R. and Portilla, J., "Robust method for texture synthesis-by-analysis based on a multiscale Gabor scheme," in [*Human Vision and Electronic Imaging*], Rogowitz, B. E. and Allebach, J. P., eds., *Proc. SPIE* **2657**, 86–97 (Jan. 29–Feb. 1 1996).

[55] Balas, B. J., "Texture synthesis and perception: Using computational models to study texture representations in the human visual system," *Vision Research* **46**, 299–309 (2006).

[56] Balas, B. J., Nakano, L., and Rozenholtz, R., "A summary-statistic representation in peripheral vision explains visual crowding," *Journal of Vision* **9**(12), 13, 1–18 (2009).

[57] Rosenholtz, R., "What your visual system sees where you are not looking," in [*Human Vision and Electronic Imaging XVI*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **7865**, 786510–1–14 (Jan. 24–27 2011).

[58] Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., and Ilie, L., "A summary statistic representation in peripheral vision explains visual search," *Journal of Vision* **12**(4), 14, 1–17 (2012).

[59] Freeman, J. and Simoncelli, E. P., "Metamers of the ventral system," *Nature Neuroscience* **14**, 1195–1204 (Sept. 2011).

[60] Zhao, X., Reyes, M. G., Pappas, T. N., and Neuhoff, D. L., "Structural texture similarity metrics for retrieval applications," in [*Proc. Int. Conf. Image Processing (ICIP)*], 1196–1199 (Oct. 2008).

[61] Zujovic, J., Pappas, T. N., and Neuhoff, D. L., "Structural similarity metrics for texture analysis and retrieval," in [*Proc. Int. Conf. Image Processing*], 2225–2228 (Nov. 2009).

[62] Zujovic, J., Pappas, T. N., and Neuhoff, D. L., "Structural texture similarity metrics for image analysis and retrieval," *IEEE Trans. Image Process.* . To appear.

[63] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.* **13**, 600–612 (Apr. 2004).

[64] Wang, Z. and Simoncelli, E. P., "Translation insensitive image similarity in complex wavelet domain," in [*IEEE Int. Conf. Acoustics, Speech, Signal Processing*], **II**, 573–576 (2005).

[65] Zujovic, J., Pappas, T. N., Neuhoff, D. L., van Egmond, R., and de Ridder, H., "A new subjective procedure for evaluation and development of texture similarity metrics," in [*Proc. IEEE 10th IVMSP Wksp.: Perception and Visual Signal Analysis*], 123–128 (June 2011).

[66] Zujovic, J., *Perceptual Texture Similarity Metrics*, PhD thesis, Northwestern Univ., Evanston, IL (Aug. 2011).

[67] Chen, J., Pappas, T. N., Mojsilovic, A., and Rogowitz, B. E., "Adaptive perceptual color-texture image segmentation," *IEEE Trans. Image Process.* **14**, 1524–1536 (Oct. 2005).

[68] Mojsilović, A., Hu, J., and Soljanin, E., "Extraction of perceptually important colors and similarity measurement for image matching, retrieval, and analysis," *IEEE Tr. Image Proc.* **11**, 1238–1248 (Nov. 2002).

[69] He, L., *A Clustering Approach for Color Texture Segmentation*, PhD thesis, Northwestern Univ., Evanston, IL (Aug. 2012).

[70] Pappas, T. N., Zujovic, J., and Neuhoff, D. L., "Image analysis and compression: Renewed focus on texture," in [*Visual Information Processing and Communication*], *Proc. SPIE* **7543** (Jan. 2010).

[71] Ramanarayanan, G., Ferwerda, J., Walter, B., and Bala, K., "Visual equivalence: towards a new standard for image fidelity," in [*ACM SIGGRAPH 2007*], ACM, New York, NY, USA (2007).

[72] Jin, G., Zhai, Y., Pappas, T. N., and Neuhoff, D. L., "Matched-texture coding for structurally lossless compression," in [*Proc. Int. Conf. Image Processing (ICIP)*], (Oct. 2012). Accepted.

[73] Yoon, S. Y. and Adelson, E. H., "Subband texture synthesis for image coding," in [*Human Vision and Electronic Imaging III*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **3299**, 489–497 (Jan. 1998).

[74] Haenselmann, T. and Effelsberg, W., "Texture resynthesis using principle component analysis," in [*Human Vision and Electronic Imaging VII*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **4662**, 440–447 (Jan. 2002).

[75] Kramm, M., "Image group compression using texture databases," in [*Human Vision and Electronic Imaging XIII*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **6806**, 680613–1–10 (Jan. 2008).

[76] Ndjiki-Nya, P., Bull, D., and Wiegand, T., "Perception-oriented video coding based on texture analysis and synthesis," in [*Proc. Int. Conf. Image Processing (ICIP))*], 2273–2276 (Nov. 2009).

[77] Ierodiaconou, S., Byrne, J., Bull, D. R., Redmill, D., and Hill, P., "Unsupervised image compression using graphcut texture synthesis," in [*Proc. Int. Conf. Image Processing (ICIP))*], 2289–2292 (Nov. 2009).

[78] Balle, J., Stojanovic, A., and Ohm, J.-R., "Models for static and dynamic texture synthesis in image and video compression," *IEEE J. Sel. Topics Signal Process.* **5**, 1353–1365 (Nov. 2011).

[79] Bosch, M., Zhu, F., and Delp, E. J., "Segmentation-based video compression using texture and motion models," *IEEE J. Sel. Topics Signal Process.* **5**, 1366–1377 (Nov. 2011).

[80] Zhang, F. and Bull, D. R., "A parametric framework for video compression using region-based texture models," *IEEE J. Sel. Topics Signal Process.* **5**, 1378–1392 (Nov. 2011).

[81] Koenderink, J., "Vision as a user interface," in [*Human Vision and Electronic Imaging XVI*], B. E. Rogowitz, T. N. P., ed., *Proc. SPIE* **7865** (Jan. 2011).

[82] Hoffman, D. D., [*Visual Intelligence: How We Create What We See*], W.W. Norton & Company, Inc. (1998).

[83] Palmer, S. E., [*Vision Science: Photons to Phenomenology*], MIT Press (1999).

[84] Pappas, T. N., Tartter, V., Seward, A. G., Genzer, B., Gourgey, K., and Kretzschmar, I., "Perceptual dimensions for a dynamic tactile display," in [*Human Vision and Electronic Imaging XIV*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **7240**, 72400K–1–12 (Jan. 2009).

[85] Zujovic, J., Pappas, T. N., Neuhoff, D. L., van Egmond, R., and de Ridder, H., "Subjective and objective texture similarity for image compression," in [*Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*], 1369–1372 (Mar. 2012).

[86] Tamura, H., Mori, S., and Yamawaki, T., "Textural features corresponding to visual perception," *IEEE Trans. Syst., Man, Cybern.* **8**, 460–473 (June 1978).

[87] Rao, A. R. and Lohse, G. L., "Towards a texture naming system: Identifying relevant dimensions of texture," *Vision Research* **36**(11), 1649–1669 (1996).

[88] Mojsilović, A., Kovačević, J., Hu, J., Safranek, R. J., and Ganapathy, S. K., "Matching and retrieval based on the vocabulary and grammar of color patterns," *IEEE Trans. Image Process.* **1**, 38–54 (Jan. 2000).

[89] Amadasun, M. and King, R., "Textural features corresponding to texture properties," *IEEE Trans. Syst., Man, Cybern.* **19**, 1264–1274 (1989).

[90] Motoyoshi, I., Nishida, S., Sharan, L., and Adelson, E. H., "Image statistics and the perception of surface qualities," *Nature* **447**, 206–209 (May 2007).

[91] Sharan, L., Li, Y., Motoyoshi, I., Nishida, S., and Adelson, E. H., "Image statistics for surface reflectance perception," *Journal of the Optical Society of America A* **25**(4), 846–865 (2008).

[92] Wijntjes, M. W. A. and Pont, S. C., "Illusory gloss on Lambertian surfaces," *J. Vision* **10**(9), 1–12 (2010).

[93] Egmond, R. V., Lemmens, P., Pappas, T. N., and de Ridder, H., "Roughness in sound and vision," in [*Human Vision and Electronic Imaging XIV*], Rogowitz, B. E. and Pappas, T. N., eds., *Proc. SPIE* **7240**, 72400B–1–12 (Jan. 2009).