Krzysztof OKARMA

West Pomeranian University of Technology, Szczecin

Image and video quality assessment with the use of various verification databases

Streszczenie. *W artykule przedstawiono wyniki weryfikacji uzyskane dla najbardziej znanych nowoczesnych wskaźników oceny jakości obrazów oraz sekwencji wideo (z wykorzystaniem analizy poklatkowej), jak również dla zaproponowanego wskaźnika złożonego, z wykorzystaniem aktualnie dostępnych baz testowych. Uzyskane wyniki wskaźników zostały porównane z ocenami subiektywnymi wyrażonymi wartościami MOS i DMOS; do weryfikacji ich zgodności użyto współczynników korelacji liniowej świadczących o przydatności poszczególnych wskaźników dla różnych rodzajów zniekształceń, jak również o zaletach wskaźnika złożonego. (**Ocena jakości obrazu i sekwencji wideo z wykorzystaniem różnych baz testowych**).*

Abstract. *In the paper the verification results obtained for the most of state-of-the-art image and video (using frame-by-frame approach) metrics, together with proposed combined one, using currently available databases are presented. Obtained values of the metrics have been compared to MOS and DMOS values and the linear correlation coefficients have been used for the verification of the usefulness of metrics for each type of distortions, demonstrating the advantages of the combined metric.*

Słowa kluczowe: ocena jakości obrazów, złożony wskaźnik jakości.
Keywords: image quality assessment, combined quality metric.

Introduction

Automatic image and video quality assessment methods may be divided into three main categories depending on the knowledge of the original (reference) image which is not affected by any distortions. The first group is called “blind” or no-reference quality assessment [1], since such methods do not require the use of the original image. Nevertheless, they are usually specialised and sensitive only to one or two chosen types of distortions. Another group of metrics know as reduced-reference methods require only the partial knowledge of the original image [2].

The most universal approach is the application of the full-reference metrics, which compare the distorted image with the original one. Such metrics are usually sensitive to many types of various distortions, such as noise, compression artifacts, transmission errors, blur and many more, and well correlated with subjective evaluations.

In recent years the rapid progress in this field has taken place and many new full-reference metrics have been provided starting from the Universal Image Quality Index proposed in 2002 [3] further extended into Structural Similarity [4]. All recently proposed metrics, e.g. based on the SVD decomposition [5] are much better than traditionally used Mean Squared Error (MSE) and similar metrics, such as e.g. PSNR, but their verification is usually performed using arbitrarily chosen database of images. A reliable comparison of the properties of some modern image and video quality metrics requires the use of several verification databases which contain the images or video files with many distortion types together with subjective quality scores.

Currently, several image quality assessment databases are available, but some of them are limited to only two or three types of distortions applied for small number of images. Another disadvantage of them is small number of human observers assessing the quality of images, what leads to relatively low reliability of delivered Mean Opinion Scores (MOS) or Differential MOS (DMOS) values. The two largest image quality assessment databases are LIVE Database delivered by Laboratory for Image and Video Engineering (LIVE) from Texas University at Austin [6] and Tampere Image Database (TID2008) containing 1700 images with 17 types of distortions [7].

Structural approach to image and video quality

Considering the poor correlation of the classical pixel-based image quality assessment methods, such as MSE or PSNR, very sensitive e.g. to image shift by one row or column, a new approach has been proposed, which is based on the comparison of the image structure within a local mask. Applying a sliding window, the overall quality index can be obtained by averaging the local values, being in fact the quality map of the assessed image.

The first metric of this type is the Universal Image Quality Index based on the three common types of distortions: the loss of contrast, luminance distortions and the loss of correlation. Due to the possible division by zero causing the instability of results, changing also the size and type of the sliding window (from 8×8 pixels rectangular to 11×11 pixels Gaussian), the modified version, known as Structural Similarity (SSIM) has been proposed [4], where the local image quality index is calculated according to:

$$(1)\ SSIM=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}\cdot\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}\cdot\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}$$

where μ, σ^2 and σ_{xy} denote the mean values, variances and the covariance (for N×N pixels window) assuming that *x* and *y* denote the reference and distorted image respectively.

The default values of the stability constants, which do not influence significantly the overall results, are equal to $C_1=(0.01\times L)^2$, $C_2=(0.03\times L)^2$ and $C_3=C_2/2$, where *L* denotes the dynamic range of pixels’ values (typically *L=255* for the 8-bit greyscale images). The three components of the SSIM index are sensitive to the luminance (*l*), structure (*s*) and contrast (*c*) respectively. The window shape and size can also be changed, leading to better quality prediction [8,9].

Such defined metric has become very popular, mainly due to its simple form and high correlation with subjective evaluations. Further research caused its extension into Multi-Scale SSIM [9], operating over a dyadic pyramid with additional weighting of the three components of the SSIM index for each scale obtained after downsampling by 2 and low-pass filtration. The local value of the metric is calculated as:

$$(2)\qquad MS-SSIM=\left(l_M\right)^{\alpha}\cdot\prod_{j=1}^{M}\left[\left(c_j\right)^{\beta_j}\cdot\left(s_j\right)^{\gamma_j}\right]$$

with the default values of the weighting exponents according to the paper [10], typically using M=5 scales.

Recently, some other modifications of the SSIM index have been proposed e.g. the R-SSIM for range image quality assessment, complex wavelet SSIM [11], gradient based SSIM or three-component SSIM [12]. Nevertheless, these metrics are not the topic of this paper.

Some other modern image quality assessment methods

One of the most interesting directions of research in the area of image quality assessment is the usage of the Singular Value Decomposition. The first idea has been presented by Eskicioglu as the M-SVD metric [13]. The singular values calculated for 8×8 pixels blocks of original and distorted images have been used for the calculation of the quality factors of each block which are proportional to the square roots of the aggregated squared differences of 8 singular values. The overall image quality score has been defined as the mean difference between the quality factors computed for each block and the middle element of the sorted vector of factors.

Since this metric has not gained popularity, another idea has been presented recently [5], known the R-SVD index defined as:

$$R-SVD=\sqrt{\frac{\sum_{i=1}^{m}\left(d_i-1\right)^2}{\sum_{i=1}^{m}\left(d_i+1\right)^2}}, \tag{3}$$

where d_i denotes the singular values of the reference matrix R'=U'·Λ·V^T, with identity matrix Λ at the diagonal. It is assumed that the U and V matrices are computed by the Singular Value Decomposition of the matrix corresponding to the original image (A=U·S·V^T) and U' and V' in the same way from the distorted image (A'= U'·S·V'^{T}).

An interesting approach based on the information theory is the Visual Information Fidelity (VIF) defined in the wavelet domain [14]. Additionally, this metric has also its pixel domain version, but its performance is slightly worse. The definition of the metric is:

$$VIF=\frac{\sum_{j=0}^{S}\sum_{i=1}^{M_j}I(c_{i,j};f_{i,j})}{\sum_{j=0}^{S}\sum_{i=1}^{M_j}I(c_{i,j};e_{i,j})}, \tag{4}$$

where *I(x;y)* denotes the mutual information between *x* and *y*. The denominator and numerator denote the information extracted by human vision from the reference and distorted images respectively, where M_j denotes the number of blocks at j-th sub-band (or scale) and S is the number of sub-bands (scales).

Since each of the metrics presented above is sensitive to many types of distortions, but in different way, their correlation with subjective evaluations is different for different types of distortions and depends on the dataset. Additionally, a high correlation is usually achieved after additional nonlinear regression, typically using the logistic function, as suggested by the Video Quality Experts Group [15]. Nevertheless, the tuning parameters of this function, obtained as a result of an additional optimisation, strongly depend on the dataset and the types of distortions, so in practical applications such high correlation is hard to obtain.

Combined Quality Metric

In order to avoid this problem the nonlinear combination of some of the metrics can be used, so the correlation with the subjective evaluations should be much more linear. In one of the recent papers [16] the definition of such Combined Quality Metric has been proposed as:

$$CQM=\left(MS-SSIM\right)^a\cdot\left(VIF\right)^b\cdot\left(R-SVD\right)^c \tag{5}$$

with near-optimal values of the exponents: *a=7*, *b=0.3* and *c=-0.15*, obtained after the optimisation of the Pearson's linear correlation coefficient (CC) for the TID2008 database [7], being currently the largest available image database containing 1700 images with 17 types of distortions assessed by 838 observed from three countries (totally 256428 comparisons of visual quality of distorted images have been performed). The obtained CC value equal to 0.86 is superior to each if the metrics applied separately.

Further modifications of the proposed CQM index have been proposed for the video quality assessment [17] and colour image quality assessment using the CIELAB colour model [18].

Image and video quality databases

The only possible method of verification of each newly developed objective image or video quality assessment method it the reliable comparison with the results of subjective evaluation. It can be performed using a database of images containing possibly large number of preferably colour images, corrupted by many types of distortions assessed by possibly large number of human observers. The subjective results are included in such database in the form of Mean Opinion Score (MOS) or Differential MOS (DMOS) values.

Apart from the TID2008 database, discussed above, the most widely used is already mentioned LIVE Database [6] containing 982 images with 779 distorted and 29 reference ones (some of them have been used multiple times for the subjective tests involving 29 observers) with 5 types of distortions: JPEG and JPEG2000 compression, white noise, Gaussian blur and transmission errors using simulated fast fading Rayleigh channel for JPEG2000 coded images.

Another relatively large database, containing 866 colour images, is the Categorical Subjective Image Quality (CSIQ) database developed in 2009 at Oklahoma State University [19]. It is built from 30 reference images corrupted by 6 types of distortions: JPEG and JPEG2000 compression, Gaussian noise, pink noise, Gaussian blur and global contrast change. All the images have been assessed by 35 observers using linear displacement strategy.

These three databases should be treated as the basic ones for the verification of objective image quality metrics. Nevertheless, there are three smaller databases, which can also be considered useful. The first one is Toyama (MICT) database published in 2000, containing 198 images obtained by lossy compression of 14 reference images using JPEG and JPEG2000 codecs and assessed by 16 college students. The second one is A57 database [20] built from 3 greyscale reference images, containing 54 test images subjected to 6 types of distortions (similar as in other databases) evaluated by 7 experts using continuous rating system. Another such database is the IRCCyN/IVC [21] released by the University of Nantes, containing 10 colour reference images and 160 distorted ones, subjected to JPEG, JPEG2000, blurring and Local Adaptive Resolution (LAR) based coding. The subjective assessment has been performed by 15 observers using the Double Stimulus Impairment Scale method. The last database is the Wireless Imaging Quality (WIQ) database [22] consisting of 7 undistorted reference images and 80 distorted test images assessed during two tests by 30 participants each using Double Stimulus Continuous Quality Scale.

The verification of the video quality assessment metrics is more troublesome, since only two databases are currently available, both delivered by LIVE. One of them, called LIVE Wireless Video Database [23], has only four types of distortions specific for wireless transmission of compressed video data and the second one (LIVE Video [24]) is more general (contains 4 distortion types: IP and wireless transmission distortions, MPEG-2 and H.264 compression). They both contain totally 160 and 150 distorted files (assessed by 30 and 29 observers) respectively, obtained from 10 original sequences in both cases.

Verification of the quality metrics

The verification of the metrics has been performed using seven image and two video quality assessment databases described above. All the images have been converted to greyscale before processing and the correlation coefficients with subjective scores have been calculated for 4 metrics discussed above and for the Combined Quality Metric. The same procedure has also been applied for the video files using the frame-by-frame approach. The calculations of the CC values have been conducted without any nonlinear mappings, removing also the results obtained for the original images. For a better comparison the CQM index has been calculated for the video files in the same way as for the images, regardless of the presence of its modified version [17], designed for the video quality assessment purposes. The obtained results are presented in Table 1 with indicated values better than those obtained for the proposed combined one for each database.

Table 1. Linear correlation coefficients of the results obtained for various metrics with the subjective evaluation (MOS/DMOS) for different databases

Metric / database	SSIM	MS-SSIM	VIF	R-SVD	CQM
Toyama	0.7175	0.7433	**0.9019**	0.8037	**0.8937**
A57	0.7531	**0.8289**	0.6141	0.3652	**0.8251**
CSIQ	0.7654	0.7708	**0.9219**	0.7411	**0.9189**
IRCCyN/IVC	0.7047	0.7679	0.8800	0.7885	**0.8943**
WIQ	0.5534	0.6089	0.7301	**0.8274**	**0.7800**
LIVE Image	**0.7364**	0.4762	**0.7327**	0.4999	**0.7214**
TID2008	0.6016	0.7843	0.7777	0.4782	**0.8600**
LIVE Video	0.5044	0.6713	0.5547	0.4486	**0.6914**
LIVE Wireless	0.8578	0.8532	0.9447	0.8287	**0.9694**

Conclusions

The results presented in the paper demonstrate the usefulness of the proposed approach based on the Combined Quality Metric. For almost each database this metric is one of the two the most linearly correlated with subjective evaluations regardless of the distortion types present in the database, what proves the universality of such approach. For two of three largest image databases (TID2008 and IVC) and both video databases the linear correlation of the combined metric is the highest one.

LITERATURE

[1] Li X., Blind Image Quality Assessment, *Proc. IEEE Conf. Image Proc.,* Rochester, USA (2002), 449-452

[2] Okarma K., Lech P., A Statistical Reduced-Reference Approach to Digital Image Quality Assessment, *Lecture Notes in Computer Science,* 5337 (2009), 43-54

[3] Wang Z., Bovik A.C., A Universal Image Quality Index, *IEEE Signal Proc. Letters*, 9 (2002), n.3, 81-84

[4] Wang Z., Bovik A.C., Sheikh H., Simoncelli P., Image Quality Assessment: From Error Measurement to Structural Similarity, *IEEE Trans. Image Proc.*, 13 (2004), n.4, 600-612

[5] Mansouri A., Mahmoudi-Aznaveh A., Torkamani-Azar F., Jahanshahi J.A., Image Quality Assessment Using the Singular Value Decomposition Theorem, *Optical Review*, 16 (2009), n. 2, 49-53

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

[7] 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*, 10 (2009), 30-45

[8] Okarma K., Influence of the 2-D sliding windows on the correlation of the digital image quality assessment results using the Structural Similarity approach with the subjective evaluation, *Przegląd Elektrotechniczny,* 86 (2010) n.7, 109-111

[9] Okarma K., Two-dimensional windowing in the Structural Similarity index for the colour image quality assessment, *Lecture Notes in Computer Science,* 5702 (2009), 501-508

[10] Wang Z., Simoncelli E., Bovik A., Multi-Scale Structural Similarity for Image Quality Assessment. Proc. 37th IEEE Asilomar Conf. on Signals, Systems and Computers (2003)

[11] Malpica W., Bovik A., Range Image Quality Assessment by Structural Similarity. Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Taipei (2009), 1149-1152

[12] Li C., Bovik A., Three-component Weighted Structural Similarity Index. *Proc. SPIE Image Quality and System Performance,* 7242 (2009)

[13] Shnayderman A., Gusev A., Eskicioglu A., An SVD-Based Gray-Scale Image Quality Measure for Local and Global Assessment. *IEEE Trans. Image Proc.,* 15 (2006), n.2, 422-429

[14] Sheikh H., Bovik A., Image Information and Visual Quality. *IEEE Trans. Image Proc.,* 15 (2006), n.2, 430-444

[15] Video Quality Experts Group, Final Report on the Validation of Objective Models of Video Quality Assessment, (2003), Available at: *http://www.vqeg.org*

[16] Okarma K., Combined Full-Reference Image Quality Metric Linearly Correlated with Subjective Assessment, *Lecture Notes in Artificial Intelligence,* 6113 (2010), 539-546

[17] Okarma K., Video Quality Assessment Using the Combined Full-Reference Approach, *Image Processing and Communications Challenges 2 - Advances in Intelligent and Soft Computing,* 84 (2010), 51-58

[18] Okarma K., Colour Image Quality Assessment Using the Combined Full-Reference Approach, *Computer Recognition Systems 4 - Advances in Intelligent and Soft Computing,* 95 (2011), 287-296

[19] Larson E., Chandler D., Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy, *Journal of Electronic Imaging*, 19 (2010), n.1, 011006

[20] Chandler D., Hemami S., VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images, *IEEE Trans. Image Proc.*, 16 (2007), n.9, 2284-2298

[21] Le Callet P., Autrusseau F., Subjective Quality Assessment IRCCyN/IVC Database, (2005), Available at: http://www.irccyn.ec-nantes.fr/ivcdb/

[22] Engelke U., Kusuma T., Zepernick H.-J., Caldera M., Reduced-Reference Metric Design for Objective Perceptual Quality Assessment in Wireless Imaging, *Signal Processing: Image Communication*, 24 (2009), n.7, 525-547

[23] Moorthy A., Seshadrinathan K., Soundararajan R., Bovik A., Wireless Video Quality Assessment: A Study of Subjective Scores and Objective Algorithms. *IEEE Trans. Circuits and Systems for Video Technology*, 20 (2010), n.4, 513-516

[24] Seshadrinathan K., Soundararajan R., Bovik A., Cormack L., Study of Subjective and Objective Quality Assessment of Video, *IEEE Trans. Image Proc.*, 19 (2010), n.6, 1427-1441

Author: *PhD eng. Krzysztof Okarma, Department of Signal Processing and Multimedia Engineering, Faculty of Electrical Engineering, West Pomeranian University of Technology, 10, 26. Kwietnia Str., 71-126 Szczecin, E-mail: okarma@zut.edu.pl*