# Image Analysis: Focus on Texture Similarity

Thrasyvoulos N. Pappas, *Fellow, IEEE,* David L. Neuhoff, *Fellow, IEEE,* Huib de Ridder,
Jana Zujovic, *Member, IEEE*

***Abstract*—Texture is an important visual attribute both for human perception and image analysis systems. We review recently proposed texture similarity metrics and applications that critically depend on such metrics, with emphasis on image and video compression and content-based retrieval. Our focus is on natural textures and structural texture similarity metrics (STSIMs). We examine the relation of STSIMs to existing models of texture perception, texture analysis/synthesis, and texture segmentation. We emphasize the importance of signal characteristics and models of human perception, both for algorithm development and testing/validation.**

***Index Terms*—Structural similarity metrics, structurally lossless compression, matched-texture coding**

## I. Introduction

THE field of image analysis has made significant strides during the last two decades, incorporating sophisticated signal processing techniques and models of human perception. One of the keys to further advances is a better understanding of texture, and in particular, texture similarity. Even though the importance of texture for human perception and image analysis is obvious, it is surprising that it has received relatively little attention in applications such as image compression, restoration, content-based retrieval (CBR), and computer vision. For example, image and video compression techniques have relied on similarity metrics that are sensitive to point-by-point deviations and thus cannot adequately model the stochastic nature of texture and how it is perceived by humans [1], [2]. Similarly, computer vision has mostly focused on object extraction rather than material perception, which is critically dependent on texture [3]. Texture analysis is important for a variety of other applications, including graphics, multimodal interfaces, and sense substitution (visual to acoustic-tactile conversion [4]). The focus of this paper is on texture similarity and metrics that incorporate knowledge of human perception. We also look at applications that make use of texture similarity metrics. Our discussion primarily considers natural textures.

First, we look at image and video compression. Compression techniques rely on image similarity metrics (typically called quality metrics), both as embedded system components that help make decisions on an image-patch-by-image-patch basis, or in order to evaluate overall system performance (e.g., by pooling metric values over a set of patches). The signal processing community has identified exploiting texture as a key to increasing compression efficiency [5]–[10]. However, one of the main obstacles in realizing this goal has been the lack of metrics that can adequately predict perceptual texture similarity, which is essential for reducing redundancy in textured regions. To accomplish this, we need metrics that allow substantial (visible upon careful observation) point-by-point differences in textured regions that appear virtually the same. This necessitates replacing the goal of perceptually lossless compression, whereby the original and compressed image are indistinguishable, with what we have called structurally lossless compression, whereby two images (or texture patches) can have visible point-by-point differences, even though neither one of them appears to be distorted and both could be considered as original images [7]. Accordingly, the texture similarity metrics that accomplish this goal are called structural texture similarity metrics (STSIMs).

Image restoration also relies on image similarity metrics, both as embedded system components and for overall performance evaluation. State-of-the-art nonlocal restoration techniques rely heavily on image self-similarity (similar patches typically occur in several image locations) to reconstruct a cleaner, more accurate overall image [11]–[13]. However, current algorithms, which are based on point-by-point similarity metrics, can exploit similarities of smooth and piecewise smooth patches, but cannot handle patches that consist of, or contain, textures.

In CBR, one may identify a number of problems. The simplest is retrieving similar patches of uniform texture. The role of texture similarity metrics is obvious in this case. A more complicated problem is finding images that correspond to the same scene as that depicted in the query image. Given the wide variations in image content (due to the details of the scene arrangement, or lighting, perspective, scale, etc.), it is unrealistic to expect that one can match complex images directly. A more tractable approach is to segment the image into perceptually uniform regions [14], [15], and then directly compare the textures of the regions. A somewhat easier alternative is to use incremental parsing to obtain image patches that are representative of image content [16],

Manuscript received June 14, 2012; revised March 6, 2013; accepted April 23, 2013. Date of publication July 9, 2013. This work was supported in part by the U.S. Department of Energy National Nuclear Security Administration (NNSA) under Grant No. DE-NA0000431. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NNSA.

T. N. Pappas is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208 USA (e-mail: pappas@eecs.northwestern.edu).

D. L. Neuhoff is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: neuhoff@umich.edu).

H. de Ridder is with the Faculty of Industrial Design Engineering, Delft University of Technology, Delft 2628 CE, The Netherlands (e-mail: h.deridder@tudelft.nl).

J. Zujovic was with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL USA. She is now with FutureWei Technologies, Santa Clara, CA 95050 USA (phone: 408-330-4736, e-mail:jana.ehmann@huawei.com).

[17]. If two images share a substantial number of similar segments or patches, then the chances of a good match are much higher. Texture similarity metrics are required for both segmentation and incremental parsing, as well as comparisons of the resulting segments or patches.

Image segmentation, in particular, imposes special requirements on texture similarity metrics. This is because the statistical characteristics of perceptually uniform regions in images of natural scenes are spatially varying due to variations in illumination, perspective view, and variations in surface properties. Thus, segmentation must be based on simple texture models and similarity metrics that can adapt to local variations.

The most difficult CBR problem is the one that is based on semantics. Texture analysis plays a critical role in this case, too, providing important "clues" [18] for content extraction. As mentioned above, one of the problems with traditional computer vision techniques has been the focus on object extraction. On the other hand, the psychophysics literature has established that the human visual system (HVS) relies on partial information about the presence of objects (visible parts and their spatial relations) in order to arrive at scene interpretation [19], [20]. Such information can be provided by segments of perceptually uniform texture and their descriptors (color and texture, as well as boundary shape, location, size, and common boundaries) [21]. Incremental parsing [16] can also be used to identify the most commonly occurring (rectangular) segments and the associated descriptors (color, texture, location, size). In either case, the segment descriptors can be used as medium-level descriptors for the extraction of semantic information. Alternatively, direct comparisons of segment textures with reference textures, utilizing a texture similarity metric, can also facilitate context analysis.

The development of STSIMs has been inspired by the introduction of the structural similarity metrics (SSIM) [22], which represent the first serious attempt to deviate from point-by-point comparisons, and by research in texture analysis-synthesis [23]. Both rely on steerable filter decompositions [24] as models of early visual processing. We will explain the limitations of SSIMs, including the complex wavelet domain implementation (CW-SSIM) [25], and discuss a new framework for structural texture similarity metrics [2], [26] that, like texture analysis-synthesis, rely on subband statistics, to completely eliminate point-by-point comparisons. We also look at how these metrics can be evaluated in the context of different applications.

In the remainder of this paper, Section II discusses texture in general and Section III reviews structural texture similarity metrics. Compression and retrieval applications are discussed in Sections IV and V, respectively, while Section VI presents approaches for evaluating metric performance. The conclusions and future research can be found in Section VII.

## II. Texture Overview

The precise definition of visual texture is difficult. However, several authors (e.g., Portilla and Simoncelli [23]) loosely define texture as "texture images are spatially homogeneous and typically contain repeated structures, often" (but not necessarily) "with some random variation (e.g., random positions, size, orientations or colors)." It is even more difficult to identify features for the perceptual or mathematical characterization of texture. For a concise review on texture perception, we refer the reader to [27], and for more detailed reviews, to [28] and [29].

Several authors have attempted to identify features for texture classification. Tamura *et al.* [30], identified six "basic features" for the perception of visual texture: coarseness, contrast, directionality, line-likeness, regularity, and roughness. Rao and Lohse [31] conducted subjective experiments and analyzed the results to identify three important perceptual dimensions for texture perception: repetitiveness versus irregularity, directional versus nondirectional, and structurally complex versus simple. Mojsilović *et al.* [32] conducted a similar investigation for a special class of color patterns (fabrics and carpets), and found five perceptual dimensions: overall color, directionality and orientation, regularity and placement rules, color purity, and pattern complexity and heaviness. These and other authors have also attempted to link such high-level features to low-level image parameters, e.g., [30], [32], [33]. However, this is difficult, partly due to the fact that texture perception is linked to semantics. The focus of this paper is on the visual similarity of two textures and low-level parameters, which can be used to quantify this similarity without taking semantics into account.

The pioneering work of Bela Julesz in the 60s and 70s aimed at understanding the statistical properties of texture that determine preattentive (effortless, instantaneous) texture discrimination [34]–[36]. Julesz's hypothesis was that discrimination could be based on Nth-order statistics, and the initial Julesz conjecture was that textures with identical second-order statistics are preattentively indistinguishable [34], [35]. He and his colleagues later proved that the conjecture was wrong [36], [37], while Victor *et al.* worked on the exact characterization of the perceptually relevant statistics [38]–[40]. Julesz then went on to emphasize the importance of local features, which he called "textons" in preattentive texture perception [41], echoing similar ideas from Beck [42]. Voorhees and Poggio [43] showed that Julesz's ideas can be applied to natural images by proposing an algorithm for the detection and comparison of textons. On the other hand, Bergen and Adelson [44] showed that low-level mechanisms, consisting of linear filtering followed by a nonlinearity (rectification) and a second stage of linear filtering can go a long way towards explaining texture discrimination. Similar models have been proposed by Malik and Perona [45] and others (for a complete list see [29]), and led to the linear-nonlinear-linear (LNL) or filter-rectify-filter (FRF) model [29]. The first stage of the LNL model consists of a multiple scale and orientation frequency decomposition, like Gabor [46] and steerable [24] filters. The use of such decompositions is motivated by human perception, as models of early visual processing [47]. Portilla and Simoncelli [23] relied on such decompositions to obtain a statistical model for texture that is parametrized by statistics that are computed on single or pairs of subband coefficients at adjacent locations, scales, and orientations. The authors adopted an analysis/synthesis methodology to demonstrate that the resulting set of parameters is necessary and sufficient for perceptual equivalence of the original and synthesized textures.

Balas [48] carried this work further, conducting subjective tests to determine the importance of the different statistics of the Portilla-Simoncelli model for the reconstruction of natural textures, viewed under preattentive conditions. Finally, recent work by Balas *et al.* [49], Rosenholtz *et al.* [50], and Freeman and Simoncelli [51] have used the model to demonstrate the dominant role of texture perception in peripheral vision.

We will revisit the Portilla-Simoncelli model in more detail in the next section. The texture similarity metrics we discuss in the next section are inspired by this model and are consistent with the ideas of the LNL model.

## III. Texture Similarity Metrics

The development of objective metrics for texture similarity is considerably more challenging than that of traditional image quality metrics because there can be substantial point-by-point deviations between textures that according to human judgment are essentially identical. Note that the metrics we discuss do not comply with the mathematical definition of a metric, so we are using the term in a loose sense.

The focus of this section is on (general purpose) structural texture similarity metrics (STSIMs). However, we also look at texture analysis/synthesis, which makes use of an implicit texture similarity metric, and image segmentation, which makes use of much simpler similarity metrics. As we will see, there is a progression, from very precise to crude metrics.

### A. Texture Analysis/Synthesis

Among several approaches for texture analysis/synthesis, we look at those that are based on multiple scale and orientation frequency decompositions. Heeger and Bergen [52] demonstrated the power of such decompositions (steerable filters) by showing that a simple technique (histogram matching), when applied in the appropriate (subband) domain, can lead to impressive texture synthesis results. However, their technique can only model and synthesize stochastic homogeneous textures. Portilla and Simoncelli [23], who also based their model on the steerable filter decomposition, used a more elaborate statistical model to synthesize a wide variety of textures. To ensure that the histogram of the synthesized texture is close to the original, they included histogram statistics – range, mean, variance, skewness, and kurtosis. The regularity of textures is captured by the subband auto-correlation statistics, while the crossband correlations represent "higher order" texture features such as edges, corners, and bars. This yields a total of 846 parameters for texture synthesis. Examples are shown in Fig. 1. Note that the reconstruction works well for the first four examples. As the authors point out, the technique has trouble with textures that contain complex repeating structures (fifth example).

Since the Portilla-Simoncelli model parameters characterize a wide class of textures, they could form a basis for a texture similarity metric, e.g., as a weighted distance between the model parameters. Of course, for texture analysis/synthesis, the model parameters are the same, so there is no explicit use of a metric. However, as discussed below, a texture similarity metric can be based on a significantly smaller parameter set.

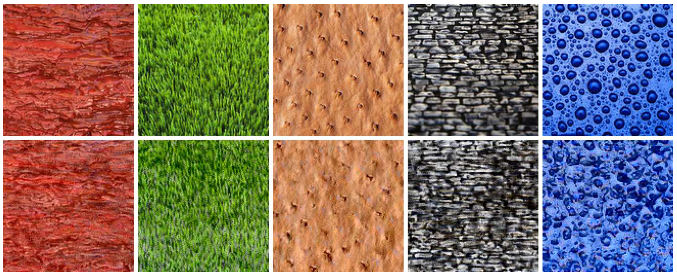


Fig. 1. Texture analysis/synthesis [23]: Original above, synthesized below

For the STSIMs we discuss next, Zujovic *et al.* [26], [53], [54] found that it is more effective to develop separate metrics for the spatial structure and the color composition of a texture. To isolate the structure of a texture, they used the grayscale component of the image. While this ignores chrominance structure, in most natural textures, the grayscale component is fairly representative of the overall structure [54]. In what follows we first review structural texture similarity metrics for the grayscale spatial structure, and then describe color composition metrics.

### B. Structural Texture Similarity Metrics - Grayscale

The development of STSIMs has been inspired, on the one hand by the introduction of SSIMs [22], [25], and on the other by research on texture analysis/synthesis [23]. One of the most important contributions of the SSIMs was the idea of replacing point-by-point comparisons with comparisons of region statistics. The goal was to give high similarity scores to images that are similar, even though they may have significant pixel-wise differences. Wang *et al.* have proposed a number of metrics, both in the space domain (SSIM) [22] and in the complex wavelet domain (CW-SSIM) [25]. However, these metrics still incorporate point-by-point comparisons between images (cross-correlations in the "structure" term), which causes them to give low similarity values to textures that are perceptually similar. In order to overcome such limitations, Zhao *et al.* [55] proposed STSIM-1, a metric that completely eliminates point-by-point comparisons by relying entirely on local image statistics. The idea was further developed by Zujovic *et al.* [53]. A comprehensive presentation of STSIMs can be found in [2]. In the following, we discuss a general framework for STSIMs and two specific metrics. An STSIM is specified by the following:

- *A subband decomposition.* The subband decomposition can be real or complex. Following Portilla and Simoncelli [23], the steerable filter decomposition is used, which like Gabor filters, is inspired by biological visual processing. A typical decomposition (used by the metrics we discuss below) has 14 subbands, corresponding to three scales, each with four orientations, plus a lowpass and a highpass subband, which are not subdivided into different orientations, as shown in Fig. 2.
- *A set of statistics,* each corresponding to one particular image, one particular subband or pair of subbands, and one particular window in that subband. The window can be local or global (the entire subband). Typical

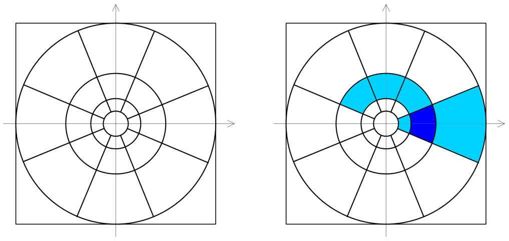


Fig. 2. Steerable filter decomposition (left), crossband correlations of dark blue band with each of the light blue bands (right)

statistics are the mean, variance, horizontal and vertical autocorrelation, and crossband correlation. For the latter, only the correlations between subbands at adjacent scales for a given orientation and between all orientations for a given scale are included, as illustrated in Fig. 2 (right). All expectations are computed as empirical averages within a window. Statistics can be computed on the complex values or the magnitudes of the coefficients. More details can be found in [2].

- *Formulas for comparing statistics,* which can take different forms depending on the values that the particular statistics take and may include statistic-dependent normalization factors. The result is a non-negative value that represents the similarity (or dissimilarity) 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.

Note that the “structure” term of the SSIM [22] is not a statistic comparison formula because it does not compute or compare statistics, one from each image. However, it is combined with similarity scores (the “luminance” and “contrast” terms) to obtain the overall SSIM value, causing the metric to sometimes take negative values.

Note also that the STSIMs follow the LNL model, with three basic stages: subband decomposition (linear), statistic comparisons (nonlinear), and pooling (linear). While there is strong perceptual justification for the overall LNL structure and selection of subband decomposition and set of statistics, the statistic comparison and pooling approaches remain quite ad hoc. We now discuss two specific metric embodiments.

### C. STSIM-2

The STSIM-2 metric [2], [26] uses the mean value, variance, and autocorrelations, computed on the complex subband coefficients, and the crossband correlation, computed on the magnitudes. The statistic comparison formulas for the mean, $c^1_{x,y}$, and variance, $c^2_{x,y}$, take the form (same as in the luminance and contrast term of the SSIMs):

$$c^i_{x,y} = \frac{2A_xA_y + C}{A_x^2 + A_y^2 + C} \tag{1}$$

where $A_x$ and $A_y$ are the statistics for images $x$ and $y$ and $C$ is a small positive constant that is included so that when the statistics are small the term will be close to 1. Since the correlation coefficients are bounded and their values lie in the unit circle of the complex plane (in contrast to the variances and the means), the comparison terms take a different form [55]):

$$c^i_{x,y} = 1 - 0.5|A_x - A_y|^p \tag{2}$$

for the corresponding statistics $A_x$ and $A_y$. Typically, $p = 1$. Both types of comparison terms produce a number in the interval $[0, 1]$, with 1 representing the highest possible similarity.

For the pooling, for each window and each subband, the similarity scores corresponding to the four single-band statistics are combined multiplicatively into one score for each subband and window location

$$q_0(x,y) = (c^1_{x,y})^{\frac{1}{4}}(c^2_{x,y})^{\frac{1}{4}}(c^3_{x,y})^{\frac{1}{4}}(c^4_{x,y})^{\frac{1}{4}} \tag{3}$$

where $c^1_{x,y}$, $c^2_{x,y}$, $c^3_{x,y}$, and $c^4_{x,y}$ are the similarity scores for the means, variances, horizontal, and vertical correlations, respectively. The resulting similarity values can then be pooled across subbands and window locations, in either order. Since the crossband correlation comparison terms involve two subbands, they are added separately.

### D. STSIM-M

This takes an entirely different approach for statistic comparisons and pooling [2]. For each window, it forms a feature vector that contains all the subband statistics and computes the distance between feature vectors. The statistics it uses are the mean, variance, horizontal and vertical autocorrelations, and the crossband correlations, all computed on the magnitudes of the subband coefficients. For the chosen steerable filter decomposition, this results in a total of 82 terms.

One of the advantages of this approach is that it can incorporate different weights for different statistics, depending on the application and database characteristics. For example, one can compute the Mahalanobis distance between the feature vectors, which if we assume that the different features are mutually uncorrelated, is the weighted Euclidean distance of the feature vectors with weights inversely proportional to the variance of each feature over the entire database. We refer to the resulting metric as STSIM-M, where “M” stands for Mahalanobis. Note that this is a dissimilarity metric that takes values between 0 and $\infty$, with 0 indicating highest similarity. This metric is better suited for comparing entire images or relatively large image patches [2].

### E. Color Composition Similarity Metrics

The most sophisticated approaches for measuring color similarity are based on dominant colors and the associated percentages, and have been used for texture retrieval and segmentation [14], [32], [56]. The idea of dominant colors is based on the fact that the human visual system cannot simultaneously perceive a large number of colors. Chen *et al.* added spatial adaptation in order to account for the nonuniformity of the statistical characteristics of natural textures [14]. The spatially adaptive dominant colors are obtained with the adaptive clustering algorithm (ACA) [57].

Zujovic *et al.* [26], [53] used the texture representation in terms of spatially adaptive dominant colors and associated percentages as the basis for color composition similarity

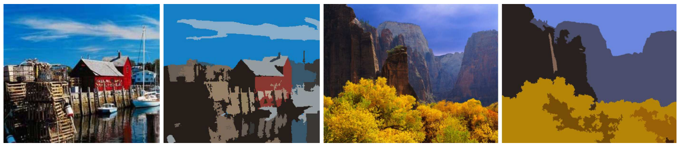


Fig. 3. Texture segmentation: Original, He [15]

metrics. As in [14], they used the Optimal Color Composition Distance (OCCD) to compare the color composition of two texture images or patches. OCCD breaks the histogram of dominant colors into fixed percentage units, finds the optimal mapping between these units, and computes the average distance between them in the *CIE L*a*b** color space [58]. This is essentially the same as the Earth mover's distance (EMD) [59].

### F. Metrics for Segmentation

As discussed in the introduction, image segmentation into perceptually uniform regions also requires texture similarity metrics. However, perceptually uniform regions of natural scenes typically have spatially varying statistical characteristics. This requires texture models with a few parameters that can be estimated from small windows, thus adapting to local variations. The segmentation algorithm exploits the fact that the spatial characteristics of the textures vary slowly within each segment and rapidly across segment boundaries. Thus, a simple similarity metric can be effective, in contrast to texture retrieval, where the metric needs to discriminate among several different textures, and texture analysis/synthesis, where small parameter changes may generate different textures.

Two recently proposed color-texture segmentation approaches utilize compact texture representations. The algorithm by Chen *et al.* [14] introduced the spatially adaptive dominant colors we mentioned above (typically four) and the percentage of each color in the neighborhood of each pixel, as well as dominant orientations (smooth, horizontal, vertical, $+45^o$, $-45^o$, and complex). The dominant orientations are based on the local median energy of the subband coefficients of the steerable filter decomposition [24]. The algorithm then uses OCCD [58] to compare the dominant colors and a simple metric for the dominant orientations [14]. This approach produces good segmentations, but the computational cost is quite high, due to the median filtering and the OCCD computation.

A more efficient approach [15] relies on a simpler texture model that exploits the fact that perceptually uniform natural textures are in the majority of cases characterized by one or two dominant colors. This was empirically observed by Depalov and Pappas in [60]. The theoretical justification for the approach in [15] is based on the assumption that the color reflected from a uniformly colored object has a fixed chrominance, and varies only in luminance. This results in a much simplified OCCD metric for color similarity, and a feature-aligned clustering approach to segmentation that is computationally efficient without any performance sacrifices relative to [14]. Examples are shown in Fig. 3.

### G. Other Approaches

There is a variety of other techniques for evaluating texture similarity. Some of these have been quite effective in the context of clustering/classification and segmentation tasks. They can be grouped into statistical and spectral methods. The statistical methods are based on calculating statistics of the gray levels in the neighborhood of each pixel (co-occurrence matrices, first and second order statistics, random field models, etc.) and then comparing the statistics of one image to those of another, while the spectral methods utilize the Fourier spectrum or a subband decomposition to characterize and compare textures. We review the most recent and most effective.

A very simple yet effective statistical technique for texture classification was proposed by Ojala *et al.* [61]; it utilizes local binary patterns (LBP) to characterize textures. An important shortcoming, however, is that it is highly localized.

The spectral methods provide a better link to human perception. The most effective rely on the energies of different subbands as features for texture segmentation, classification, and retrieval. Do and Vetterli [62] use wavelet coefficients as features and show that their distribution can be modeled as a generalized Gaussian density, which requires the estimation of two parameters. They then base the classification on the Kullback-Leibler distance between two feature vectors. Manjunath and Ma [63] use Gabor filters to model early HVS processing.

Finally, we want to mention MPEG-7 texture descriptors [64], which include the homogeneous texture descriptor that consists of the means and variances of the absolute values of the Gabor coefficients; the edge histogram descriptor that is based on local edge histograms; and the texture browsing descriptor that attempts to capture higher-level perceptual attributes such as regularity, directionality, and coarseness. Ojala *et al.* [61] have shown that the MPEG-7 descriptors are rather limited and provide only crude texture retrieval results.

The techniques we reviewed in this subsection have been shown to be quite effective in evaluating texture similarity in the context of clustering and segmentation tasks. However, there has been very little work towards evaluating their effectiveness in providing texture similarity scores that are consistent across texture content, agree with human judgments of texture similarity, and can be used in different applications.

## IV. Structurally Lossless Image Compression

State-of-the-art lossy image compression techniques generally rely on a transformation (DCT, subband or wavelet decomposition) to efficiently encode the image values. Such techniques work well in smooth regions, where the energy is concentrated in low-frequency coefficients, but are not as efficient in textured regions and transition regions (i.e., regions containing edges), where there is much energy in high frequencies. In video compression, transition regions can be efficiently encoded using motion compensation, whereby an image region (block) is encoded as the sum of a previously encoded block and a residual. However, due to the stochastic nature of texture, this approach has not worked well for textured regions because, first, it is difficult to find good matches with traditional similarity metrics, and second, even if a good match is found with a texture similarity metric (see Section III), the residual would either cost more bits to encode than starting from scratch or would result in significant distortion. Accordingly, the key to efficient compression of textured regions is to find previously encoded texture patches that can be reused with little modification and, consequently, very few bits. This requires a texture similarity metric that agrees well with perception. It also requires the use of texture blending techniques [65] to ensure that transitions between similar texture patches are not noticeable. The most obvious approach is to reuse texture as is, but simple transformations (scaling, perspective) can also be used.

As mentioned in the introduction, this requires a reconsideration of the goals of image compression. One commonly considered goal has been perceptually lossless compression in which, as we saw, the original and compressed image are visually indistinguishable in a side-by-side comparison. The development of perceptual similarity metrics to accomplish this goal [66], [67] was a significant advance over previous techniques that made only implicit use of HVS characteristics. However, in many applications severe bandwidth limitations dictate the need for further compression. With present methods, this can be done at the expense of significant compression artifacts or a significant reduction in image resolution [68]. Hemami *et al.* conducted systematic studies to quantify perceptual distortion in suprathreshold (visible artifacts) applications [69], [70]. In contrast, structurally lossless compression aims to exploit the perceptual redundancy of texture for substantial bitrate reductions that do not affect the visual quality and overall perception of the image. Exploiting the stochastic nature of texture and the human eye's insensitivity to its point-by-point variations, may result in substantial point-by-point changes that may be perceptible when the original and compressed images are viewed side-by-side, but are not noticeable when the reproduction is viewed by itself. In fact, the quality of the two images should be comparable, so that it is not obvious which is the original. 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".

### A. *Matched-Texture Coding*

In this subsection, we review a new compression method, called *matched-texture coding (MTC)* [72], whose goal is to obtain structurally lossless compression using the idea suggested above of encoding textured image patches by pointing to previously encoded perceptually similar patches. The rest of the image, i.e., the non-textured regions, is encoded with a baseline method, such as JPEG.

Accordingly, the key to this method is a texture similarity metric that enables good judgments to be made as to what constitutes good matches for textured patches. In addition to the metric and baseline coding method, key components of MTC are a procedure for identifying which candidates are to be tested, a similarity threshold, a mechanism for signifying the chosen candidate to the decoder, and a method for blending [65] the chosen candidate with the reproductions that surround it, so as to avoid blocking artifacts.

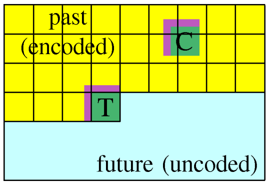


Fig. 4. Context for candidate (C) and target (T) blocks in side matching

There are two basic versions of MTC – Direct Matching and Side Matching – as well as combinations. In both, for a chosen $N$ the method successively encodes nonoverlapping $N \times N$ blocks. Given such a *target* block to be encoded, MTC first seeks a candidate in the already coded region of the image that sufficiently matches the target. If it does not find one, it divides the target into four square subblocks and recursively repeats the process on the subblocks. The smallest size for the target block is $16 \times 16$; if no suitable candidate can be found, then it is encoded with the baseline coder, e.g., JPEG applied to each $8 \times 8$ subblock.

*1) Direct Matching (DM):* DM-MTC operates like motion compensation in video coding. For each target block, it seeks the best candidate, as assessed by the STSIM metric, among all possible candidates in the previously encoded region of the image, and if a quality threshold is met, it encodes the location of this candidate.

*2) Side Matching (SM):* The key idea of SM-MTC is that if the pixels in the left and upper border (called the *context*) of a candidate in the already encoded region closely match the corresponding context of the target block, then there is a good chance that the candidate itself matches the target. To find a suitable target match, SM-MTC searches for the $K$ closest side matches, and selects the one that results in the best target match, according to STSIM. If that meets a quality threshold, it encodes its index; this typically requires much fewer bits than the DM-MTC approach. As illustrated in Fig. 4, the context is typically taken to be an L-shaped region. To enable good blending, mean-squared error is used as the SM metric [65], in contrast to the STSIM used for target matching. Thus, SM serves a dual role, to identify candidates for target encoding and to facilitate smooth blending. Moreover, SM enables better target matching because textures tend to be locally uniform, and thus among candidates with roughly equal STSIM quality, the ones with good point-by-point context matches tend to be more similar to the target. Of course, SM can also work for smooth and transition blocks.

Finally, we should also mention that blending of JPEG-coded blocks with previously MTC-coded blocks requires special care, and that context matching and blending are also needed in the DM version. More details can be found in [72].

Taking into account compression, reproduced image quality, and computational complexity, the best results so far have been attained with a combination of DM and SM [72]. A representative result is shown in Fig. 5 which compares MTC with JPEG at 0.34 b/pixel. For 44% of the image, JPEG, is

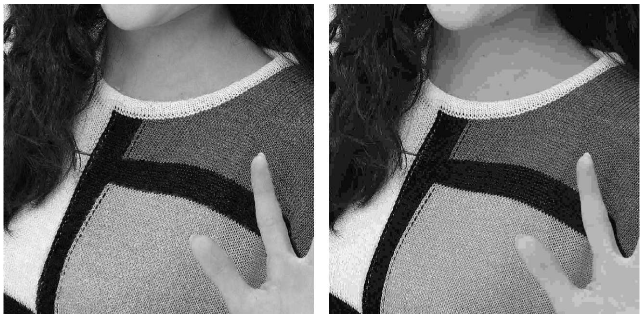


Fig. 5. Coding of "woman" at 0.34 bpp: (a) MTC, (b) JPEG

replaced by Block Matching. This allows the allocation of more bits to the JPEG-encoded blocks, which in turn provide better candidates for encoding the MTC blocks.

It is interesting to note that MTC fixes the most serious shortcoming of conventional compression techniques, namely, their inability to exploit the complementarity of smooth and textured regions. In smooth regions, the signal is simple but a strict fidelity metric must be met. In textured regions, the signal is complex but a lot more forgiving due to strong masking effects. Conventional techniques (e.g., JPEG) can handle the first case very efficiently, but signal complexity causes them to be inefficient in the latter. However, it is the complexity of the texture, which allows substantial point-by-point changes without affecting visual quality, that enables the efficient encoding of texture patches by MTC, provided of course that a suitable texture similarity metric is available.

A conceptual precursor to the SM approach is side-matched vector quantization [73]. There are also close links to non-local restoration techniques [11]–[13] and fractal image coding [74], both of which exploit image self-similarity. However, in fractal coding, all that needs to be encoded is the specification of the similarity transformations. All of these precursors rely on pointwise similarity (local shape) and could greatly benefit from the incorporation of STSIMs.

### B. *Other Texture-Based Approaches*

There have been a variety of other approaches for exploiting texture in image compression. The most common approach is to identify textured segments of the image that can be replaced by synthesized texture. The segmentation info and the parametric description of the texture is what needs to be encoded for these segments. A similarity metric is, of course, also needed, but sometimes is implicit in the texture reconstruction method. One of the earliest attempts for such compression-by-synthesis was presented by Popat and Picard [75], for both lossy and lossless compression.

Ballé *et al.* [9] have used a statistical texture model to encode blocks of homogeneous texture. The image is divided into nonoverlapping blocks, and each block that is classified as texture is decomposed into a lowpass component, which is encoded with standard compression techniques, and a highpass component, which is encoded by texture synthesis; the texture model parameters are sent as side information. All other blocks are encoded with standard techniques. Note that this texture coding approach can only handle homogeneous stochastic textures; it cannot be used to encode periodic textures or blocks partially covered with texture. Similar approaches relying on segmentation and texture synthesis have been proposed by Bosch *et al.* [8], Zhang and Bull [10], as well as other "compression-by-synthesis" papers [5], [6].

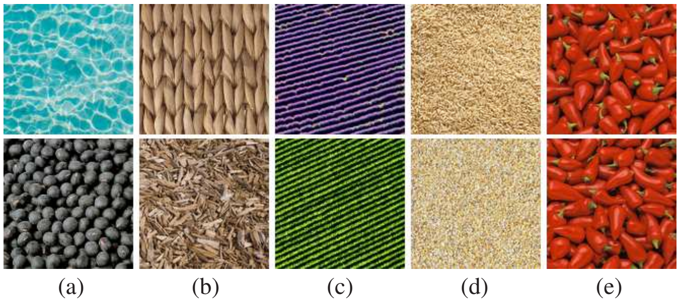


Fig. 6. Examples of texture pairs: (a) dissimilar textures, (b) similar color, different structure, (c) similar structure, different color, (d) similar in all perceptual dimensions, (e) "identical"

## V. Content-Based Retrieval

As discussed in the introduction, a texture similarity metric is critical for CBR problems. We have identified two problems of interest.

The first is the retrieval of identical textures [2], whereby one is interested only in exact matches, that is, samples of the same texture. An example of identical textures is shown in Fig. 6(e). This problem is important when searching for images that contain a particular texture (concrete wall, gray-shingle roof) or images that correspond to the same scene. However, it is also of interest in its own right, e.g., for retrieving textures from a database (of carpets, fabrics, marble tiles, etc.). To construct a database that contains groups of identical textures, all one has to do is cut patches from larger perceptually uniform textures, as explained in [2]. The advantage of such a database is that the ground truth is known, and therefore, metric testing does not require any subjective tests. Of course, this is true to the extent that the textures from which the identical pieces are sampled are perceptually uniform.

The second problem is the retrieval of similar textures [54]. This is considerably more difficult, because texture is a multidimensional attribute and similarity can be defined along one or more perceptual dimensions (color, scale, orientation, regularity, shape of texture elements, etc.). When textures are similar along one perceptual dimension and dissimilar along other perceptual dimensions, it is difficult to quantify their overall similarity. For example, the textures in Fig. 6(b) have similar colors, while the textures in Fig. 6(c) have similar structure (orientation, scale, periodicity). Whether the textures in each pair are similar, or which of the pairs is more similar than the other can be quite subjective. It is only when textures are similar along all perceptual dimensions, as in the textures in Fig. 6(d), that subjects consistently classify them as similar. Thus, an important problem for texture retrieval is distinguishing between similar and dissimilar pairs.

## VI. Testing Texture Similarity Metrics

In Sections I, IV, and V, we discussed applications that make use of STSIMs. Each application imposes its own requirements on metric performance. In image compression it is important to ensure a monotonic relationship between measured and perceived similarity. However, this monotonicity is needed only when the images we compare are fairly similar; when the images are dissimilar, it is sufficient that the metric simply gives a low value. For compression, it is also important that the similarity metric give consistent values across different types of images, so that we can establish a uniform quality criterion. In CBR, as discussed above, it is important to distinguish between similar and dissimilar pairs. The ordering of the retrieved images may also be desired, but only at the high end of the scale, i.e., identifying the most similar images. At the bottom of the scale, it is sufficient to simply declare that a texture pair is dissimilar, and there is no need to compare it to other dissimilar pairs. Finally, at the high end of the scale, there may be a need for thresholds, to establish if two textures are sufficiently similar or identical. In both applications, similarity can be quantified only over a small range at the top of the scale, while below a certain threshold, it suffices to simply declare texture pairs as dissimilar.

From a perceptual perspective, Zujovic *et al.* [26], [54] found that, when judging texture similarity, subjects are consistent at the high end of the similarity scale, where images exhibit similarity in every texture attribute (scale, directionality, color, regularity, etc.). When two textures are similar in some respect (e.g., color composition, directionality) but different in another (e.g., scale, regularity), the subject-to-subject agreement is poor because each subject puts different weights on different texture attributes in determining overall texture similarity. It thus makes no sense to require monotonic metric behavior in this range. Overall, human subjects are not able to make consistent judgments when asked to order pairs of dissimilar textures. On the other hand, Zujovic *et al.* [26], [54] found that human subjects give consistently low scores for dissimilar textures.

Thus, both human perception and application requirements agree that a monotonic relationship is desired in the region of very similar textures, and that in the rest of the similarity range, the metric should be able to distinguish between similar and dissimilar textures. Fig. 7 is a schematic illustration of good metric behavior according to these requirements [26], [54]. It plots subjective rankings versus objective metric values. The subjective similarity scores on the $x$-axis are for all possible pairs of texture patches in some hypothetical database. For the sake of argument, we assume here that it is possible to derive consistent subjective similarity scores for all image pairs. In reality, this would be difficult, if not impossible. Observe that a monotonic relationship is desired only in the region of very similar textures (to the right of $T_1$; this also includes identical textures). The similar range (to the right of $T_2$) is where subjects agree that textures are similar but do not assign consistent similarity scores; in this range we do not expect a monotonic relationship, but expect high metric scores. In the region of dissimilar textures (to the left of $T_3$), the subjects agree that textures are dissimilar but, again, do not assign consistent similarity scores; the only constraint here is that the metric yields low values.

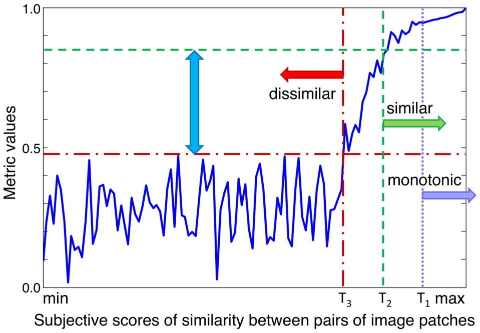


Fig. 7. Desired metric behavior (metric values vs. subjective similarity scores)

| | Known-item-search | | | Similar texture retrieval | | |
|---|---|---|---|---|---|---|
| Metric | P@1 | MRR | MAP | P@1 | MRR | MAP |
| PSNR | 0.04 | 0.07 | 0.06 | 0.14 | 0.23 | 0.17 |
| S-SSIM | 0.09 | 0.11 | 0.06 | 0.41 | 0.49 | 0.24 |
| CW-SSIM | 0.39 | 0.46 | 0.40 | 0.84 | 0.90 | 0.64 |
| CW-SSIM global | 0.27 | 0.36 | 0.28 | 0.72 | 0.82 | 0.54 |
| STSIM-2 | 0.74 | 0.80 | 0.74 | **0.86** | **0.90** | **0.69** |
| STSIM-2 global | 0.93 | 0.95 | 0.89 | 0.84 | 0.89 | 0.62 |
| STSIM-M | **0.96** | **0.97** | **0.92** | 0.84 | 0.89 | 0.62 |
| Do and Vetterli | 0.84 | 0.89 | 0.80 | 0.79 | 0.85 | 0.56 |
| Ojala *et al.* | 0.90 | 0.92 | 0.86 | 0.57 | 0.68 | 0.39 |

TABLE I
INFORMATION RETRIEVAL STATISTICS

Zujovic *et al.* [26], [54] have identified three distinct domains for metric testing:

- identifying identical textures
- distinguishing similar and dissimilar textures
- metric monotonicity at the high end of the scale

Each domain requires a different database (and associated ground truth) and different testing procedures.

For the retrieval of identical textures, as discussed in the Section V, the database can be constructed by cutting patches from larger perceptually uniform textures. Then the metric is tested on its ability to distinguish identical from nonidentical textures. For retrieval of similar textures, the database needs to have clusters of similar textures, and the metric should be able to separate similar (within cluster) and dissimilar (across clusters) textures. To form such similarity clusters, Zujovic *et al.* have devised an efficient procedure called Visual Similarity by Progressive Grouping (ViSiProG) [26], [54]. In ViSiProG, each subject is asked to form small groups of similar textures one at a time, in a step-by-step fashion, picking similar images (typically a set of nine) out of a small set of images, and repeating the process with a new set that contains the group and a new set of images, progressively refining the similarity group. The results of several groups from several subjects are combined to form a similarity matrix for the entire database, which can be analyzed by spectral clustering [76] to form the final similarity clusters.

For the CBR problems, a number of statistical evaluation measures have been developed to test system performance. These include precision at one (measures in how many cases the first retrieved document is relevant), mean reciprocal rank (measures how far away from the first retrieved document is the first relevant one [77]), and mean average precision [78]. Table I shows representative results. For the known-item-search case, the database included 1180 texture patches taken

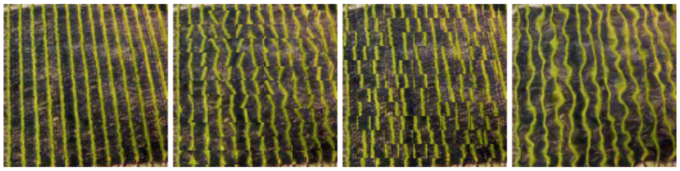


Fig. 8. Examples of texture distortions: Original, rotation, translation, warping

from 425 perceptually distinct images, each considered to be a homogeneous texture [2]. For the retrieval of similar textures, the tests were performed on 11 clusters with a total of 120 texture patches, which the subjects selected from a total of 505 patches. Texture pairs that belong to the same cluster were considered similar and those that belong to different clusters were considered dissimilar. All texture patches were of size $128 \times 128$. Note that in the known-item-search, global methods have significantly better performance than the local ones, while the opposite is true in the similar texture retrieval case. This can be explained by the fact that global metrics provide more robust estimates, which are important in differentiating identical from different textures, while local metrics provide better accuracy when comparing different degrees of similarity. Note also that the best metric performance in the known-item-search case is better than the performance in the similar texture retrieval case. This should be expected because the former is not as well defined as the latter.

Another measure of performance that evaluates a metric's ability to establish an absolute threshold, based on which identical and nonidentical or similar and dissimilar textures can be distinguished, is through the use of the receiver operating characteristic (ROC). This type of performance is quite distinct from the retrieval statistics we discussed above, and is important for both retrieval and compression applications. Overall, the STSIM-2 and STSIM-M metrics outperform existing metrics according to the different criteria. More details can be found in [2], [26].

For testing metric monotonicity, it is important to collect a set of images that covers fine differences in distortion level at the high end of the similarity scale, which can then be ranked by human subjects, thus providing ground truth for metric testing. Obtaining such a set would be a difficult in the context of real applications like structurally lossless compression discussed in Section IV. Thus, Zujovic *et al.* [79] chose to generate synthetic textures that model distortions that occur in such applications. Due to the nature of structurally lossless compression, the synthesized distortions take the form of variations in natural textures, that is, variations in position, orientation, and color [23]. Thus, Zujovic *et al.* implemented the following distortions: random rotation of small local patches, random shifts of small local patches, and image warping, whereby the images are distorted according to the random deviations of the control points of the underlying mesh. Examples are shown in Figure 8. The idea is that the severity of each type of distortion can be easily controlled by varying the distortion parameters (probabilistic distribution of rotations, shifts, and mesh points), so that the monotonicity of a metric can be assessed.

| | PSNR | SSIM | CWSSIM | STSIM2 |
|---|---|---|---|---|
| Borda's rule | 0.72 | 0.74 | 0.84 | **0.88** |

TABLE II
PEARSON'S $r$ FOR DIFFERENT ANALYSIS METHODS

For the subjective experiments conducted in [79], three levels were selected for each of the three types of distortions, and subjects were asked to rank the distorted images from best to worst, compared to the original image. In more recent experiments, we found it beneficial to also include the original texture in the set of distorted textures, so the subjective similarity values can be anchored. In order to determine metric performance across content, the subjects were also asked to rank the worst distortions for each of the original textures.

The ranking data can then be analyzed in a number of ways to obtain a subjective similarity score for each (distorted) image. The simplest is Borda's rule, which bases the similarity score on the mean rank of each image. A second approach is to use Thurstonian scaling [80]. A third alternative is to treat the ranks as distances between images and to use multidimensional scaling (MDS) [81], [82] to obtain the similarity scores.

Once the subjective similarity scores are obtained, the metric performance can be evaluated using Pearson's correlation coefficient, which evaluates absolute metric performance (correlation between metric and subjective scores), and Spearman rank correlation coefficient, which describes how well a metric ranks the distorted images compared to the subjective rankings. Table II shows Pearson's $r$, averaged over 10 textures for Borda's rule [26], [79]. The Spearman results and other approaches are comparable.

## VII. Conclusions

We have reviewed algorithms for measuring texture similarity that incorporate knowledge of human perception, and their importance for further advances in image analysis applications. We examined image compression and content-based retrieval in considerable detail; however, texture similarity metrics, and a better understanding of texture in general, are critical for a variety of other applications, including image restoration, computer vision, graphics, sense substitution, and multimodal interfaces.

Our main focus was on structural texture similarity metrics (STSIMs). We argued that they follow naturally from work on texture analysis/synthesis [23], which relies on a multiple frequency and orientation subband decomposition and statistical analysis of the subband coefficients. The selection of such models is based on both perceptual principles and signal characteristics, and is motivated by and supports an ecological approach to visual perception, whereby the visual system relies on informative statistical cues [83], rather than solving a complicated (and compute intensive) "inverse" problem to figure out the properties of the surfaces it is looking at.

The development of texture similarity metrics is closely linked to the methods for their testing and validation. Based on both human perception and application requirements, we argued that texture similarity can be consistently quantified only over a small range at the top of the similarity scale, where textures are similar in every respect (scale, directionality, color, regularity, etc.), while in the rest of the similarity range, it suffices to simply declare texture pairs as similar or dissimilar. Quantifying texture perception beyond this narrow range requires a better understanding of the importance of the different texture attributes. Thus, a promising direction for current research is quantifying texture similarity along specific texture attributes, such as surface reflectance (gloss) [83], [84] and roughness [85]. Other attributes, such as directionality, regularity/periodicity, and scale are also under consideration.

Other directions for future research include the development of similarity metrics that can provide consistent measurements between smooth, textured, and transition regions, and exploring the relationship – and integration with – of visual texture to other modalities (tactile and acoustic).

## References

[1] A. C. Brooks, X. Zhao, and T. N. Pappas, “Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions,” *IEEE Tr. Image Proc.*, vol. 17, no. 8, pp. 1261–1273, Aug. 2008.

[2] J. Zujovic, T. N. Pappas, and D. L. Neuhoff, “Structural texture similarity metrics for image analysis and retrieval,” *IEEE Tr. Image Proc.* To appear.

[3] E. H. Adelson, “On seeing stuff: The perception of materials by humans and machines,” in *Human Vision and Electronic Imaging VI*, Proc. SPIE, vol. 4299, San Jose, CA, Jan. 2001, pp. 1–12.

[4] T. N. Pappas, V. Tartter, A. G. Seward, B. Genzer, K. Gourgey, and I. Kretzschmar, “Perceptual dimensions for a dynamic tactile display,” in *Human Vision and Electronic Imaging XIV* (B. E. Rogowitz and T. N. Pappas, eds.), vol. 7240 of *Proc. SPIE*, (San Jose, CA), pp. 72400K-1–12, Jan. 2009.

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

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

[7] T. N. Pappas, J. Zujovic, and D. L. Neuhoff, “Image analysis and compression: Renewed focus on texture,” in *Visual Information Processing and Communication*, Proc. SPIE, Vol. 7543, San Jose, CA, Jan. 2010.

[8] M. Bosch, F. Zhu, and E. J. Delp, “Segmentation-based video compression using texture and motion models,” *IEEE J. Sel. Topics Signal Proc.*, vol. 5, no. 7, pp. 1366–1377, Nov. 2011.

[9] J. Balle, A. Stojanovic, and J.-R. Ohm, “Models for static and dynamic texture synthesis in image and video compression,” *IEEE J. Sel. Topics Signal Proc.*, vol. 5, no. 7, pp. 1353–1365, Nov. 2011.

[10] F. Zhang and D. R. Bull, “A parametric framework for video compression using region-based texture models,” *IEEE J. Sel. Topics Signal Proc.*, vol. 5, no. 7, pp. 1378–1392, Nov. 2011.

[11] A. Buades, B. Coll, and J.-M. Morel, “A review of image denoising algorithms, with a new one,” *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” *IEEE Tr. Image Proc.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[13] C. Kervrann and J. Boulanger, “Local adaptivity to variable smoothness for exemplar-based image regularization and representation,” *Int. Journal of Computer Vision*, vol. 79, no. 1, pp. 45–69, 2008.

[14] J. Chen, T. N. Pappas, A. Mojsilovic, and B. E. Rogowitz, “Adaptive perceptual color-texture image segmentation,” *IEEE Tr. Image Proc.*, vol. 14, no. 10, pp. 1524–1536, Oct. 2005.

[15] L. He, *A Clustering Approach for Color Texture Segmentation*. PhD dissertation, EECS Dept, Northwestern Univ., Evanston, IL, Aug. 2012.

[16] S. Bae and B.-H. Juang, “Multidimensional incremental parsing for universal source coding,” *IEEE Tr. Image Proc.*, vol. 17, no. 10, pp. 1837–1848, Oct. 2008.

[17] ——, “IPSILON: incremental parsing for semantic indexing of latent concepts,” *IEEE Tr. Image Proc.*, vol. 19, no. 7, pp. 1933–1947, July 2010.

[18] J. J. Koenderink, “Vision and information,” in *Perception Beyond Inference: The Information Content of Visual Processes*, L. Albertazzi, G. J. van Tonder, and D. Vishwanath, Eds. MIT Press, 2010, pp. 27–57.

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

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

[21] T. N. Pappas, J. Chen, and D. Depalov, “Perceptually based techniques for image segmentation and semantic classification,” *IEEE Commun. Mag.*, vol. 45, no. 1, pp. 44–51, Jan. 2007.

[22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Tr. Image Proc.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[23] J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statictics of complex wavelet coefficients,” *Int. J. Computer Vision*, vol. 40, no. 1, pp. 49–71, Oct. 2000.

[24] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multi-scale transforms,” *IEEE Tr. Info. Th.*, vol. 38, no. 2, pp. 587–607, Mar. 1992.

[25] Z. Wang and E. P. Simoncelli, “Translation insensitive image similarity in complex wavelet domain,” in *IEEE Int. Conf. Acoustics, Speech, Signal Proc.*, vol. II, Philadelphia, PA, 2005, pp. 573–576.

[26] J. Zujovic, “Perceptual texture similarity metrics,” Ph.D. dissertation, EECS Dept., Northwestern Univ., Evanston, IL, Aug. 2011.

[27] M. S. Landy, “Texture perception,” in *Encyclopedia of Neuroscience*, 3rd ed., G. Adelman and B. H. Smith, Eds. Amsterdam: Elsevier, 2004.

[28] J. R. Bergen, “Theories of visual texture perception,” in *Spatial Vision*, ser. Vision and Visual Dysfunction, D. Regan, Ed. Cambridge, MA: CRC Press, 1991, vol. 10, pp. 114–134.

[29] M. S. Landy and N. Graham, “Visual perception of texture,” in *The Visual Neurosciences*, L. M. Chalupa and J. S. Werner, Eds. Cambridge, MA: MIT Press, 2004, pp. 1106–1118.

[30] H. Tamura, S. Mori, and T. Yamawaki, “Textural features corresponding to visual perception,” *IEEE Tr. Syst., Man, Cybern.*, vol. 8, no. 6, pp. 460–473, June 1978.

[31] A. R. Rao and G. L. Lohse, “Towards a texture naming system: Identifying relevant dimensions of texture,” *Vision Research*, vol. 36, no. 11, pp. 1649–1669, 1996.

[32] A. Mojsilović, J. Kovačević, J. Hu, R. J. Safranek, and S. K. Ganapathy, “Matching and retrieval based on the vocabulary and grammar of color patterns,” *IEEE Tr. Image Proc.*, vol. 1, no. 1, pp. 38–54, Jan. 2000.

[33] M. Amadasun and R. King, “Textural features corresponding to texture properties,” *IEEE Tr. Syst., Man, Cybern.*, vol. 19, pp. 1264–1274, 1989.

[34] B. Julesz, “Visual pattern discrimination,” *IRE Tr. Info. Th.*, vol. 8, pp. 84–92, Feb. 1962.

[35] B. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frisch, “Inability of humans to discriminate between visual textures that agree in second-order statistics – revisited,” *Perception*, vol. 2, no. 4, pp. 391–405, 1973.

[36] B. Julesz, E. N. Gilbert, and J. D. Victor, “Visual discrimination of textures with identical third-order statistics,” *Biological Cybernetics*, vol. 31, no. 3, pp. 137–140, 1978.

[37] T. Caelli and B. Julesz, “On perceptual analyzers underlying visual texture discrimination: part I,” *Biol. Cyber.*, vol. 28, no. 3, pp. 167–175, 1978.

[38] J. D. Victor, C. Chubb, and M. M. Conte, “Interaction of luminance and higher-order statistics in texture discrimination,” *Vision Res.*, vol. 45, pp. 311–328, 2005.

[39] T. Maddness, Y. Nagai, J. D. Victor, and R. R. L. Taylor, “Multilevel isotrigon textures,” *J. Opt. Soc. Am. A*, vol. 24, no. 2, pp. 278–293, Feb. 2007.

[40] G. Tracik, J. S. Prentice, J. D. Victor, and V. Balasubramanian, “Local statistic in natural scenes predict the saliency of synthetic textures,” *Proc. Nat. Acad. Sciences USA*, vol. 107, no. 42, pp. 18 149–18 154, Oct. 2010.

[41] B. Julesz, “Textons, the elements of texture perception and their interactions,” *Nature*, vol. 290, pp. 91–97, 1981.

[42] J. Beck, “Similarity grouping and peripheral discriminability under uncertainty,” *American Journal of Psychology*, vol. 85, no. 1, pp. 1–19, 1972.

[43] H. Voorhees and T. Poggio, “Computing texture boundaries,” *Nature*, vol. 333, no. 6171, pp. 364–367, May 1988.

[44] J. R. Bergen and E. H. Adelson, “Early vision and texture perception,” *Nature*, vol. 333, no. 6171, pp. 363–364, May 1988.

[45] J. Malik and P. Perona, “Preattentive texture discrimination with early vision mechanisms,” *J. Opt. Soc. Am. A*, vol. 7, pp. 923–932, 1990.

[46] M. Porat and Y. Y. Zeevi, “Localized texture processing in vision: Analysis and synthesis in Gaborian space,” *IEEE Tr. Biomed. Eng.*, vol. 36, no. 1, pp. 115–129, 1989.

[47] R. L. D. Valois and K. K. D. Valois, *Spatial Vision*. New York: Oxford University Press, 1990.

[48] B. J. Balas, “Texture synthesis and perception: Using computational models to study texture representations in the human visual system,” *Vision Research*, vol. 46, pp. 299–309, 2006.

[49] B. J. Balas, L. Nakano, and R. Rozenholtz, “A summary-statistic representation in peripheral vision explains visual crowding,” *Journal of Vision*, vol. 9, no. 12, pp. 13, 1–18, 2009.

[50] R. Rozenholtz, J. Huang, A. Raj, B. J. Balas, and L. Ilie, “A summary statistic representation in peripheral vision explains visual search,” *Journal of Vision*, vol. 12, no. 4, pp. 14, 1–17, 2012.

[51] J. Freeman and E. P. Simoncelli, “Metamers of the ventral system,” *Nature Neuroscience*, vol. 14, no. 9, pp. 1195–1204, Sept. 2011.

[52] D. J. Heeger and J. R. Bergen, “Pyramid-based texture analysis/synthesis,” in *Proc. Int. Conf. Image Proc. (ICIP), vol. III*, Washington, DC, Oct. 1995, pp. 648–651.

[53] J. Zujovic, T. N. Pappas, and D. L. Neuhoff, “Structural similarity metrics for texture analysis and retrieval,” in *Proc. Int. Conf. Image Proc.*, Cairo, Egypt, Nov. 2009, pp. 2225–2228.

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

[55] X. Zhao, M. G. Reyes, T. N. Pappas, and D. L. Neuhoff, “Structural texture similarity metrics for retrieval applications,” in *Proc. Int. Conf. Image Proc. (ICIP)*, San Diego, CA, Oct. 2008, pp. 1196–1199.

[56] W. Y. Ma, Y. Deng, and B. S. Manjunath, “Tools for texture/color based search of images,” in *Human Vision and Electronic Imaging II*, Proc. SPIE, Vol. 3016, San Jose, CA, Feb. 1997, pp. 496–507.

[57] T. N. Pappas, “An adaptive clustering algorithm for image segmentation,” *IEEE Tr. Signal Proc.*, vol. SP-40, no. 4, pp. 901–914, Apr. 1992.

[58] A. Mojsilović, J. Hu, and E. Soljanin, “Extraction of perceptually important colors and similarity measurement for image matching, retrieval, and analysis,” *IEEE Tr. Image Proc.*, vol. 11, no. 11, pp. 1238–1248, Nov. 2002.

[59] Y. Rubner, C. Tomasi, and L. J. Guibas, “The earth mover’s distance as a metric for image retrieval,” *Int. Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.

[60] D. Depalov and T. N. Pappas, “Analysis of segment statistics for semantic classification of natural images,” in *Human Vision and Electronic Imaging XII*, Proc. SPIE Vol. 6492, San Jose, CA, Jan. 29 – Feb. 1 2007, pp. 6492OD–1–6492OD–11.

[61] T. Ojala, T. Menp, J. Viertola, J. Kyllnen, and M. Pietikinen, “Empirical evaluation of MPEG-7 texture descriptors with a large-scale experiment,” in *Proc. $2^{nd}$ Int. Wksp. Texture Anal. Synthesis*, 2002, pp. 99–102.

[62] M. N. Do and M. Vetterli, “Texture similarity measurement using Kullback-Leibler distance on wavelet subbands,” in *Proc. Int. Conf. Image Proc.*, vol. 3, Vancouver, BC, Canada, Sept. 2000, pp. 730–733.

[63] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of image data,” *IEEE Tr. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.

[64] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, “Color and texture descriptors,” *IEEE Tr. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, June 2001.

[65] A. A. Efros and W. T. Freeman, “Image quilting for texture synthesis and transfer,” in *Proc. 28th Int. Conf. Comp. Graphics Inter. Techn. (SIGGRAPH-01)*, Los Angeles, CA, Aug. 2001, pp. 341–346.

[66] M. P. Eckert and A. P. Bradley, “Perceptual quality metrics applied to still image compression,” *Signal Proc.*, vol. 70, pp. 177–200, 1998.

[67] T. N. Pappas, R. J. Safranek, and J. Chen, “Perceptual criteria for image quality evaluation,” in *Handbook of Image and Video Processing*, 2nd ed., A. C. Bovik, Ed. Academic Press, 2005, pp. 939–959.

[68] S. Bae, T. N. Pappas, and B.-H. Juang, “Subjective evaluation of spatial resolution and quantization noise tradeoffs,” *IEEE Tr. Image Proc.*, vol. 18, no. 3, pp. 495–508, Mar. 2009.

[69] M. G. Ramos and S. S. Hemami, “Suprathreshold wavelet coefficient quantization in complex stimuli: psychophysical evaluation and analysis,” *J. Opt. Soc. Am. A*, vol. 18, no. 10, pp. 2385–2397, Oct. 2001.

[70] D. M. Chandler and S. S. Hemami, “Effects of natural images on the detectability of simple and compound wavelet subband quantization distortions,” *J. Opt. Soc. Am. A*, vol. 20, no. 7, pp. 1164–1180, July 2003.

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

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

[73] T. Kim, “Side match and overlap match vector quantizers for images,” *IEEE Tr. Image Proc.*, vol. 1, no. 2, pp. 170–185, Apr. 1992.

[74] A. E. Jacquin, “Image coding based on a fractal theory of iterated contractive image transformations,” *IEEE Tr. Image Proc.*, vol. 1, no. 1, pp. 18–30, Jan. 1992.

[75] K. Popat and R. W. Picard, “Novel cluster-based probability model for texture synthesis, classification, and compression,” in *Proc. SPIE Visual Communications*, Cambridge, MA, 1993.

[76] U. von Luxburg, “A tutorial on spectral clustering,” *Statistics and Computing*, vol. 17, pp. 395–416, 2007.

[77] E. M. Voorhees, “The TREC-8 question answering track report,” in *In Proc. 8th Text Retrieval Conf. (TREC-8)*, E. M. Voorhees and D. K. Harman, Eds., Nov. 17–19, 1999, pp. 77–82.

[78] E. M. Voorhees, “Variations in relevance judgments and the measurement of retrieval effectiveness,” *Information Processing & Management*, vol. 36, no. 5, pp. 697–716, Sept. 2000.

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

[80] L. L. Thurstone, “A law of comparative judgment,” *Psychological review*, vol. 34, no. 4, p. 273, 1927.

[81] W. S. Torgerson, *Theory and methods of scaling*. New York, NY: Wiley, 1958.

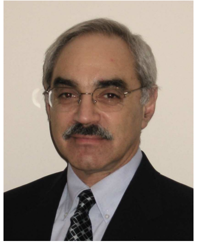
[82] J. B. Kruskal and M. Wish, *Multidimensional scaling*. Beverly Hills, CA: Sage Publications, 1977.

[83] L. Sharan, Y. Li, I. Motoyoshi, S. Nishida, and E. H. Adelson, “Image statistics for surface reflectance perception,” *J. Opt. Soc. Am. A*, vol. 25, no. 4, pp. 846–865, 2008.

[84] I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson, “Image statistics and the perception of surface qualities,” *Nature*, vol. 447, pp. 206–209, May 2007.

[85] R. V. Egmond, P. Lemmens, T. N. Pappas, and H. de Ridder, “Roughness in sound and vision,” in *Human Vision and Electronic Imaging XIV*, Proc. SPIE, vol. 7240, San Jose, CA, Jan. 2009, pp. 72 400B–1–12.

**Thrasyvoulos N. Pappas** received the S.B., S.M., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, in 1979, 1982, and 1987, respectively. From 1987 until 1999, he was a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. Since 1999 he has been with the Department of Electrical Engineering and Computer Science at Northwestern University. His research interests are in image and video quality and compression, image and video analysis, content-based retrieval, perceptual models for multimedia processing, model-based halftoning, and tactile and multimodal interfaces. Dr. Pappas is a Fellow of the IEEE and SPIE. He has served as an elected member of the Board of Governors of the Signal Processing Society of IEEE (2004-07), editor-in-chief of the IEEE Transactions on Image Processing (2010-12), chair of the IEEE Image and Multidimensional Signal Processing Technical Committee (2002-03), and technical program co-chair of ICIP-01 and ICIP-09. Since 1997 he has been co-chair of the SPIE/IS&T Conference on Human Vision and Electronic Imaging.



**David L. Neuhoff** received the B.S.E. from Cornell University, Ithaca, NY, USA, in 1970, and the M.S. and Ph.D. in Electrical Engineering from Stanford University, Stanford, CA, USA, in 1972 and 1974, respectively. Since graduation he has been a faculty member at the University of Michigan, where he is now the Joseph E. and Anne P. Rowe Professor of Electrical Engineering. From 1984 to 1989 he was an Associate Chair of the EECS Department, and since September 2008 he is again serving in this capacity. He spent two sabbaticals at Bell Laboratories, Murray Hill, NJ, and one at Northwestern University. His research and teaching interests are in communications, information theory, and signal processing, especially data compression, quantization, image coding, image similarity metrics, source-channel coding, halftoning, sensor networks, and Markov random fields. He is a Fellow of the IEEE. He co-chaired the 1986 IEEE International Symposium on Information Theory, was technical co-chair for the 2012 IEEE Statistical Signal Processing (SSP) workshop, has served as an associate editor for the IEEE Transactions on Information Theory, has served on the Board of Governors and as president of the IEEE Information Theory Society.

**Huib de Ridder** received the M.Sc. degree in Psychology from the University of Amsterdam , Amsterdam, The Netherlands, in 1980 and the Ph.D. degree in technical sciences from Eindhoven University of Technology, Eindhoven, The Netherlands, in 1987. Dr. de Ridder is Professor of Informational Ergonomics in the faculty of Industrial Design Engineering at Delft University of Technology, The Netherlands. From 1982 till 1998 he was affiliated with the Vision Group of the Institute for Perception Research (IPO), Eindhoven, The Netherlands, where his research focused on both fundamental and applied visual psychophysics. From 1987 till 1992 his research on the fundamentals of image quality metrics was supported by a personal fellowship from the Royal Netherlands Academy of Arts and Sciences (KNAW). In November 1998 he moved to Delft University of Technology where he became full professor in December 2000. His research focuses on human behavior at both perceptual and cognitive level covering topics like user understanding (e.g., intention tracking in user-product interaction, interaction with embedded intelligence, engagement), information presentation (e.g., picture perception, form perception, image quality, sound design), ambient intelligence, user experience and social connectedness. Since 2012 he has been co-chair of the SPIE/IS&T Conference on Human Vision and Electronic Imaging.

**Jana Zujovic** (M'09) received the Diploma in electrical engineering from the University of Belgrade in 2006, and the M.S. and Ph.D. degrees in electrical engineering and computer science from Northwestern University, Evanston, IL, in 2008 and 2011, respectively. From 2011 until 2013, she was working as a postdoctoral fellow at Northwestern University. Currently she is employed as a senior research engineer at FutureWei Technologies, Santa Clara, CA. Her research interests include image and video analysis, image quality and similarity, content-based retrieval and pattern recognition.