 doi:10.20944/preprints202102.0189.v1

Article

An improved SPSIM index for image quality assessment

Mariusz Frackiewicz*, Grzegorz Szolc and Henryk Palus

Department of Data Science and Eng., Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland
* Correspondence: Mariusz.Frackiewicz@polsl.pl; Tel.: +48 32 2371066

Abstract: Objective Image Quality Assessment (IQA) measures are playing an increasingly important role in the evaluation of digital image quality. New IQA indices are expected to be strongly correlated with subjective observer evaluations expressed by MOS/DMOS scores. One such recently proposed index is the SuperPixel-based SIMilarity (SPSIM) index, which uses superpixel patches instead of the rectangular pixel grid.The authors in this paper have been proposed three modifications of SPSIM index. For this purpose, the color space used by SPSIM was changed and the way SPSIM determines similarity maps was modified using methods derived from the algorithm for computing the MDSI index. The third modification was a combination of the first two. These three new quality indices were used in the assessment process. The experimental results obtained on many color images from five image databases demonstrated the advantages of the proposed SPSIM modifications.

Keywords: image quality assessment; image databases; superpixels; color image; color space; image quality measures

---

1. Introduction

The quantitative domination of the acquired color images over grey level images results in the development not only of color image processing methods but also of Image Quality Assessment (IQA) methods. Among these methods, the most developed are the methods comparing the processed images to the original ones, which are called Full-Reference IQA (FR-IQA). In this way, they assess the image quality after performing various image processing operations. The assessments obtained for each IQA measure can be compared with the subjective Mean Opinion Score (MOS) or Difference Mean Opinion Score (DMOS) ratings from human observers. The result of this comparison will be the correlation coefficients: the higher, the better is the measure. In conclusion, the aim of research in the field of IQA is to find measures to objectively and quantitatively assess image quality according to subjective judgment by HVS (Human Visual System).

Low-complexity measure named Mean Squared Error (MSE) has been dominated for a long time among IQA measures used in color image processing due to its simplicity. Similarly popular was the use of other signal theory based measures such as Mean Absolute Error (MAE) and Peak Signal-to-Noise Ratio (PSNR). Over time, a weak link has been noted between these measures and human visual perception. The work [1] shows the drawback in the PSNR measure on the example of the Lighthouse image. Additional noise was added to this image and its copy and was located in the lower or upper part of the image. In the first case the noise was invisible to the observer, in the second case it definitely reduced the image quality assessment. Meanwhile, the PSNR values for both images were the same. Comparison of images on a pixel-to-pixel basis does not allow the influence of the pixel's neighborhood on the perceived color. This is a disadvantage of all those the measures mentioned above.

Since 2004 the Structural Similarity Index Measure (SSIM) [2] is used as the IQA method, that reflects the perceptual change of information about the structure of objects in the scene. This measure uses the idea of pixel inter-dependencies, which are adjacent to each other in the image space. The SSIM value in compared images is created as a combination of three similarities of luminance, contrast

 doi:10.20944/preprints202102.0189.v1

and structure. It is calculated by averaging the results obtained in separate local windows (e.g. 8x8 or 11x11 pixels). In 2011, it was proposed Gradient Similarity Index Measure (GSIM) [3] to take into account the similarity of the gradient in the images, which expresses their edges. These ideas have thus become a starting point for new perceptual measures of image quality and many such IQA measures were proposed during the last decade. In paper [4] a quality index named Feature SIMilarity Index (FSIMc) was proposed. In this case the local quality of evaluated image is characterized by two low-level features maps based on phase congruency (PC) and gradient magnitude (GM). FSIMc, a color version of FSIM, is a result of IQ chrominance components incorporation. Another example of a gradient approach is a simple index called Gradient Magnitude Similarity Deviation (GMSD) [5]. The local image quality is expressed in this case by the gradient magnitude similarity map, then the standard deviation of this map is calculated and used as the final quality index GMSD.

Zhang L. et al. [6] described a measure named Visual Saliency-based Index (VSI) that uses local quality map of the distorted image based on saliency changes. In addition, visual saliency serves as a weighting function conveying the importance of the particular regions in the image. Recently Nafchi et al. [7] proposed new IQA model named a Mean Deviation Similarity Index (MDSI), which uses a new gradient similarity for evaluation of local distortions and chromaticity similarity for color distortions. The final computation of the MDSI score requires a pooling method for both similarity maps; here a specific deviation pooling strategy was used. A Perceptual SIMilarity (PSIM) index was presented in [8] by computing the micro- and macrostructural similarities, described as usual by gradient magnitude maps.Additionally, this index uses color similarity and realizes perceptual-based pooling. IQA models that have been developed in recent years are complex and include more and more properties of the human visual system. Such model was developed by Shi C. and Lin Y. and named Visual saliency with Color appearance and Gradient Similarity (VCGS) [9]. This index is based on the fusion of data from three similarity maps defined by visual salience with color appearance, gradient and chrominance. IQA models increasingly take into account the fact that humans pay more attention to the overall structure of an image rather than local information about each pixel.

Many IQA methods firstly transform the RGB components of image pixels into other color spaces, that are more related to color perception. These are generally spaces belonging to the category "luminance-chrominance" spaces such as YUV, YIQ, YCrCb, CIELAB etc. Such measures are developed in recent years e.g. MDSI. They achieve the higher correlation coefficients with MOS during tests on the publicly available image databases.

Similarity assessment idea which is presented above was used by Sun et al. to propose another IQA index named SuperPixel-based SIMilarity index (SPSIM) [10]. The idea of a superpixel dates back to 2003 [11]. A group of pixels with similar characteristics (intensity, color, etc.) can be replaced by a single superpixel. Superpixels provide a convenient and compact representation of images useful for computationally complex tasks. They are usually much smaller than pixels, so algorithms operating on superpixel images can be faster. Superpixels preserve most of the image edges. There are many different methods for decomposing image into superpixels, of which the fast SLIC (Simple Linear Iterative Clustering) method has gained the most popularity. Typically, image quality measures make comparisons of features extracted from pixels (e.g. MSE, PSNR) or rectangular patches (e.g. SSIM, HPSI). Such patches usually have no visual meaning, whereas superpixels, unlike artificially generated patches, have a visual meaning and are matched to the image content.

In this paper, we consider new modifications of SPSIM measure with improved correlation coefficients with MOS. This paper is organized as follows: After introduction Section 2 presents a relative new image quality measure named SPSIM. Section 3 introduces two modifications of SPSIM measure. In Section 4 and Section 5 the results of experimental tests on different IQA databases are presented. Section 6 concludes this paper. Finally, in the Section 6 the paper is concluded.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0189.v1

2. Related Work

The most popular quality index widely used in the image processing is the MSE and the version for color images is defined as:

$$MSE = \frac{1}{3MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[(R_{ij}-R^*_{ij})^2+(G_{ij}-G^*_{ij})^2+(B_{ij}-B^*_{ij})^2\right], \tag{1}$$

where: MN represents the image resolution, R_{ij}, G_{ij}, B_{ij} are the color components of the pixel (i,j) in the original image and R^*_{ij}, G^*_{ij}, B^*_{ij} are the color components of this pixel in the processed image.

For the purposes of the rest of this article, the two quality indices mentioned in Section 1 and the relatively new ones will be presented in more details: MDSI [7] and SPSIM [10].

Many IQA measures work as follows: determine local distortions in the images, build similarity maps and implement a pooling strategy based on a mean, weighted mean, standard deviation etc. stands to compute a final quality score. An example of this approach to IQA index modelling is the Mean Deviation Similarity Index (MDSI) mentioned in the previous section [7]. The calculation of MDSI starts with the conversion RGB color space components of the input images to a luminance component:

$$L = 0.2989R + 0.5870G + 0.1140B \tag{2}$$

and two chromaticity components:

$$\begin{bmatrix} H \\ M \end{bmatrix} = \begin{bmatrix} 0.30 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{bmatrix}\begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{3}$$

This index is based on computation the gradient similarity (GS) for structural distortions and chromaticity similarity (CS) color distortions. The local structural similarity map is typically determined from gradient values. Classically, structural similarity maps are derived from gradient values calculated independently for the original and the distorted image. In the case of the MDSI measure, the classical approach has been extended by using the gradient value map for the combined values of the luminance channels of both images:

$$F = 0.5(L_r + L_d), \tag{4}$$

where: F - fused luminance image, r - reference image, d - distorted image. Formulae for proposed structural similarity are presented as follows:

$$GS_{rf}(x) = \frac{2G_r(x)G_f(x)+C_2}{G_r^2(x)+G_f^2(x)+C_2}, \tag{5}$$

$$GS_{df}(x) = \frac{2G_d(x)G_f(x)+C_2}{G_d^2(x)+G_f^2(x)+C_2}, \tag{6}$$

$$\widehat{GS}(x) = GS(x) + [GS_{df}(x) - GS_{rf}(x)]. \tag{7}$$

The simple Prewitt operator is used to calculate the gradient magnitude. The authors of the MDSI index have also modified the method of determining the local chromaticity similarity. The previously discussed IQA measures, which used the chrominance of a digital image, such as FSIM or VSI, determined the chromaticity similarity separately for each of the two chrominance channels. In

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0189.v1

the case of MDSI, it was proposed to determine the color similarity for both chrominance channels simultaneously, using the following formula:

$$\widehat{CS}(x) = \frac{2(H_r(x)H_d(x) + M_r(x)M_d(x)) + C_3}{H_r^2(x) + H_d^2(x) + M_r(x)^2 + M_d(x)^2 + C_3}, \tag{8}$$

where $C2$ and $C3$ are constants used for numerical stability. Such joint color similarity map $\widehat{CS}(x)$ can be combined with $\widehat{GS}(x)$ map as following weighted mean:

$$\widehat{GCS}(x) = \alpha\widehat{GS}(x) + (1-\alpha)\widehat{CS}(x), \tag{9}$$

where α controls the relative importance of similarity maps $\widehat{GS}(x)$ and $\widehat{CS}(x)$.

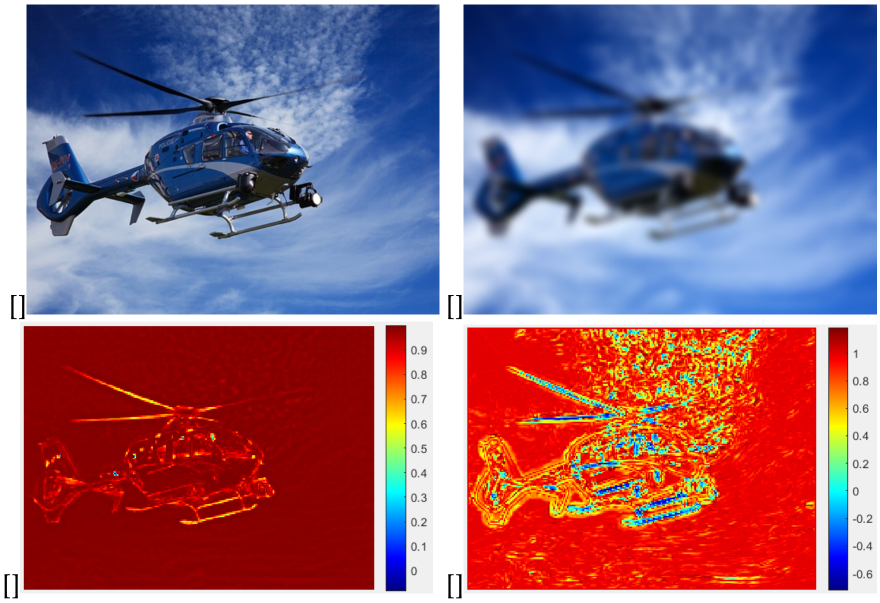

Figure 1. Similarity maps for MDSI index: (a) reference image, (b) distorted image, (c) color similarity map $\widehat{CS}(x)$, (d) similarity map $\widehat{GCS}(x)$

The final computational step is to transform the resulting $\widehat{GCS}$ map (Figure 1) into an MDSI score using a pooling strategy based on specific deviation technique:

$$MDSI = \left[\frac{1}{N}\sum_{i=1}^{N}|\widehat{GCS}_i^{1/4} - (\frac{1}{N}\sum_{i=1}^{N}|\widehat{GCS}_i^{1/4})|\right]^{1/4}. \tag{10}$$

Suggestions for the selection of several parameters whose values affect the performance of the MDSI index can be found in the original article [7]. Unlike other quality measures, in SPSIM (SuperPixel-based SIMilarity) [10] the feature extraction is based on superpixel segmentation. Superpixels are groups of interconnected pixels with similar properties, such as colour, intensity or structure. The result of such pixel grouping is a mosaic with a much smaller number of so-called superpixels, which allows faster further processing. An important advantage of segmentation using superpixels in comparison to other oversegmentation algorithms is the possibility to determine a priori the number of generated superpixels. In addition, segmentation using superpixels allows better separating the perceptually important regions from the image.

Among the algorithms that generate superpixels we can distinguish between graph-based, gradient-based, clustering-based, watershed-based algorithms etc. [12]. Many of these methods are methods provided for image segmentation with parameters leading to oversegmentation. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0189.v1

shape and size of superpixels can vary depending on the used algorithm. Each pixel is contained in one and only one superpixel. Superpixel generation algorithms control the number and properties of superpixels, such as compactness or minimum size. One of the most popular and fastest algorithms for superpixel segmentatation is the k-means based SLIC (Simple Linear Iterative Clustering) algorithm [13]. It is characterised by the fact that the output superpixels are of similar shape and size. Its undoubted advantages include the fact that segmentation only requires the determination of the desired number of superpixels in the output image. Therefore, SLIC algorithm is used in the described here SPSIM quality index. Algorithm SLIC generates for each superpixel the mean CIELab color value and the Local Binary Pattern (LBP) features. Superpixels are generated only on reference image and then applied to both the reference and distorted images. Algorithm for SPSIM index calculation is based on superpixel luminance similarity, superpixel chrominance similarity, and pixel gradient similarity. The values of the similarity maps are calculated in YUV color space instead of RGB color space where Y - luminance and U, V - chrominance components. If we use for a superpixel containing a pixel i the symbol s_i then we can write following formulae for the luminance L_i and luminance similarities $M_L(i)$:

$$L_i = \frac{1}{|s_i|} \sum_{j \in s_i} Y(j), \quad M_L(i) = \frac{2L_r(i)\, L_d(i) + T_1}{L_r^2(i) + L_d^2(i) + T_1}, \tag{11}$$

where $Y(j)$ is the luminance of the pixel j, $L_r(i)$ and $L_d(i)$ are average luminance values for the superpixel s_i in both reference and distorted images. Similar formulae can be constructed for both the U and V chrominance components:

$$U_i = \frac{1}{|s_i|} \sum_{j \in s_i} U(j), \quad M_U(i) = \frac{2U_r(i)\, U_d(i) + T_1}{U_r^2(i) + U_d^2(i) + T_1}, \tag{12}$$

$$V_i = \frac{1}{|s_i|} \sum_{j \in s_i} V(j), \quad M_U(i) = \frac{2V_r(i)\, V_d(i) + T_1}{V_r^2(i) + V_d^2(i) + T_1}. \tag{13}$$

Then chrominance similarity M_C can be determined as below:

$$M_C(i) = M_U(i) M_V(i). \tag{14}$$

A formula similar to the above may describe the gradient similarity M_G:

$$M_G(i) = \frac{2G_r(i)\, G_d(i) + T_2}{G_r^2(i) + G_d^2(i) + T_2}, \tag{15}$$

where the gradient magnitude G consists of two components determined by a simple Prewitt operator and T_1, T_2 are constants chosen by the authors of the algorithm to take account of a contrast type error. Further details of the determination of T_1 and T_2 are described in [10].

The formula for determining the similarity of a superpixel i in both images can be written as:

$$M(i) = M_G(i) \left[M_L(i)\right]^{\alpha} e^{\beta(M_C(i)-1)}, \tag{16}$$

where α and β parameters defining the weights of the luminance and chrominance components.

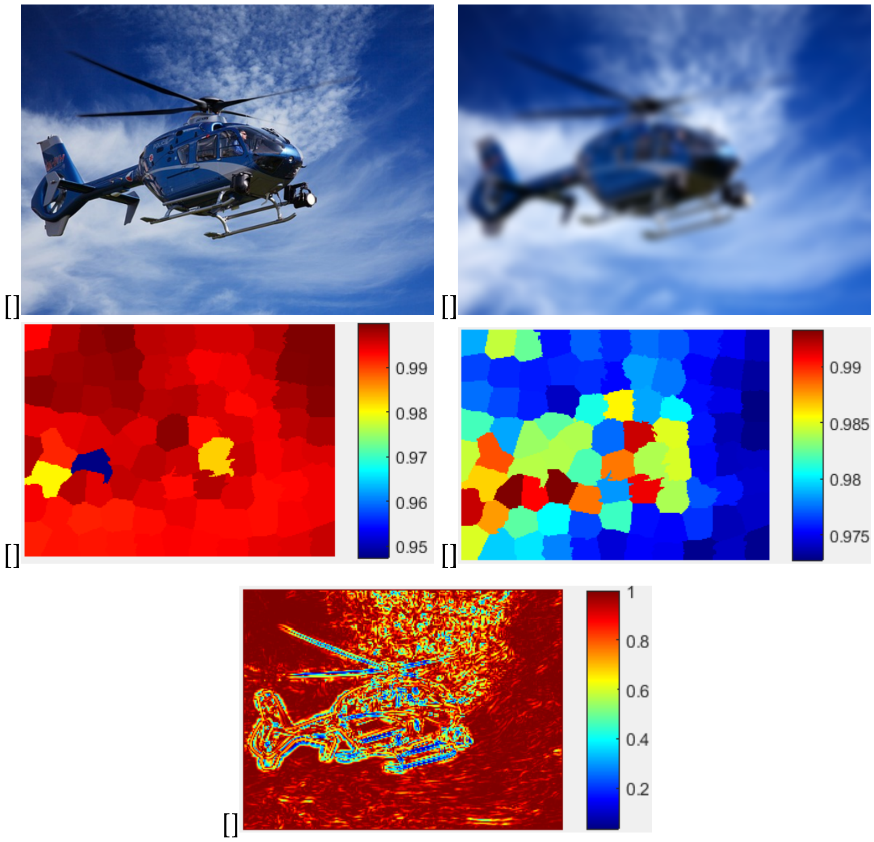

Figure 2. Local similarity maps for SPSIM index (100 superpixels): (a) reference image, (b) distorted image, (c) luminance similarity map, (d) chrominance similarity map, (e) gradient similarity map

Finally, SPSIM index is weighted sum of $M(i)$ and weights which are calculated using a standard deviation std and kurtosis $Kurt$ of the superpixels:

$$TC(i) = \frac{std(S(i))}{Kurt\,[S(i)] + 3}\ , \tag{17}$$

$$w(i) = exp(0.05 \cdot abs(TC_d(i) - TC_r(i))), \tag{18}$$

$$SPSIM = \frac{\sum_{i=1}^{N} M(i)w(i)}{\sum_{i=1}^{N} w(i)}. \tag{19}$$

Figures 2 - 3 presents local similarity maps for two different superpixel resolutions: 100 and 400. The images show that for more superpixels, the local similarity maps are more detailed.

Good results obtained by the SPSIM index show that superpixels adequately reflect the operation of the human visual system and are suitable as a framework for image processing. SPSIM is not the only IQA model that uses superpixel segmentation. Last year a measure based on local SURF matching and SuperPixel Difference (SSPD) was presented [14]. In this complex model applied to Depth Image-Based Rendering (DIBR) synthesised views, the calculation of gradient magnitude differences at the superpixel level was used as one of three information channels. The score obtained from these superpixel comparisons is one element of the final score expressed by the SSPD index.

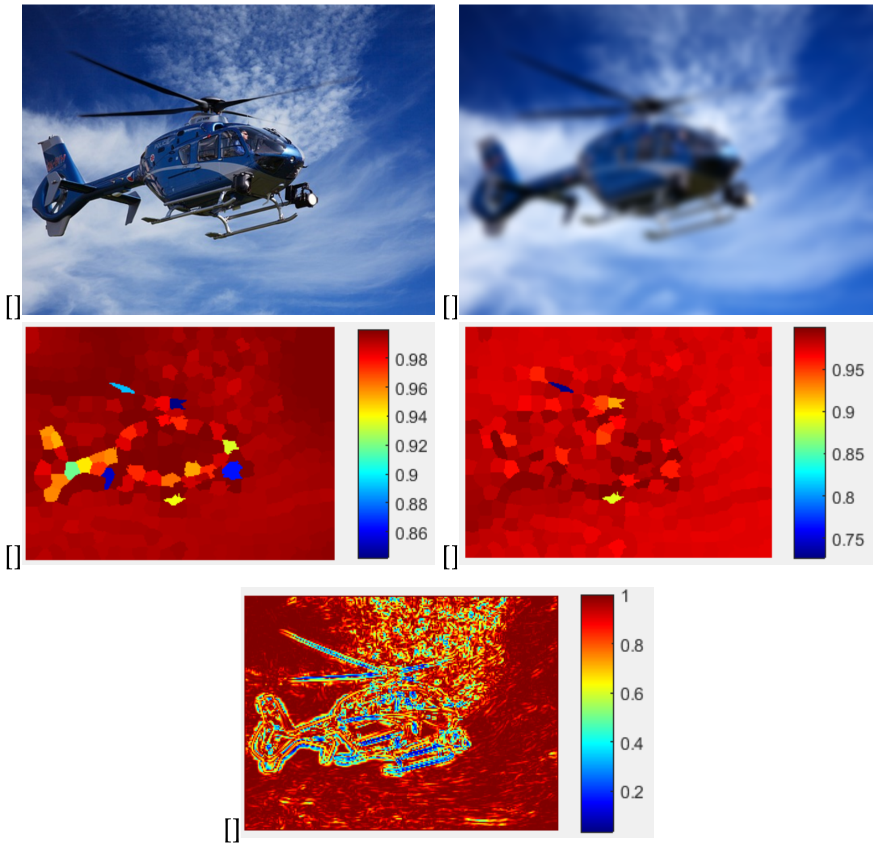

Figure 3. Local similarity maps for SPSIM index (400 superpixels): (a) reference image, (b) distorted image, (c) luminance similarity map, (d) chrominance similarity map, (e) gradient similarity map

3. The Proposed Modifications of SPSIM

In 2013 [15] a modification of the SSIM index was proposed including an analysis of color loss based on a color space that represents color information better than the classical RGB space. A study was carried out to determine which color space best represents chromatic changes in an image. A comparison was made between LIVE images, obtained by modifying the index using YCbCr, HSI, YIQ, YUV and CIELab color spaces. From among the analysed colour spaces, the best convergence with subjective assessment was shown for the quality measure based on YCbCr space. YCbCr is a color space proposed by the International Telecommunication Union, which was designed for processing JPEG format digital images [16] and MPEG format video sequences. Conversion RGB to YCbCr is expressed by the formulae:

$$Y = Min(Max(0, round(0.299R + 0.587G + 0.114B)), 255), \tag{20}$$

$$Cb = Min(Max(0, round(-0.1687R - 0.3313G + 0.5B + 128)), 255), \tag{21}$$

$$Cr = Min(Max(0, round(0.5R - 0.4187G + 0.0813B + 128)), 255). \tag{22}$$

The Y component represents image luminance, while the Cb and Cr components represent the chrominance. The high convergence of the color version of the SSIM index using the YCbCr color space with the subjective assessment allows us to assume that the use of this color space for another IQA indices should also improve the prediction performance of the image quality assessment. Therefore, it is possible to modificate the SPSIM index algorithm consisting in replacing the YUV space used in it by the YCbCr space while keeping the other steps of the algorithm unchanged.

Another proposal to improve image quality assessment using the SPSIM index is to exploit the advantages of the previously described MDSI quality measure. For both SPSIM and MDSI measures,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0189.v1

local similarity of chrominance and image structure are used to determine their values. Both of these components differently for each of the index. In the case of the SPSIM index, the chrominance similarity for each of the two channels is calculated separately, whereas for the MDSI index, the chrominance similarity is calculated simultaneously. Similarly for structural similarity, the MDSI index takes into account additionally a map of the gradient values for the combined luminance channels, whereas the map of local structural similarities in SPSIM is calculated for each of the images separately. Therefore, the second proposal for modification of the SPSIM index consists in to determine the local chrominance and image structure similarity maps based on approaches known from the MDSI index. Such approach employs a combination of two methods: SPSIM and MDSI.

Both modifications of the SPSIM algorithm proposed above were separately implemented and compared with other quality measures. Furthermore a solution combining both proposed modifications to SPSIM has been also implemented. The effectiveness of the resulting modifications to SPSIM is presented in the next section.

4. Experimental Tests

Four benchmark databases LIVE (2006), TID2008, CSIQ (2010) and TID2013 were selected for the initial experiment, which are characterised by a large number of reference images, a variety of distortions and different levels of their occurrence in the images.

The LIVE image dataset contains 29 raw images that have been subjected to five types of distortions: JPEG compression, JPEG2000 compression, white noise, Gaussian blur and bit errors occurring during transmission of the compressed bit stream. Each type of distortion was assessed by an average of 23 participants. The majority of images have the resolution 768x512 pixels [17].

The TID2008 image database contains 1700 distorted images, which were generated using 17 types of distortions, 4 levels per distortion superimposed on 25 reference images. Subjective ratings were obtained based on 256 428 comparisons made by 838 observers. All images have a resolution of 512x384 pixels [18].

The CSIQ database contains 30 reference images and 866 distorted images using six types of distortions. The distortions used used include JPEG compression, JPEG2000 compression, Gaussian noise, contrast reduction and Gaussian blur. For each original image, between four to five levels of distortions of each type. The average values subjective ratings were calculated on the basis of 5000 ratings obtained from 35 interviewed subjects. The resolution of all images is 512x512 pixels [19].

Image database TID2013 is an extension version of the earlier image collection TID2008. For the same reference images, the number of distortion types has been increased to 24, and the number of distortion levels to 5. The database contains a total of 3000 distorted digital images. The research group, on the basis of which the average subjective ratings of the images were determined, was also increased [20].

Below is a summary table of the most relevant information about the selected IQA benchmark databases (Table 1). The information on the KADID-10k database will be further discussed in Sect.5.

Table 1. Comparison of the selected IQA databases

Database name	No. of original images	No. of distortion types	Ratings per image	Environment	No. of distorted images
LIVE	29	5	23	lab	779
TID2008	25	17	33	lab	1700
CSIQ	30	6	5~7	lab	866
TID2013	25	24	9	lab	3000
KADID-10k	81	25	30	crowdsourcing	10125

The values of the individual IQA measures should be compared with the values of the subjective ratings for the individual images. Four criteria are usually used to assess the monotonicity, linearity and accuracy of such prediction: the Spearman's rank order correlation coefficient (SROCC), the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0189.v1

Kendall's rank order correlation coefficient (KROCC), Pearson's linear correlation coefficient (PLCC) and root mean squared error (RMSE). The formulae for such comparisons are presented below:

$$SROCC = 1 - \frac{6 * \sum_{i=1}^{N} d_i^2}{N(N^2-1)} \quad , \tag{23}$$

$$KROCC = \frac{N_c - N_d}{0.5 * (N-1) * N} \quad , \tag{24}$$

$$PLCC = \frac{\sum_{i=1}^{N}(p_i - \overline{p})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{N}(p_i - \overline{p})^2 (s_i - \overline{s})^2}} \quad , \tag{25}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - s_i)^2} \quad . \tag{26}$$

Tables 2-5 contain the values of the Spearman, Kendall and Pearson correlation coefficients and RMSE errors for the tested IQA measures. In each column of the table, the top three results are shown in bold. The right two columns contain the arithmetic average and weighted average values with respect to the number of images in each database. Taking into account the weighted averages of the calculated correlation coefficients and the RMSE errors, we can see that modifications of the SPSIM index give very good results. A superpixel number of 400 was assumed for the SPSIM index and its three modifications.

Table 2. Comparison of SROCC values for selected IQA measures

	CSIQ	LIVE	TID2008	TID2013	$\overline{SROCC}$	$\overline{SROCC_w}$
PSNR	0.8087	0.8730	0.5245	0.6869	0.7233	0.6829
SSIM	0.8718	0.9226	0.6780	0.7214	0.7985	0.7550
FSIMc	0.9309	**0.9645**	0.8840	0.8510	0.9076	0.8847
GMSD	**0.9570**	0.9603	0.8907	0.8044	0.9031	0.8675
VSI	0.9422	0.9524	0.8979	0.8965	0.9223	0.9100
MDSI	**0.9568**	**0.9667**	**0.9208**	0.8899	**0.9336**	0.9167
PSIM	**0.9620**	0.9623	0.9119	0.8926	**0.9322**	0.9158
VCGS	0.9442	0.9558	0.8975	0.8926	0.9225	0.9087
SPSIM	0.9434	0.9607	0.9104	0.9043	0.9297	0.9182
SPSIM (YCbCr)	0.9445	0.9606	0.9127	**0.9054**	0.9308	**0.9195**
SPSIM (MDSI)	0.9425	**0.9630**	**0.9131**	**0.9052**	0.9310	**0.9195**
SPSIM (YCbCr_MDSI)	0.9434	0.9625	**0.9150**	**0.9067**	**0.9319**	**0.9208**

Table 3. Comparison of KROCC values for selected IQA measures

	CSIQ	LIVE	TID2008	TID2013	$\overline{KROCC}$	$\overline{KROCC_w}$
PSNR	0.5989	0.6801	0.3696	0.4958	0.5361	0.4987
SSIM	0.6776	0.7474	0.4876	0.5286	0.6103	0.5648
FSIMc	0.7684	**0.8363**	0.6991	0.6665	0.7426	0.7100
GMSD	**0.8122**	0.8268	0.7092	0.6339	0.7455	0.7021
VSI	0.7850	0.8058	0.7123	0.7183	0.7554	0.7365
MDSI	**0.8123**	**0.8395**	**0.7515**	0.7123	**0.7789**	**0.7521**
PSIM	**0.8265**	0.8298	**0.7395**	0.7161	**0.7780**	**0.7514**
VCGS	0.7899	0.8141	0.7171	0.7166	0.7594	0.7387
SPSIM	0.7859	0.8268	0.7294	0.7249	0.7668	0.7469
SPSIM (YCbCr)	0.7884	0.8267	0.7336	**0.7272**	0.7690	0.7495
SPSIM (MDSI)	0.7859	**0.8311**	0.7362	**0.7282**	0.7704	0.7509
SPSIM (YCbCr_MDSI)	0.7877	0.8307	**0.7393**	**0.7306**	**0.7721**	**0.7530**

 doi:10.20944/preprints202102.0189.v1

Table 4. Comparison of PLCC values for selected IQA measures

	CSIQ	LIVE	TID2008	TID2013	$\overline{PLCC}$	$\overline{PLCC_w}$
PSNR	0.7857	0.8682	0.5405	0.6788	0.7183	0.6796
SSIM	0.8579	0.9212	0.6803	0.7459	0.8013	0.7651
FSIMc	0.9191	**0.9613**	0.8762	0.8769	0.9084	0.8928
GMSD	**0.9541**	**0.9603**	0.8788	0.8590	0.9131	0.8897
VSI	0.9279	0.9482	0.8762	0.9000	0.9131	0.9033
MDSI	**0.9531**	**0.9659**	**0.9160**	0.9085	**0.9359**	**0.9236**
PSIM	**0.9642**	0.9584	**0.9077**	0.9080	**0.9346**	**0.9218**
VCGS	0.9301	0.9509	0.8776	0.9000	0.9147	0.9044
SPSIM	0.9335	0.9576	0.8927	0.9090	0.9232	0.9139
SPSIM (YCbCr)	0.9346	0.9564	0.8946	**0.9099**	0.9238	0.9148
SPSIM (MDSI)	0.9327	0.9592	0.9049	**0.9165**	0.9283	0.9208
SPSIM (YCbCr_MDSI)	0.9334	0.9583	**0.9051**	**0.9173**	**0.9285**	**0.9213**

Table 5. Comparison of RMSE values for selected IQA measures

	CSIQ	LIVE	TID2008	TID2013	$\overline{RMSE}$	$\overline{RMSE_w}$
PSNR	0.1624	13.5582	1.1290	0.9103	3.9400	2.4196
SSIM	0.1349	10.6320	0.9836	0.8256	3.1440	1.9776
FSIMc	0.1034	**7.5296**	0.6468	0.5959	**2.2189**	1.3936
GMSD	**0.0786**	**7.6214**	0.6404	0.6346	2.2437	1.5080
VSI	0.0979	8.6817	0.6466	0.5404	2.4916	1.4181
MDSI	**0.0795**	**7.0790**	**0.5383**	0.5181	**2.0537**	1.4792
PSIM	**0.0696**	7.7942	**0.5632**	0.5193	2.2366	**1.2470**
VCGS	0.0964	8.4557	0.6433	0.5404	2.4339	1.3855
SPSIM	0.0942	7.8711	0.6047	0.5167	2.2717	1.3629
SPSIM (YCbCr)	0.0934	7.9766	0.5996	**0.5143**	2.2960	1.3959
SPSIM (MDSI)	0.0947	7.7210	0.5712	**0.4958**	**2.2207**	**1.3483**
SPSIM (YCbCr_MDSI)	0.0942	7.8048	**0.5705**	**0.4935**	2.2407	**1.3573**

The performance of the analysed new quality measures was measured using a PC with an Intel Core i5-7200U 2.5GHz processor and 12GB of RAM. The computational scripts were performed using the Matlab R2019b platform. For each image from the TID2013 set, the computation time of each quality index was measured. The performance of an index is measured as the average time to evaluate the image quality. Table 6 shows the results obtained for the studied IQA measures.

Table 6. Computation times for algorithms of new quality measures

IQA index	$\overline{time}$ [ms]
MDSI	2.14
PSIM	5.33
VCGS	47.68
SPSIM	16.07
SPSIM (YCbCr)	16.80
SPSIM (MDSI)	15.86
SPSIM (YCbCr_MDSI)	17.44

5. Additional Tests on Large-Scale IQA Database

The IQA databases previously used by us contained a limited number of images, most of them had TID2013 database i.e. 3000 images. The use of online crowdsourcing for image assessment allowed to build larger databases. In 2019, a large-scale image database called KADID-10k (Konstanz Artificially

 doi:10.20944/preprints202102.0189.v1

Distorted Image quality Database) [21], which contains more than 10,000 images with subjective scores of their quality (MOS) was created and published. This database contains a limited number of image types (81), a limited number of artificial distortion types (25) and a limited number of levels for each type of distortion (5), resulting in 10,125 digital images. Recently, the KADID-10k database is frequently used to training, validation and testing of deep learning networks used in image quality assessment [22]. Further use of pretrained convolutional neural networks (CNNs) in IQA algorithms seems very promising.

High perceptual quality images from the public website have been rescaled to a resolution 512x384 (Figure 4). The artificial distortions used during the creation of the KADID-10k database include blurs, noises, spatial distortions etc. A novelty was the use of idea of crowdsourcing for subjective IQA in this database. Details of crowdsourcing experiments are described in the article [21].

Figure 4. KADID-10k image database: reference images

Large-scale IQA database KADID-10k we used in tests of SPSIM index and its modifications with different number of superpixels. The results obtained are presented in Table 7. The best results for a given number of superpixels are highlighted in bold. In each case, the best results were obtained for the SPSIM modification based on a combination of color space change to YCbCr and application of the similarity maps used in MDSI.

6. Conclusions

The article shows that there is still much room for improvement in the field of FR-IQA indices. Among the many measures of quality, the SPSIM index using superpixel patches and local similarity maps has recently appeared. In this article, we show that simple SPSIM modifications (change of color space, different way of definition of similarity maps inspired by MDSI index) lead to improved correlation with MOS scores. While changing the color space slightly improves these correlations, a different way of defining similarity maps improves them more. The combination of both modifications marked as SPSIM(YCbCr_MDSI) gives the best results.

Table 7. The impact of the number of superpixels on: SROCC, KROCC, PLCC and RMSE values.

	SROCC	KROCC	PLLC	RMSE
No. of SP: 100				
SPSIM	0.8672	0.6782	0.8675	0.5386
SPSIM (YCbCr)	0.8674	0.6788	0.8675	0.5385
SPSIM (MDSI)	0.8705	0.6827	0.8696	0.5346
SPSIM (YCbCr_MDSI)	**0.8709**	**0.6836**	**0.8698**	**0.5342**
No. of SP: 200				
SPSIM	0.8715	0.6836	0.8718	0.5303
SPSIM (YCbCr)	0.8719	0.6843	0.8719	0.5301
SPSIM (MDSI)	0.8750	0.6883	0.8741	0.5259
SPSIM (YCbCr_MDSI)	**0.8755**	**0.6892**	**0.8744**	**0.5254**
No. of SP: 400				
SPSIM	0.8743	0.6871	0.8744	0.5254
SPSIM (YCbCr)	0.8746	0.6879	0.8745	0.5251
SPSIM (MDSI)	0.8777	0.6917	0.8766	0.5209
SPSIM (YCbCr_MDSI)	**0.8781**	**0.6927**	**0.8769**	**0.5203**
No. of SP: 1600				
SPSIM	0.8755	0.6897	0.8749	0.5242
SPSIM (YCbCr)	0.8762	0.6909	0.8755	0.5231
SPSIM (MDSI)	0.8791	0.6945	0.8774	0.5193
SPSIM (YCbCr_MDSI)	**0.8799**	**0.6958**	**0.8781**	**0.5180**
No. of SP: 4000				
SPSIM	0.8759	0.6908	0.8751	0.5239
SPSIM (YCbCr)	0.8769	0.6923	0.8760	0.5222
SPSIM (MDSI)	0.8797	0.6959	0.8778	0.5186
SPSIM (YCbCr_MDSI)	**0.8807**	**0.6974**	**0.8787**	**0.5168**

Author Contributions: Conceptualization, M.F. and H.P.; methodology, M.F.; software, G.S.; validation, M.F. and H.P.; investigation, M.F. and G.S.; resources, M.F.; data curation, M.F.; writing—original draft preparation, M.F.; writing—review and editing, M.F.; visualization, M.F.; supervision, H.P.

Acknowledgments: This work was supported by the Polish Ministry for Science and Education under internal grant $02/070/BK_20/0005(BK-324/RAU4/2020$ for the Department of Data Science and Engineering, Silesian University of Technology, Gliwice, Poland.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nadenau, M.; Winkler, S.; Alleysson, D.; Kunt, M. Human vision models for perceptually optimized image processing–a review. *Proceedings of the IEEE* **2000**, *32*.
2. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **2004**, *13*, 600–612.
3. Liu, A.; Lin, W.; Narwaria, M. Image quality assessment based on gradient similarity. *IEEE Transactions on Image Processing* **2011**, *21*, 1500–1512.
4. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing* **2011**, *20*, 2378–2386.
5. Xue, W.; Zhang, L.; Mou, X.; Bovik, A.C. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing* **2013**, *23*, 684–695.
6. Zhang, L.; Shen, Y.; Li, H. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing* **2014**, *23*, 4270–4281.
7. Nafchi, H.Z.; Shahkolaei, A.; Hedjam, R.; Cheriet, M. Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator. *IEEE Access* **2016**, *4*, 5579–5590.
8. Gu, K.; Li, L.; Lu, H.; Min, X.; Lin, W. A fast reliable image quality predictor by fusing micro- and macro-structures. *IEEE Transactions on Industrial Electronics* **2017**, *64*, 3903–3912.

 doi:10.20944/preprints202102.0189.v1

9. Shi, C.; Lin, Y. Full Reference Image Quality Assessment Based on Visual Salience With Color Appearance and Gradient Similarity. *IEEE Access* **2020**, *8*, 97310–97320.
10. Sun, W.; Liao, Q.; Xue, J.H.; Zhou, F. SPSIM: A superpixel-based similarity index for full-reference image quality assessment. *IEEE Transactions on Image Processing* **2018**, *27*, 4232–4244.
11. Ren, X.; Malik, J. Learning a classification model for segmentation. Proc. IEEE Int. conf. on Comp.Vis. IEEE, 2003, Vol. 1, pp. 10–17.
12. Stutz, D.; Hermans, A.; Leibe, B. Superpixels: An evaluation of the state-of-the-art. *Computer Vision and Image Understanding* **2018**, *166*, 1–27.
13. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2012**, *34*, 2274–2282.
14. Mahmoudpour, S.; Schelkens, P. Synthesized view quality assessment using feature matching and superpixel difference. *IEEE Signal Processing Letters* **2020**, *27*, 1650–1654.
15. Jadhav, M.; Dandawate, Y.; Pisharoty, N. Performance evaluation of structural similarity index metric in different color spaces for HVS based assessment of quality of colour images. *International Journal of Engineering and Technology* **2013**, *5*, 1555–1562.
16. ITU-T Recommendation T.871: Information technology - Digital compression and coding of continuous-tone still images: JPEG File Interchange Format (JFIF).
17. Sheikh, H.; Sabir, M.; Bovik, A. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing* **2006**, *15*, 3440–3451.
18. Ponomarenko, N.; Lukin, V.; Zelensky, A.; Egiazarian, K.; Carli, M.; Battisti, F. TID2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics* **2009**, *10*, 30–45.
19. Larson, E.; Chandler, D. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging* **2010**, *19*, 011006.
20. Ponomarenko, N.; Jin, L.; Ieremeiev, O.; Lukin, V.e.a. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication* **2015**, *30*, 57–77.
21. Lin, H.; Hosu, V.; Saupe, D. Kadid-10k: A large-scale artificially distorted iqa database. 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2019, pp. 1–3.
22. Varga, D. Composition-preserving deep approach to full-reference image quality assessment. *Signal, Image and Video Processing* **2020**, *14*, 1265–1272.