PROBABILISTIC MEASURE OF COLOUR IMAGE PROCESSING FIDELITY

Eugeniusz Kornatowski — Krzysztof Okarma *

In the paper a probabilistic approach to quality assessment of image processing algorithms is proposed. Presented scalar measure can be used for any colour space and gives very similar results regardless on the image content. It can be an interesting supplement to existing image quality metrics in applications where the details of the processing algorithm are known. Its good correlation with subjective evaluation is comparable to some other recently proposed measures such as Universal Image Quality Index, Structural Similarity and SVD-based measure.

K e y w o r d s: colour image quality assessment, image processing fidelity, statistical image processing, objective measures

1 INTRODUCTION

In many cases for various image processing algorithms there is a need of objective output image quality assessment which should be well correlated with Human Visual System (HVS). In recent years some novel image quality metrics have been proposed but their disadvantages are quite strong dependence on the colour space and the necessity of using the original image during calculations. Another idea is based on blind image quality assessment (without requirement of using original image) but such methods are often very complicated and not always lead to satisfactory results.

In many applications the assessment of output image quality can be successfully replaced by measuring the quality of the processing algorithm corresponding to the amount of distortions introduced into processed image. Such approach is independent on the image content but requires the exact knowledge of the processing algorithm.

2 QUALITY MEASURES IN COLOUR IMAGE PROCESSING

Traditional approach to objective image quality assessment is based on some well-known scalar metrics such as Mean Square Error (MSE) and some similar measures such as Root Mean Square Error (RMSE), Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR) *etc* [1]. Generally they are defined for greyscale images only so colour information is not utilised at all. A possible extension is the calculation of such metrics for each RGB channel independently and treating their mean (or weighted mean) value as the overall image quality measure. On the other hand using some other colour spaces such as YUV or CIE L*a*b* is also possible. In that case one of the main disadvantages of that approach is the dependence on the colour space used during calculations.

Another idea can be subjective image quality measurement with statistical analysis of the results obtained from questionnaires filled by observers (*eg* Mean Opinion Score - MOS). However performing such evaluations can be troublesome especially for large number of observers needed for reliable statistical analysis.

Conventional objective metrics are poor correlated with HVS so some other proposals have been presented in recent years. One of the most popular seems to be the Universal Image Quality Index proposed by Wang and Bovik [2]. Such measure models image distortions as the combination of three elements: loss of correlation, luminance distortion and loss of contrast. Assuming $x_{i,j}$ and $y_{i,j}$ are the values of luminance for the pixel (i,j) of original and distorted image respectively, it is defined as the local index for single image block ($N \times N$ pixels - usually $N = 8$) as

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \frac{2\bar{x}\bar{y}}{(\bar{x})^2 + (\bar{y})^2} \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} = \\ = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)\,[(\bar{x})^2 + (\bar{y})^2]}\,, \tag{1}$$

where $\bar{x}$ and $\bar{y}$ are the mean values and σ stands for the standard deviation in the original and distorted image blocks respectively [2]. The overall quality index is defined as the mean value of metrics (1) obtained for all blocks using sliding window approach. Extension of such approach for assessment of colour images is discussed in the paper [3].

In the paper [4] the definition of Universal Image Quality Index has been extended into Structural Similarity (SSIM) introducing the possibility of choosing the importance exponent for each of three factors in (1) with additional stability enhancement for regions where $\bar{x}$ or σ_x^2

* Szczecin University of Technology, Chair of Signal Processing and Multimedia Engineering, 71-126 Szczecin, Poland, korn@ps.pl, okarma@ps.pl

are close to zero. The modification of (1) based on using two coefficients preventing instability can be expressed as

$$SSIM = \frac{(2\bar{x}\bar{y} + C_1)\ (2\sigma_{xy} + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_1)\ [(\bar{x})^2 + (\bar{y})^2 + C_2]}\ , \quad (2)$$

where C_1 and C_2 are small values chosen experimentally as suggested by authors of the paper [4].

Another recently published idea is based on the Singular Value Decomposition (SVD) of the matrix corresponding to the image block of fixed size (*eg* 8×8). Numerical quality measure can be calculated as [5]

$$M_{SVD} = \frac{\sum_{i=1}^{K} |D_i - D_{mid}|}{K}\ , \quad (3)$$

where K is the total number of $N \times N$ pixels blocks and D_{mid} stands for the middle element of the sorted vector D calculated as

$$D_i = \sqrt{\sum_{k=1}^{N} (s_k - \hat{s}_k)^2}\ , \quad (4)$$

where s_k and $\hat{s}_k$ are the singular values for original and distorted image block respectively.

Local measures obtained for each block are then averaged in order to achieve single scalar metric.

Extension of proposed measure into colour image assessment in YUV colour model has been discussed in the paper [6]. Analysis of the usefulness of some popular colour spaces for the colour image quality assessment using the metrics described above is presented in [7].

Another approach based on full image transforms (DFT, DWT and DCT) using luminance layer from YUV colour space has been also presented recently [8]. The idea is based on the calculation of the mean of four standard deviations of the differences between transform's magnitude coefficients for four bands using original and distorted image.

3 PROPOSED APPROACH

As an alternative for the methods discussed above using Probabilistic Image Fidelity (PIF) criterion is proposed. The idea of such measure will be presented for greyscale images at first. It is based on the assumption that some of nonlinear digital signals or image processing algorithms can be described by polynomials as shown *eg* by Yli-Harja *et al* [9]. Such polynomials illustrate the relation between the distribution functions of input and output signals, assuming independence of the random variables (greyscale levels for monochrome images). In the paper [9] some example polynomials for some weighted median filtering algorithms are presented. In the general case for the filter window consisting of $N = 2v+1$ pixels we obtain

$$G_v(x) = \sum_{i=v+1}^{2v+1} c_i\, F^i(x)\ , \quad (5)$$

where G and F denote the distribution function of input and output signals' probability distribution respectively, x denotes discrete random variable (for image processing algorithms $x \in \langle 0, L\rangle$ is an integer number corresponding to the luminance level).

An interesting, although evident, case is the filter with $v = 0$ (one-element window) leading to

$$G_0(x) = F(x)\ . \quad (6)$$

Such case is equivalent to the algorithms which do not introduce any distortions in the processed signal. However, for higher orders of the polynomial (5) the output signal projection's error increases. In purpose to estimate the quality of processing in the probabilistic fidelity sense, the difference between the curves (5) and (6) should be calculated.

For some DSP algorithms functions G and F can be determined analytically [9]. However generally, *eg* for any of lossy compression algorithms, analytical definition of (5) can be very difficult to find. For that reason an experimental method is proposed, based on the examination of the linearity for the function (5). Deviation of that function in comparison to the linear one can be used to determine the probabilistic processing fidelity.

For the signal with given probability distribution, probability density function of the random variable x is denoted as $f(x)$. For discrete random variable the probability function $P(x = x_k)$ is defined and the distribution function can be expressed as

$$F(x) = \sum_{x_k < x} P(X = x_k)\ . \quad (7)$$

Assuming the particular case for processing a greyscale image represented by $L+1$ luminance levels represented by integer numbers ($0 \le x \le L,\ x \in N$) we obtain

$$P(x = x_k) = \frac{1}{L} \quad and \quad F(x) = \frac{x}{L} \quad (8)$$

and

$$G(x) = \sum_i C_i \left(\frac{x}{L}\right)^i\ . \quad (9)$$

Equation (9) represents the same curve as (5) for the arguments $0 \le F \le 1$ but in that case the input argument of G is a random variable $0 \le x \le L$ with uniform distribution. On the basis of (9) we can conclude that for input signal with uniform distribution, the output distribution function depending on the random variable characterizes also the dependence of the output distribution function on the input one. The necessary condition is scaling the input range of function G (proportional coefficient $1/L$ in (9). Testing an algorithm using input signal with uniform distribution the universal expression $G(F)$ can be obtained, correct also for the input signal with any distribution.

The idea described above has been verified for the standard 2-D median filter with mask of 3×3 pixels where function $G(F)$ is given as [9]

$$G(x) = 126\, F^5(x) - 420\, F^6(x) + 540\, F^7(x) - 315\, F^8(x) + 70\, F^9(x)\;. \tag{10}$$

Obviously in purpose to achieve good convergence of curves, which can be obtained analytically and experimentally, good quality of uniform distribution generator is required. Experimental computation of $G(F)$ is the basis for further calculations leading to determining PIF measure for analyzed algorithm of filtering, compression *etc.*

The general form of proposed criterion PIF can be expressed as

$$PIF = 1 - const\, \frac{\int_0^1 \left[G_v(F) - G_0(F)\right]^2 dF}{\int_0^1 \left[G_0(F)\right]^2 dF}\;. \tag{11}$$

Assuming that $PIF = 1$ for $v = 0$ and $PIF = 0$ for $v = \infty$ it is related to the ideal maintenance and complete degradation of the output image respectively. Considering also the formula

$$\int_0^1 \left[G_0(F)\right]^2 dF = \frac{1}{3} \tag{12}$$

the final form of the expression (11) can be derived as

$$PIF = 1 - 12 \int_0^1 \left[G(F) - F\right]^2 dF\;. \tag{13}$$

To estimate the value of PIF experimentally the following steps should be performed:

- determining the $G(F)$ function experimentally
- numerical calculation of PIF value (13)

Analysed algorithm of digital image processing should be applied to the input noise image of uniform distribution, then the output probability density function should be determined. It requires previous calculation of the histogram and then, according to (7), the discrete distribution function can be calculated.

Obtained PIF coefficient corresponds to the measure of image's structure preserving after processing and is independent on the processed image. The only but very important requirement is using the generator with uniform probability distribution with the best possible properties for testing purposes.

A characteristic property of the PIF measure defined in the formula (13) is the fact that it is not sensitive on some types of disturbances which do not change the probability function of the input signal or change it imperceptibly. An example of such algorithm can be the whirl effect illustrated in Fig. 1f.

Calculating the PIF coefficient value for that algorithm we obtain 1 regardless of visible changes of the image. In purpose to maintain also such types of disturbances the extension of the definition [13] is proposed. The modified formula can be expressed as:

$$RPIF = \frac{R+1}{2}\, PIF = \frac{R+1}{2} \left(1 - 12 \int_0^1 \left[G(F) - F\right]^2 dF\right), \tag{14}$$

where R denotes the correlation between input and output images (equivalent to Matlab's corr2 function scaled to the range $< 0\,, 1 >$ in purpose to ensure the same dynamic range for proposed criterion). Similar approach is used also by Wang and Bovik [2] in their Universal Image Quality Index (1) where the first factor also represents the loss of correlation.

Calculation of RPIF criterion is then performed by two-pass algorithm: calculation of the PIF coefficient (13) using test image of noise with uniform distribution and then determining of the correlation coefficient $\frac{R+1}{2}$ for analyzed image.

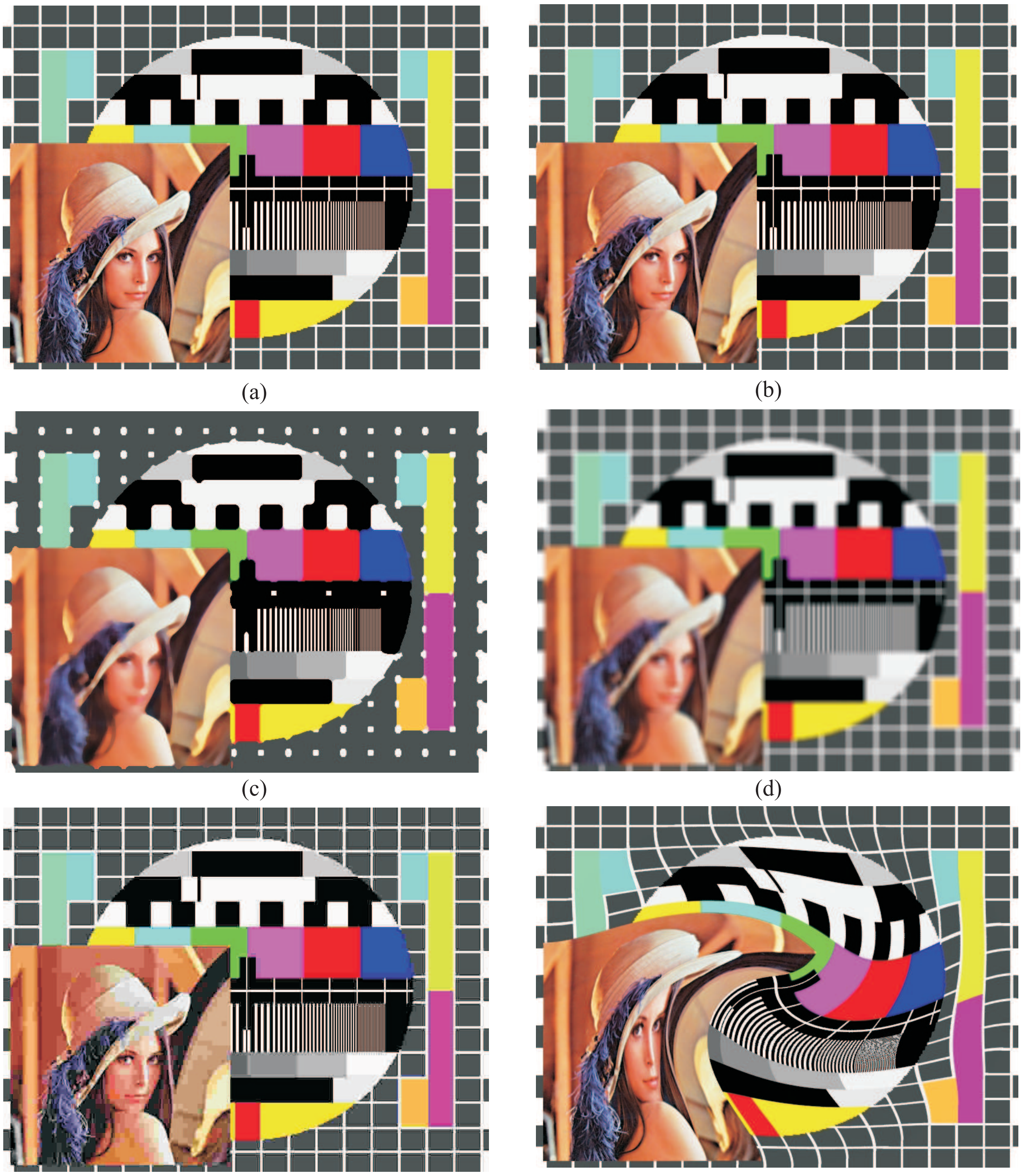
Extension of proposed measure for the assessment of full colour images is possible using different colour spaces. For any colour space the following formula is proposed:

$$RPIF_N = \left(\prod_{i=1}^{N} RPIF_i\right)^{1/N}, \tag{15}$$

which is better correlated with subjective evaluation than *eg* arithmetic mean value from coefficients obtained for all channels, as confirmed by several tests using a set of commonly used images performed by the authors. Results presented in this paper are obtained using formula (15) in RGB colour space. Calculations of the RPIF coefficient value for each RGB channel lead to the same results but slight differences caused by the correlation factor have been noticed.

4 COMPARISON OF RESULTS

With regard to the fact that some conventional image quality metrics are not always consistent with the subjective evaluation some comparative experiments have been performed. For the tests a set of images *eg* image shown in Fig. 1a has been used. All presented values of image quality metrics have been calculated for algorithms chosen in order to ensure questionless subjective evaluation results. In the experiments lossy JPEG compression algorithm, low-pass and median filtration using various mask sizes have been used. Compression ratios have been estimated as file sizes after compression relative to achieved using lossless JPEG algorithm. Degrading the quality of image



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

Fig. 1. Example results of processing: (a) Original image; (b) Median filtered image - mask 3×3; (c) Median filtered image - mask 11×11; (d) Low-pass filtered image - mask 11×11; (e) JPEG compressed image - compression ratio 1:15; (f) Whirl effect applied to original image.

Table 1. Results obtained for median and low-pass filtering, and JPEG compression

Quality	Median filter mask			Low-pass filter mask			JPEG compression		
measure	3×3	7×7	11×11	3×3	7×7	11×11	1:2	1:10	1:20
PSNR (R)	20.879	15.837	12.559	19.962	16.357	14.631	44.415	28.162	22.013
PSNR (G)	20.724	15.776	12.550	19.879	16.316	14.617	46.707	28.701	22.558
PSNR (B)	20.684	15.758	12.455	19.841	16.304	14.592	41.694	27.236	21.189
M_{SSIM} (R)	0.963	0.875	0.705	0.898	0.709	0.543	0.990	0.853	0.733
M_{SSIM} (G)	0.958	0.866	0.701	0.895	0.705	0.543	0.994	0.839	0.729
M_{SSIM} (B)	0.942	0.848	0.676	0.878	0.689	0.524	0.983	0.817	0.698
M_{SVD} (R)	14.329	40.022	126.585	38.386	80.931	110.758	2.874	14.418	55.055
M_{SVD} (G)	15.465	42.086	128.086	39.679	81.919	110.411	2.262	12.774	49.578
M_{SVD} (B)	17.810	44.892	135.088	40.169	82.275	110.821	4.254	18.083	63.457
$RPIF_{\mathrm{N}}$	0.766	0.454	0.350	0.539	0.256	0.167	0.999	0.940	0.870

using such algorithms the results of subjective evaluation are easy to forecast. Some results of processing are shown in Fig. 1b-e.

More detailed comparison of obtained results for various image quality measures is presented in Table 1 (PSNR values are expressed in dBs). The values of M_{SSIM} metric have been calculated without using any smoothing window assuming $C_1 = (0.01 \times 255)^2$ and $C_2 = (0.03 \times 255)^2$ as suggested in the paper [4].

The values of Universal Image Quality Index obtained for the test images have been incorrect (outside the dynamic range $\langle -1, 1 \rangle$) because of many dark regions in the images. For such regions the values of $\bar{x}$, $\bar{y}$, σ_x^2 and σ_y^2 may be very close to zero so the algorithm becomes unstable so only SSIM measure has been used in comparisons. However achieved results are sensitive to the values of C_1 and C_2, e.g. results for JPEG compressed image with ratio 1:15 for $C_1 = (0.05 \times 255)^2$ and $C_2 = (0.05 \times 255)^2$ are equal to $M_{SSIM(R)} = 0.8359$, $M_{SSIM(G)} = 0.8426$ and $M_{SSIM(B)} = 0.8072$ for RGB channels respectively while values obtained with the same values of C_1 and C_2 as presented in Table 1 are about 10% worse ($M_{SSIM(R)} = 0.7547$, $M_{SSIM(G)} = 0.7453$ and $M_{SSIM(B)} = 0.7116$).

All measures have noticeable different values for each RGB channel what complicates the evaluation using single scalar measure. In some other colour spaces there is the same situation and additionally different weights should be applied *eg* 50% for luminance and 25% for chrominance channels.

Some of the measures have different dynamic ranges, what also complicates the comparisons. Even one of the latest measure based on SVD has different scale and the dynamic ranges of Universal Image Quality Index and SSIM are $\langle -1, 1 \rangle$.

5 CONCLUSION

Proposed method of image processing algorithms quality assessment is almost independent on colour space and can be treated as an interesting supplement to existing measures well correlated with HVS like M_{SSIM} and M_{SVD}.

Using RPIF measure for colour image quality assessment requires the knowledge of processing algorithm and its accuracy depends on the quality of random generator used during calculations. However, presented results illustrate the advantages of proposed measure over many others commonly used in image processing applications. Presented approach can be useful especially for development of some new colour image processing algorithms to measure their quality.

References

[1] ESKICIOGLU, A.—FISHER, P.—CHEN, S.: Image Quality Measures and Their Performance, IEEE Trans. Comm. **43** No. 12 (1995), 2959–2965.

[2] WANG, Z. BOVIK,—A.: A Universal Image Quality Index, IEEE Signal Processing Letters **9** No. 3 (2002), 81–84.

[3] OKARMA, K.: Quality Measurement of Colour Images using Modified Universal Image Quality Index, Proc. of XXVIII Int. Conf. on Fundamentals of Electrotechnics and Circuit Theory IC-SPETO'2005, Gliwice-Ustroń, Poland, pp. 415–418.

[4] WANG, Z.—BOVIK, A.—SHEIKH, H.—SIMONCELLI, E.: Image Quality Assessment From Error Measurement to Structural Similarity, IEEE Trans. Image Process. **13** No. 4 (2004), 600–612.

[5] SHNAYDERMAN, A.—GUSEV, A.—ESKICIOGLU, A.: A Multidimensional Image Quality Measure Using Singular Value Decomposition, Proc. SPIE Image Quality and Syst. Perf. **5294** No. 1 (2003), 82–92.

[6] SHNAYDERMAN, A.—ESKICIOGLU, A.: Assessment of Full Color Image Quality with Singular Value Decomposition, 17-th IS&T/SPIE Annual Symposium on Electronic Imaging, Image Quality and System Performance, San Jose, CA, Proc. SPIE 5668 (2005).

[7] OKARMA, K.: Application of Various Color Spaces in Modern Digital Image Quality Assessment Method, Proc. of XXX Int. Conf. on Fundamentals of Electrotechnics and Circuit Theory IC-SPETO'2007, Gliwice-Ustroń, Poland, pp. 217–218.

[8] GIRSHTEL, E.—SLOBODYAN, V.—WEISSMAN, J.—ESKICIOGLU, A.: Comparison of Three Full-Reference Color Image Quality Measures, 18-th IS&T/SPIE Annual Symposium on Electronic Imaging, Image Quality and System Performance, San Jose, CA, Proc. SPIE 6059 (2006).

[9] YLI-HARJA, O.—ASTOLA, J.—NEUVO, Y.: Analysis of the Properties of Median and Weighted Median Filters Using Threshold Logic and Stack Filter Representation, IEEE Trans. Signal Process **39** No. (2) (1991) 395–410.

Received 5 March 2007

Eugeniusz Kornatowski graduated from the Faculty of Electrical Engineering, Szczecin University of Technology (1981), in branch automatic control engineering and electric metrology). He obtained his PhD at the Faculty of Electrical Engineering in 1989 (with honors). Since 1981 he has been working at Szczecin University of Technology. Between 1981 and 1989 he was at the Chair of Theoretical Electrotechnics and Computer Science; currently, he is an Assistant Professor in the Chair of Signal Processing and Multimedia Engineering. His research and education activities are within the area of digital signal processing, in particular nonlinear signal processing algorithms and objective criterions of their quality.

Krzysztof Okarma graduated from Szczecin University of Technology (Faculty of Electrical Engineering), Poland, in 1999. Just after that he joined the Chair of Signal Processing and Multimedia Engineering at the same faculty. In 2001 he received also the MSc Eng degree from the Faculty of Computer Science at the same University and in 2003 he received his PhD degree (with honors) in Electrotechnics from the Faculty of Electrical Engineering at Szczecin University of Technology. Currently, he is an Assistant Professor in the Chair of Signal Processing and Multimedia Engineering and a Visiting Professor in Higher School of Technology and Economics in Szczecin. His research interests include signal processing, digital image and video processing and quality assessment, as well as image and video transmission and coding.