2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

Dynamic Texture Similarity Criterion

Radek Richtr[1,2] and Michal Haindl[1,3]
[1]Institute of Information Theory and Automation of the CAS,
[2]Faculty of Information Technology, CTU, Prague, Czech Republic
[3]Faculty of Management, University of Economics, Jindřichův Hradec, Czech Republic
Email: {richtrad,haindl}@utia.cas.cz

***Abstract*—Dynamic texture similarity ranking is a challenging and still unsolved problem. Evaluation of how well are various dynamic textures similar to humans perception view is extremely difficult even for static textures and requires tedious psycho-physical experiments. Human perception principles are largely not understood yet and the dynamic texture perception is further complicated with a distinct way of perceiving spatial and temporal domains, which complicates any similarity criterion definition. We propose a novel dynamic texture criterion based on the Fourier transformation and properties of dynamic texture spatio-temporal frequencies. The presented criterion correlates well with performed psycho-physical tests while maintaining sufficient diversity and descriptiveness.**

I. Introduction

Neither static nor dynamic (or temporal, DT) rigorous mathematical texture definition exists. Dynamic textures (DT) can be vaguely defined as spatially repetitive motion patterns exhibiting homogeneous temporal properties. Examples might be smoke, haze, fire or liquids, also waving trees or straws or some moving mechanical objects.

Mutual similarity assessment and similarity ranking of two or more visual textures is a difficult problem due to real material textures complex dependencies on 16 physical observation parameters [1]. Evaluation of how well various texture models conform to 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 a fair comparison of distinct models, material recognition, etc. This problem is not satisfactorily solved even for simpler static textures [2], [3]. Currently the only reliable, but extremely impractical and expensive option, is to exploit the methods of visual psycho-physics. The psycho-physical methods [1] 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. Few published static texture criteria allow to test selected texture properties such as the texture regularity [4], etc. Others claim to test general texture quality [5]–[7]. Our recent test [2] on our texture fidelity benchmark (http://tfa.utia.cas.cz) of several state-of-the-art image quality measures and several recently published static texture criteria confirms their insufficient reliability and low robustness. The evaluated criteria were - the structural similarity (SSIM) index [8], the visual information fidelity (VIF) methods [9], the visual signal-to-noise- ratio [10] (VSNR), the mean-squared error (MSE) [11], the complex wavelet - structural similarity (CW-SSIM) index [12], and the structural texture similarity measure (STSIM-1, STSIM-2, STSIM-M) [6], and all are severely restricted to only gray-scale textures. The results have demonstrated [2] that the standard image quality criteria (MSE, VSNR, VIF, SSIM, CW-SSIM) do not correlate well with the human quality assessment of textures at all. Although, the STSIM texture criteria have significantly higher correlation with human ranking, they do not successfully solve this problem. Our textural qualitative criterion based on the generative Markov texture model statistics ζ [3] is fully multispectral and slightly outperforms the best alternative - the STSIM fidelity criterion. All previous texture similarity criteria can be formally generalized also for dynamic textures if they are applied to the corresponding frame couples of compared dynamic textures and subsequently combined these partial results. However, we can expect admissible ranking only for some oversimplified tests such as identical dynamic textures which differ only in additive noise level. Thus a novel more robust criterion which we present here is clearly needed.

II. Similarity Criterion

The proposed DT similarity criterion is based on the three-dimension Fourier transformation for each spectral band (see Fig.1). The Fourier transformation of a function $f(x_1, x_2, x_3)$ finds the spatial frequencies $\xi = (\xi_1, \xi_2, \xi_3)$. The 3-dimension Fourier transformation for the $f(x_1, x_2, x_3)$ function can be written as:

$$\mathcal{F}\left\{f(\xi_1, \xi_2, \xi_3)\right\} = \int_{R^3} e^{-2\pi i \xi x} f(x_1, x_2, x_3) dx_1, dx_2, dx_3 \ .$$

The harmonics are the complex exponential $e^{\pm 2\pi i x \xi}$ with three spatial frequencies ξ. In three-dimensions a given ξ defines a family $\xi \cdot x =$ integer of parallel planes (of zero phase) in the $x = (x_1, x_2, x_3)$ space. The normal to any of the planes is the vector ξ and adjacent planes are a distance $1/||\xi||$ apart. The exponential is periodic in the direction ξ with period $1/||\xi||$ [13]. The combination of 2D and 1D Fourier transformation is used to detect dynamics of significant local and global spatial frequencies. The crucial part of video similarity perception is its structures dynamic behavior [14]. This behavior is mainly described by a complex exponential with the normal ξ with a dominant temporal component ξ_3. Components ξ_3 similar to $\xi_{1,2}$ are less recognizable. For simplicity (and because of the separability of the individual

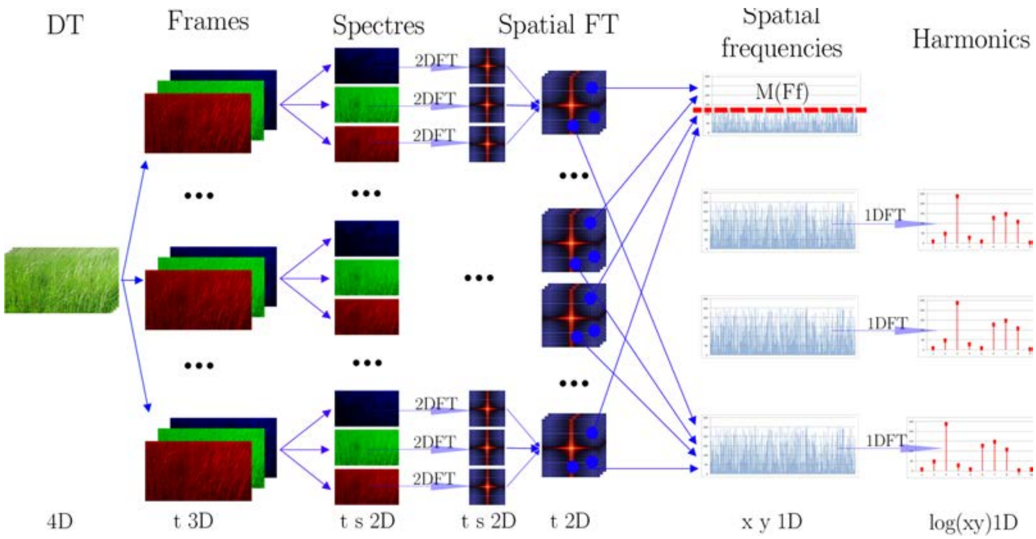

Fig. 1. The similarity criterion flowchart.

vectors information) we assume that variations of harmonics with similar normal vectors ξ are strongly dependent and can be represented with one normal with a dominant temporal component. Provided there is a group of representative normal vectors ξ we assume a new parallel harmonic. This simplifies the criterion evaluation with negligible loss of generality. In addition to omitting unnecessary information, it may be possible to neglect certain frequencies to accelerate the calculation of the last criterion evaluation step to maintain asymptotic complexity $O\left(T\log_2(NM)\log\left(T\log_2(NM)\right)\right)$ where $N \times M$ is a frame size and T the temporal length. During extensive tests not reported in this article, we have concluded that a sufficient subset of crucial spatial frequencies is roughly equal to $w_f log_2(NM)$, where w_f is a small constant in the range $1 \leq w_f \leq 10$ for each compared DT.

A. DT Similarity

The DT similarity criterion is based on the comparison of harmonic frequencies time series related to significant spatial frequencies in the compared dynamic textures. Let's denote a set of harmonics (of Y_s or Y_c DT) $\xi^{\zeta;\bullet}_{Y}$ for δ spectral band:

$$\xi^{\zeta;\bullet}_{Y^\delta} = \int_{R^1} \exp\{-2\pi i x_3 \xi_3\}\, \mathcal{F}\{f^\delta(\xi_1,\xi_2)\}\, dx_3 \quad , \tag{1}$$

for any of magnitudes $|\mathcal{F}\{f(\xi_1,\xi_2)\}|$ from both compared textures Y_s and Y_c which is between $w_f log_2(NM)$-th largest vales and zero otherwise. Where $\bullet$ denotes all corresponding index values (e.g., temporal frames) and $\zeta = (\xi_1,\xi_2)$ is some fixed spatial frequencies ξ_1,ξ_2. $w_f \geq 1$ is the parameter set to speed up the process and to better separate the noise from crucial spatial frequencies. We set experimentally this value to $w_f = 2$. Recall that due to steadily growing audiovisual data resolution, this step sharply reduces the criterion computational time complexity.

Whereas the harmonics set expresses the behavior of one spacial frequency alongside temporal dimension, it is essential to use only those harmonics that carry essential visual information. If the spatial frequency temporal behavior can be considered significant, it is used in the criterion. Note that the significant frequencies designation characterizes the given frequencies through all the compared DTs (i.e., Y_s and Y_c). This detection ensures that the missing or modified frequencies in the synthesized texture are detected and evaluated.

B. Criterion

Let us denote function $\Phi_{Y_s,Y_c} \in \langle -NM\delta\beta; 0\rangle$ with a measure L between two DTs Y_s and Y_c as:

$$\uparrow \Phi_{Y_s,Y_c} = \sum_{\forall\delta} \sum_{\forall\xi_1 \in Y_s^\delta \cap Y_c^\delta} \sum_{\forall\xi_2 \in Y_s^\delta \cap Y_c^\delta} \phi^{\xi_1,\xi_2}_{Y_s^\delta,Y_c^\delta} \quad , \tag{2}$$

where $\uparrow$ is the desired value orientation, β is the Fourier transformation maximum, $\phi^{\xi_1;\xi_2}_{Y_s^\delta,Y_c^\delta}$ for a δ spectral band is:

$$\phi^{\xi_1,\xi_2}_{Y_s^\delta,Y_c^\delta} = -M^\delta_{x_3}\left(\mathcal{F}\{f_{x_3}(\xi_1,\xi_2)\}\right)\, |L(\xi^{\xi_1,\xi_2,\bullet}_{Y_s^\delta}, \xi^{\xi_1,\xi_2,\bullet}_{Y_c^\delta})| \quad , \tag{3}$$

and Y_s and Y_c are source and comparison textures, respectively. $L(u,v)$ is the measure between impulse characteristic u and v (a distance between two harmonics sets) like eq. (4), (5) or(6). $M_{x_3}(\gamma)$ is a function which return the maximal value for the argument vector γ and scale value to the range $\langle 0; 1\rangle$ over all frame's maxima. Φ_{Y_s,Y_c} is a set of all measure values $\phi^{\bullet;\bullet}_{Y_s,Y_c}$ between DTs Y_s, Y_c across all spatial frequencies. Physical meaning of $\phi^{\xi_1,\xi_2}_{s,c}$ represents different dynamics of the spatial frequency $\xi = (\xi_1,\xi_2)$, i.e., periodicity and magnitude variability between textures Y_s and Y_c. For the purpose of comparing dynamic textures where one was created by a modification of the other and thus it shares major common part of the criterion function, some additional information can be obtained. In the case of the synthesized dynamic texture Y_c, which was created by modifying the original DT Y_s, the loss of information between the synthesized and the original texture can be expressed in terms of dynamics as the absence of some spatial frequencies (see Fig.2). This is the case of frequent synthesis with inappropriate sample or insufficiently learned model. The human perception consequence of this is a loss of structure, missing sharp edges and details, as well as low frequencies that DT synthesizing methods are unable to represent. The lost frequencies, both spatial and temporal, term is:

$$L^{lost}(\xi^{\zeta,\bullet}_{Y_s},\xi^{\zeta,\bullet}_{Y_c}) \;=\; \sum_{\forall \xi_3 \in Y_s \cap Y_c} l^{l}(\xi^{\zeta,\xi_3}_{Y_s},\xi^{\zeta,\xi_3}_{Y_c}) \;, \qquad (4)$$

$$l^{l}(s_i,c_i) \;=\; \begin{cases} s_i - c_i & \text{if } s_i \geq c_i \\ 0, & \text{otherwise} \end{cases} ,$$

where s_i (original) and c_i (compared) are i-th values from the given harmonics ξ.

However, due to synthesis or inpainting methods, some other phenomena that reduce visual quality of the synthesized DTs exist. The problem is not the missing information, but false frequencies introduced by these methods and not related to the original texture. False synthetic DT periodicities are visually disturbing the perception of the quality of DTs. This is the apparent repetition of the same samples as well as the visible boundaries between data samples from which the new texture composed. The false frequencies, both spatial and temporal, term is:

$$L^{false}(\xi^{\zeta,\bullet}_{Y_s},\xi^{\zeta,\bullet}_{Y_c}) \;=\; \sum_{\forall \xi_3 \in Y_s \cap Y_c} l^{f}(\xi^{\zeta,\xi_3}_{Y_s},\xi^{\zeta,\xi_3}_{Y_c}) \;, \qquad (5)$$

$$l^{f}(s_i,c_i) \;=\; \begin{cases} c_i - s_i & \text{if } c_i \geq s_i \\ 0, & \text{otherwise} \end{cases} .$$

The newly created false frequencies are harmonics $e^{\pm 2\pi i x \xi}$ with normal $\xi = (\xi_1,\xi_2,\xi_3)$ where $\xi_{(1,2,3)}$ varies to any combination as the patches are R^n space with noticeable error (which are usually minimized by the synthesizing method). Thus $L^{false}(s,c)$ can be efficiently computed on any harmonics ξ^{x_1,x_2,x_3}_{Y} from which we, for simplicity and to streamline the computation process, pick $\xi^{\zeta,\bullet}_{Y}$ and ξ^{x_1,x_2,x_3}_{Y} where $x_{(1,2,3)}$ are maximized mutual value difference (i.e., ξ yielding an oblique harmonics).

Both criteria (4), (5) complement the absolute difference in texture behavior and allow more accurate behavior comparison of different DT synthesis approaches and more detailed analysis of their weaknesses and strengths. In the proposed criterion we use their combination:

$$\begin{aligned} L^{\alpha}(\xi^{\zeta,\bullet}_{Y_s},\xi^{\zeta,\bullet}_{Y_c}) \;=\;& \alpha_l\, L^{lost}(\xi^{\zeta,\bullet}_{Y_s},\xi^{\zeta,\bullet}_{Y_c}) \\ &+ \alpha_f\, L^{false}(\xi^{\zeta,\bullet}_{Y_s},\xi^{\zeta,\bullet}_{Y_c}) \;, \qquad (6) \end{aligned}$$

where α_l and α_f are weight parameters between 0 and 1. After our extensive tests, it seems that for the purpose of computed similarity of DTs the values of $\alpha_l = \alpha_f = 1$ are sufficient. For applications comparing the original and modified (synthesized or inpainted) textures the parameter α_l seems to be more important and determining, thus $\alpha_l \leq \alpha_f$.

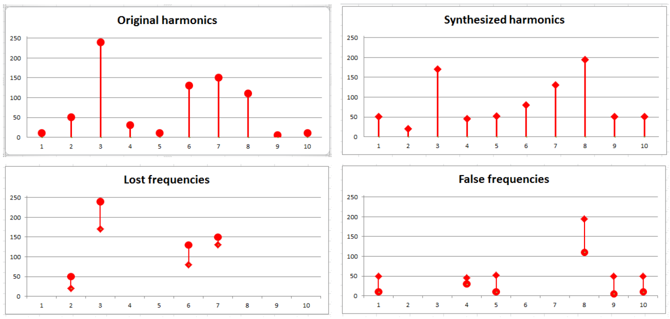

Fig. 2. Illustration of harmonics of two DTs examples; the first row shows harmonics of the original DT and its synthesis, respectively; the second row shows the lost a false (respectively) frequencies visualization.

III. Experimental Dynamic Textures

The proposed similarity criterion was tested on our dynamic texture database and the DynTex [15] database. To determine the texture class, the DynTex class labels (sea, flowers, calm water, etc.) were used also for our own database. Our DTs have a noticeably higher resolution (including Full HD and greater) and time duration in the order of tens of seconds or minutes. Fig. 3 illustrates representative GRASS and FOLIAGE Full HD DTs with all to all computed similarity criterion (only a representative part is shown for explanation purposes). The dataset consists of visually similar textures, yet with different dynamics. The database contains textures with pronounced (strong wind, rain) and low dynamics as well as textures with similar dynamics (directional wind) but very different structure.

Due to the high variability of dynamics and structure, while maintaining a similar class, this is an interesting and representative dataset allowing effective comparison and demonstration of strong and weak characteristics of the criterion. The criterion was tested also on other datasets with similar results.

The submitted dataset has been subjected to detailed psycho-physical testing with more than 150 users who evaluated the likeness of individual DTs. This user evaluation (Tab. I) is therefore used for the proposed criterion validation. These results show both pairs rated as extremely similar (21-14), and couples that have been rated as highly different (11-12, 10-12). DTs 23-22 and 02-12 have similar structures and similar dynamics but very different color, but these couples were evaluated as relatively similar.

The computed similarity criterion values are in Tab.II while the STSIM-1 and STSIM-2 values are in Tab.III. For the purpose of DTs similarity we compute STSIM-1 and STSIM-2 as the average for every frame, which gives slightly better results than using one representative frame only.

IV. Results

A. *Psycho-Physical Validation*

The human ratings (see Tab. I) of over 150 participants was used to validate the proposed criterion. The observers evaluated similarity using a rating between 0 to 5. Table II shows the comparison between our criterion values and values

Fig. 3. Representative selected frames from several GRASS and FOLIAGE dynamic textures marked with labels as indicated in Tabs. I - IV

obtained from testing. Since the difference may also have negative values, the goal is to minimize the absolute value.

Note, that the difference is usually very close to zero and high values occur rarely. The total mean variation over the values provided by the psycho-physical tests is only $0,1389$ (or about 0.5 on the scale 0 to 5 that was used for the psycho-physical tests).

The values of commonly used texture similarity criteria (STSIM-1 and STSIM-2) and values provided by psycho-physical tests participant are compared in Tab. IV. The metrics give a noticeably worse result with absolute values often near (and higher) to $0,5$ which represent difference greater than 2 on the user used evaluation scale. The extremal value $> 0,76$ occurs only once in both cases and values near zero occurs rarely. The total mean variation over the values provided by the psycho-physical tests is $0,3498$ for STSIM-1 and $0,332$ for STSIM-2, respectively.

B. Criterion Validation

Additional tests were performed to verify the accuracy of the criterion. Small randomly placed sub-parts were extracted from high resolution textures (see Fig.5) and subsequently used to compute the criterion. The average similarity values of these sub-parts within one particular DT were compared with the average values of the criterion of those sub-parts with similarly extracted pieces from other but similar DT (see Fig.4). The similarity ratio between intra-DT and inter-DT values is calculated as:

$$\psi^j_{Y_s,Y_\bullet} = \frac{n-1}{(m-1)n} \frac{\sum_{i\neq s,i=1}^{m} \sum_{k=1}^{n} \Phi_{Y_{s_k},Y_{i_k}}}{\sum_{i\neq j,i=1}^{n} \Phi_{Y_{s_i},Y_{s_j}}} ,$$

where Y_{s_i} is i-th sup-part from the source DT Y_s. n denotes the number of sub-parts and m is number of used DTs. The average ratio per DT can be subsequently express as $\Psi_{Y_s,\bullet} = \frac{1}{n}\sum_{i=1}^{n} \psi^i_{Y_s,\bullet}$ and the mean ratio $\Psi_{Y_\bullet,\bullet} = \frac{1}{m}\sum_{i=1}^{m} \Psi_{Y_i,\bullet}$.

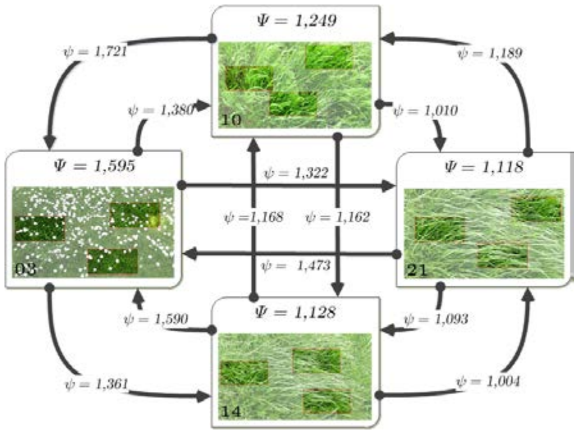

Fig. 4. Visualization of similarity ratios between four grass DTs. Values of $\Psi^j_{Y_s,Y_c}$ are indicated by arrows. Values inside boxes indicate $\Psi_{Y_s,\bullet}$ ratio between inter-DT and all intra-DT similarities and $\Psi_{Y_\bullet,\bullet} = 1.29$.

The values $\psi = 1$ represent extremely similar dynamics of compared DTs. Since our criterion does not acquire absolute values smaller than 1, the ratio $\Psi_{Y_s,Y_\bullet} > 1$ determines significantly higher mutual similarities of the inner sub-parts than similarity to sub-parts of other DTs. The ratio of similarities was computed on more than 300 pairs of DTs with the average value of $1,35$. Let us mention that in some significant cases, the ratio of the distance of values may be less than 1. These cases are usually caused either by different dynamics (e.g., leaves versus branches, lee, etc.) within the DT or by the texture being composed of multiple DT textures.

V. Conclusion

A novel fully spectral dynamic textural similarity criterion is presented and successfully validated with several state-of-the-art criteria, which were straightforwardly extended to temporal domain. Due to the absence of reliable DT benchmarking we performed extensive psycho-physical measurements of

TABLE I

RESULTS OF PSYCHO-PHYSICAL TEST ON PAIRS OF DTS GRASS. RANGE FROM 0 (DIFFERENT) TO 1 (SIMILAR).

	0	1	2	3	4	10	11	12	13	14	20	21	22	24	30	31	34
0		0,46	0,68	0,65	0,4	0,26	0,44	0,45	0,39	0,38	0,65	0,61	0,18	0,45	0,53	0,48	0,48
1			0,48	0,48	0,31	0,17	0,28	0,48	0,44	0,22	0,46	0,22	0,2	0,26	0,73	0,31	0,55
2				0,61	0,55	0,19	0,44	0,4	0,41	0,62	0,5	0,41	0,14	0,4	0,4	0,21	0,5
3					0,35	0,17	0,17	0,75	0,4	0,24	0,51	0,31	0,24	0,28	0,55	0,28	0,66
4						0,46	0,35	0,4	0,77	0,53	0,84	0,55	0,35	0,33	0,6	0,77	0,51
10							0,4	0,2	0,28	0,66	0,28	0,46	0,66	0,46	0,24	0,46	0,2
11								0,17	0,31	0,48	0,2	0,31	0,44	0,4	0,26	0,4	0,24
12									0,48	0,22	0,6	0,26	0,17	0,28	0,46	0,26	0,66
13										0,46	0,6	0,48	0,26	0,33	0,6	0,48	0,68
14											0,37	0,88	0,51	0,51	0,31	0,77	0,26
20												0,44	0,26	0,24	0,57	0,44	0,64
21													0,35	0,35	0,44	0,77	0,37
22														0,71	0,28	0,33	0,17
24															0,57	0,48	0,42
30																0,4	0,68
31																	0,35

TABLE II

PRESENTED DYNAMIC TEXTURE SIMILARITY CRITERION VALUES; BOTTOM TRIANGLE: THE PROPOSED TEXTURE SIMILARITY CRITERION VALUES SCALED TO RANGE $\langle 0;1 \rangle$; UPPER TRIANGLE: DIFFERENCE BETWEEN TEST RESULTS (TAB. I) AND OUR CRITERION. POSITIVE VALUES SUGGEST UNDERESTIMATION OF THE CRITERION AND SUBSEQUENTLY NEGATIVE VALUES SUGGEST OVERSTATEMENT. THE TOTAL MEAN VARIATION OVER THE VALUES IS 0,1431.

	0	1	2	3	4	10	11	12	13	14	20	21	22	24	30	31	34
0	X	0,1	0,35	0,32	0,07	0,08	0,18	0,1	0,08	0,11	0,29	0,32	-0,06	0,15	0,21	0,18	0,15
1	0,38	X	0,05	0,14	-0,04	**0**	0,02	0,01	0,13	-0,02	0,06	-0,08	-0,04	-0,06	0,39	-0,01	0,2
2	0,35	0,45	X	0,29	0,19	0,01	0,17	0,09	0,08	0,22	0,12	0,16	-0,09	0,07	0,06	-0,11	0,14
3	0,34	0,35	0,33	X	-0,15	0,1	-0,02	0,13	-0,02	0,02	-0,04	0,13	0,07	-0,13	0,09	-0,08	0,16
4	0,34	0,36	0,37	0,51	X	0,33	0,11	-0,11	0,39	0,28	0,35	0,3	0,12	-0,04	0,19	0,42	0,06
10	0,19	0,19	0,19	0,08	0,15	X	0,2	0,13	0,12	0,46	0,15	0,28	0,49	0,3	0,08	0,29	0,07
11	0,28	0,28	0,29	0,21	0,25	0,21	X	-0,03	0,06	0,23	-0,04	0,06	0,22	0,13	0,02	0,13	**0**
12	0,37	0,49	0,32	0,63	0,52	0,08	0,22	X	0,1	-0,05	**0**	0,01	0,01	-0,13	-0,01	-0,12	0,15
13	0,33	0,33	0,34	0,43	0,4	0,18	0,26	0,4	X	0,19	0,18	0,23	0,01	**0**	0,21	0,15	0,29
14	0,29	0,25	0,42	0,23	0,27	0,22	0,27	0,28	0,28	X	0,1	**0,54**	0,28	0,24	0,05	**0,5**	-0,01
20	0,38	0,41	0,39	0,56	0,51	0,15	0,25	0,61	0,43	0,28	X	0,21	0,05	-0,18	0,11	-0,02	0,15
21	0,3	0,31	0,26	0,19	0,27	0,2	0,26	0,27	0,27	0,36	0,25	X	0,13	0,1	0,2	**0,5**	0,13
22	0,25	0,25	0,25	0,19	0,24	0,19	0,23	0,18	0,26	0,24	0,23	0,24	X	0,48	0,06	0,1	-0,04
24	0,32	0,33	0,34	0,43	0,38	0,18	0,28	0,43	0,34	0,29	0,43	0,26	0,24	X	0,2	0,15	0,03
30	0,39	0,35	0,35	0,48	0,42	0,17	0,26	0,49	0,4	0,27	0,47	0,25	0,24	0,39	X	0,04	0,24
31	0,32	0,34	0,34	0,38	0,37	0,19	0,28	0,4	0,35	0,29	0,47	0,29	0,24	0,35	0,37	X	-0,02
34	0,35	0,37	0,37	0,51	0,46	0,14	0,26	0,53	0,41	0,28	0,51	0,25	0,23	0,4	0,46	0,38	X

TABLE III

STSIM-I AND STSIM-II CRITERIA VALUES (AVERAGE OF THE FIRST 60 FRAMES OF EACH DT) OF THE REPRESENTATIVE DT SUBSETS; UPPER TRIANGLE: STSIM-I VALUES; BOTTOM TRIANGLE: STSIM-II VALUES; NOTE LOW DEVIATION OF VALUES.

	0	1	2	3	4	10	11	12	13	14	20	21	22	24	30	31	34
0	X	0,77	0,65	0,79	0,76	0,78	0,79	0,8	0,74	0,78	0,78	0,77	0,73	0,79	0,76	0,8	0,78
1	0,75	X	0,74	0,76	0,74	0,75	0,76	0,79	0,74	0,73	0,74	0,74	0,76	0,77	0,77	0,73	0,75
2	0,71	0,74	X	0,63	0,62	0,63	0,64	0,66	0,64	0,61	0,62	0,65	0,69	0,65	0,71	0,6	0,61
3	0,76	0,75	0,71	X	0,77	0,78	0,8	0,8	0,77	0,77	0,77	0,74	0,75	0,81	0,77	0,79	0,79
4	0,74	0,74	0,7	0,75	X	0,75	0,78	0,77	0,76	0,72	0,76	0,69	0,75	0,77	0,75	0,75	0,78
10	0,75	0,74	0,7	0,76	0,74	X	0,79	0,78	0,77	0,78	0,76	0,75	0,75	0,79	0,77	0,78	0,79
11	0,76	0,75	0,71	0,77	0,75	0,76	X	0,8	0,79	0,77	0,78	0,74	0,77	0,81	0,79	0,79	0,81
12	0,76	0,76	0,71	0,76	0,75	0,75	0,76	X	0,76	0,76	0,79	0,76	0,75	0,79	0,76	0,77	0,78
13	0,74	0,74	0,71	0,75	0,74	0,75	0,76	0,75	X	0,73	0,75	0,68	0,75	0,8	0,8	0,75	0,8
14	0,75	0,73	0,69	0,75	0,73	0,75	0,75	0,74	0,73	X	0,75	0,79	0,7	0,78	0,75	0,8	0,77
20	0,75	0,74	0,7	0,75	0,74	0,74	0,75	0,75	0,74	0,74	X	0,73	0,73	0,78	0,75	0,77	0,77
21	0,75	0,74	0,7	0,74	0,71	0,74	0,73	0,74	0,71	0,75	0,73	X	0,66	0,73	0,71	0,76	0,72
22	0,74	0,75	0,73	0,75	0,74	0,74	0,75	0,74	0,74	0,72	0,73	0,71	X	0,79	0,81	0,72	0,78
24	0,76	0,75	0,71	0,77	0,75	0,76	0,77	0,76	0,76	0,75	0,75	0,74	0,76	X	0,81	0,8	0,78
30	0,75	0,75	0,73	0,76	0,74	0,75	0,76	0,75	0,77	0,74	0,74	0,73	0,77	0,77	X	0,76	0,79
31	0,76	0,74	0,69	0,76	0,74	0,75	0,76	0,75	0,74	0,76	0,74	0,74	0,73	0,76	0,75	X	0,79
34	0,75	0,74	0,7	0,76	0,75	0,76	0,77	0,76	0,76	0,75	0,75	0,73	0,76	0,77	0,76	0,76	X

TABLE IV

DIFFERENCE BETWEEN TEST RESULTS (TAB. I) AND STSIM-I (UPPER TRIANGLE), STSIM-II (BOTTOM TRIANGLE); POSITIVE VALUES SUGGEST UNDERESTIMATION OF THE CRITERION AND SUBSEQUENTLY NEGATIVE VALUES SUGGEST OVERSTATEMENT. THE AVERAGE DIFFERENCE OF ABSOLUTE VALUES ARE $0,349$ FOR STSIM-I AND 0.332 FOR STSIM-II. NOTE THAT BOTH CRITERIA OFTEN SEEM OVERVALUED DUE TO DTS COLOR SIMILARITIES.

0	0	1	2	3	4	10	11	12	13	14	20	21	22	24	30	31	34
0	0	-0,31	0,03	-0,14	-0,36	-0,52	-0,35	-0,35	-0,35	-0,4	-0,13	-0,16	-0,55	-0,34	-0,23	-0,32	-0,3
1	-0,29	0	-0,26	-0,28	-0,43	-0,58	-0,48	-0,31	-0,3	-0,51	-0,28	-0,52	-0,56	-0,51	-0,04	-0,42	-0,2
2	-0,03	-0,26	0	-0,02	-0,07	-0,44	-0,2	-0,26	-0,23	0,01	-0,12	-0,24	-0,55	-0,25	-0,31	-0,39	-0,11
3	-0,11	-0,27	-0,1	0	-0,42	-0,61	-0,63	-0,05	-0,37	-0,53	-0,26	-0,43	-0,51	-0,53	-0,22	-0,51	-0,13
4	-0,34	-0,43	-0,15	-0,4	0	-0,29	-0,43	-0,37	0,01	-0,19	0,08	-0,14	-0,4	-0,44	-0,15	0,02	-0,27
10	-0,49	-0,57	-0,51	-0,59	-0,28	0	-0,39	-0,58	-0,49	-0,12	-0,48	-0,29	-0,09	-0,33	-0,53	-0,32	-0,59
11	-0,32	-0,47	-0,27	-0,6	-0,4	-0,36	0	-0,63	-0,48	-0,29	-0,58	-0,43	-0,33	-0,41	-0,53	-0,39	-0,57
12	-0,31	-0,28	-0,31	-0,01	-0,35	-0,55	-0,59	0	-0,28	-0,54	-0,19	-0,5	-0,58	-0,51	-0,3	-0,51	-0,12
13	-0,35	-0,3	-0,3	-0,35	0,03	-0,47	-0,45	-0,27	0	-0,27	-0,15	-0,2	-0,49	-0,47	-0,2	-0,27	-0,12
14	-0,37	-0,51	-0,07	-0,51	-0,2	-0,09	-0,27	-0,52	-0,27	0	-0,38	0,09	-0,19	-0,27	-0,44	-0,03	-0,51
20	-0,1	-0,28	-0,2	-0,24	0,1	-0,46	-0,55	-0,15	-0,14	-0,37	0	-0,29	-0,47	-0,54	-0,18	-0,33	-0,13
21	-0,14	-0,52	-0,29	-0,43	-0,16	-0,28	-0,42	-0,48	-0,23	0,13	-0,29	0	-0,31	-0,38	-0,27	0,01	-0,35
22	-0,56	-0,55	-0,59	-0,51	-0,39	-0,08	-0,31	-0,57	-0,48	-0,21	-0,47	-0,36	0	-0,08	-0,53	-0,39	-0,61
24	-0,31	-0,49	-0,31	-0,49	-0,42	-0,3	-0,37	-0,48	-0,43	-0,24	-0,51	-0,39	-0,05	0	-0,24	-0,32	-0,36
30	-0,22	-0,02	-0,33	-0,21	-0,14	-0,51	-0,5	-0,29	-0,17	-0,43	-0,17	-0,29	-0,49	-0,2	0	-0,36	-0,11
31	-0,28	-0,43	-0,48	-0,48	0,03	-0,29	-0,36	-0,49	-0,26	0,01	-0,3	0,03	-0,4	-0,28	-0,35	0	-0,44
34	-0,27	-0,19	-0,2	-0,1	-0,24	-0,56	-0,53	-0,1	-0,08	-0,49	-0,11	-0,36	-0,59	-0,35	-0,08	-0,41	0

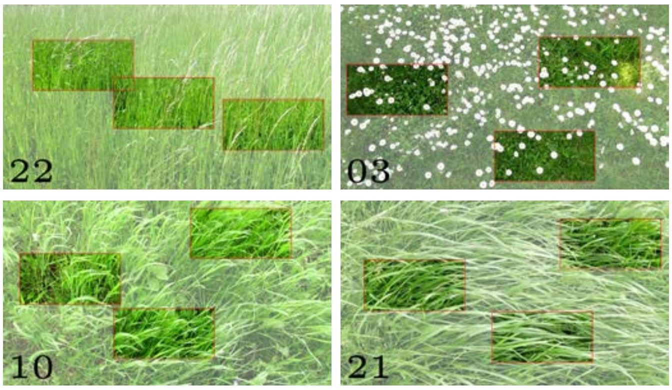

Fig. 5. Illustration of randomly chosen sub-parts of four DTs. Three sub-parts of each DT are marked and used in the validation experiment.

dynamic texture similarities with more than 150 participants. The results demonstrate that the generalized textural criteria STSIM-1, or STSIM-2 can not adequately measure the similarity between dynamic textures. The presented criterion significantly corresponds well with the values obtained from psychophysical tests. The validation of our criterion has clearly shown that our criterion is discriminatory and sufficiently diversifying within a given dynamic texture class. The presented criterion evaluates the spatial frequencies temporal behaviour but not possible color variations. Further research will cover also this aspect and possible scale and illumination invariance.

REFERENCES

[1] M. Haindl and J. Filip, *Visual Texture*, ser. Advances in Computer Vision and Pattern Recognition. London: Springer-Verlag London, January 2013.

[2] M. Haindl and M. Kudělka, "Texture fidelity benchmark," in *Computational Intelligence for Multimedia Understanding (IWCIM), 2014 International Workshop on*. Los Alamitos: IEEE Computer Society CPS, November 2014, pp. 1 – 5. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7008812&isnumber=7008791

[3] M. Kudělka and M. Haindl, "Texture fidelity criterion," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, September 2016, pp. 2062 – 2066. [Online]. Available: http://2016.ieeeicip.org/

[4] W.-C. Lin, J. Hays, C. Wu, Y. Liu, and V. Kwatra, "Quantitative evaluation of near regular texture synthesis algorithms," in *CVPR*. IEEE Computer Society, 2006, pp. 427–434.

[5] Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann, "Empirical evaluation of dissimilarity measures for color and texture," *Computer vision and image understanding*, vol. 84, no. 1, pp. 25–43, 2001.

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

[7] J. Zujovic, T. N. Pappas, D. L. Neuhoff, R. van Egmond, and H. de Ridder, "Effective and efficient subjective testing of texture similarity metrics," *JOSA A*, vol. 32, no. 2, pp. 329–342, 2015.

[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. [Online]. Available: http://dx.doi.org/10.1109/TIP.2003.819861

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

[10] D. M. Chandler and S. 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.

[11] Z. Wang and A. 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.

[12] Z. Wang and E. 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.

[13] R. W. Goodman, *Discrete Fourier and wavelet transforms: an introduction through linear algebra with applications to signal processing*. World Scientific Publishing Co Inc, 2016.

[14] M. Fahle and T. Poggio, "Visual hyperacuity: spatiotemporal interpolation in human vision," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 213, no. 1193, pp. 451–477, 1981.

[15] R. Pteri, S. Fazekas, and M. J. Huiskes, "Dyntex: A comprehensive database of dynamic textures." *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1627–1632, 2010.