Improved image quality measure through particle swarm optimization

Emil Dumic, Sonja Grgic, Mislav Grgic
Department of Wireless Communications
University of Zagreb, Faculty of Electrical Engineering and Computing
Zagreb, Croatia
emil.dumic@fer.hr

***Abstract*— In this paper we present new approach to the objective image quality evaluation based on discrete wavelet transform (DWT) and particle swarm optimization (PSO). DWT was applied on image difference (difference between original and degraded image) that is decomposed into approximation and detail subbands. DWT coefficients were computed using Coiflet wavelet filter banks. The coefficients were used to compute new image quality measure (IQM) that is defined as perceptual weighted difference between coefficients of original and degraded image. Weighting factors for wavelet subbands have been experimentally determined using PSO optimization algorithm to achieve the best possible correlation with results of subjective (perceptual) image quality evaluation. Experimental results demonstrate that the proposed technique has high correlation with results of subjective test and low computational time important for real-time applications.**

Keywords - Discrete Wavelet Transform, MSE, PSNR, VIF, VSNR, SSIM, MSSIM, Image Quality Measure, DMOS

I. INTRODUCTION

Discrete wavelet transform (DWT) can be used in various image processing applications, such as image compression and coding [1]. In this paper we examine how DWT can be used in the image quality evaluation, which has become crucial for the most image processing applications. Quality of image can be evaluated using different measures. The best way to do this is by making visual experiment, under controlled conditions, in which human observers grade which image provides better quality. Such experiments are time consuming and costly. Much easier approach is to use some objective measure that evaluates the numerical error between the original image and tested one. In real world, there is no perfect way for objective assessment of image quality [2]. However, there is no current standard and objective definition of image quality.

This paper is organized as follows. Section II describes used subjective databases that were used for creating and testing new image quality measure. Section III describes existing objective measures. In section IV our image quality measure is described in more detail. Section V explains performance measures used in comparing objective measures. Section VI compares different objective image quality measures with results of subjective assessment and finally section VII gives the conclusion.

II. SUBJECTIVE IMAGE QUALITY MEASURE

A. LIVE database

To be able to compare several later described objective methods, we used subjective quality results from [3]. Subjective quality evaluation was based on ITU-R recommendation BT.500-11 [4]. Details of the subjective testing can be found in [5]. Briefly they are as follows: 29 high-resolution 24 bits per pixel RGB color images (typically 768x512) were degraded using five degradation types: JP2K - JPEG2000 compression; JPEG - JPEG compression; WN - white noise in the RGB components; Gblur - Gaussian blur; Fastfading - transmission errors in the JPEG2000 bit stream using a fast-fading Rayleigh channel model.

Each of these 29 images had versions with 7-9 different qualities for JPEG and JPEG2000 and 6 images with different qualities for white noise, Gaussian blur and Fastfading. About 20-29 observers had to grade image quality on a continuous scale with 5 grades ("Bad", "Poor", "Fair", "Good" and "Excellent"). In this way observers evaluated total of 982 images, out of which 203 were reference and 779 degraded images. The experiments were conducted in seven sessions, separately for each type of distortion: two sessions for JPEG2000, two for JPEG, and one each for white noise, gaussian blur, and fastfading transmission errors.

Raw scores (after outlier removal and subject rejection described in [5]) for each subject were converted in difference scores between test and reference image:

$$d_{i,j} = r_{iref(j)} - r_{i,j} \tag{1}$$

where $r_{iref(j)}$ denotes the raw quality score assigned by the i-th subject to the reference image corresponding to the j-th distorted image and $r_{i,j}$ score for the i-th subject and j-th image. Difference scores were converted to Z-scores:

$$z_{i,j} = \frac{d_{i,j} - \overline{d}_i}{\sigma_i} \tag{2}$$

where $\overline{d}_i$ is the mean of the raw score differences overall images ranked by the subject i, and σ_i is the standard deviation. Z-scores are used to make scores more equal, because each observer uses different part of grading scale. Finally, a Difference Mean Opinion Score (DMOS) value for each distorted image was computed by shifting Z-scores to the full range (1 to 100).

B. *New generated database*

LIVE image database was used in the optimization process of IQM. To be able to assess IQM, in the evaluation process different database than the one utilized in optimization process should be used. Therefore, we constructed new subjective image quality database - VCL@FER. This database consists of 23 different images with 4 different types of degradation and 6 degradation levels per image and per degradation type. We used four degradation types:

- JP2K - JPEG2000 compression;
- JPEG - JPEG compression;
- WN - white noise in the RGB components;
- Gblur - Gaussian blur.

VCL@FER image database contains 575 different images for evaluation (6*4*23 and 23 original images). 118 subjects had to grade 11307 images (about 96 images per subject), where every image was evaluated between 16 and 36 times (on average 20.04 grades per image). Every subject had to grade different types of degradation in one session, e.g. sessions were not divided on degradation types (like in LIVE database). This will influence on objective measure performances. Similar algorithm like in LIVE database, (2), was used to calculate Mean Opinion score (MOS) except we didn't calculate difference scores between test and reference image (1). Z-scores were also shifted to the full range 0-100, (3):

$$mos_i = \frac{100}{\max(z_i) - \min(z_i)} \cdot (z_i - \min(z_i)) \tag{3}$$

After that, all grades that fell outside of the 2.5σ calculated for the same image were removed from the further calculation. After calculating new σ all grades were in new 2.5 σ. This way we removed 167 grades (1.5%). At the end we calculate mean MOS for every image, type of degradation and level of degradation.

III. EXISTING OBJECTIVE MEASURES

In this paper we compared new IQM measure with several commonly used objective quality measures, which were applied to a luminance channel only, because they give better correlation results with subjective testing. We tested following measures:

- MSE (Mean Squared Error);
- PSNR (Peak Signal-to-Noise Ratio);
- SSIM (Structural Similarity);
- MSSIM (Multiscale SSIM);
- VIF (Visual Information Fidelity);
- VSNR (Visual Signal-to-Noise Ratio);
- IQM (Image Quality Measure - our proposed measure).

MSE represents the power of noise, or difference between original and tested image:

$$MSE = \frac{\sum_i \sum_j (a_{i,j} - b_{i,j})^2}{x \cdot y} \tag{4}$$

In (4) $a_{i,j}$ and $b_{i,j}$ are corresponding pixels from original and tested image, x and y describe height and width of an image.

PSNR is the ratio between the maximum possible power of a signal and the power of noise expressed in terms of the logarithmic decibel:

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \tag{5}$$

In (5) 255 is maximum possible amplitude for 8-bit image.

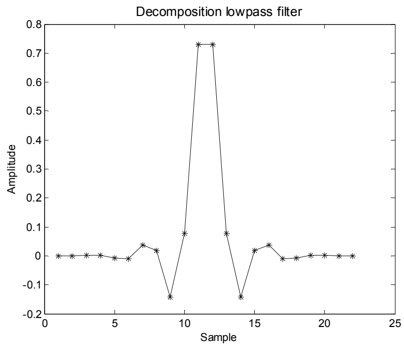
SSIM a novel method for measuring the similarity between two images [6]. It is computed from three image measurement comparisons: luminance, contrast and structure. It can give results between 0 and 1, where 1 means excellent quality and 0 means poor quality. Similar like SSIM, MSSIM (Multiscale SSIM) method is a convenient way to incorporate image details at different resolutions [7]. This is a novel image synthesis-based approach which helps calibrating the parameters (like viewing distance). Visual Information Fidelity Criterion (VIF) [8] quantifies the Shannon information that is shared between the reference and the distorted images relative to the information contained in the reference image itself. It uses Natural Scene Statistics (NSS) modeling in concern with an image degradation model and an HVS model. Results of this measure can be between 0 and 1, where 1 means perfect quality and near 0 means poor quality.

Visual Signal-to-Noise Ratio (VSNR) [9] operates in two stages. Firstly, threshold for distortions of a degraded image is determined, to decide if it is below or above human sensitivity of error detection. If the distortions are above threshold, Euclidean distances in distortion-contrast space of multiscale wavelet decomposition are determined. Finally, VSNR is calculated from a linear sum of these distances. Higher VSNR means that tested image is less degraded.

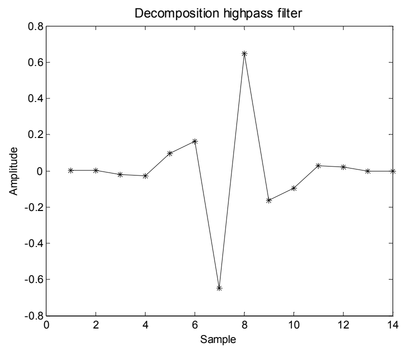
IV. IMAGE QUALITY MEASURE

Some of the existing objective measures described in previous section did not take into account HVS in the sense that eye will see and grade image quality according to the type of an error, as well as location of an error in subband space. Because of that, our method calculates image quality using wavelet decomposition and grades quality depending on the wavelet subband in which error occurs. Experiments on image databases have shown that different types of image degradation produce different error distributions in wavelet subbands. For example, for JPEG and JPEG2000 compressed images errors will be placed in the higher wavelet subbands (HH subband, level 2 and higher) while images with Gblur and Fastfading degradations will also have errors in lower subbands. White noise has equally distributed errors in all subbands.

In our research error image of luminance component (difference between original and degraded image) is firstly transformed using DWT. We used Coiflet wavelet filter because it gives best results on both image databases. Coif22_14 [10] (22 coefficients in decomposition lowpass and 14 in decomposition highpass filter), has even lengths and linear phase. Coefficients of this wavelet filter are presented in Fig. 1, while 2D DWT is shown in Fig. 2.



(a)



(b)

Figure 1. Wavelet coefficients: (a) Coif22_14 decomposition lowpass filter, (b) Coif22_14 decomposition highpass filter

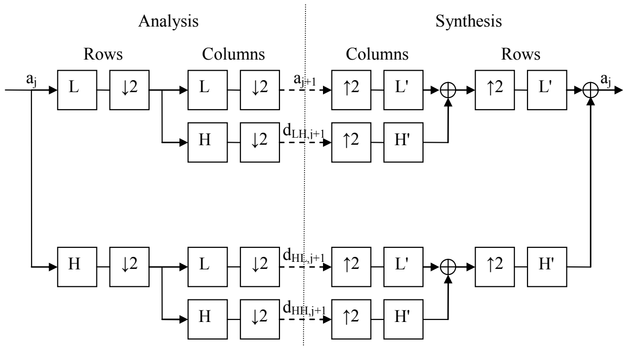


Figure 2. Wavelet decomposition and reconstruction: L - lowpass analysis filter (scaling function); H - highpass analysis filter (wavelet function); L' and H' are lowpass and highpass reconstruction filters; a is approximation coefficient and d is detail coefficient; ↓2 and ↑2 denotes downsampling and upsampling by factor 2

After decomposing difference image into 2 level decomposition, error distance in each wavelet subband can be computed using the following equation:

$$E = \sum_{i} \sum_{j} | e_{i,j} | \tag{6}$$

In (6) $e_{i,j}$ are coefficients from difference image, in the same subband.

Weighting factors for level 2 decomposition have been experimentally determined using PSO optimization algorithm [11] and subjective scores from LIVE database. Results of PSO optimization algorithm are weighting factors determined to give the best possible correlation results between subjective scores and IQM. They are presented in Table 1, according to indexing of DWT bands, Fig. 3.

TABLE I. WEIGHTING FACTORS $W_{\lambda,\theta}$ FOR 2 LEVEL COIF22_14 DWT

	Level (λ)	
Orientation (θ)	1	2
1	-	1
2	-0.188	-1.4
3	-3	8.85
4	-0.188	-1.4

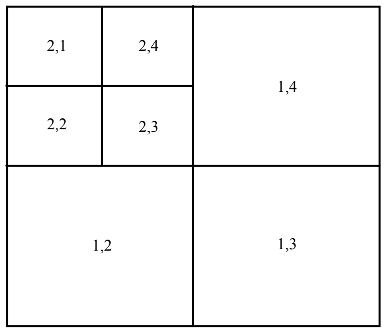


Figure 3. Indexing of DWT bands. Each band is identified by a level and orientation (λ, θ). This example shows a two level transform.

Final measure IQM is then calculated as:

$$IQM = \frac{1}{\dim_1 \cdot \dim_2} \cdot \sum_{\lambda=1}^{2} \sum_{\theta=1}^{4} w_{i,j} \cdot E_{i,j} \tag{7}$$

In (7) $w_{i,j}$ are weighting factors in related subband (normalized to the approximation weighted factor, $w_{2,1}$), E is error distance calculated according to (6) and $\dim_{1,2}$ are dimensions of the whole image. They're used (normalization) so that final IQM measure won't have high values and they don't influence on the correlation results very much. From Table 1 it can be seen that all subbands have to be included in IQM measure, some have to be calculated using negative weighting factor (experimentally they give better results). Also, higher decomposition levels made our measure overfitted to LIVE database so it gave somewhat worse results on our subjective database (and only slightly better results on LIVE database). After that, IQM was evaluated using another database VCL@FER.

V. PERFORMANCE MEASURES

To be able to compare different image quality measures and DMOS, we used several different measures of performance:

- Pearson's product-moment correlation coefficient;
- RMSE (Root Mean Square Error);
- Spearman's rank-order correlation coefficient.

Pearson's product-moment correlation coefficient is calculated as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)\cdot s_x \cdot s_y}, \mathrm{i} = 1,\ldots,n \tag{8}$$

where in (8) x_i and y_i are sample values, (x are results for different objective measures and y are results for DMOS), $\bar{x}$ and $\bar{y}$ are sample mean, s_x and s_y are standard deviation (calculated using n − 1 in the denominator). Pearson's correlation reflects the degree of linear relationship between two variables, from −1 to 1, where 0 means that there is no relationship and ±1 means perfect fit.

RMSE (Root Mean Square Error) is calculated as:

$$RMSE = \sqrt{\frac{1}{n-k}\cdot(x-y)^2} \tag{9}$$

In (9) n is the number of tested images modified by a correction for degrees of freedom (k=5 in our case, we have 5 parameters in fitted function, (13)), x is DMOS or MOS measure depending on the database used and y fitted objective measure after nonlinear regression.

Spearman's correlation coefficient is a measure of a monotone association that is used when the distribution of the data makes Pearson's correlation coefficient undesirable or misleading. Spearman's coefficient is not a measure of the linear relationship between two variables. It assesses how well an arbitrary monotonic function can describe the relationship between two variables, without making any assumptions about the frequency distribution of the variables [12].

VI. RESULTS

SSIM, MSSIM, VIF, VSNR quality measures were calculated using software from [13]. We calculated Pearson's correlation coefficient before and after nonlinear regression. The nonlinearity chosen for regression for each of the methods tested was a 5-parameter logistic function (a logistic function with an added linear term), as it was proposed in [14]. For LIVE database coefficients of logistic function can be found in [15].

Overall results and results for degradations separately are given in Figs. 4-8. Results for MSE are not shown because they're the same for MSE and PSNR. Also, VIF performance measures were calculated using $\log_{10}$ of VIF grades in LIVE database (although without log it gives pretty similar results) and in our database VIF performance measures were calculated directly from VIF grades.

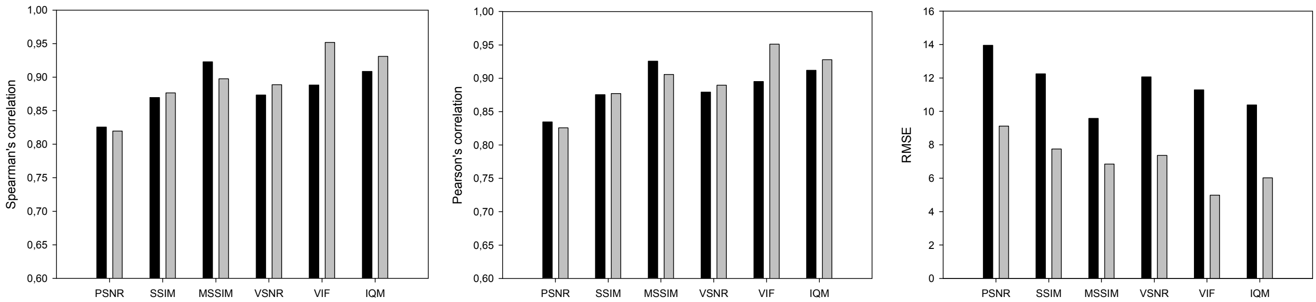


Figure 4. Comparison of objective measures using Spearman's correlation, Pearson's correlation and RMSE (after linearization) for all images in databases. Black bars denote our image database and gray bars denote LIVE database.

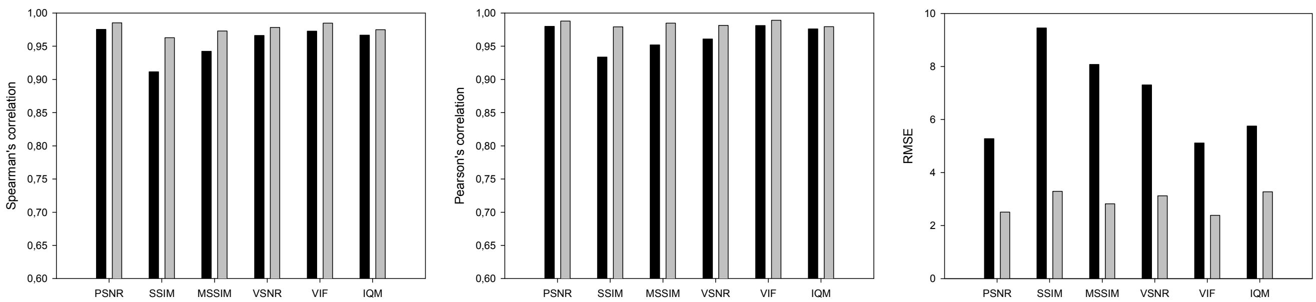


Figure 5. Comparison of objective measures using Spearman's correlation, Pearson's correlation and RMSE (after linearization) for AWGN degraded images. Black bars denote our image database and gray bars denote LIVE database.

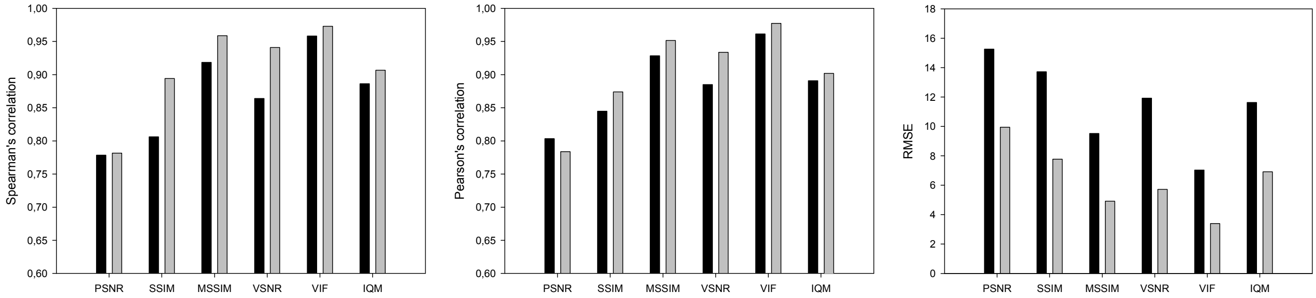


Figure 6. Comparison of objective measures using Spearman's correlation, Pearson's correlation and RMSE (after linearization) for Gaussian blur degraded images. Black bars denote our image database and gray bars denote LIVE database.

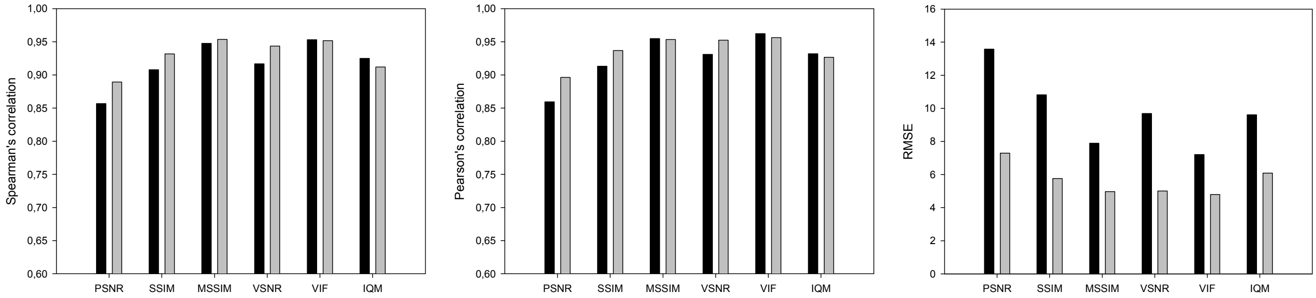


Figure 7. Comparison of objective measures using Spearman's correlation, Pearson's correlation and RMSE (after linearization) for JP2K degraded images. Black bars denote our image database and gray bars denote LIVE database.

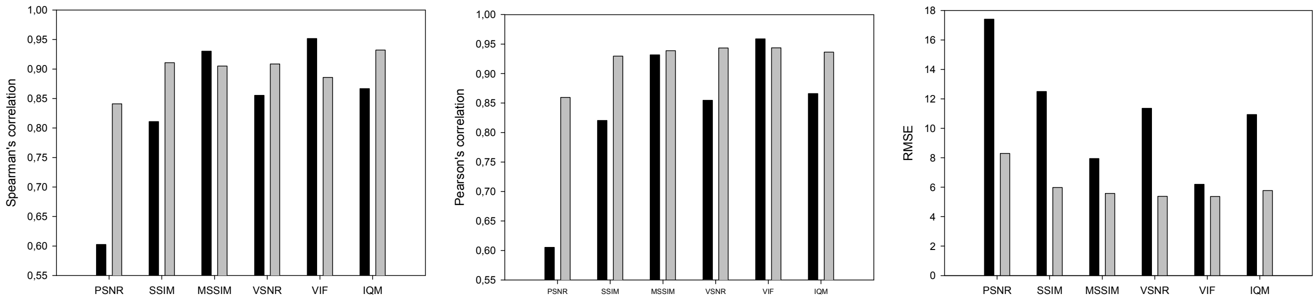


Figure 8. Comparison of objective measures using Spearman's correlation, Pearson's correlation and RMSE (after linearization) for JPEG degraded images. Black bars denote our image database and gray bars denote LIVE database.

Our measure (IQM) gives good results in both databases (about 0.9 or higher), except for JPEG degradation in VCL@FER database (0.87).

It can be generally concluded that for all degradation types performance measures are lower for VCL@FER database, because we used MOS subjective measure, while in LIVE database DMOS measure is calculated. DMOS measure is more adjusted to the full reference measures (which use both original and degraded image), but in real life it is usually only degraded image that is known so people grade images based on their perception about how good or bad image should look like. Pearson's and Spearman's correlation give also similar results, if grades are linearized, especially for higher correlations, so only Spearman's correlation could be used to compare objective measures. In LIVE database, VIF measure gives best results for all degradations together, as well as for degradations separately. However, in our database MSSIM gives best results for all degradations together, not VIF measure. This is probably because we used all degradation types in one subjective measurement, unlike LIVE database in which subjective measurements were done separately, for each type of degradation. It means that comparison of objective image quality measures depends very much on image database used for testing and grading procedure. Good measure is one that works well in different image databases with different types and levels of degradation.

Average time required to calculate each of the objective measures described before are given in Table II. SSIM, MSSIM, VIF and VSNR measures were calculated over entire LIVE database using [13] and then averaged. PSNR was calculated directly from (5). Same computer configuration was used for calculating all objective measures: Intel Core 2 Quad CPU Q6600 (2.40 GHz), 4 GB RAM, Windows Vista 64. Probably it is possible to speed up algorithms, using MEX-compiler from C/ C++ or Fortran source code instead of Matlab .m files.

TABLE II. AVERIGE TIME REQUIRED TO CALCULATE OBJECTIVE MEASURES

Measure	Time (seconds)
PSNR	0.0050
SSIM	0.1478
MSSIM	0.3104
VIF	3.6940
VSNR	0.6297
IQM	0.1737

VII. CONCLUSION

In this paper we proposed new image quality measure based on DWT in different wavelet subbands. We examined how different objective measures correlate with subjective DMOS and MOS measures in two different subjective databases, also presenting new objective measure. Our image quality measures take into account properties of human visual system and provide better correlation with DMOS than some other quality measures. It works good in both image databases used in testing. Proposed IQM could be considered as a good starting point for evaluation and fair comparison of different types of image degradation, especially in applications where image quality evaluation should be performed in real-time.

In the future IQM measure could be computed adaptively, depending on the type of the degradation. Based on IQM and properties of wavelet domain, the development of new no reference image quality measure could be considered as well.

ACKNOWLEDGMENT

The work described in this paper was conducted under the research projects: "Picture Quality Management in Digital Video Broadcasting" (036-0361630-1635), and "Intelligent Image Features Extraction in Knowledge Discovery Systems" (036-0982560-1643), supported by the Ministry of Science, Education and Sports of the Republic of Croatia.

REFERENCES

[1] S. Grgic, M. Grgic, B. Zovko-Cihlar, "Performance analysis of image compression using wavelets", IEEE Transactions on Industrial Electronics, Vol. 48, Issue 3, 682-695 (June 2001)

[2] Video Quality Experts Group, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Multimedia Quality", http://www.vqeg.org/ (September 2008)

[3] H.R. Sheikh, Z.Wang, L. Cormack, A.C. Bovik, "LIVE Image Quality Assessment Database Release 2", http://live.ece.utexas.edu/research/quality

[4] ITU-R BT.500-11 "Methodology for the subjective assessment of the quality of television pictures", International Telecommunication Union/ITU Radiocommunication Sector (January 2002)

[5] H.R. Sheikh, "Image Quality Assessment Using Natural Scene Statistics," Ph.D. dissertation, University of Texas at Austin (May 2004)

[6] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Trans. on Image Proc., Vol. 13, No. 4, 600-612 (April 2004)

[7] Z. Wang, E.P. Simoncelli, A.C. Bovik, "Multi-scale structural similarity for image quality assessment", in 37th Proc. IEEE Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, CA (Nov. 2003)

[8] H.R. Sheikh and A.C. Bovik, "Image information and visual quality", IEEE Trans. Image Processing, Dec. 2003

[9] D.M. Chandler and S.S. Hemami, "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images", IEEE Transactions on Image Processing, Vol. 16, Issue 9, 2284-2298 (September 2007)

[10] D. Wei, H.T. Pai, A.C. Bovik, "Antisymmetric Biorthogonal Coiflets for Image Coding", Proceedings of the IEEE International Conference on Image Processing (ICIP), Vol. 2, 282-286 (October 1998)

[11] J. Kennedy and R. C. Eberhart, "Particle swarm optimization", Proc. IEEE Int'l. Conf. on Neural Networks, IV, 1942–1948 (1995)

[12] J. Hauke and T. Kossowski, "Comparison of values of Pearson’s and Spearman’s correlation coefficient on the same sets of data", Proceedings of the MAT TRIAD 2007 Conference, Bedlewo, Poland (March 2007)

[13] Visual Quality Assessment Package Version 1.1, Available at: http://foulard.ece.cornell.edu/gaubatz/metrix_mux/

[14] H.R. Sheikh, M.F. Sabir, A.C. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms", IEEE Transactions on Image Processing, Vol. 15, Issue 11, 3440-3451 (November 2006)

[15] Dumic, E., Grgic, S., Grgic, M., "New Image-quality Measure based on Wavelets", Journal of Electronic Imaging, Vol. 19, No. 1, Article ID 011018, 19 pages (January 2010)