TEXTURE FIDELITY CRITERION

Miloš Kudělka Jr. *Michal Haindl*

The Institute of Information Theory and Automation
of the Czech Academy of Sciences
182 08 Prague, Czech Republic

ABSTRACT

Visual texture fidelity evaluation is important but still unsolved problem. Evaluation of how well various texture models conform with human visual perception of their original measured pattern is required not only for assessing the visual dissimilarities between a model output and the original measured texture, but also for optimal settings of model parameters, for fair comparison of distinct models, or visual scene understanding. We propose a novel texture fidelity criterion based on the fully multi-spectral generative underlying Markovian texture model, which correlates well with human texture quality ranking verified on the texture fidelity benchmark.

Index Terms— Image texture, texture fidelity, visual quality criterion, Markovian texture model

1. INTRODUCTION

Evaluation of how well various texture models conform with human visual perception is important not only for assessing the similarities between a model output and the original measured texture, but also for optimal settings of model parameters, for fair comparison of distinct models, etc. Few published criteria allow to test selected texture properties such as the texture regularity [1], etc. Others claim to test general texture quality [2, 3, 4]. We have recently tested [5] several state-of-the-art image quality measures and several recently published texture criteria. The tested criteria were - the mean-squared error (MSE) [6], the visual signal-to-noise-ratio [7] (VSNR), the structural similarity (SSIM) index [8], the complex wavelet - structural similarity (CW-SSIM) index [9], the visual information fidelity (VIF) methods [10], and the structural texture similarity measure (STSIM-1, STSIM-2, STSIM-M) [3]. All of these criteria consider only gray-scale images. The results have demonstrated [5] that the standard image quality criteria (MSE, VSNR, VIF, SSIM, CW-SSIM) cannot be used for texture quality validation because they do not correlate with human quality assessment at all. Similarly the STSIM texture criteria, while performing better, do not successfully solve this problem and new measures are clearly needed. Among these published criteria, the STSIM texture criteria have significantly higher correlation with human ranking [5], thus we use them for comparison with our novel proposed texture fidelity criterion. Currently the only reliable, but extremely impractical and expensive option, is to exploit the methods of visual psycho-physics. The psychophysical methods [11] require a lengthy process of experiment design, tightly controlled laboratory condition, and representative panel of human testing subjects. Such testing obviously cannot be performed on a daily basis.

Thanks to the Czech Science Foundation Project No. GAČR 14-10911S for funding.

1.1. Structural Texture Similarity Measure

The STSIM criteria are based on a set of statistics computed for each texture subband factor. They are extensions of the CW-SSIM criterion in three versions, STSIM-1, STSIM-2 and STSIM-M [3]. STSIM-1 is created from CW-SSIM [9] by replacing the 'structural' term with terms that compare first-order auto-correlations of corresponding subband coefficients $\rho^m_Y(0,1)$ in the horizontal and $\rho^m_Y(1,0)$ in the vertical direction. $Y, \widetilde{Y}$ are the target and a compared textures, respectively. Y_r 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,

$$\text{STSIM-1}^m(Y,\widetilde{Y}) = \tag{1}$$
$$= \left(l^m_{Y,\widetilde{Y}}\right)^{1/4} \left(c^m_{Y,\widetilde{Y}}\right)^{1/4} \left(c^m_{Y,\widetilde{Y}}(0,1)\right)^{1/4} \left(c^m_{Y,\widetilde{Y}}(1,0)\right)^{1/4} ,$$
$$l^m_{Y,\widetilde{Y}} = \left(\frac{2\mu_Y\mu_{\widetilde{Y}} + C_1}{\mu^2_Y + \mu^2_{\widetilde{Y}} + C_1}\right) ,$$
$$c^m_{Y,\widetilde{Y}}(0,1) = 1 - 0.5\left|\rho^m_Y(0,1) - \rho^m_{\widetilde{Y}}(0,1)\right|^p ,$$
$$\rho^m_Y(0,1) = \frac{E\left\{\left[Y^m_r - \mu^m_Y\right]\left[Y^m_{r_1,r_2+1} - \mu^m_Y\right]^*\right\}}{(\sigma^m_Y)^2} ,$$

where $l^m_{Y,\widetilde{Y}}$ is a luminace, and C_1 is a small positive constant. STSIM-2 adds cross-band correlation coefficient $\rho^{m,n}_{|Y|}(0,0)$

between subbands m and n

$$\text{STSIM-2}(Y,\widetilde{Y}) = \tag{2}$$
$$= \frac{\sum_{m=1}^{N_b} \text{STSIM-1}^m(Y,\widetilde{Y}) + \sum_{i=1}^{N_c} c_{Y,\widetilde{Y}}^{m_i,n_i}(0,0)}{N_b + N_c} \ ,$$
$$c_{Y,\widetilde{Y}}^{m,n}(0,0) = 1 - 0.5\left|\rho_{|Y|}^{m,n}(0,0) - \rho_{|\widetilde{Y}|}^{m,n}(0,0)\right|^p \ ,$$
$$\rho_{|Y|}^{m,n}(0,0) = \frac{E\left\{\left[Y_r^m - \mu_{|Y|}^m\right]\left[Y_r^n - \mu_{|Y|}^n\right]\right\}}{\sigma_{|Y|}^m \sigma_{|Y|}^n} \ ,$$

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

STSIM-M (STSIM-Mahalanobis) chooses another approach. Rather than combining aforementioned terms into a single measure, it uses them to create feature vectors f_Y and $f_{\widetilde{Y}}$ and then calculates the *Mahalanobis distance* between the feature vectors:

$$\text{STSIM-M}(Y,\widetilde{Y}) = \sqrt{\sum_{i=1}^{N_p} \frac{(f_{Y,i} - f_{\widetilde{Y},i})^2}{\sigma_{fi}^2}} \ , \tag{3}$$

where σ_{fi}^2 is the standard deviation of the i-th feature across all feature vectors in the set. Therefore, to compute the distance between two textures, STSIM-M requires statistics based on the whole set and the results are relative only to the set, which is unfavourable for our cause and therefore the STSIM-M was not included in our tests.

2. TEXTURE FIDELITY CRITERION

The proposed texture fidelity criterion is derived from the multi-spectral generative textural model from the wide-sense Markovian random field family.

2.1. Textural Model

Let us assume that multispectral texture image is composed of d spectral planes (usually $d = 3$ for colour images). $Y_r = [Y_{r,1}, \ldots, Y_{r,d}]^T$ is the multispectral 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. The spectral planes are modelled by 3-dimensional CAR Causal Auto-regressive Random (3DCAR) model. The CAR representation assumes that the multispectral texture pixel Y_r can be modelled as a linear combination of its neighbours:

$$Y_r = \gamma Z_r + \epsilon_r \ , \qquad Z_r = [Y_{r-s}^T : \forall s \in I_r]^T \tag{4}$$

where Z_r is the $d\eta \times 1$ data vector with multiindices r, s, $\gamma = [A_1, \ldots, A_\eta]$ is the $d \times d\,\eta$ unknown parameter matrix with square sub-matrices A_s. Some selected contextual causal or unilateral neighbour index shift set is denoted I_r and $\eta = cardinality(I_r)$. A unilateral neighbourhood I_r (the left upper orientation) is defined as $I_r \subset I_r^U = \{s : s_1 < r_1 \ or \ (s_1 = r_1, s_2 < r_2)\}$ and similarly ([12]) its subset - the causal neighborhood. The neighborhood order is based on the Euclidean distance from r. The white noise vector ϵ_r has normal density with zero mean and unknown covariance matrix Σ, same for each pixel. The texture is analysed in a chosen direction, where multiindex t changes according to the movement on the image lattice. Parameter estimation of a 3DCAR model using either the maximum likelihood, or the least square or Bayesian methods can be found analytically. The Bayesian parameter estimates $\hat{\gamma}$ of the 3DCAR model using the normal-gamma parameter prior, given the known history of the CAR process $Y^{(t-1)} = \{Y_{t-1}, Y_{t-2}, \ldots, Y_1, Z_t, Z_{t-1}, \ldots, Z_1\}$ for the given pixel position can be computed using statistics [12]:

$$\hat{\gamma}_{t-1}^T = V_{zz(t-1)}^{-1} V_{zy(t-1)} \ , \tag{5}$$
$$\hat{\Sigma}_{r-1} = \frac{\lambda_{(r-1)}}{\beta(r) - d\eta + d + 1} \ , \tag{6}$$
$$V_{t-1} = \begin{pmatrix} \sum_{u=1}^{t-1} Y_u Y_u^T & \sum_{u=1}^{t-1} Y_u Z_u^T \\ \sum_{u=1}^{t-1} Z_u Y_u^T & \sum_{u=1}^{t-1} Z_u Z_u^T \end{pmatrix} + V_0$$
$$= \begin{pmatrix} V_{yy(t-1)} & V_{zy(t-1)}^T \\ V_{zy(t-1)} & V_{zz(t-1)} \end{pmatrix} \ ,$$
$$\lambda_{t-1} = V_{yy(t-1)} - V_{zy(t-1)}^T V_{zz(t-1)}^{-1} V_{zy(t-1)} \ ,$$

where

$$\beta(r) = \beta(0) + r - 1 \ ,$$

$\beta(0)$ is an initialization constant, and the positive definite matrix V_0 represents a prior knowledge, see [12] for details. Moreover, the parameter estimate can be efficiently computed for all pixel positions using a numerically robust recursive formula [12], which is advantageous for texture segmentation applications. Finally, the optimal contextual neighbourhood I_r can be found analytically by maximising the corresponding posterior probability [12].

The parameter estimates (5),(6) can also be evaluated recursively [12]. The posterior probability density [12] of the model is:

$$p(Y_r \,|\, Y^{(r-1)}, \hat{\gamma}_{r-1}) =$$
$$\frac{\Gamma(\frac{\beta(r)-\eta+3}{2})}{\Gamma(\frac{\beta(r)-\eta+2}{2})\,\pi^{\frac{1}{2}}\,(1 + Z_r^T V_{zz(r-1)}^{-1} Z_r)^{\frac{1}{2}}\,|\lambda_{(r-1)}|^{\frac{1}{2}}}$$
$$\left(1 + \frac{(Y_r - \hat{\gamma}_{r-1} Z_r)^T \lambda_{(r-1)}^{-1} (Y_r - \hat{\gamma}_{r-1} Z_r)}{1 + Z_r^T V_{zz(r-1)}^{-1} Z_r}\right)^{-\frac{\beta(r)-\eta+3}{2}} \tag{7}$$

And the conditional mean value predictor of the one-step-ahead predictive posterior density (7) for the normal-gamma parameter prior is

$$E\left\{Y_r \,|\, Y^{(r-1)}\right\} = \hat{\gamma}_{r-1} Z_r \ . \tag{8}$$

We used functional neighborhood I_r created by manually selecting 10 neighbors (5 at the border and 5 between the border pixels and the main one) from the unilateral part of the hierarchical neighborhood of the order 150, where the furthest neighbor is at the Euclidean distance of 10 pixels from the main pixel. Results for unilateral part of the classical hierarchical neighborhood were not satisfying and the order of such neighborhood cannot be too high because of the texture size restriction. With such functional neighborhood we could significantly rise the neighborhood order while possibly keeping some information from the whole neighborhood without worsening the time complexity.

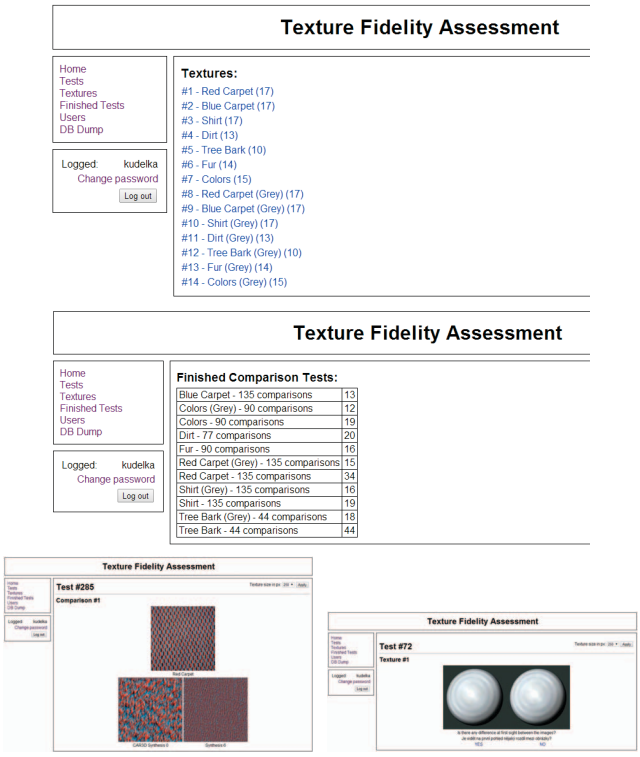

Fig. 1. The benchmark website (http://tfa.utia.cas.cz), home page (upper row), partial assessment result (second row), three modifications of the red carpet texture (bottom left), and a test screen.

2.2. Criterion

Our criterion ζ measures cross-prediction error when using data from the original texture Y and estimated parameters $\widetilde{\gamma}$ from the synthetic texture $\widetilde{Y}$

$$\zeta(Y, \widetilde{Y}) = \frac{1}{|I|} \sum_{\forall r \in I} |Y_r - \widetilde{\gamma}_{r-1} Z_r|. \qquad (9)$$

3. TEST DATA

The criterion is verified on the texture fidelity benchmark [5] which has been created to help the validation of texture fidelity criteria being developed (Fig. 1). The benchmark is a web based service (**http://tfa.utia.cas.cz**) designed for performance evaluation, mutual comparison, and ranking of various texture fidelity measures. The benchmark contains six measured (three man-made textile and three natural) and one synthetic color texture (Fig. 2) together with their gray scale versions as the target textures. The benchmark textures are mathematically synthesized using various mathematical models and variable quality constraints. Textures are synthesized using random field type of models [11], mainly variants of the auto-regressive Markov random field models, or Gaussian mixture models. Each measured texture has up to seventeen synthesized versions.

The color texture (Fig. 2 - bottom right, no. 7) has 33 471 distinct colors and it was deliberately manually created to have a histogram with numerous local extrema. Textures were mathematically synthesized using various mathematical models and variable quality constraints. The models used were either random field type of models, mainly variants of the wide-sense Markov random field models [13], or Gaussian mixture models [14]. Synthetic variants of these textures are ranked by benchmark users and the collected data serve to create modelling quality ranks.

The benchmark supports rapid verification and development of new fidelity criteria approaches and contains seven color, variable texture quality, series (see Fig. 2) together with their grey-scale counterparts.

Fig. 2. Target color benchmark textures (no. 1 - 4 upper row, 5 - 7 bottom row).

4. RESULTS

Single texture fidelity criteria are evaluated based on the consistency between human and a criterion-based ranking. The Spearman's rank correlation coefficient is used to compare human quality ranks obtained from our benchmark with the ranks created from the results of all the STSIM-1 and STSIM-2 variants and also from our proposed criterion. Worse al-

texture no.	ζ	STSIM-1	STSIM-2
1	0.866	0.483	-0.539
2	0.797	0.505	0.400
3	0.801	0.860	0.571
4	0.363	0.857	0.813
5	0.721	0.636	0.370
6	0.600	0.767	0.767
7	0.670	0.420	0.174
8	0.566	0.608	-0.385
10	0.909	0.892	0.654
12	0.794	0.782	0.321
14	0.456	0.662	0.530
$\oslash$	**0.686**	0.679	0.334

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

texture no.	ζ	STSIM-1	STSIM-2
1	132.5659	0.0006	0.0003
2	30.7192	0.0017	0.0006
3	4.8728	0.0012	0.0003
4	2.6779	0.0011	0.0003
5	1.5416	0.0002	0.0002
6	1.9117	0.0015	0.0001
7	12.2320	0.0007	0.0001
8	0.6327	0.0006	0.0003
10	0.3471	0.0012	0.0003
12	0.1025	0.0002	0.0001
14	0.7608	0.0007	0.0002
median var	1.9117	0.0007	0.0003

Table 2. Criteria variances for all 14 individual benchmark texture sets.

ternative results using standard image quality criteria (MSE, VSNR, VIF, SSIM, CW-SSIM) can be consulted in [5].

In our previous article [5] we tested all the variants of STSIM-1 and STSIM-2 and only the results of the global variants of both measures are presented here.

The correlations for all 14 benchmark textures (color and gray-scale) can be seen in Tab. 1. Average correlation of our criterion ζ is slightly better than the best of both tested STSIM variants (STSIM-1) and significantly better than the STSIM-2 criterion. As mentioned in our article [5], the STSIM scores for individual textures have only small variance with respect to the significant visual differences between synthetic textures. Tab. 3 illustrates this problem on criteria values for the texture number 10 (Fig. 2 - gray-scale version of no. 3). This textile texture was chosen because both criteria, our and STSIM-1, have the highest rank correlation from all benchmark textures and our criterion loses its multi-spectral

texture no. 10			
synthesis exp.	ζ	STSIM-1	STSIM-2
1	4.5820	0.8713	0.9181
2	5.7472	0.8725	0.8917
3	5.5038	0.8341	0.9126
4	4.3384	0.9013	0.9298
5	4.8765	0.8807	0.9124
6	4.8806	0.9018	0.9258
7	4.5872	0.8754	0.9296
8	4.0194	0.9216	0.9338
9	4.8738	0.9278	0.8977
10	3.8918	0.9370	0.9429
11	4.5845	0.9243	0.9190
12	4.2380	0.9266	0.8989
13	4.5893	0.9206	0.9103
14	3.7684	0.9505	0.9404
15	3.8319	0.9659	0.9597
16	3.6978	0.9575	0.9519
var	**0.3471**	0.0012	0.0003

Table 3. Criteria values and their variances for the gray-scale textile texture no. 10. The first column is a list of various quality modeling experiments.

advantage. It is also the case, that the synthetic textures in this set are not that different concerning other texture sets, so we chose one more set to present. The variance of our criterion (0.3471) is higher by several orders with respect to the magnitude of the scores and we consider this lower sensitivity to be a disadvantage of all STSIM measures. This conclusion is supported by evidence of all remaining benchmark sets criteria variances in Tab. 2.

5. CONCLUSIONS

The proposed texture fidelity criterion is the only genuine multi-spectral textural qualitative criterion based on the generative Markovian texture model statistics. It can be computed analytically and outperforms the best alternative - the STSIM fidelity criterion. The criterion correlates well with human texture quality ranking which is verified on the texture fidelity benchmark using the Spearman's rank correlation. Criterion values for individual textures also show a higher variance which we consider to be an advantage for quality ranking between subtly distinguishable synthetic texture realizations. The criterion can be easily exploited also in texture classification applications.

Finally, most authors, including [3], separate color from the structural information when testing texture fidelity. We believe that the color cannot be omitted in a universal texture fidelity criterion without compromising its performance and reliability.

6. REFERENCES

[1] Wen-Chieh Lin, James Hays, Chenyu Wu, Yanxi Liu, and Vivek Kwatra, "Quantitative evaluation of near regular texture synthesis algorithms," in *CVPR*. 2006, pp. 427–434, IEEE Computer Society.

[2] Yossi Rubner, Jan Puzicha, Carlo Tomasi, and Joachim M Buhmann, "Empirical evaluation of dissimilarity measures for color and texture," *Computer vision and image understanding*, vol. 84, no. 1, pp. 25–43, 2001.

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

[4] Jana Zujovic, Thrasyvoulos N Pappas, David L Neuhoff, René van Egmond, and Huib de Ridder, "Effective and efficient subjective testing of texture similarity metrics," *JOSA A*, vol. 32, no. 2, pp. 329–342, 2015.

[5] M. Haindl and M.Jr. Kudělka, "Texture fidelity benchmark," in *Computational Intelligence for Multimedia Understanding (IWCIM), 2014 International Workshop on*, Los Alamitos, November 2014, pp. 1 – 5, IEEE Computer Society CPS.

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

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

[8] 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.

[9] Zhou Wang and Eero P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *In Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP 05). IEEE International Conference on*, 2005, pp. 573–576.

[10] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430–444, 2006.

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

[12] Michal Haindl, "Visual data recognition and modeling based on local markovian models," in *Mathematical Methods for Signal and Image Analysis and Representation*, Luc Florack, Remco Duits, Geurt Jongbloed, Marie-Colette Lieshout, and Laurie Davies, Eds., vol. 41 of *Computational Imaging and Vision*, chapter 14, pp. 241–259. Springer London, 2012, 10.1007/978-1-4471-2353-8_14.

[13] Michal Haindl, "Visual data recognition and modeling based on local markovian models," in *Mathematical Methods for Signal and Image Analysis and Representation*, pp. 241–259. Springer, 2012.

[14] Michal Haindl, Jiri Grim, Petr Somol, Pavel Pudil, and Mineichi Kudo, "A gaussian mixture-based colour texture model," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. IEEE, 2004, vol. 3, pp. 177–180.