A SOFT MEASURE FOR IDENTIFYING STRUCTURE FROM RANDOMNESS IN IMAGES

Aous Thabit Naman and David Taubman

School of Electrical Engineering and Telecommunications,
The University of New South Wales, Australia.

ABSTRACT

This paper presents a novel measure for identifying strong structure features, such as edges, from randomness, such as regions predominated by noise, within an image. The proposed structural measure is localized in space and scale; for a given scale, it gives values close to one in the vicinity of strong structures and close to zero in regions predominated by noise. The proposed structural measure is a primitive operation that can be used in a wide variety of image analysis techniques to identify regions which has structure; for example, motion estimation is more meaningful in structured regions than in regions filled with noise. The first innovation in this work is in converting an image into a ternary feature map that are rather resistant to noise and changes in illumination. The second is the structural measure, which is derived from the degree of non-uniformity amongst the magnitudes of the DFT coefficients obtained over a small window within the ternary maps. In this work, we show that the proposed structural measure is robust and gives a good indication of the strength of structure when compared to alternate strategies; moreover, we show that the computational cost of the proposed structural measure is reasonable.

Index Terms— Image Analysis, Image Processing, Feature Extraction, Image Texture Analysis

1. INTRODUCTION

Many image processing algorithms can benefit from knowing that the region being processed has structural features, such as edges and geometrical shapes. For example, image registration is more meaningful between two structured regions in the images being registered. Denoising is another obvious application which stands to benefit from a knowledge of the structured image regions.

In this paper, we propose a structural feature measure that is local in space and scale, returning a value in the range 0 (primarily noise) to 1 (strongly structured) at each location within each detail band of a Laplacian pyramid [1]. This is important, because different scales in a multi-resolution representation reveal different features and high frequency bands are often quite sparse – i.e., predominantly noise – while also carrying useful information. The proposed measure has no dependence on signal strength (e.g., through thresholds) or assumptions concerning the signal or noise power spectra. The method can also be applied to image compression formats that are inherently multi-resolution, such as JPEG2000 [2]. We have found these to be important features in our work on multi-resolution motion estimation and motion detection [3, 4]. We are unaware of alternate measures of structure that possess such features.

This paper introduces two innovations; the first involves representing each detail band of the image as a ternary feature map. The proposed ternary feature maps share some concepts with local binary patterns, first introduced by Ojala *et al.* [5]. Both approaches represent the image using a small alphabet, where the value taken by a given pixel depends on its neighborhood; the resulting set is somewhat illumination-independent and noise-resistant.

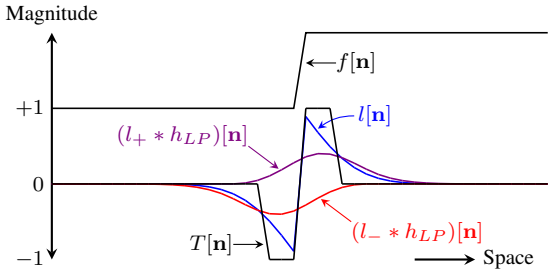


Fig. 1: A typical response of $T[\mathbf{n}]$ along an edge in the image $f[\mathbf{n}]$. The figure also shows some of the intermediate values.

The second innovation of this work is the structural measure itself, which involves a Short-Time Fourier Transform (STFT) of the ternary feature map at each scale, followed by a measure that is inspired by the concept of compression coding gain [2, 6]; the idea is that a large coding gain is indicative of a strongly structured region. The proposed structure also lends itself to an efficient implementation.

The rest of the paper is organized as follows: Section 2 explains the proposed ternary feature maps; Section 3 describes the proposed structural measure; Section 4 discusses how the structural measure is extended to multiple resolutions; and Section 5 compares the structural measure against alternative approaches.

2. TERNARY FEATURE MAPS

In general, the proposed ternary feature map $T[\mathbf{n}]$ is calculated from an image that has its local mean removed; that is, for a given image $f[\mathbf{n}]$, we first estimate the local mean from $(G_\sigma * f)[\mathbf{n}]$ for each location $\mathbf{n}$ in $f[\mathbf{n}]$, where G_σ is a symmetric finite impulse response (FIR) low-pass filter. Then, the proposed ternary feature map $T[\mathbf{n}]$ is calculated from $l[\mathbf{n}] = f[\mathbf{n}] - (G_\sigma * f)[\mathbf{n}]$. We can think of $l[\mathbf{n}]$ as the detail band at one level of a Laplacian pyramid, as discussed in Section 4.

Figure 1 shows a typical response of $l[\mathbf{n}] = f[\mathbf{n}] - (G_\sigma * f)[\mathbf{n}]$ along an edge in $f[\mathbf{n}]$. For regions that are dominated by noise, we expect the mean $(G_\sigma * f)[\mathbf{n}]$ to be the mean of the underlying signal, not the noise; therefore, $l[\mathbf{n}]$ is also dominated by noise, since the mean $(G_\sigma * f)[\mathbf{n}]$ is removed from $l[\mathbf{n}]$.

The ternary feature map $T[\mathbf{n}]$ is derived from two quantities, $l_+[\mathbf{n}] = \max\{l[\mathbf{n}], 0\}$ and $l_-[\mathbf{n}] = \min\{l[\mathbf{n}], 0\}$, using

$$T[\mathbf{n}] = \begin{cases} 1, & l[\mathbf{n}] > (l_+ * h_{\mathrm{T}})[\mathbf{n}] \\ -1, & l[\mathbf{n}] < (l_- * h_{\mathrm{T}})[\mathbf{n}] \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

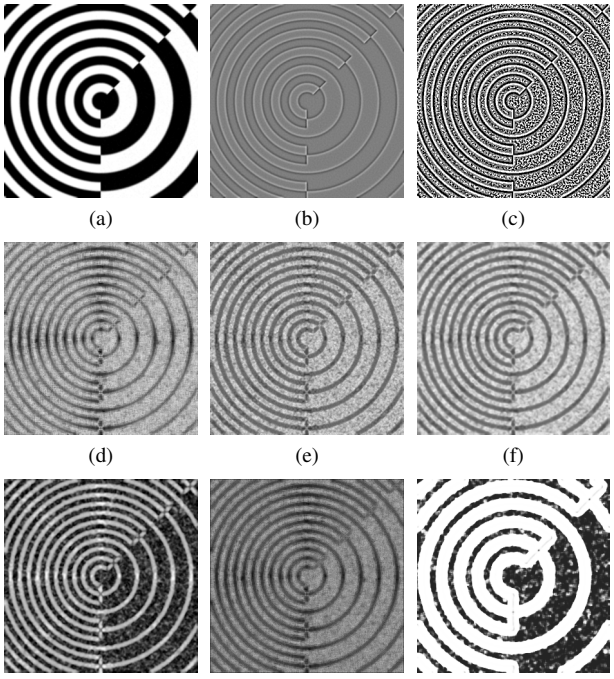


Fig. 2: Results and alternative approaches for a synthetic image, where the gray color represent 0. (a) A synthetic image (PSNR = 32dB). (b) Its mean-removed image $l[\mathbf{n}]$. (c) Its ternary feature map $T[\mathbf{n}]$. (d) $A'[\mathbf{n}]$ when $c_{\mathbf{k}}[\mathbf{n}] = c'_{\mathbf{k}}[\mathbf{n}]$ and $w[\mathbf{p}] = 1$. (e) $A'[\mathbf{n}]$ when $c_{\mathbf{k}}[\mathbf{n}] = c'_{\mathbf{k}}[\mathbf{n}]$ and $w[\mathbf{p}]$ as defined in (3). (f) $A'[\mathbf{n}]$ when $c_{\mathbf{k}}[\mathbf{n}]$ as defined in (4) and $w[\mathbf{p}]$ as defined in (3). (g) The image's structural measure $A[\mathbf{n}]$. (h) $A'[\mathbf{n}]$ when the DCT is used instead of DFT and $w[\mathbf{p}] = 1$. (i) Detected structure using the IRPR measure (Subsection 5.1) for a synthetic image that has rings twice as wide as the image in (a).

where $h_{\mathrm{T}}[\mathbf{n}]$ is another symmetric low-pass FIR filter. We can think of $(l_+ * h_{\mathrm{T}})[\mathbf{n}]$ and $(l_- * h_{\mathrm{T}})[\mathbf{n}]$ as the local positive and negative means, that serve as spatially varying thresholds, as shown in Figure 1. These thresholds help to suppress noise and ringing artefacts in the neighborhood of strong image features.

Figures 2c and 3b provide examples of ternary feature maps. The former is derived from a synthetic image, Figure 2a, to which a small amount of noise has been added (PSNR = 32dB). Figure 2b shows the mean-removed image $l[\mathbf{n}]$ for this synthetic case. The second example is based on a real image from the Kodak test set, shown in Figure 3a.

The low-pass filters, G_σ and h_T, are design parameters that can be optimized for the task at hand. In this paper, all filters have support 7×7; G_σ is a separable Gaussian filter with $\sigma = 1.5$, while h_T is also a separable filter formed from the one-dimensional triangular transfer function $H_T(z) = z^{-3} + 2z^{-2} + 3z^{-1} + 4 + 3z + 2z^2 + z^3$. Both G_σ and h_T are normalized such that their DC-gain is 1.

Ternary feature maps preserve strong image features, such as edges, cleaning most of the noise near these features; for regions dominated by unstructured noise, the ternary features also lack structure. Ternary feature maps are rather robust to changes in illumination and somewhat resistant to noise. For these reasons, ternary feature maps are interesting in their own right and we have found other applications for them [3, 4].

3. STRUCTURAL MEASURE

The proposed structural measure $A[\mathbf{n}]$ is calculated from $T[\mathbf{n}]$. The idea is to estimate the structure within a neighborhood $\mathcal{P}_{\mathbf{n}}$ of location $\mathbf{n}$ by taking a Discrete Fourier Transform (DFT) of $T[\mathbf{p}]$ over the samples $\mathbf{p} \in \mathcal{P}_{\mathbf{n}}$ and then evaluating the degree of non-uniformity amongst the magnitudes of the DFT coefficients. In practice, $\mathcal{P}_{\mathbf{n}} = [n_1 - H, n_1 + H] \times [n_2 - H, n_2 + H]$, so that a $(2H + 1) \times (2H + 1)$ point DFT is used. The DFT basis functions are pre-weighted using a non-separable raised-cosine weighting function, which helps to avoid directional bias. Specifically, the DFT magnitude coefficients associated with location $\mathbf{n}$ are given by

$$c'_{\mathbf{k}}[\mathbf{n}] = \left| \sum_{-H \le p_1, p_2 \le H} T[\mathbf{p} + \mathbf{n}] w[\mathbf{p}] \exp\left(-j \frac{2\pi \mathbf{p}^t \mathbf{k}}{2H + 1}\right) \right| \quad (2)$$

where $\mathbf{k}$ belongs to the set $\mathcal{K}$, discussed shortly, and the weighting function $w[\mathbf{p}]$ is given by

$$w[\mathbf{p}] = \begin{cases} \frac{1}{2} + \frac{1}{2} \cdot \cos\left(\pi \frac{\sqrt{p_1^2 + p_2^2}}{H+1}\right), & p_1^2 + p_2^2 \le (H+1)^2 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Equivalently, the cosine-weighted DFT coefficients produced at each location may be interpreted as the output from a Short-Time Fourier Transform of the ternary feature map.

The set $\mathcal{K}$ contains one coefficient from each complex-conjugate pair of DFT coefficients. $\mathcal{K}$ does not include the DC coefficient. We have found that $H = 3$ gives a good compromise between locality and resistance to noise; in this case, the number of elements in $\mathcal{K}$ is 24.

The squared magnitude coefficients $(c'_{\mathbf{k}}[\mathbf{n}])^2$ may be interpreted as crude estimates of the local power spectrum associated with the ternary feature map – actually, these correspond to the periodogram of the raised-cosine weighted ternary features. It is well known that the periodogram estimate of the power spectrum of a random process is very noisy. With this in mind, we improve the robustness of the algorithm by averaging four sets of coefficient magnitudes from nearby locations, yielding

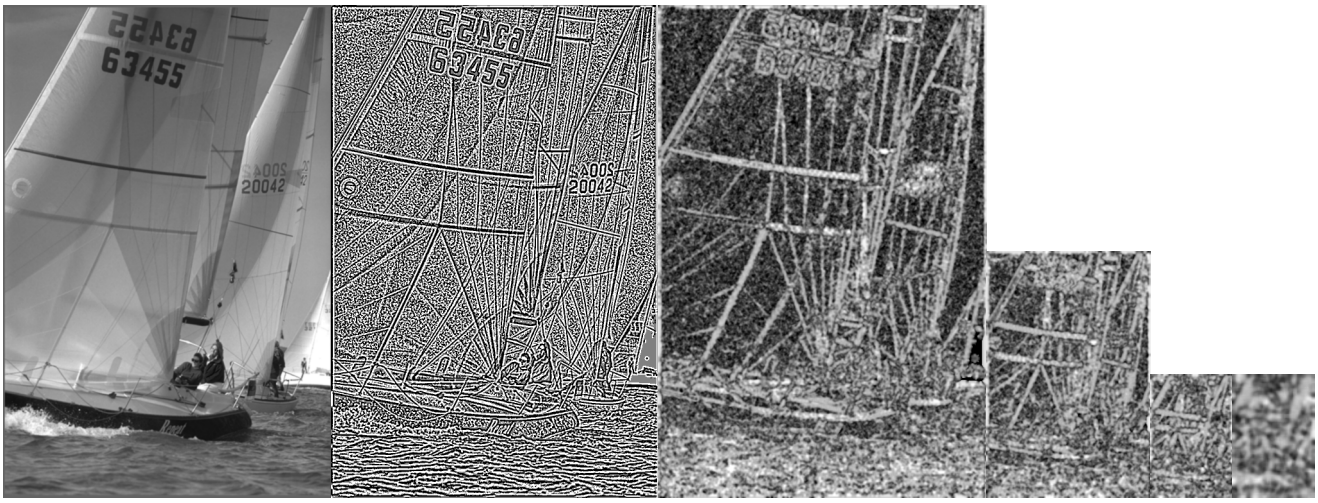
$$c_{\mathbf{k}}[\mathbf{n}] = \frac{1}{4} \cdot \sum_{i,j \in \{-1,1\}} c'_{\mathbf{k}}[n_1 + i, n_2 + j] \quad (4)$$

We then form an estimate of the uniformity of the $c_{\mathbf{k}}[\mathbf{n}]$ values, over $\mathbf{k}$, given by

$$A'[\mathbf{n}] = \left(\frac{1}{|\mathcal{K}|} \cdot \sum_{\mathbf{k} \in \mathcal{K}} c_{\mathbf{k}}[\mathbf{n}] \right)^2 \Bigg/ \frac{1}{|\mathcal{K}|} \cdot \sum_{\mathbf{k} \in \mathcal{K}} (c_{\mathbf{k}}[\mathbf{n}])^2 \quad (5)$$

When the transform coefficients $c_{\mathbf{k}}[\mathbf{n}]$ are roughly equal, we expect $A'[\mathbf{n}]$ to be close to 1; conversely, if the transform coefficients $c_{\mathbf{k}}[\mathbf{n}]$ are widely distributed, we expect $A'[\mathbf{n}]$ to approach the minimum value of $\frac{1}{|\mathcal{K}|}$. We interpret these two cases as unstructured noise and structured imagery, respectively. This interpretation is motivated by the notion that the DFT should have a high coding gain (highly skewed coefficient distribution) for structured content, but no coding gain (uniform coefficient distribution) for unstructured noise.

The middle row in Figure 2 shows the effect of the various steps on the $A'[\mathbf{n}]$ values. Both Figures 2d and 2e involve no averaging (i.e., $c_{\mathbf{k}}[\mathbf{n}] = c'_{\mathbf{k}}[\mathbf{n}]$). The image in Figure 2e incorporates the raised-cosine weighting function, whereas Figure 2d is generated with $w[\mathbf{p}] = 1$; the anisotropic nature of the unweighted results



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

Fig. 3: The structural measure at a few scales. (a) Original image. (b) Its ternary feature map. (c) Its structural measure. (d) Its structural measure at half resolution. (e) Its structural measure at quarter resolution. (f) Its structural measure at one-eighth resolution zoomed by 2.

are revealed by the presence of thicker lines along the vertical and horizontal features in Figure 2d. The image in Figure 2f incorporates both averaging and raised-cosine weight; evidently, the result is less noisy without noticeable loss of spatial resolution.

Figure 4 shows the probability density function (PDF) of $A'[\mathbf{n}]$ obtained experimentally using the Kodak test image set. We notice that more than 99.9% of the area under the probability curve lies in the range $[0.25, 0.91]$ of $A'[\mathbf{n}]$. For this reason, we define

$$A[\mathbf{n}] = \min\left\{\max\left\{\frac{0.91 - A'[\mathbf{n}]}{0.66}, 0\right\}, 1\right\} \qquad (6)$$

The max and min operators above ensure that the value of $A[\mathbf{n}]$ stays within the interval $[0, 1]$, where values close to 1 indicate strong structures, while those close to 0 indicate high randomness. This is our proposed measure for the degree of structure, in the neighborhood of location $\mathbf{n}$. Figure 2g shows this final result for the structural measure for the synthetic image in Figure 2a. Similarly, Figure 3c shows the structural measure obtained at each location in the Kodak test image of Figure 3a.

From an implementation point of view, the costliest step in calculating the structural measure lies in (2), which conceptually involves $(2H+1)^2$ complex multiplications and additions for each element of $\mathcal{K}$.[1] Fortunately, however, no multiplications are actually required to implement the summation in (2), since $T[\mathbf{n}] \in \{-1, 0, 1\}$ and the other two terms $(w[\mathbf{p}]\exp(\cdot))$ can be combined into one term that is indexed by $\mathbf{k}$ and $\mathbf{p}$. With this in mind, we need at most $2{\cdot}|\mathcal{K}|{\cdot}(2H+1)^2$ real additions per location, each of which can potentially be implemented in low precision fixed-point arithmetic. The fundamental complexity of a fixed-point addition operation is vastly lower than that of multiplication. Moreover, $T[\mathbf{n}]$ is frequently equal to 0, so that many of these additions can be skipped.

4. EXTENSION TO MULTIPLE RESOLUTIONS

Since structural features are of different sizes, it is useful to estimate the structural measure at different resolutions. Here, we employ the

[1] Generation of the ternary features is very simple, while the total number of operations in (4) and (5) is only on the order of $|\mathcal{K}|$.

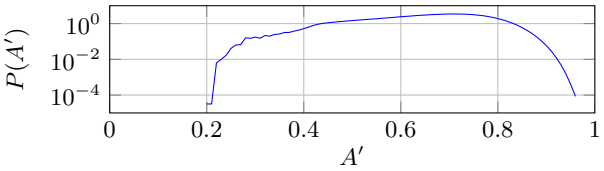


Fig. 4: Empirically-obtained probability density function of $A'[\mathbf{n}]$.

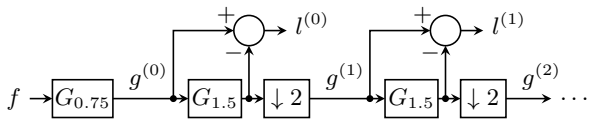


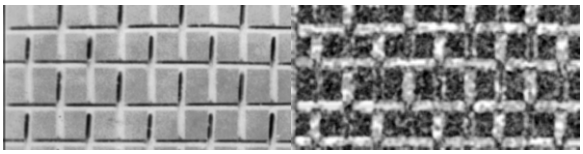
Fig. 5: The Laplacian pyramid generation in this work. The symbol $\downarrow 2$ denotes downsampling by 2, vertically and horizontally.

Laplacian pyramid shown in Figure 5 to obtain detail bands $l^{(d)}[\mathbf{n}]$, where $d = 0, 1, \ldots$. This figure shows that an image f is pre-filtered by $G_{0.75}$ (the same G_σ filter of Section 2 with $\sigma = 0.75$). The purpose of this pre-filtering is to produce $g^{(0)}[\mathbf{n}]$ that has its high frequency filtered in a way similar to all the other $g^{(d)}[\mathbf{n}]$, where $d > 0$; that is, without this step, $l^{(0)}[\mathbf{n}]$ would have more high frequency contents than the other $l^{(d)}[\mathbf{n}]$, where $d > 0$.

Each detail band $l^{(d)}[\mathbf{n}]$, referred to as a mean-removed image in Section 2, can be used to obtained a structural measure at that scale. Figures 3 and 6 show the structural measure at different scales for one image from the Kodak test set and one from the Brodatz texture dataset [7]. At a fine spatial resolution, the structural measure is rather sparse, as can be seen in Figures 3d and 6b.

5. ALTERNATIVE APPROACHES

Here, we present two alternative approaches for the proposed structural measure that serve pedagogical reasons as well.



(a) (b)

(c) (d) (e)

Fig. 6: Multiple scales of the structural measure $A[\mathbf{n}]$. (a) The original image taken from Brodatz texture dataset. (b) $A[\mathbf{n}]$ at full resolution. (c) $A[\mathbf{n}]$ at half the resolution. (d) $A[\mathbf{n}]$ at quarter the resolution zoomed by a factor of 2. (e) $A[\mathbf{n}]$ at one-eighth the resolution zoomed by a factor of 4.

5.1. Inter-Resolution Power Ratio (IRPR)

One initially appealing alternative to the proposed structural measure is to exploit the fact that the power spectrum of a typical image rolls off at least with the square of the frequency, while that of additive white noise is flat.

We define the IRPR as the ratio of the power of the samples in $l^{(d)}[\mathbf{n}]$ in a region around location $\mathbf{n}$ to the power of the samples in the corresponding region in $l^{(d+1)}[\mathbf{n}]$. Since $l^{(d+1)}[\mathbf{n}]$ is from a coarser scale than $l^{(d)}[\mathbf{n}]$, we upsample and interpolate $l^{(d+1)}[\mathbf{n}]$ to the spatial density of $l^{(d)}[\mathbf{n}]$; we write $\tilde{l}^{(d+1)}[\mathbf{n}]$ for this upsampled and interpolated $l^{(d+1)}[\mathbf{n}]$. The IRPR r is then defined by

$$r[\mathbf{n}] = \sum_{\mathbf{p}\in\mathcal{R}_\mathbf{n}} \left(l^{(d)}[\mathbf{p}]\right)^2 \Big/ \sum_{\mathbf{p}\in\mathcal{R}_\mathbf{n}} \left(\tilde{l}^{(d+1)}[\mathbf{p}]\right)^2 \qquad (7)$$

where $\mathcal{R}_\mathbf{n}$ is some region around location $\mathbf{n}$; here, we use a 5×5-pixel region centered around location $\mathbf{n}$.

Figure 7 shows the conditional PDFs of the IRPR given that the region under consideration is structured, denoted by $P(r|\mathcal{S})$, and is unstructured, denoted by $P(r|\mathcal{N})$. The PDF for the structured case $P(r|\mathcal{S})$ is obtained from the Kodak test image set, from regions that have their local variance higher than a certain small threshold. The unstructured case $P(r|\mathcal{N})$ is obtained from an empty image filled with white Gaussian noise. Our IRPR measure then becomes estimating the conditional probability of being structured given the observed IRPR, $P(\mathcal{S}|r)$; that is,

$$P(\mathcal{S}|r) = \frac{P(r|\mathcal{S})}{P(r|\mathcal{S}) + P(r|\mathcal{N})} \qquad (8)$$

assuming that $P(\mathcal{S}) = P(\mathcal{N}) = \frac{1}{2}$, which is a reasonable and commonly-applied assumption.

While the IRPR measure works reasonably well for a synthetic image, as shown in Figure 2i, it performs poorly for a real image, as shown in Figure 8. This poor performance is due to the noise being non-stationary non-white in real images; noise statistics in real images are influenced by demosaicing, Wiener filtering (used in most image capturing devices to reduce noise), among other reasons. This noise is hard to model and, more importantly, does not have the required spectral properties for this approach to work. Another reason is that the IRPR measure for a given scale depends on the next coarser scale; this reduces the spatial resolution of the IRPR measure compared to the proposed structural measure.

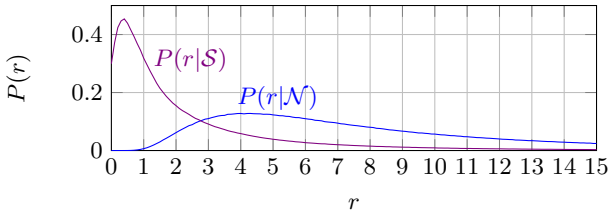


Fig. 7: Empirically-obtained probability density function for IRPR for both structured regions $\mathcal{S}$ and regions predominated by noise $\mathcal{N}$.

(a)

(b) (c)

Fig. 8: A comparison between the structural measure and the power ratio. (a) The original image. (b) Its structural measure. (c) Its IRPR measure.

5.2. Discrete Cosing Transform instead of the DFT

Here, we consider the use of the Discrete Cosine Transform (DCT) in place of the DFT, leaving other aspects of the proposed structual measure intact. The DCT alternative is initially appealing, because it avoids complex arithmetic, including complex magnitude operations. However, this approach has the disadvantage of being anisotropic. This can be explained by remembering that the DCT of a signal is equivalent to the DFT of another signal that is obtained by symmetric extension of the first signal. In two dimensions, symmetric extension changes the orientation of non-vertical and non-horizontal features, creating artificial angles. Figure 2h shows $A'[\mathbf{n}]$ when the DCT is used instead of the DFT in (2); the loss of isotropy should be apparent.

6. CONCLUSIONS

In this work, we have presented two novel and innovative ideas: the ternary feature maps and the structural measure. Ternary feature maps are images whose samples take on values from the set $\{-1, 0, 1\}$; they preserve the predominate features of the original image while being somewhat robust to changes in illumination and noise. The structural measure utilizes the ternary feature map to produce a multi-scale soft isotropic measure of the strength of structure in the original image. The proposed method for the structural measure works better than alternative approaches. We expect a wide variety of applications to benefit from both the ternary feature maps and the structural measure.

7. REFERENCES

[1] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *Communications, IEEE Transactions on*, vol. 31, no. 4, pp. 532 – 540, apr 1983.

[2] D.S. Taubman and M.W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, Boston, 2002.

[3] A.T. Naman, R. Xu, and D. Taubman, "Inter-frame prediction using motion hints," *20th Proc. IEEE Int. Conf. Image Proc. 2013*, September 2013, Accepted for publication.

[4] R. Xu and D. Taubman, "Robust dense block-based motion estimation using a 2-bit transform on a laplacian pyramid," *20th Proc. IEEE Int. Conf. Image Proc. 2013*, September 2013, Accepted for publication.

[5] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," *Proc. of the 12th IAPR Int. Conf. on Pattern Recognition*, vol. 1, pp. 582–585, October 1994.

[6] V.K. Goyal, "Theoretical foundations of transform coding," *Signal Processing Magazine, IEEE*, vol. 18, no. 5, pp. 9 –21, September 2001.

[7] P. Brodatz, *Textures: A photographic album for artists and designers*, Dover Publications, New York, 1966.