STUDIA INFORMATICA 2014
Volume 34 Number 4 (114)

Przemysław SKUROWSKI
Silesian University of Technology, Institute of Informatics

TOWARDS THE NEW CONCEPT OF LINEAR IMAGE QUALITY ASSESSMENT. REVIEW OF PHASE CORRELATION FORMULAS

Summary. The article reports the progress in a development of an image quality assessment (IQA) method based on a new concept – phase correlation. The primary idea stems from the classical observation that structural information of the media/image information is stored within the phase part of the Fourier spectra. The paper describes the review and selection process of a correlation formula for future full reference IQA method. The results were verified with reference database against the human visual quality responses given in mean opinion score (MOS) scale, whereas the key goal was to keep the linearity of the results as it was observed in preliminary results. There were tested correlation models linear and circular raw and weighted in using various weighting schemes.

Keywords: image quality, phase correlation, circular correlation

NOWA KONCEPCJA LINIOWEJ METODY OCENY JAKOŚCI OBRAZÓW. ANALIZA FORMUŁ KORELACJI FAZ

Streszczenie. Artykuł opisuje prace związane z rozwojem metody oceny jakości obrazów, bazującej na nowej koncepcji – korelacji faz. Podstawowy pomysł wynika z klasycznej obserwacji, że informacja strukturalna w mediach/obrazach reprezentowana jest przez część fazową widma Fourierowskiego. Artykuł opisuje przegląd metod pomiaru korelacji i ich ocenę w celu dobrania formuły korelacyjnej dla przyszłej metody oceny obrazów. Wyniki zostały zweryfikowane względem ludzkich ocen, podanych w skalach MOS, zawartych w referencyjnych bazach danych, gdzie jako główny cel badawczy postawiono poszukiwanie liniowej zależności, tak jak zaobserwowano to we wstępnych wynikach. Przetestowano modele korelacji liniowej i cyklicznej w postaci prostej i ważonej z wykorzystaniem różnych schematów doboru wag.

Słowa kluczowe: jakość obrazów, korelacja faz, korelacja cykliczna

1. Introduction

Image quality assessment (IQA) methods plays pivotal role for image compression and reproduction. The key problem for IQA algorithms is to provide relevance to the human visual perception. It is not a trivial task [1] as human quality perception is not a linear process in terms of classical signal processing measures such as MSE or SNR so they perform poorly [2] when used for evaluating image quality.

The key for the article is discovery [3] by Oppenheim and Lim, that the most of the media information structure is stored within the phase spectrum (Fig. 1). Long lasting knowledge on the significance of signal phase has not influenced the development of image quality assessment (IQA) methods for quite a long time. The most of image distortion or quality objective measure methods are based mainly on amplitude spectra using some human related weighting function [4–6], alternatively there is used multiresolution approach [7] or sophisticated statistics [8]. Therefore measuring image similarity or distortion by phase comparing still remains promising field of research that recently gained some interest from the scientists [9–12]

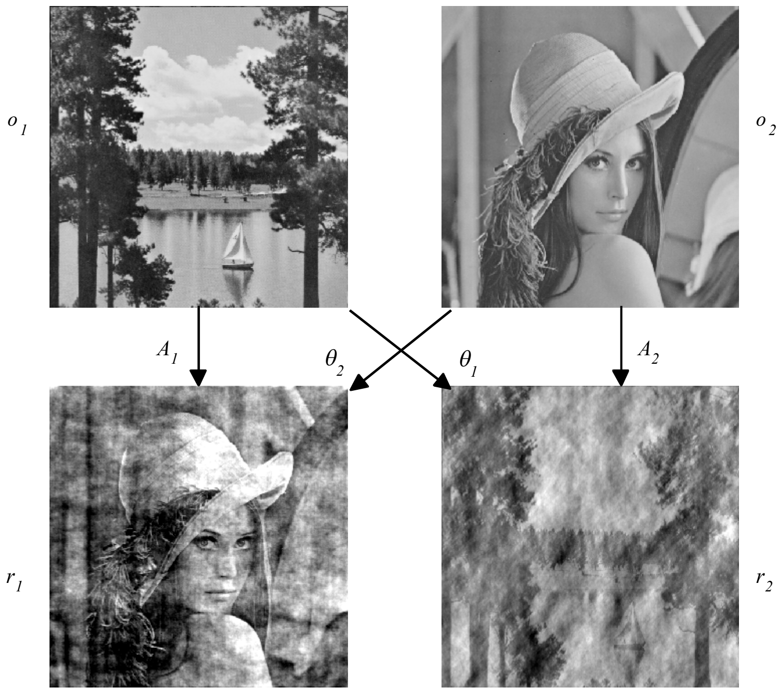

Fig. 1. Swapping amplitude (A) and phase (θ) of Fourier spectra for 'lake' and 'lena' images
Rys. 1. Zamiana amplitud (A) i faz (θ) widm Fouriera dla obrazów 'lake' i 'lena'

Authors former results [13] demonstrated (see Fig. 2) interesting and noteworthy property that the Pearson's phase correlation coefficients (R^P) between the phase spectra of an original and distorted images are linearly correlated with human subjective quality of experience (QoE) judgments given using the MOS scale. That is highly desired property, alas, the simple

R^P measure appeared to have good accuracy but still unsatisfactory precision (see Fig. 2). Another advantage is the fact, that the resulting measure is conceptually simple and easy to implement in any development environment.

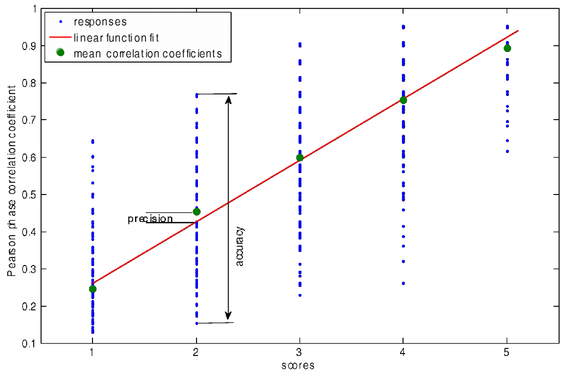

Fig. 2. Preliminary results
Rys. 2. Wstępne wyniki

The aim of the work is to overcome low precision limitation by selection of a more appropriate correlation method resulting in precise responses. As a key criterion there was assumed correspondence of the results to the human visual quality responses from reference databases given in mean opinion score (MOS) scale, another intention was to keep the property of linearity of the results as it is desired feature and the most notable property of the proposed method. Following types of correlation were analyzed and tested: linear, circular – simple and weighted by spectral amplitude to choose the best *phase correlation coefficient* (PCC) to use in image quality measurement.

2. Background

2.1. Spectral representation

The discrete two dimensional Fourier transform [14] is given as:

$$O(u,v)=F\left(o(m,n)\right)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} o(m,n)\,e^{-j(2\pi u/M)m}\,e^{-j(2\pi v/N)n}, \tag{1}$$

which returns as a result complex spectra consisting of real ($\Re$) and imaginary ($\Im$) parts (spectra). With Euler's formula these parts can be also considered as an amplitude A and phase θ spectra which are given as:

$$\begin{aligned} A(u,v) &= |O(u,v)| = \sqrt{\Re(O(u,v))^2+\Im(O(u,v))^2} \\ \theta(u,v) &= \arctan\left(\frac{\Im(O(u,v))}{\Re(O(u,v))}\right) \end{aligned}, \tag{2a,b}$$

where: $F(\ )$ is Fourier transform; $o(\ )$, $O(\ )$ are image in spatial and spectral form respectively; M, N are spatial sizes of images; m, n are spatial coordinates; u, v are frequency coordinates.

2.2. Linear correlation

Initially for the measurement of spectral phase similarity we used one of the most canonical tools Pearson Correlation Coefficient [15] which can be interpreted as normalized common (of both variables) variance from the respective mean values:

$$R_{xy}^{P}=\frac{s_{xy}}{s_x s_y}\ , \tag{3}$$

where: s_{xy} is the covariance; s_x, s_y are the standard deviations for random variables x and y.

Another agreement measure between two variables is Lin's Concordance Correlation Coefficient (CCC) [16, 17]. It is especially preferred in life sciences, it measures how close relationship between two variables is to a 45-degree line from the origin (of a slope 1):

$$R_{xy}^{L}=\frac{2\, s_{xy}}{s_x^2+s_y^2+(\bar{x}-\bar{y})^2}\ , \tag{4}$$

Where mean values, STD deviations and covariance estimators are respectively computed as:

$$\bar{x}=\frac{1}{N}\sum_{i=1}^{N} x_i, \quad s_x=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2}, \quad s_{xy}=\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})\ . \tag{5a-c}$$

2.3. Circular correlation

The phase of complex number varies in the cyclic range $[0, 2\pi)$ therefore the distance between two values given as a simple absolute difference might be incorrect (see Fig. 3), therefore it has been the natural next step in our analysis. For the need of analysis of circular data the whole sub-discipline in statistics is established [18]. The first approach to measurement of agreement between circular data was proposed by Fisher [19]. Although, we neglected it, as it appeared to be slow during our initial tests (in its convenient basic form) and to provide much worse results than the more convenient, modern version of circular correlation proposed by Jammalamadaka [18] which adopts the classical linear measure:

$$R_{\alpha\beta}^{J}=\frac{\sum_{i=1}^{N}\sin(\alpha_i-\bar{\alpha})\sin(\beta_i-\bar{\beta})}{\sqrt{\sum_{i=1}^{N}\sin^2(\alpha_i-\bar{\alpha})}\sqrt{\sum_{i=1}^{N}\sin^2(\beta_i-\bar{\beta})}}\ , \tag{6}$$

where: α, β are angular random variables, which can describe the phase angle of a complex number, therefore their mean values can be computed as:

$$\bar{\alpha}=\arg\left(\frac{1}{N}\sum_{i=1}^{N}e^{j\alpha_i}\right). \tag{7}$$

Using above we can note dispersion measures (covariance and STD deviation) as:

$$s_\alpha=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\sin^2(\alpha_i-\bar{\alpha})},\quad s_{\alpha\beta}=\frac{1}{N}\sum_{i=1}^{N}(\alpha_i-\bar{\alpha})(\beta_i-\bar{\beta}). \tag{8a,b}$$

Now, we can rewrite (6), as a simple corollary of above:

$$R^J_{\alpha\beta}=\frac{s_{\alpha\beta}}{s_\alpha s_\beta}, \tag{9}$$

Using the same concept we can try to adopt Lin's formula (4) of Concordance CC to the circular domain:

$$R^{cL}_{\alpha\beta}=\frac{2s_{\alpha\beta}}{s^2_\alpha+s^2_\beta+\sin^2(\bar{\alpha}-\bar{\beta})}, \tag{10}$$

Although, please treat it as an author's utility function, which has not been studied comprehensively for its properties elsewhere.

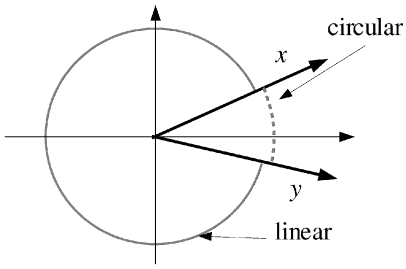

Fig. 3. Circular and linear distance between angular values *x* and *y*
Rys. 3. Odległość cykliczna i liniowa pomiędzy wartościami kątowymi *x* i *y*

2.4. Weighted correlation measures

The participation of phase of each harmonics in an image is related to its amplitude, therefore it was obvious in the next step to incorporate amplitude weighting into measurement. In such a case it is just necessary to substitute certain parts in respective formulas with their weighted counterparts (weighted mean, STD deviation and covariance):

$$\overline{wx}=\sum_{i=1}^{N}w_i x_i,\quad ws_x=\sqrt{\sum_{i=1}^{N}w_i(x_i-\overline{wx})^2},\quad ws_{xy}=\sum_{i=1}^{N}w_i(x_i-\overline{wx})(y_i-\overline{wy}), \tag{11a-c}$$

where w_i is weighting function. For the circular measures it is:

$$\begin{aligned}&\overline{w\alpha}=\arg\left(\sum_{i=1}^{N}w_i e^{j\alpha_i}\right),\quad ws_\alpha=\sqrt{\sum_{i=1}^{N}w_i\sin^2(\alpha_i-\overline{w\alpha})},\\&ws_{\alpha\beta}=\sum_{i=1}^{N}w_i\sin(\alpha_i-\overline{w\alpha})\sin(\beta_i-\overline{w\beta})\end{aligned}. \tag{12a-c}$$

The choice of appropriate weighting function will be discussed in the experimental part.

3. Results

The experimental part consists of the evaluation and tuning of correlation models for the needs of the IQA. Color RGB images at the current stage of development require to be converted to the grayscale, as it is a common step for many methods [5, 8], for this purpose the Matlab `rgb2gray` ($L = 0.2989*R + 0.5870*G + 0.1140*B$) was used.

The analysis consisted of two steps. In the initial step there was analyzed performance of simple correlation formulas then a weighting scheme is selected as a next step.

In order to evaluate the results, there were two basic criteria assumed:

- Relevance to the human judgment (monotonicity) which was measured with the Spearman (SROCC) and Kendall (KROCC) rank correlation coefficients – these correlation coefficients are assumed as the criteria since the most of IQAs are nonlinearly related to the human responses. This criterion relates to the precision.
- Linearity was verified using Pearson CC (verified better with CoD for linear model).
- The accuracy of a model (with special emphasize on linearity) – described as a *Goodness of Fit* (GoF) of a regression model between human responses and a measure result. It is measured with a *Coefficient of Determination* given as explained part of the variance – it can be computed effectively as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \sum_i (y_i - f_i)^2 / \sum_i (y_i - \bar{y})^2 , \tag{13}$$

  where: y_i is measured value; f_i predicted value; $\bar{y}$ mean value; SS_{res}, SS_{tot} residual and total sum of squares respectively.

3.1. The data

The results were evaluated using two reference databases – LIVE database [20] and TID2008 database [21]. Both of them are commonly applied for the evaluation of IQA methods – LIVE as a classic one and TID as newer and more comprehensive, containing some "odd" distortions. The LIVE contains DMOS scale evaluations of five image distortions at various degrees for 29 reference images collected in 25000 evaluations. The TID2008 database consists of 17 distortions at 4 levels for 25 ref images collected in 256428 MOS evaluations. Distortions in the TID2008 were grouped by the database authors into several categories. The IQA results can vary heavily depending on these categories – please see [21] where various IQAs are demonstrated against various distortion classes.

Table 1

Types of image distortions in reference databases

No	Type (in case of different nomenclature: LIVE/TID)	# in LIVE	# in TID
1	White noise / Additive Gaussian noise	174	100
2	Additive noise in color components		100
3	Spatially correlated noise		100
4	Masked noise		100
5	High frequency noise		100
6	Impulse noise		100
7	Quantization noise		100
8	Gaussian blur	174	100
9	Image denoising		100
10	JPEG compression	233	100
11	JPEG2000 compression	227	100
12	JPEG transmission errors		100
13	Fast Fading / JPEG2000 transmission errors	174	100
14	Non eccentricity pattern noise		100
15	Local block-wise distortions of different intensity		100
16	Mean shift (intensity shift)		100
17	Contrast change		100

3.2. Results and analysis of simple phase correlation formulas (PCC)

The observed results are plotted in respective Figs. 4-6 for visual examination and gathered in Tab. 2 to compare numerical results. One can note several interesting observations and conclusions on these.

Table 2

Results of evaluation criteria for PCC

	LIVE		TID2008	
Criterion:	Linear CC	Circular CC	Linear CC	Circular CC
SROCC	0.9519	0.9496	0.5291	0.5116
KROCC	0.8158	0.8120	0.3774	0.3644
Pearson	0.9545	0.9393	0.5796	0.5955
R^2(lin)	0.9111	0.8823	0.3360	0.3547
R^2(pow)	0.9115	0.8898	0.4235	0.4055
R^2(exp)	0.8998	0.8949	0.3066	0.3275
R^2(log)	0.9111	0.8814	0.4347	0.4173
R^2(sigm)	0.9087	0.8941	0.4251	0.3910

- The concordance and Pearson based correlation formulas provided almost identical results (to the fourth digit after the decimal point) in our cases, both for linear and for circular formulas for the provided datasets, therefore formulas (3) PCC_P and (8) PCC_J will be used as the default correlation versions – linear and circular.

- Circular models, due to uniform circular PDF, appeared to have volatility of the mean value (from ≈0 to ≈π) which resulted in an occasional change of sign of the resulting value. As a temporary countermeasure we used absolute value of PCC.
- The results for the LIVE database appear to be very promising, having both evaluation criteria at a reasonably high level. The rank correlation coefficients are high and the coefficients of determination provide that model explain approx. 80% of variance. Moreover, the linear model is on par with more flexible functions, whereas the competing fits were elongated very much so they resemble the linear function.
- The results for TID2008 database are ambiguous, as one can expect since the base is known to be 'harder case' for the metrics. Correlation is notably lower and regression models – logarithmic and power – perform far better than the desired linear function. On the other hand the linear model seems to be appropriate for the all but the lowest quality subset of images.

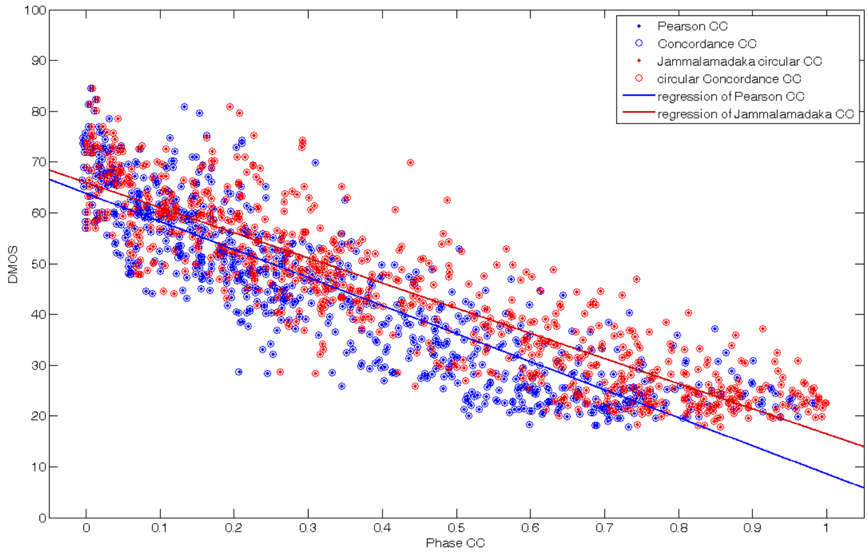

Fig. 4. Results plot PCC versus DMOS (all four types) for LIVE database
Rys. 4. Wykres wyników PCC vs DMOS (wszystkie cztery rodzaje) dla bazy LIVE

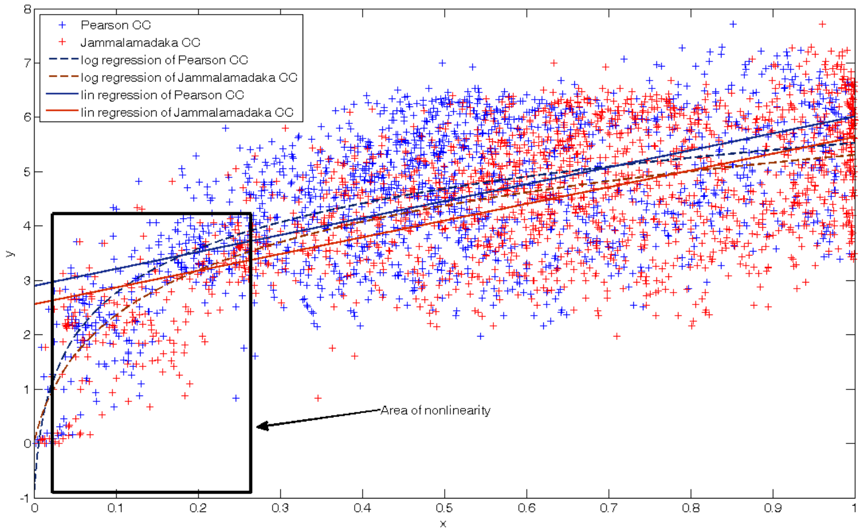

Fig. 5. Results plot PCC versus MOS (two basic types) for TID2008 database
Rys. 5. Wykres wyników PCC vs MOS (dwa podstawowe rodzaje) dla bazy TID2008

The last observation suggests the need for some qualitative analysis of the results which was necessary to perform. In the scatter plot (Fig. 6), with color distinction of every type of distortion and with the distinction of distortion categories according to [21] – actual (•) and exotic (□), we cannot observe any special distinction of distortion type in the nonlinear relationship area. We can suspect the simple PCCs describe strong distortion poorly and require improvement. One of such potential improvements is amplitude weighting.

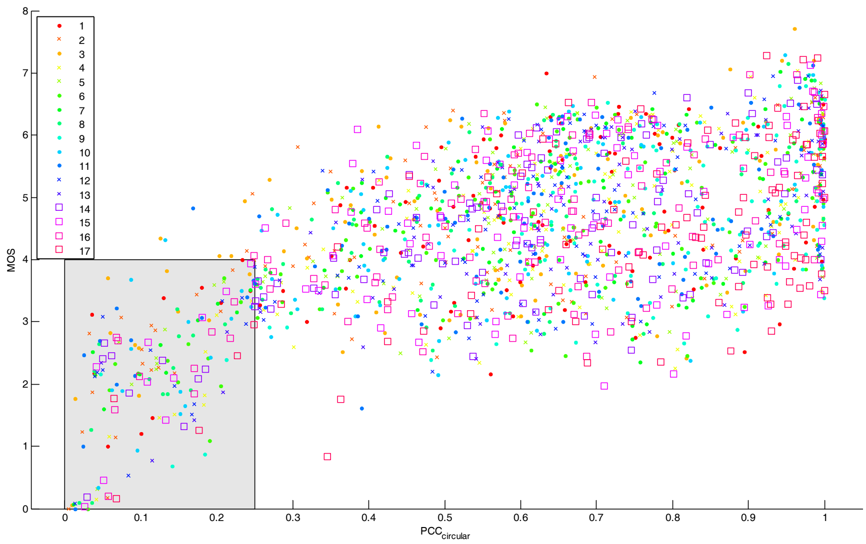

Fig. 6. Different distortion categories plot (in TID) PCC vs MOS for Jammalamadaka CC
Rys. 6. Wykres różnych kategorii zniekształceń (w TID) PCC vs MOS dla Jammalamadaka CC

3.3. Selection of the amplitude weighting (WPCC)

The intuitively obvious extension of the basic formulas (PCC) of the proposed approach is to incorporate weighting of the phases by the amplitudes (WPCC). The phase of each of harmonics can contribute to the image structure with different degree which is proportional to the amplitude. Since it is unknown whether the amplitude of source (A_1) or the distorted (A_2) image determine quality it was necessary to analyze several schemes, these are:

$$\begin{aligned}&\mathrm{SRC}=A_1,\ \ \mathrm{DST}=A_2,\ \ \mathrm{MAX}=\max\{A_1(n),A_2(n)\},\\&\mathrm{MIN}=\min\{A_1(n),A_2(n)\},\ \ \mathrm{MEAN}=(A_1+A_2)/2\end{aligned}. \tag{14a-e}$$

All WPCC results were qualitatively (visually) examined versus MOS/DMOS responses as demonstrated in Figs. 4-6 and verified against fitted regression models. Resulting correlations are demonstrated in the Tab. 3, the results of a special interest are underlined and provided in figures further in this paragraph. In the computation of the goodness of fit for all the regression schemes in the tables 4 and 5, there were excluded reference images – one can note a dot in (0,1) position in plots in Fig 7. In order to clearly identify weighing type we used following naming scheme WPCC$_P$-*type* and WPCC$_J$-*type* respectively for the linear and circular formulas where types are identified as in above equations (14).

Table 3

Correlation coefficients for tested weighting schemes of linear and circular WPCCs

		Linear PCC			Circular PCC		
	Measure:	SROCC	KROCC	Pearson	SROCC	KROCC	Pearson
LIVE database	<u>SRC</u>	0.9562	0.8241	0.9188	0.9576	0.8277	0.8708
	DST	0.8601	0.6895	0.7814	0.8737	0.7026	0.7176
	<u>MAX</u>	0.9437	0.8000	0.9098	0.9465	0.8057	0.8669
	MIN	0.8771	0.7140	0.8091	0.9077	0.7459	0.7630
	<u>MEAN</u>	0.9419	0.7959	0.8980	0.9477	0.8073	0.8526
TID database	<u>SRC</u>	0.6657	0.4831	0.7327	0.6091	0.4375	0.6867
	DST	0.5365	0.3682	0.5392	0.4972	0.3413	0.4554
	<u>MAX</u>	0.6268	0.4465	0.6815	0.5837	0.4137	0.6616
	MIN	0.6085	0.4296	0.6325	0.5844	0.4100	0.6186
	<u>MEAN</u>	0.6419	0.4588	0.6936	0.5963	0.4235	0.6717

Table 4

Goodness of Fit for different regression models for WPCCs in LIVE database

PCC	Criterion:	SRC	DST	MAX	MIN	MEAN
Linear	R^2 (linear)	0.8441	0.6105	0.8277	0.6546	0.8064
	R^2 (power2)	0.9073	0.7435	0.8924	0.7668	0.8880
	R^2 (exp)	0.8993	0.6948	0.8823	0.7357	0.8721
	R^2 (log)	0.8419	0.6077	0.8255	0.6512	0.8039
	R^2 (sigmoid)	**0.9093**	0.7349	0.8935	0.7636	0.8885
Circular	R^2 (linear)	0.7583	0.5150	0.7515	0.5821	0.7270
	R^2 (power2)	**0.8764**	0.7554	0.8681	0.7938	0.8683
	R^2 (exp)	0.8364	0.6270	0.8259	0.7062	0.8149
	R^2 (log)	0.7557	0.5076	0.7490	0.5382	0.7241
	R^2 (sigmoid)	0.8693	0.7257	0.8583	0.7751	0.8578

Table 5

Goodness of Fit for different regression models for WPCCs in TID database

PCC	Criterion:	SRC	DST	MAX	MIN	MEAN
Linear	R^2 (linear)	0.5369	0.2907	0.4645	0.4000	0.4810
	R^2 (power2)	0.5347	0.2897	0.4703	0.3941	0.4807
	R^2 (exp)	0.5070	0.2917	0.4380	0.3881	0.4552
	R^2 (log)	0.5434	0.2903	0.4796	0.4002	0.4900
	R^2 (sigmoid)	**0.5435**	0.2996	0.4789	0.4135	0.4911
Circular	R^2 (linear)	0.4716	0.2074	0.4378	0.3827	0.4512
	R^2 (power2)	0.4645	0.2547	0.4302	0.3898	0.4421
	R^2 (exp)	0.4684	0.2305	0.4244	0.3949	0.4412
	R^2 (log)	0.4711	0.2062	0.4395	0.3815	0.4512
	R^2 (sigmoid)	**0.4730**	0.2594	0.4375	0.4005	0.4511

Furthermore, during the analysis of the resulting WPCC values we can observe one particular case – the best theoretically motivated measure – source amplitude weighted circular PCC ($WPCC_J$-SRC). It has very high rank correlation but slightly odd behavior regarding the regression model in TID database. The visual inspection in Fig. 8 revealed that the measure is oversensitive to a single class of distortion – contrast change (number 17) marked with 'o' sign in plots – this observation is deliberated in the discussion of results.

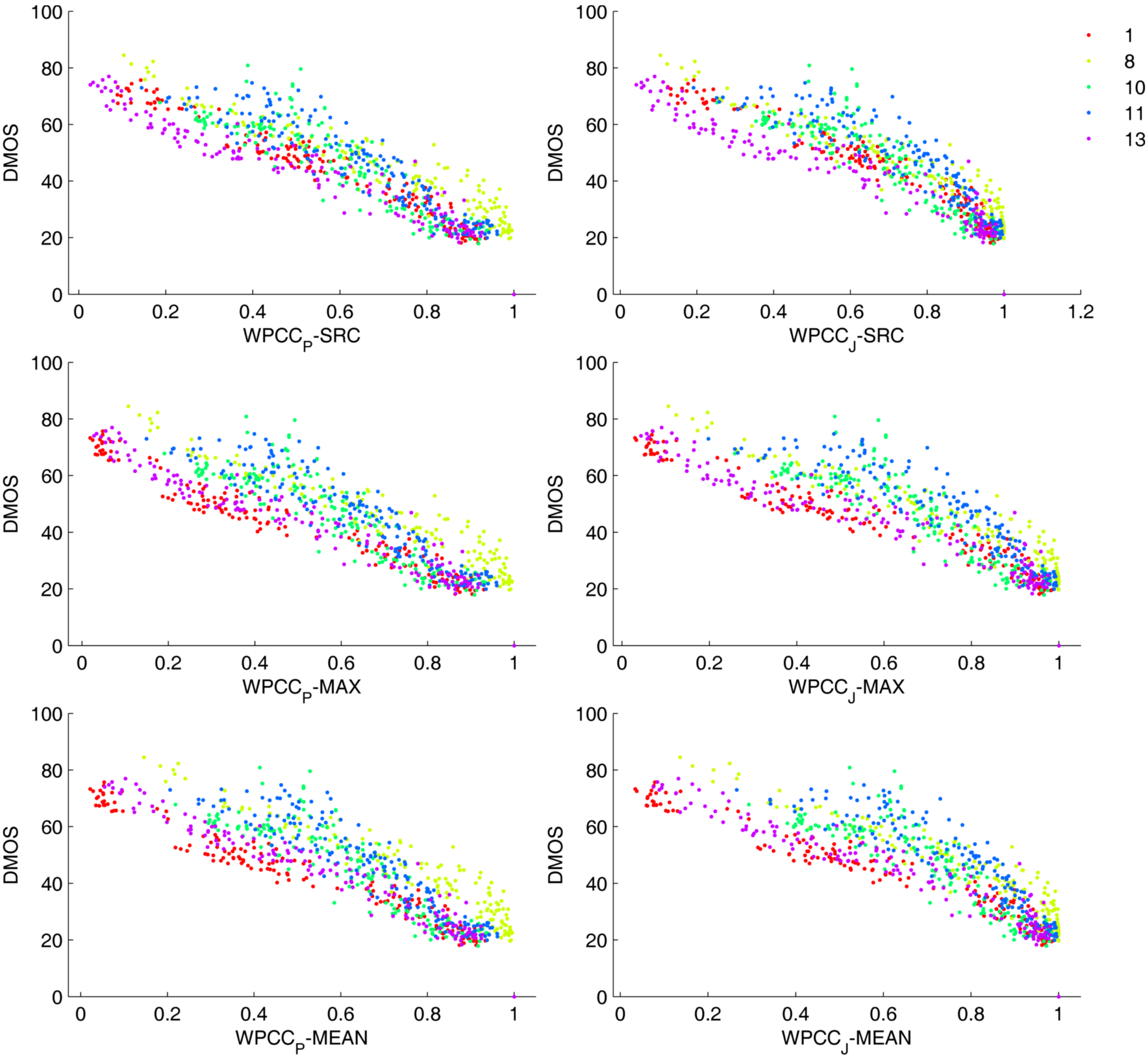

Fig. 7. Result plots of different weighting (SRC, MAX, MEAN) schemes in WPCC versus DMOS for LIVE database, categories as in Tab. 1

Rys. 7. Wykres wyników różnych metod ważenia (SRC, MAX, MEAN) dla WPCC vs DMOS dla bazy LIVE, kategorie jak w tab. 1

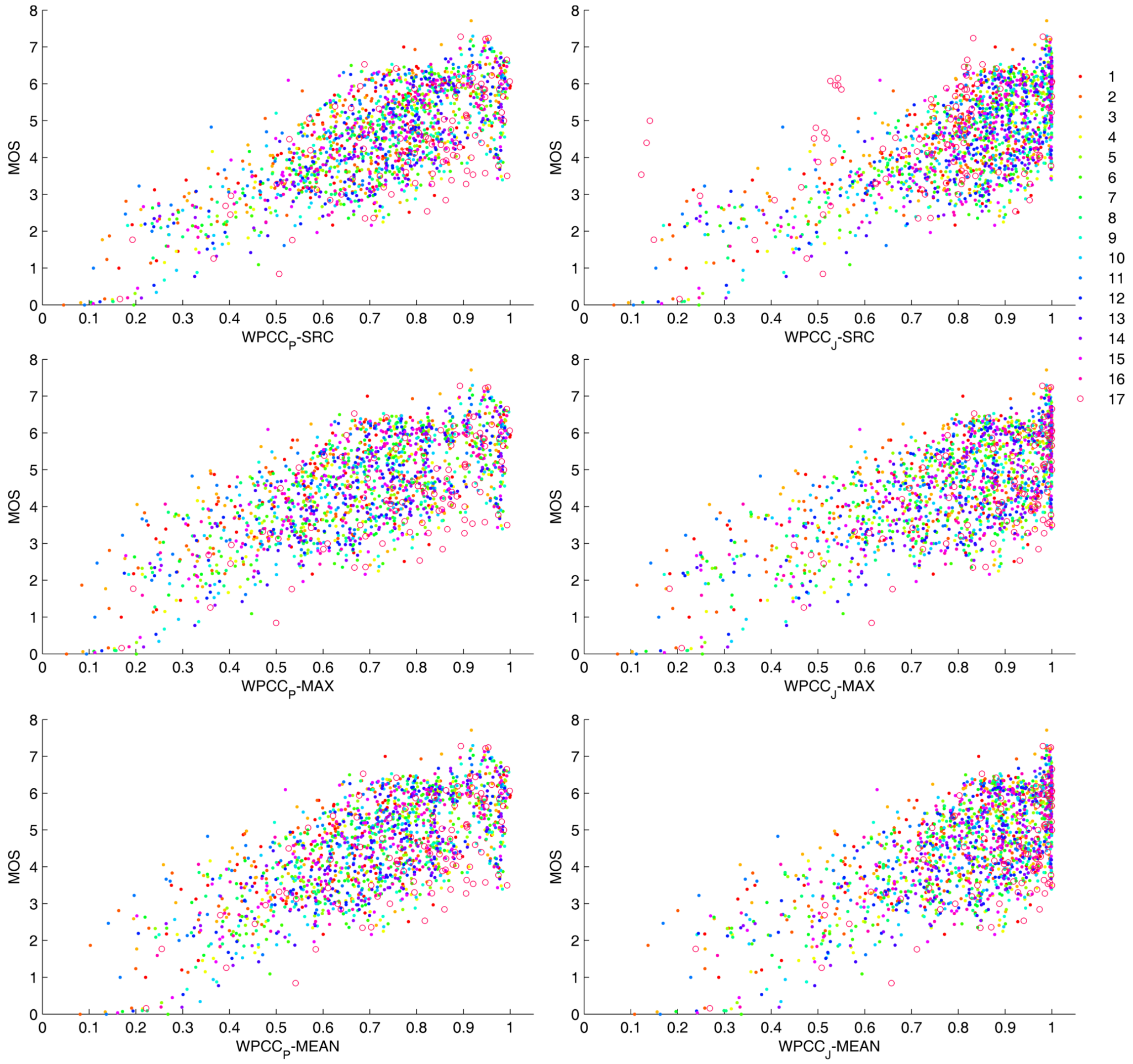

Fig. 8. Result plots of different weighting (SRC, MAX, MEAN) schemes in WPCC versus MOS for TID database, categories as in Tab. 1

Rys. 8. Wykres wyników różnych metod ważenia (SRC, MAX, MEAN) dla WPCC vs MOS dla bazy TID, kategorie jak w Tab. 1

3.4. Discussion of results

As one can note from the Tab. 3 versus Tabs. 4-5 the weighting models which provide the high rank correlation result also in good fit of the regressive models. Therefore there will be no contradiction in selection of the most appropriate weighting formula for the WPCC.

Both linear and circular CC with all three interesting weighting schemes provide similar results which are nearly linearly related to the human responses. Even, the more flexible regression models like a sigmoid or power function, are elongated such they resemble straight line (see Fig. 9). Therefore, according to the Occam razor principle, one should choose the simplest of the possible solutions – the linear model in this case.

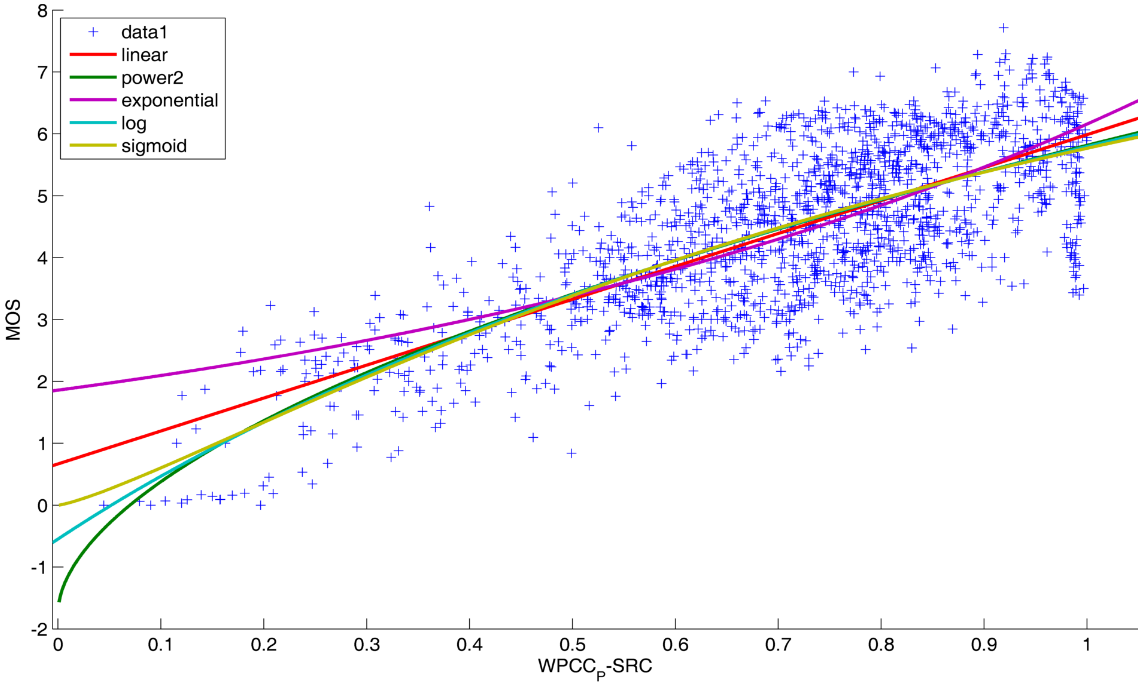

Fig. 9. Results plot WPCC$_P$-SRC versus MOS for TID2008 database
Rys. 9. Wykres wyników WPCC$_P$-SRC vs MOS dla bazy TID2008

Qualitative analysis of the resulting values of PCC suggest that they exhibit mainly the structure of information, although the information on contrast and mean color are poorly represented by phase (see example in Fig. 1). Thus certain types of distortions can be measured improperly. Average color value, as it is stored in Fourier DC component, always has 0 phase for real valued functions, they always fit each other and they are very significant due to the amplitude weighting. On the other hand the contrast change can be overrated (see results of class 17 distortion for WPCC$_J$-SRC in Fig. 8) as it changes a lot of harmonics. These issues should be addressed in further development of the WPCC based IQA method.

Further improvement is possible by taking viewing conditions into account. For this purpose we used simple heuristic approach which was adopted from the default implementation of single scale SSIM[1]. To adopt the image to the approximate proper viewing scale, which is approximately 3-5 times of image height or width, low pass filtering (LPF) is used with normalized uniform kernel of size f, in next step the input image is downsampled (↓) by the same f factor. This process and computing of f factor is described with formulas:

$$\begin{aligned} f &= \max\{1, \operatorname{round}(\frac{\min(M,N)}{256})\} \\ o_{scaled} &= [o * (\frac{1}{f}\operatorname{rect}(f))] \downarrow f \end{aligned}, \qquad (15)$$

where M, N is size of image. It results (Tab. 6) in notable precision improvement of results but in the most of interesting cases we lose the linearity. For the TID database the linear

[1]Available at SSIM webage https://ece.uwaterloo.ca/~z70wang/research/ssim/

regression improved performance but performance of other models was improved even more. For the LIVE database the ROCC coefficients were slightly improved but most of CoD coefficients were downgraded. Luckily the most clear results are for the $WPCC_P$-SRC which returned the best results both for TID and LIVE databases.

Table 6

Results of evaluation criteria for selected WPCCs for scale adopted images.

database:	TID2008				LIVE			
measure:	$WPCC_P$-SRC		$WPCC_J$-SRC		$WPCC_P$-SRC		$WPCC_J$-SRC	
criterion:	original	adopted	original	adopted	original	adopted	original	adopted
SROCC	0.6657	0.7658	0.6091	0.7489	0.9562	0.9587	0.9576	0.9601
KROCC	0.4831	0.5811	0.4375	0.5612	0.8241	0.8301	0.8277	0.8341
Pearson	0.7327	0.7935	0.6867	0.7334	0.9188	0.8434	0.8708	0.7439
R^2(lin)	0.5369	0.6296	0.4716	0.5379	0.8441	0.7114	0.7583	0.5534
R^2(pow)	0.5347	0.6291	0.4645	0.5937	0.9073	0.9158	0.8764	0.8647
R^2(exp)	0.5070	0.6316	0.4684	0.5937	0.8993	0.8321	0.8364	0.6737
R^2(log)	0.5434	0.6284	0.4711	0.5356	0.8419	0.6962	0.7557	0.4815
R^2(sigm)	0.5435	0.6360	0.4730	0.6018	0.9093	0.9029	0.8693	0.8076

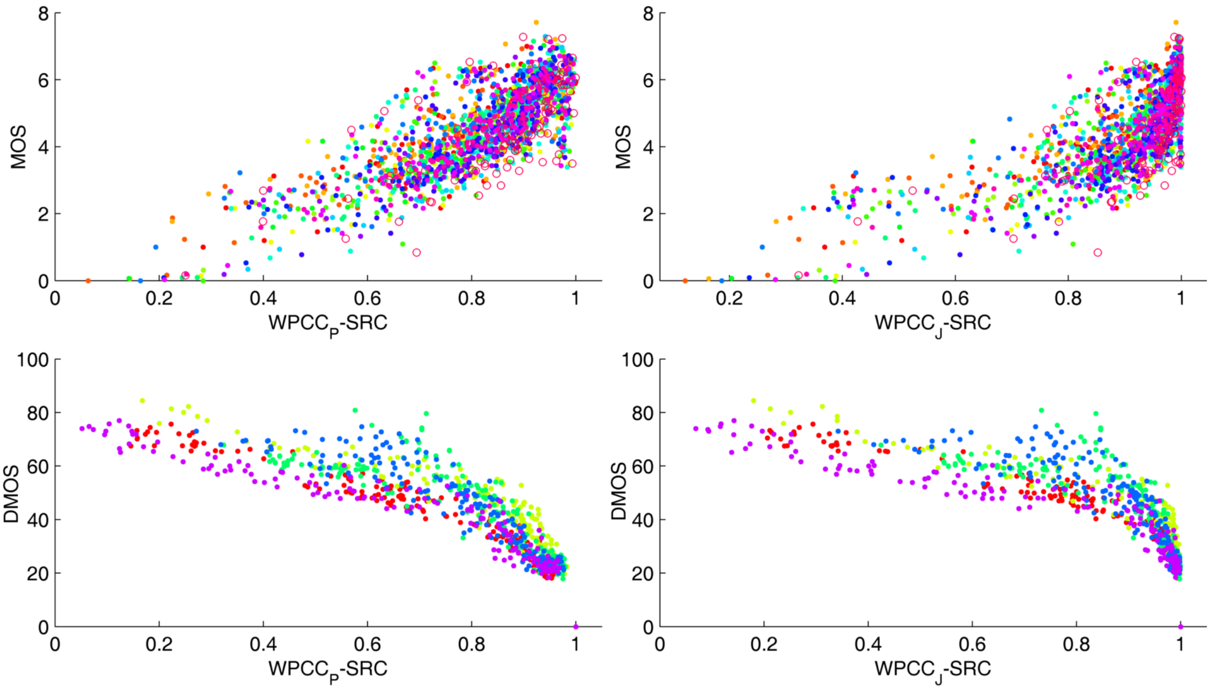

Fig. 10. Viewing conditions adopted results for selected WPCCs vs MOS for TID (first row) and LIVE (second) . Marker categories as in Figs. 7, 8.

Rys. 10. Wyniki po adaptacji do warunków obserwacji dla wybranych WPCC vs MOS dla baz TID (pierwszy wiersz) i LIVE (drugi). Oznaczenia markerów jak na rys. 7, 8.

The results of scale adoption are ambiguous, but the conclusion based on them is quite straightforward – including viewing conditions can result in significant improvement but one needs the knowledge on them first – what was the viewing distance, size of image and its resolution. Using of more sophisticated approaches should be taken into consideration. The difference of scaling adoption results between databases stems probably from the fact that such a rough approximate rule as used in our case worked especially well for the TID database which was gathered massively in 'uncontrolled' way with no special attention for the viewing distance [22]. Meanwhile LIVE respondents were asked to keep the fixed distance from the screen so the results are already well tuned.

To compare the results of the PCC to the results of other methods (Tab. 7) we employed the TID2008 database as it is more comprehensive and demanding one. The various measure results (sorted according SROCC) computed using Metrixmux [23], are provided along with the database [21, 24] and with measures [8, 11]. Additional results annotated with asterisk (*) are computed with scale adoption according to Eq. (15).

Table 7

Spearman and Kendall ROCCs for WPCC against some of existing IQAs

Measure	SROCC	KROCC	Pearson	Measure	SROCC	KROCC	Pearson
FSIMc	0.884	0.699	0.834	UQI	0.600	0.435	0.652
FSIM	0.880	0.695	0.830	PSNRHVS	0.594	0.476	0.576
MSSSIM	0.853	0.654	0.784	XYZ	0.577	0.434	0.482
SSIM*	0.775	0.577	0.740	IFC	0.569	0.426	0.212
$WPCC_P$-SRC*	**0.766**	**0.581**	**0.793**	PSNRHVSM	0.559	0.449	0.550
VIF	0.750	0.586	0.778	PSNRY	0.553	0.402	0.519
$WPCC_J$-SRC*	**0.749**	**0.561**	**0.733**	SNR	0.523	0.374	0.493
VSNR	0.705	0.534	0.293	MSE	0.525	0.369	0.293
$WPCC_P$-SRC	**0.666**	**0.483**	**0.733**	PSNR	0.525	0.369	0.489
VIFP	0.655	0.495	0.638	WSNR	0.488	0.393	0.463
SSIM	0.645	0.468	0.754	LINLAB	0.487	0.381	0.258
NQM	0.624	0.461	0.608	DCTUNE	0.476	0.372	0.286
$WPCC_J$-SRC	**0.609**	**0.437**	**0.687**				

Comparing the relationship between various IQAs (see Fig. 11), one can note that, except for the proposed WPCC, only the Universal Quality Index (UQI) [25] can be visually considered as linearly related to the human responses but it has poorer correlation to human judgments. Moreover, the single scale $WPCC_J$-SRC*, $WPCC_P$-SRC* and SSIM* which use relatively simple mathematical apparatus, provide results which are on par or outperform much more complex multiresolution/multiscale measures such as VIF, VIFP, VSNR. Significantly better results are obtained using MSSSIM and FSIM which are multiscale measures and as such are much more complex both conceptually and computationally (2-5 times). The complexity issue is important especially for the latter which employs pretty complex apparatus to compare information structure.

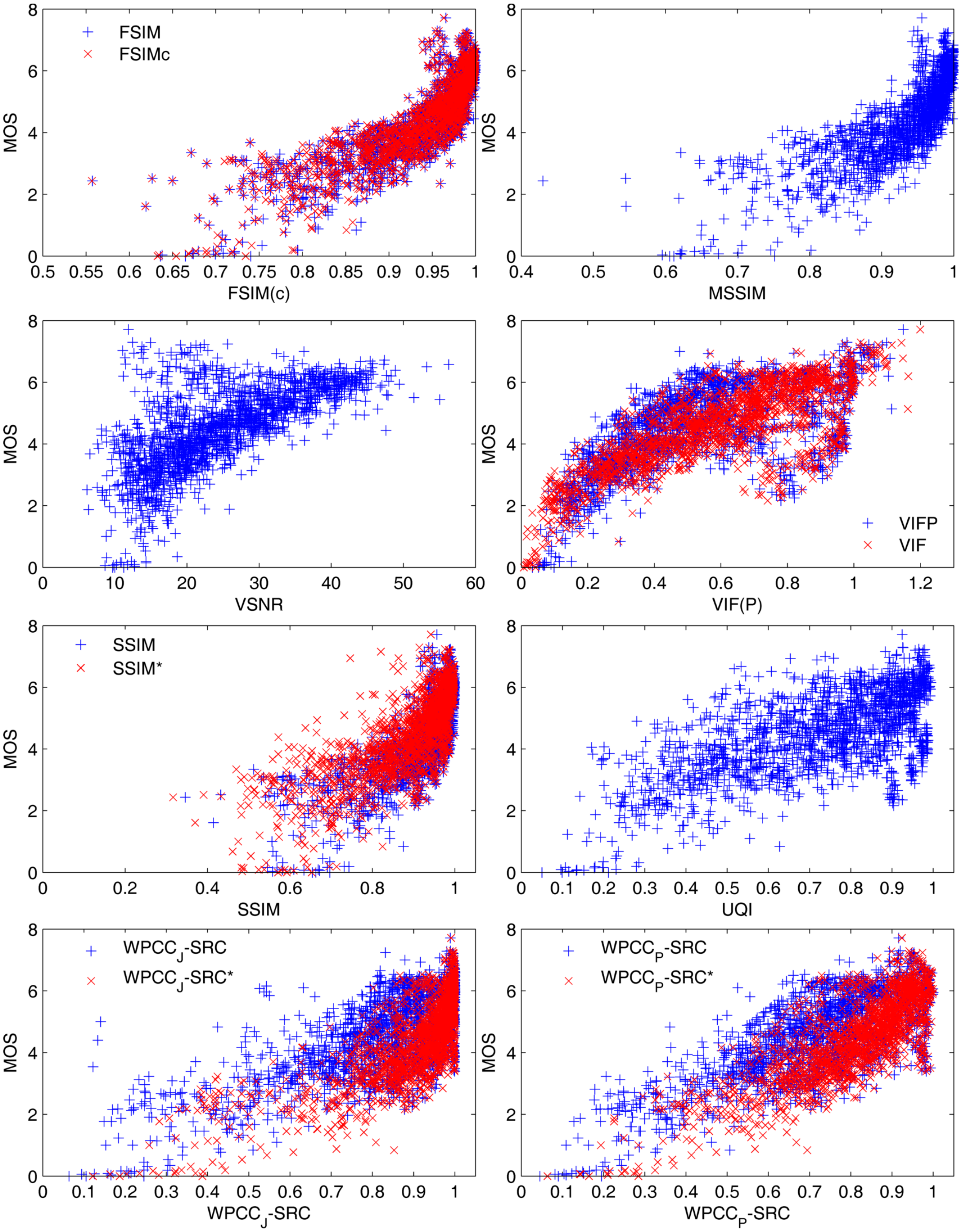

Fig. 11. Result plots of selected (best) different IQAs and selected WPCCs for TID2009 database
Rys. 11. Wyniki dla wybranych (najlepszych) metod IQA i wybranych WPCC dla bazy TID2009

4. Conclusions

The results of the proposed method are promising and at the current stage of development that already can be used to measure visual quality. If one would like to use WPCC 'as is' as a standalone quality measure then the $WPCC_P$-SRC (source weighted linear Pearson CC) seems to be the most appropriate choice due to the highest values of evaluation criteria. The $WPCC_J$-SRC seems to be better theoretically motivated and also provides good results, although it seems to be oversensitive (probably due to high accuracy) to contrast changes so there is some work need to be done to solve these issues. Further extensions would include computations of WPCC in color and possibly incorporating viewing conditions by additional weighting by contrast sensitivity functions (CSF) such as in [6, 14] and/or using multiresolution approach. Another prospective improvement can be using sliding window approach to evaluate measure by spatial pooling of WPCC distortion maps and some form of a spatial weighting.

The proposed WPCC solution is intended to become one day a part of more complex measures – the final concept is not fully established yet – but on the basis of SSIM concept it should be a function involving at least three components – color value, contrast and structure. WPCC would fit into such a form of measure very well as the last part which describes the structure similarity.

BIBLIOGRAPHY

1. Wang, Z., Bovik, A.C., Lu, L.: Why is image quality assessment so difficult? [in:] 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2002, vol. 4, p. IV–3313÷IV–3316.
2. Wang, Z., Bovik, A.C.: Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. IEEE Signal Process. Mag., vol. 26, no. 1, 2009, p. 98÷117.
3. Oppenheim, A.V., Lim, J.S.: The importance of phase in signals. Proc. IEEE, vol. 69, no. 5, 1981, p. 529÷541.
4. Campbell, F.W.: The human eye as an optical filter. Proc. IEEE, vol. 56, no. 6, 1968, p. 1009÷1014.
5. Mantiuk, R., Daly, S.J., Myszkowski, K., Seidel, H.-P.: Predicting visible differences in high dynamic range images: model and its calibration. [in:] Proc. SPIE 5666, Human Vision and Electronic Imaging X, 2005, vol. 5666, p. 204÷214.
6. Zhang, X., Wandell, B.A.: A spatial extension of CIELAB for digital color-image reproduction. J. Soc. Inf. Disp., vol. 5, no. 1, 1997, p. 61.
7. Wang, Z., Simoncelli, E.P.: Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. [in:] Proc. SPIE 5666, Human Vision and Electronic Imaging X, 2005, vol. 5666, p. 149÷159.

8. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process., vol. 13, no. 4, 2004, p. 600÷612.
9. Liu, Z., Laganiere, R.: On the Use of Phase Congruency to Evaluate Image Similarity. [in:] 2006 IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings, 2006, vol. 2, p. II 937 ÷ II 940.
10. Shao, X., Zhong, C.: Phase congruency assesses hyperspectral image quality. [in:] Proc. SPIE 8514, Satellite Data Compression, Communications, and Processing VIII, 2012, vol. 8514, p. 851408÷851408÷11.
11. Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: A Feature Similarity Index for Image Quality Assessment. IEEE Trans. Image Process., vol. 20, no. 8, Aug. 2011, p. 2378÷2386.
12. Narwaria, M., Lin, W., McLoughlin, I.V., *et al.*: Fourier Transform-Based Scalable Image Quality Measure. IEEE Trans. Image Process., vol. 21, no. 8, Aug. 2012, p. 3364÷3377.
13. Skurowski, P., Gruca, A.: Image Quality Assessment Using Phase Spectrum Correlation, [in:] Proceedings of the International Conference on Computer Vision and Graphics: Revised Papers, Lecture Notes in Computer Science, vol. 5337, Springer-Verlag, Berlin, Heidelberg 2009, p. 80÷89.
14. Bracewell, R.N.: The Fourier transform and its applications. 3rd ed., McGraw Hill, Boston 2000.
15. Rice, J.A.: Mathematical statistics and data analysis. 3rd ed., Thomson/Brooks/Cole, Belmont, CA, 2007.
16. Carrasco, J.L., Jover, L.: Estimating the Generalized Concordance Correlation Coefficient through Variance Components. Biometrics, vol. 59, no. 4, Dec. 2003, p. 849÷858.
17. Lin, L.I.-K.: A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics, vol. 45, no. 1, Mar. 1989, p. 255.
18. Jammalamadaka, S.R.: Topics in circular statistics. World Scientific, River Edge N.J 2001.
19. Fisher, N.I., Lee, A.J.: A correlation coefficient for circular data. Biometrika, vol. 70, no. 2, Jan. 1983, p. 327÷332.
20. Sheikh, H.R., Wang, Z., Cormack, L., Bovik, A.C.: LIVE Image Quality Assessment Database Release 2. [Online]. Available: http://live.ece.utexas.edu/research/quality.
21. Ponomarenko, N., Lukin, V., Zelensky, A., *et al.*: TID2008 – A Database for Evaluation of Full Reference Visual Quality Assessment Metrics. Adv. Mod. Radioelectron., vol. 10, no. 4, 2009, p. 30÷45.
22. Winkler, S.: Analysis of Public Image and Video Databases for Quality Assessment. IEEE J. Sel. Top. Signal Process., vol. 6, no. 6, 2012, p. 616÷625.
23. Gaubatz, M.: MeTriX MuX Visual Quality Assessment Package. [Online]. Available: http://foulard.ece.cornell.edu/gaubatz/metrix_mux/.
24. Ponomarenko, N., Battisti, F., Egiazarian, K., *et al.*: Metrics performance comparison for color image database. [in:] 4th International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, 2009, p. 14÷16.
25. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Process. Lett., vol. 9, no. 3, 2002, p. 81÷84.

Omówienie

Artykuł omawia postępy prac nad autorską metodą obiektywnej oceny jakości obrazów, bazującą na korelacji faz w widmie Fourierowskim. Przetestowano formuły korelacji liniowej Pearsona i konkordancji Lina oraz ich cyklicznych wersji – zaproponowanej przez Jammalamadaka podstawowej wersji korelacji, oraz wzoru konkordancji zaadaptowanej podobnie przez autora. Okazało się, że dla faz widm konkordancje dają identyczne wyniki z podstawowymi korelacjami.

W pracy wykazano, dla ogólnie uznanych zestawów danych pomiarowych (LIVE i TID2008), że degradacja struktury obrazu – wyrażona korelacją liniową (3), bądź cykliczną (6) w dziedzinie widma fazowego transformaty Fouriera pomiędzy obrazem oryginalnym i zniekształconym – jest prawie liniowo skorelowana z ludzkim postrzeganiem spadku jakości. Wymaga to zastosowania korelacji ważonej względnym udziałem poszczególnych składowych fazowych, które wyrażone są poprzez odpowiednie amplitudy w widmie. Szczególnie dobre rezultaty uzyskano przy wykorzystaniu jako wag amplitud obrazu źródłowego (14a) i średnich amplitud obrazu źródłowego i zniekształconego (14e).

Wyniki obliczeń eksperymentalnych wykazały, że ważona korelacja faz (WPCC) posiada wysoką korelację rangową (Spearmana, Kendalla) z ocenami respondentów – MOS (tab. 7), na poziomie uznanych miar, takich jak VIF czy SSIM. Ponadto, zależność pomiędzy wartościami miary jest bliska liniowej, gdzie dużo bardziej elastyczna funkcja sigmoidalna używana powszechnie do opisu zależności miara-MOS opisywała przeciętnie zaledwie od 1% (TID) do 7% (LIVE) wariancji więcej (wg współczynnika determinacji R^2) – patrz tab. 3-4. Kolejnym udoskonaleniem było wprowadzenie heurystycznej adaptacji do warunków obserwacji, co skutkuje uzyskaniem współczynnika korelacji Spearmana względem baz na poziomie metod wielorozdzielczych, które są dużo bardziej złożone zarówno koncepcyjnie, jak i obliczeniowo.

We wnioskach wskazano użycie jako samodzielnej miary jakości liniowego współczynnika korelacji Pearsona ważonego amplitudą obrazu źródłowego, nie dyskwalifikując jednocześnie miary cyklicznej. Wskazano przyszłe zastosowania w bardziej złożonych miarach, biorących pod uwagę nie tylko strukturę obrazu oraz dalsze kierunki rozwoju miary – uwzględniające kolor czy filtrację przestrzenną.

Address

Przemysław SKUROWSKI: Silesian University of Technology, Institute of Informatics, ul. Akademicka 16, 44-100 Gliwice, Poland, przemyslaw.skurowski@polsl.pl.