2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW) | 979-8-3503-0261-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSPW59220.2023.10193606

# TEXTURE QUALITY CRITERIA COMPARISON

*Michal Haindl*

Institute of Information Theory and Automation
Czech Academy of Sciences
Pattern Recognition Department
182 08 Prague, Czechia

*Nahidbanu Shaikh*

Faculty of Information Technology
Czech Technical University
Department of Theoretical Computer Science
160 00 Prague, Czechia

## ABSTRACT

Visual scene recognition or modeling predominantly uses visual textures representing an object's material properties. However, the single material texture varies in scale and illumination angles due to mapping an object's shape. We present a comparative study of thirteen possible texture quality criteria and show the superior performance of two multispectral measures derived from the Markovian descriptive model.

***Index Terms***— Texture quality criteria, Spearman correlation, Human quality ranking, Texture quality benchmark.

## 1. INTRODUCTION

A human observer recognizes a visual scene using shape and material attributes. Unfortunately, the surface material's appearance vastly changes under variable observation conditions, negatively affecting its automatic and reliable recognition in numerous artificial intelligence applications. Consequently, most material modeling or recognition attempts apply unnaturally restricted observation conditions [1]. Image quality criteria can use three types of knowledge [2] - knowledge about the reference image (full/reduced-reference FR/RR), knowledge about the degradation procedure, and knowledge about the human visual system (HVS). However, our understanding of the human visual system, and thus the computational HVS modeling used in image quality assessment (IQA) criteria, needs to be more sketchy and complete [3]. The quality assessment criteria can be categorized into full-reference (FR), reduced-reference (RR), and no-reference (NR) groups. The other main categorization is based on the human evaluator's involvement in a perceptual (subjective) and no human evaluator (objective).

Evaluation of how well various texture models $Y$ differ from actual texture measurement $X$ and thus conform with human visual perception is essential for assessing the similarities between model output and the original measured texture and for optimal settings of model parameters, for fair comparison of distinct models, etc. Currently, the only reliable but extraordinarily impractical and expensive option is to exploit the methods of visual psychophysics. Perceptual assessment is naturally preferable and accurate because the human visual system is the ultimate receiver of visual information in most visual applications. However, the psycho-physical methods [4] require carefully designed experiments on human subjects using highly controlled visual stimuli and viewing conditions. Thus it is a lengthy process of experiment design, tightly controlled laboratory conditions, and a representative panel of human testing subjects. Such testing cannot be directly embedded into a practical daily performed system. Nevertheless, the perceptual assessment is necessary for the verification quality criteria development because the performance of any automatic visual quality assessment criterion can be confirmed by measuring its correlation with human perception.

An automatic texture fidelity verification is needed as the only practical solution to evaluate the quality of texture-generating algorithms, database texture retrieval, enhancement, compression, denoising, coding, transmission, and many others in practical applications and algorithm development. Texture quality modeling belongs mainly to the full-reference image quality criteria category because the target or measured material is available. However, the primary difference from most FR image quality criteria is that those texture quality subset criteria cannot require pixelwise correspondence between tested and target images. Several survey papers were published [5, 6, 7, 3, 8, 2, 9, 10, 11]. However, they are mostly restricted to images where the pixelwise correspondence with an optimal pattern exists. [5] list several reduced/no-references quality assessment methods. [6] reviews perceptual visual quality metrics, just-noticeable distortion, visual attention, quality assessment databases, and common feature and artifact detection. Few studies target texture quality assessment [12, 13, 14, 15, 16, 17, 18].

This paper's contribution and novelty is a joint test of texture quality assessment criteria for textural models to simulate realistic visual scene recognition conditions. Moreover, we present a comparative analysis with thirteen alternative textural quality criteria. For this analysis, we use the unique UTIA Texture Fidelity Benchmark [12].

Thanks to the Czech Science Foundation project 19-12340S for funding.

## 2. IMAGE AND TEXTURE QUALITY MEASURES

The significant difference between image quality and texture quality measures is in the non-existing pixelwise correspondence for textures. Thus simple measures like the L measures cannot be used directly on identically indexed pixels. $X, Y$ denotes the original (measured) and compared (e.g., synthetic) multispectral texture, $y_{r_3}, x_{r_3}$ $r_3-$th monospectral texture component ($x_{\bullet,\bullet,i} = x_i$), $\mu_X, \mu_Y$ mean values, and $\sigma_{x_{r_3}}, \sigma_{y_{r_3}}$ standard deviations, $\sigma_{x_{r_3} y_{r_3}}$ is the sample cross-correlation of $x_{r_3}$ and $y_{r_3}$ after removing their means, $\bullet$ are all corresponding indices, $r = \{r_1, r_2, r_3\}$ is a multiindex with the row, column, and spectral indices. Most of the quality measures consider only gray-scale images. In this case, we generalize them to multispectral images using their average overall spectra (denoted in (1)-(3) but further avoided to simplify notation). Although the signal fidelity measures (SNR1, MSE, PSNR) have clear physical meaning; they do not correlate well with human visual perception [19, 20] and violate the non-pixelwise requirement. The signal-to-noise-ratio (SNR) is defined

$$\text{SNR1}(\mathbf{X},\mathbf{Y}) = \frac{1}{d}\sum_{i=1}^{d} \frac{E\{x_i\}}{E\{|x_i - y_i|\}} \ , \qquad (1)$$

$$\text{SNR2}(\mathbf{X},\mathbf{Y}) = \frac{1}{d}\sum_{i=1}^{d} E\left\{\frac{E\{x_i\}}{|y_i|}\right\} \ . \qquad (2)$$

The mean-squared error (MSE) [20] multispectral criterion is

$$\text{MSE}(\mathbf{X},\mathbf{Y}) = \frac{1}{MNd}\sum_{r_1=1}^{M}\sum_{r_2=1}^{N}\sum_{r_3=1}^{d}(\mathbf{X_r} - \mathbf{Y_r})^2 \ , \qquad (3)$$

where $M$ number of rows, $N$ number of columns, $d$ number of spectra. The peak signal-to-noise ratio (PSNR) [21] is the ratio between the maximum possible power of an image and the power of corrupting noise.

$$\text{PSNR} = 10\ \log_{\mathbf{10}}\left(\frac{\max\{y_{\bullet,\bullet,\bullet}, x_{\bullet,\bullet,\bullet}\}^2}{MSE}\right) \ . \qquad (4)$$

PSNR is measured in decibels (dB).

The structural similarity indices (SSIM, MSSIM, STSIM) are based on the assumption that structural information about an image can be described by a function (usually a simple multiplication) of three terms: luminance $l$, contrast $c$ and structure $s$. The structural similarity (SSIM) index [22] is:

$$\text{SSIM}(\mathbf{x_i},\mathbf{y_i}) = l(\mathbf{x_i},\mathbf{y_i})c(\mathbf{x_i},\mathbf{y_i})s(\mathbf{x_i},\mathbf{y_i}) = \qquad (5)$$
$$\left(\frac{2\mu_{\mathbf{x_i}}\mu_{\mathbf{y_i}} + C_1}{\mu_{\mathbf{x_i}}^2 + \mu_{\mathbf{y_i}}^2 + C_1}\right)\left(\frac{2\sigma_{\mathbf{x_i}}\sigma_{\mathbf{y_i}} + C_2}{\sigma_{\mathbf{x_i}}^2 + \sigma_{\mathbf{y_i}}^2 + C_2}\right)\left(\frac{\sigma_{\mathbf{x_i}\mathbf{y_i}} + C_3}{\sigma_{\mathbf{x_i}}\sigma_{\mathbf{y_i}} + C_3}\right) \ ,$$

where $C_1, C_2, C_3$ are small positive constants that stabilize each term. Single terms rightwards in (5) measure image luminance, contrast, and structure similarity, respectively. SSIM considers different distortions to have the same importance as visual perception, which is only a rough approximation of their different importance in reality.

$$MSSIM\,(x_i, y_i) = \frac{1}{n}\sum_{j=1}^{n} SSIM\left({}^j x_i, {}^j y_i\right) \ , \qquad (6)$$

where $n$ is the number of subblocks in images where SSIM is locally evaluated. The STSIM criteria [23] (STSIM-1, STSIM-2, and STSIM-M) are based on statistics computed for each texture subband factor. STSIM-1 is created from CW-SSIM [24] by replacing the 'structural' term with terms that compare first-order auto-correlations of corresponding subband coefficients $\rho_X^m(0,1)$ in the horizontal and $\rho_X^m(1,0)$ in the vertical direction. $x_{r,i}$ is a pixel at location $r \in I$, where $I$ is discrete two dimensional rectangular lattice, the multiindex $r = [r_1, r_2]$ is composed of $r_1$ row and $r_2$ column index, respectively. In the equations for a single subband $m$, the $p$ is typically set to 1,

$$\text{STSIM-1}^m(x_i, y_i) = \left(l^m_{x_i,y_i}\right)^{\frac{1}{4}}\left(c^m_{x_i,y_i}\right)^{\frac{1}{4}}\left(c^m_{x_i,y_i}(0,1)\right)^{\frac{1}{4}}$$
$$\left(c^m_{x_i,y_i}(1,0)\right)^{\frac{1}{4}} \ , \qquad (7)$$
$$c^m_{x_i,y_i}(0,1) = 1 - 0.5\left|\rho^m_{x_i}(0,1) - \rho^m_{y_i}(0,1)\right|^p \ ,$$
$$\rho^m_{x_i}(0,1) = \frac{E\left\{\left[x^m_{r,i} - \mu^m_{x_i}\right]\left[x^m_{r-1,r_2+1,i} - \mu^m_{x_i}\right]^*\right\}}{(\sigma^m_{x_i})^2} \ ,$$

where $l^m_{x_i,y_i}, c^m_{x_i,y_i}$ are defined in (5). STSIM-2 adds crossband correlation coefficient $\rho^{m,n}_{|X|}(0,0)$ between subbands $m, n$

$$\text{STSIM-2}(X,Y) = \frac{\sum_{m=1}^{N_b}\text{STSIM-1}^m(X,Y) + \sum_{i=1}^{N_c} c^{m_i,n_i}_{X,Y}}{N_b + N_c} \ ,$$
$$c^{m,n}_{X,\,Y} = 1 - 0.5\left|\rho^{m,n}_{|X|}(0,0) - \rho^{m,n}_{|Y|}(0,0)\right|^p \ ,$$
$$\rho^{m,n}_{|X|}(0,0) = \frac{E\left\{\left[X^m_r - \mu^m_{|X|}\right]\left[X^n_r - \mu^n_{|X|}\right]\right\}}{\sigma^m_{|X|}\sigma^n_{|X|}} \ , \qquad (8)$$

where $N_b$ is the number of subbands and $N_c$ is the number of possible crossband correlations.

Visual information fidelity (VIF) method [25] is defined as the ratio of the summed mutual information

$$\text{VIF} = \frac{I(C;F|x)}{I(C;E|x)} = \frac{\sum_{i=1}^{N} I(c_i;f_i|x_i)}{\sum_{i=1}^{N} I(c_i;e_i|x_i)} \ , \qquad (9)$$

$$I(c_i;e_i|x_i) = \frac{1}{2}\log\frac{|x_i^2 C_U + \sigma_n^2\mathbf{I}|}{|\sigma_n^2\mathbf{I}|} = \frac{1}{2}\sum_{j=1}^{M}\log\left(1 + \frac{x_i^2\lambda_j}{\sigma_n^2}\right),$$
$$I(c_i;f_i|x_i) = \frac{1}{2}\log\frac{|g_i^2 x_i^2 C_U + (\sigma_v^2 + \sigma_n^2)\mathbf{I}|}{|(\sigma_v^2 + \sigma_n^2)\mathbf{I}|}$$
$$= \frac{1}{2}\sum_{j=1}^{M}\log\left(1 + \frac{g_i^2 x_i^2\lambda_j}{\sigma_v^2\sigma_n^2}\right) \ ,$$

where $E$ and $F$ are models in a wavelet domain for what human visual system (HVS) captures from original and test images, respectively, $N$ is the number of subbands, $C_U$ is a covariance matrix (without considering noise and scale factors) of $E$, $\lambda_j$ is the $j$-th eigenvalue of $C_U$, $x$ is a realization of an original image, and $g$ is an attenuation factor.

Assuming Gaussianity of the wavelet decomposition $\tilde{X}, \tilde{Y}$ of textures $X, Y$, IFC is the mutual information between the source and the distorted images [26]:

$$IFC(X,Y) = \frac{1}{2} \sum_{\forall d} \sum_{\forall r \in I} \log_2 \left( 1 + \frac{g^2_{r,d} s^2_{r,d} \sigma^2_U}{\sigma^2_V} \right) \quad (10)$$

where $d$ is a subband of the wavelet decomposition $\tilde{X}, \tilde{Y}$, $g_i$ is the attenuation factors that capture the loss of signal energy in a $i-$th subband to the blur distortion, $\sigma^2_U$ is the variance of a Gaussian scalar RF with mean zero, $\sigma^2_V$ is white Gaussian RF variance. $\sigma^2_U = 1$ is assumed to be unity, $s^2_{r,d}$ is estimated from localized sample variance estimation.

Deep image structure and texture similarity index (DISTS) [18] combines the quality measurements from different convolution layers using a weighted sum:

$$\begin{aligned} DISTS(X,Y,\alpha,\beta) &= 1 - \sum_{i=0}^{m} \sum_{j=1}^{n_i} \Big( \alpha_{ij} l(\tilde{X}^{(i)}_j, \tilde{Y}^{(i)}_j) \\ &+ \beta_{ij} s(\tilde{X}^{(i)}_j, \tilde{Y}^{(i)}_j) \Big) \ , \quad (11) \\ l(\tilde{X}^{(i)}_j, \tilde{Y}^{(i)}_j) &= \frac{2\mu^{(i)}_{\tilde{X}_j} \mu^{(i)}_{\tilde{Y}_j} + c_1}{\left(\mu^{(i)}_{\tilde{X}_j}\right)^2 + \left(\mu^{(i)}_{\tilde{Y}_j}\right)^2 + c_1} \ , \quad (12) \\ s(\tilde{X}^{(i)}_j, \tilde{Y}^{(i)}_j) &= \frac{2\sigma^{(i)}_{\tilde{X}_j \tilde{Y}_j} + c_2}{\left(\sigma^{(i)}_{\tilde{X}_j}\right)^2 + \left(\sigma^{(i)}_{\tilde{Y}_j}\right)^2 + c_2} \ , \quad (13) \end{aligned}$$

where $\alpha_{ij}, \beta_{ij}$ are positive learnable weights, satisfying $\sum_{i=0}^{m} \sum_{j=1}^{n_i} (\alpha_{ij} + \beta_{ij}) = 1$. DISTS is insensitive to mild local and global geometric distortions, an advantageous property for textures.

The visual signal-to-noise ratio (VSNR) [21] is a two-stage approach. In the first stage, contrast thresholds for the detection of distortions are computed in order to determine whether the distortions in the distorted image is visible. The threshold contrast is used in the second step to compute contrast detection thresholds.

$$VSNR = 20 \log_{10} \left( \frac{C(X)}{\alpha d_{pc} + (1-\alpha) \frac{d_{gp}}{\sqrt{2}}} \right) \quad (14)$$

where $C(X)$denotes the RMS contrast of the original image $X$ given by $C(X) = \sigma_{L(X)} / \mu_{L(X)}$.

Markovian texture fidelity criteria ($\zeta(X,Y)$, CPM) were proposed in [13, 14]. The criterion $\zeta$ measures cross-prediction error when using data from the original texture

| texture no. | $\zeta$ | CPM | STSIM-1 | STSIM-2 |
|---|---|---|---|---|
| 1 | 0.866 | 0.765 | 0,598 | 0,539 |
| 2 | 0.797 | 0.706 | 0,301 | 0,512 |
| 3 | 0.801 | 0.723 | -0,179 | -0,130 |
| 4 | 0.363 | 0.797 | -0,808 | -0,725 |
| 5 | 0.721 | 0.782 | -0,620 | -0,626 |
| 6 | 0.600 | 0.859 | -0,846 | -0,793 |
| 7 | 0.670 | 0.534 | -0,436 | -0,514 |
| 8 | 0.566 | 0.711 | 0,475 | 0,414 |
| 10 | 0.909 | 0.797 | -0,069 | -0,007 |
| 12 | 0.794 | 0.745 | -0,802 | -0,802 |
| 14 | 0.456 | 0.560 | -0,704 | -0,646 |
| $median\{\rho\}$ | 0.721 | **0.745** | -0,436 | -0,514 |
| $\sigma_\rho$ | 0,17 | **0,09** | 0,52 | 0,51 |

**Table 1**. Spearman's rank correlation between the human rank and the single criteria results. Textures 1–7 are color images, the remaining are gray-scale.

$X$ and estimated parameters $\widetilde{\gamma}$ from the 3-dimensional causal auto-regressive (3DCAR) texture model $Y$:

$$\zeta(X,Y) = \frac{1}{|I|} \sum_{\forall r \in I} |X_r - \widetilde{\gamma}_{r-1} Z_r| \ , \quad (15)$$

where $\widetilde{\gamma}_{r-1} = [A_{1,r-1}, \ldots, A_{\eta,r-1}]$ is the $d \times d\,\eta$ estimated parameter matrix with square sub-matrices $A_{s,r-1}$ and $Z_r$ is the $d\eta \times 1$ data vector with multiindices $r, s$, see [27] for details. CPM measure [14] uses twice downsampled textures by a factor of two, which are subsequently upsampled back to the original size and combined these images with the original one together, essentially creating a 9-spectral $(d = 9)$ image. The CPM measures the difference between prediction and cross-prediction:

$$\mathrm{CPM}(X,Y) = \max \left\{ \beta(X,\gamma,\tilde{\gamma}), \tilde{\beta}(Y,\gamma,\tilde{\gamma}) \right\}, \quad (16)$$

$$\beta(X,\gamma,\tilde{\gamma}) = \frac{1}{2^l d} \sum_{i=1}^{d} \frac{\sum_{\forall r \in I_{\{s\}}} (\tilde{\gamma}_{r-1} Z_r - \gamma_{r-1} Z_r)}{|I_{\{s\}}|} \ ,$$

$$\tilde{\beta}(Y,\gamma,\tilde{\gamma}) = \frac{1}{2^l d} \sum_{i=1}^{d} \frac{\sum_{\forall r \in I_{\{s\}}} \left(\gamma_{r-1} \tilde{Z}_r - \tilde{\gamma}_{r-1} \tilde{Z}_r\right)}{|I_{\{s\}}|} \ ,$$

where $l$ is the number of bits per spectral band, $I_{\{s\}} \subset I$ is some window identical on both textures, $\gamma$ are parameters from the original texture $X$, while $\tilde{Y}, \tilde{Z}_r, \widetilde{\gamma}$ are data and parameters from the synthetic (or compared) $i-$th texture. Due to the pixel range normalization $\mathrm{CPM}(X,Y) \in \langle 0; 1 \rangle$ with 0 being the best value.

A universal objective image quality index (IQI) [28] models an image distortion as a combination of loss of correlation, luminance distortion, and contrast distortion:

$$IQI(X,Y) = \frac{4\sigma_{XY} \mu_X \mu_Y}{(\sigma^2_X + \sigma^2_Y)(\mu^2_X \mu^2_Y)} \ . \quad (17)$$

| Pearson correlation | MSE | PSNR | SSIM | MSSIM | SNR1 | SNR2 | IQI | VIF | VSNR | IFC | NQM | STSIM1 | STSIM2 | Spearman correlation | $\sigma$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PSNR | -0,193 | | | | | | | | | | | | | 0,058 | 0,32 |
| SSIM | -0,819 | 0,064 | | | | | | | | | | | | -0,125 | 0,35 |
| MSSIM | -0,767 | -0,026 | 0,861 | | | | | | | | | | | -0,127 | 0,29 |
| SNR1 | -0,918 | -0,021 | 0,849 | 0,852 | | | | | | | | | | -0,228 | 0,31 |
| SNR2 | -0,322 | -0,015 | 0,373 | 0,324 | 0,643 | | | | | | | | | -0,301 | 0,26 |
| IQI | -0,848 | 0,061 | 0,577 | 0,608 | 0,583 | 0,242 | | | | | | | | -0,203 | 0,38 |
| VIF | -0,368 | 0,050 | 0,227 | 0,390 | 0,257 | 0,199 | 0,349 | | | | | | | -0,176 | 0,17 |
| VSNR | -0,013 | -0,392 | 0,201 | 0,415 | 0,231 | 0,193 | 0,422 | 0,140 | | | | | | 0,069 | 0,40 |
| IFC | -0,404 | 0,019 | 0,321 | 0,483 | 0,319 | 0,206 | 0,378 | 0,997 | 0,201 | | | | | -0,164 | 0,20 |
| NQM | -0,861 | 0,027 | 0,707 | 0,815 | 0,811 | 0,313 | 0,930 | 0,478 | 0,635 | 0,481 | | | | -0,037 | 0,42 |
| STSIM1 | -0,397 | 0,026 | 0,308 | 0,121 | 0,475 | 0,114 | 0,050 | 0,155 | 0,016 | 0,140 | 0,194 | | | -0,436 | 0,52 |
| STSIM2 | -0,363 | 0,019 | 0,287 | 0,120 | 0,423 | -0,078 | -0,064 | 0,155 | 0,070 | 0,144 | 0,036 | 0,968 | | -0,514 | 0,51 |
| DISTS | 0,298 | -0,075 | -0,463 | -0,163 | -0,269 | 0,001 | -0,248 | -0,356 | 0,101 | -0,278 | -0,072 | -0,586 | -0,652 | 0,551 | 0,28 |
| MSE | | | | | | | | | | | | | | 0,067 | 0,30 |

**Table 2**. The quality criteria Pearson's correlation over all fourteen material series between measured and modeled texture (uncorrelated criteria on the significance level 0,05) and Spearman's correlation with the human ranking and its standard deviation.

The noise quality measure (NQM) [29] takes into account the variation, in contrast, sensitivity with distance, image dimensions, and spatial frequency, variation in the local luminance mean, contrast interaction between spatial frequencies, and contrast masking effects:

$$NQM(X,Y) = 10\log_{10}\left(\frac{\sum_{\forall r} X_r^2}{\sum_{\forall r}(X_r - Y_r)^2}\right) \quad . \qquad (18)$$

## 3. RESULTS

We used the texture fidelity benchmark [12] for the validation of texture quality criteria. The benchmark uses six natural and one synthetic color texture together with their grayscale versions. Textures were mathematically synthesized using various mathematical models and variable quality constraints. The benchmark contains quality-ranked texture series by human observers, so ground truth is available. Spearman's rank correlation measures the prediction monotonicity, while Pearson correlation measures linearity and consistency. Tab. 1 shows the Spearman rank correlation for a single benchmark textures series, the median correlation, and the standard deviation for two Markovian and two STSIM criteria, respectively. The Markovian criteria better agree with the human quality ranking and minor standard deviation. Tab. 2 illustrates alternative tested criteria Pearson correlation and their Spearman correlation. The table shows a strong linear relation between (MSE, SSIM, SNR1, IQI, NQM), (SSIM, MSSIM, SNR1), (MSSIM, NQM), (IQI, NQM), (VIF, IFC), and (STSIM1, STSIM2) groups of criteria. The Spearman correlation between human and criteria ranking (Tab. 2) shows only STSIM2 and DISTS to be correlated with human ranking on the 5% significance level.

## 4. CONCLUSION

The results indicate that both Markovian criteria, based on the Markovian descriptive model, are the most robust textural criteria for texture quality evaluation $\zeta$ and CPM criteria are both correlated with human ranking on the 1% significance level and can be efficiently, recursively, and adaptively learned. The significant advantage of Markovian criteria is their multispectral nature, contrary to the majority of possible alternative quality measures. Only the alternative criteria STSIM2 and DISTS are correlated but only on the 5% significance level. The other advantage of Markovian criteria is their more minor standard deviation than STSIM2 and DISTS.

## 5. REFERENCES


[1] Sean Bell, Paul Upchurch, Noah Snavely, and Kavita Bala, "Material recognition in the wild with the materials in context database," in *Proc. IEEE CVPR*, 2015, pp. 3479–3487.

[2] Lihuo He, Fei Gao, Weilong Hou, and Lei Hao, "Objective image quality assessment: a survey," *International Journal of Computer Mathematics*, vol. 91, no. 11, pp. 2374–2388, 2014.

[3] Damon M Chandler, "Seven challenges in image quality assessment: past, present, and future research," *International Scholarly Research Notices*, vol. 2013, 2013.

[4] Michal Haindl and Jiří Filip, *Visual Texture*, Advances in Computer Vision and Pattern Recognition. Springer-Verlag London, London, January 2013.

[5] Zhou Wang and Alan C Bovik, "Reduced-and no-reference image quality assessment," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 29–40, 2011.

[6] Weisi Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[7] Anush Krishna Moorthy and Alan Conrad Bovik, "Visual quality assessment algorithms: what does the future hold?," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 675–696, 2011.

[8] Rajiv Soundararajan and Alan C Bovik, "Survey of information theory in visual quality assessment," *Signal, Image and Video Processing*, vol. 7, no. 3, pp. 391–401, 2013.

[9] Redzuan Abdul Manap and Ling Shao, "Non-distortion-specific no-reference image quality assessment: A survey," *Information Sciences*, vol. 301, pp. 141–160, 2015.

[10] Shaoping Xu, Shunliang Jiang, and Weidong Min, "No-reference/blind image quality assessment: a survey," *IETE Tech. Review*, vol. 34, no. 3, pp. 223–245, 2017.

[11] Guangtao Zhai and Xiongkuo Min, "Perceptual image quality assessment: a survey," *Science China Information Sciences*, vol. 63, no. 11, pp. 1–52, 2020.

[12] M. Haindl and M.Jr. Kudělka, "Texture fidelity benchmark," in *IWCIM 2014*, Los Alamitos, November 2014, pp. 1 – 5, IEEE Computer Society CPS.

[13] Miloš Kudělka and Michal Haindl, "Texture fidelity criterion," in *2016 IEEE ICIP*. September 2016, pp. 2062 – 2066, IEEE.

[14] Miloš Kudělka and Michal Haindl, "Multispectral texture fidelity measure," in *Proc. SITIS 2018*, Los Alamitos, November 2018, pp. 658 – 663, IEEE CPS.

[15] Fei Zhou, Rongguo Yao, Bozhi Liu, and Guoping Qiu, "Visual quality assessment for super-resolved images: Database and method," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3528–3541, 2019.

[16] Zhisen Tang, Yuanlin Zheng, Wei Wang, and Kaiyang Liao, "Image quality assessment by combining masking texture and perceptual color difference model," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 14, no. 7, pp. 2938–2956, 2020.

[17] J. Rajevenceltha and Vilas H. Gaidhane, "An efficient approach for no-reference image quality assessment based on statistical texture and structural features," *Engineering Science and Technology, an International Journal*, vol. 30, pp. 101039, 2022.

[18] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli, "Image quality assessment: Unifying structure and texture similarity," *IEEE Trans. PAMI*, vol. 44, no. 5, pp. 2567–2581, 2022.

[19] Dihong Tian and Ghassan AlRegib, "Fqm: a fast quality measure for efficient transmission of textured 3d models," in *Proceedings of the 12th annual ACM international conference on Multimedia*, 2004, pp. 684–691.

[20] Zhou Wang and Alan C Bovik, "Mean squared error: love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[21] Damon M Chandler and Sheila S Hemami, "Vsnr: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.

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

[23] J. Zujovic, T.N. Pappas, and D.L. Neuhoff, "Structural texture similarity metrics for image analysis and retrieval," *IEEE Trans. Image Processing*, vol. 22, no. 7, pp. 2545–2558, 2013.

[24] Zhou Wang and Eero P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *Proc. IEEE ICASSP '05*, 2005, pp. 573–576.

[25] Hamid R Sheikh and Alan C Bovik, "Image information and visual quality," *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[26] Hamid R Sheikh, Alan C Bovik, and Gustavo De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Processing*, vol. 14, no. 12, pp. 2117–2128, 2005.

[27] Michal Haindl, "Visual data recognition and modeling based on local markovian models," in *Mathematical Methods for Signal and Image Analysis and Representation*, vol. 41, chapter 14, pp. 241–259. Springer London, 2012, 10.1007/978-1-4471-2353-8_14.

[28] Zhou Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, March 2002.

[29] Niranjan Damera-Venkata, Thomas D Kite, Wilson S Geisler, Brian L Evans, and Alan C Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 636–650, 2000.