On the Usefulness of Combined Metrics for 3D Image Quality Assessment

Krzysztof Okarma

Department of Signal Processing and Multimedia Engineering,
Faculty of Electrical Engineering,
West Pomeranian University of Technology, Szczecin
26. Kwietnia 10, 71-126 Szczecin, Poland
okarma@zut.edu.pl

Abstract. Due to the growing popularity of 3D imaging technology which can be observed in recent years, one of the most relevant challenges for image quality assessment methods has become their extension towards reliable evaluation of stereoscopic images. Since known 2D image quality metrics are not necessarily well correlated with subjective quality scores of 3D images and the exact mechanism of the 3D quality perception is still unknown, there is a need of developing some new metrics better correlated with subjective perception of various distortions in 3D images. Since a promising direction of such research is related with the application of the combined metrics, the possibilities of their optimization for the 3D images are discussed in this paper together with experimental results obtained for the recently developed LIVE 3D Image Quality Database.

1 Motivation

Analysis of stereoscopic 3D images is currently a very active area of research which has developed a number of algorithms and working systems. A natural consequence of the development of stereovision algorithms is extremely important need to develop methods to assess the quality of stereoscopic 3D images e.g. due to some limitations of communication is stereovision based systems. However, this task is not a straightforward extension of the methods used to assess the quality of 2D images, because apart from the quality assessment of each image of the pair, it is necessary to assess the quality in terms of the perception of 3D scenes. As in the case of a reliable quality assessment of color images [1], as well as video sequences [2], it is still an open field of research with a growing number of publications in leading journals and scientific conferences recognized worldwide by researchers interested in the field of image processing and analysis.

Research activities on the quality of stereoscopic images are conducted in several centers in the world although they are mostly related to image quality for cinema and television purposes where cinema production a complex process of obtaining and processing stereoscopic images is used. In addition, the captured image is processed in computer post-production in order to highlight the

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_17

3D effects in a certain part of the image. In the case of television production, even when its real-time implementation, it can be treated as a fully controlled process without any unpredictable components with limited possibility of distortions. Nevertheless, in many applications e.g. related to telerobotics, some image distortions may appear and their character may be completely different than it takes place in the movies or 3DTV.

Until recent years, most of the known image quality metrics was limited to 2D images in grayscale, which was the effect of availability of test databases necessary to verify the proposed objective metrics in compliance with subjective scores. In the past few years there has been a rapid development of quality assessment methods for images and video sequences partly due to the development of new databases of test images and video sequences, containing also the subjective scores. Among these databases one can also find some images containing several types of distortions occurring at the same time such as LIVE Multiply Distorted Image Quality Database [3], and video sequence databases developed by the Laboratory for Image and Video Engineering (LIVE), EPFL/PoliMi, IRCCyN/IVC or MMSP.

Nevertheless, the most important from the point of view of the development of methods for assessing the quality of stereoscopic images are the databases of 3D images and 3D video sequences. The first of this kind of database, published in 2008, is IRCCyN/IVC 3D Image Quality Database [4] containing blurry images as well as JPEG and JPEG2000 compressed ones. However, in subsequent years, some other databases have appeared containing a larger number of distortions, such as MMSP 3D Image Quality Assessment Database and MMSP 3D Video Quality Assessment Database for different interocular distances between the cameras, IRCCyN/IVC DIBR Image Quality Database for different depth-image based rendering algorithms. In addition, over the past two years three new datasets have been made available: IRCCyN/IVC NAMA3DS - COSPAD1 3D Video Quality Database, IVP Anaglyph Image Database and the most relevant LIVE 3D Image Quality Database [5,6] used in this paper. In addition, there are also some other databases available but they do not contain any results of subjective evaluations, such as e.g. RMIT 3D Video Library, which are useless for the development of 3D image quality metrics.

2 Recent Attempts to 3D Image Quality Assessment

Many new ideas related to image and video quality assessment, also for 3D images, are presented each year at the International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM) usually held in Scottsdale, Arizona. Analyzing the published proceedings from recent years, growing interest of 3D image quality assessment can be easily noticed. An example of such metrics, proposed by You et al. [7], is based on the integration of disparity information and 2D metrics by calculation of two quality maps which are further combined locally. The mean is taken as an intermediate quality of the distorted image and finally is combined with the quality of the disparity image.

Some other state-of-the-art 3D image quality assessment metrics have also been proposed by various researchers during last five years. An early example is the metric proposed by Gorley and Holliman [8] based on asymmetric compression of stereoscopic signals and asymmetric distortions. Benoit et al. also proposed the fusion of 2D quality metrics and the depth information [4] using the combination of well-known Structural Similarity (SSIM) [9] and C4 metric exploiting an implementation of an elaborated model of the human visual system. One of their (still up-to-date) conclusions is that "the depth information can improve the quality metric but the relation with image 'naturalness', 'viewing experience' and 'presence' has still to be investigated in depth, depending also in the different 3D display technology used".

Considering the necessity of changing parameters of the video transmission "on-the-fly", the reduced-reference metric for 3D depth map transmission using the extracted edge information has been proposed by Hewage et al. [10]. The authors of this paper noticed that edges and contours of the depth map can represent different depth levels and hence can be used in quality evaluations. Nevertheless, since this metric is based on the PSNR, assuming the binary character of edge information, it may be considered as a good starting point for further extensions. Similarly, the metric proposed by Yang et al. [11] is also based on the PSNR which is supplied by the Stereo Sense Assessment (SSA) but the main advantage of this method is the possibility of fast calculation. An alternative method proposed by Zhu and Wang [12] uses a multi-channel vision model based on 3D wavelet decomposition which is more perceptually consistent than e.g. PSNR or MSE. The main idea is based on the wavelet-based perceptual decomposition of the video sequences into three domains (temporal, spatial and view domain) and further contrast conversion, masking and nonlinear summation. Another idea has been proposed by Shen et al. [13] for estimation of stereo image quality based on a multiple channel Human Visual System (HVS) for use in image compression. Proposed metric utilizes the contourlet transform but the results presented by the authors are quite preliminary.

Considering the demands of objective no-reference ("blind") 3D image quality assessment metrics which do not require the knowledge of an original undistorted image, the first attempts in this field have also been made. An example metric based on the assumption that the perceived distortion and disparity of any stereoscopic display is strongly dependent on local features, such as edge (non-plane) and non-edge (plane) areas, has been proposed by Akhter et al. [14]. The metric has been developed for JPEG coded stereoscopic images and is based on segmented local features of artifacts and disparity.

3 Application of Combined Metrics

Due to the presence of various approaches to the problem of 3D image quality assessment, currently there is no single universal metric which could be applied for most 3D images, even containing only some most typical distortions. Since many of the existing solutions mentioned above use quite complicated models utilizing both the image information and the reference depth, they may not be attractive for many technical systems e.g. assuming real-time video transmission. For those reasons the goal if this paper is related to the verification of the usefulness of relatively simple approach based on multi-method fusion leading to combined metrics, previously successfully applied for the 2D images.

The idea of the combined metrics is based on the nonlinear combination of three or more image quality assessment methods of different kinds. The main motivation of such an approach is the necessity of nonlinear mapping of the raw quality scores which is conducted by many researchers, typically using the logistic function. Avoiding such mapping is possible by including the nonlinearity inside the computational procedure by weighted product of various metrics leading to combined (hybrid) metrics.

The first such approach denoted as Combined Quality Metric (CQM) [15] has been proposed as the combination of three metrics: Multi-Scale Structural Similarity [16], Visual Information Fidelity [17] and R-SVD metric [18] with weights optimized using the largest available image quality assessment database known as Tampere Image Database (TID2008). Further replacement of the R-SVD metric by Feature Similarity (FSIM) [19] has led to the Combined Image Similarity Index [20] achieving high correlation with subjective scores for the most relevant datasets. Similar approach has also been presented in the paper [21] where a deeper analysis of many combinations has been conducted. Nevertheless, the proposed multi-metric fusion is based on the machine learning approach decreasing the universality of this particular solution. Another modification of the combined metric (Extended Hybrid Image Similarity - EHIS) has been analyzed in the paper [22] which is based on the modifications of the FSIM metric together with Riesz-based Feature Similarity [23] proposed earlier by the the FSIM inventors.

Since the main goal of the paper is related to the verification of usefulness of combined metrics, the experiments have been focused on the calculation of correlation coefficients with subjective evaluations for the metrics utilizing image data without the use of depth information which are available only for reference images. Such assumption makes it possible to compare the consistency of all metrics with subjective perception of image distortions as the depth information remains unchanged for the images included in LIVE 3D Image Quality Database used for the experiments. The dataset has been chosen due to rather poor definitions of stereoscopic distortions in all currently available databases and the significant differences between them.

One of the main advantages of the LIVE 3D database is its division into two independent sets (Phase I and Phase II) containing symmetrical and asymmetrical distortions respectively. Due to such construction of the database the combined metrics can be independently analyzed for those two kinds of distortions since much better correlation of existing 2D metrics with provided Differential Mean Opinion Scores (DMOS) should be expected for the 'symmetrical' set (Phase I).

4 Details of Experiments and Results

In the conducted experiments some well-known 2D image quality metrics have been used together with the combined ones such as mentioned CQM [15], CISI [20] and EHIS [22]. Since the color image quality assessment is a separate problem, not discussed in this paper, all the metrics (except FSIMc being the color version of Feature Similarity) have been applied for the images converted to grayscale according to the widely used ITU BT.601-7 recommendation:

$$Y = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B \ , \tag{1}$$

applied e.g. in MATLAB's *rgb2gray* function.

In the first part of experiments, conducted independently for both parts of the database, the Pearson's Linear Correlation Coefficients (PCC) with DMOS values have been calculated indicating the quality prediction accuracy (for the average raw scores obtained for the left and right images without any nonlinear mapping), as well as Spearman Rank Order Correlation Coefficients (SROCC) for verification of metrics' prediction monotonicity. The results obtained for some well-known metrics and three combined ones assuming the default weights proposed in respective papers (typically obtained as the result of optimization for the well-known TID2008 database) are presented in Table 1. Conducting the optimization of the weighting exponents for all combined metrics further increase of the PCC and SROCC values can be gained and the obtained results are shown in Table 2.

Analyzing the values of the correlation coefficients presented in Table 1 it can be noticed that the direct application of the combined metrics for the 'symmetrical' subset (Phase I) does not lead to satisfactory results, especially for the EHIS metric. Nevertheless, for the asymmetrically distorted Phase II images the best results have been achieved for the CISI metric with default weights [20] both by means of prediction accuracy and monotonicity.

Considering the different perception of distortions in the 2D and 3D images, the optimization of weights have been conducted, using MATLAB's *fminsearch* and *fminunc* functions, for the combined metrics and meaningful increase of the correlation with subjective scores has been achieved for all of them as shown in Table 2. Nevertheless, it is worth noticing that the optimization of weights using the Phase II subset leads to slightly worse results for the 'symmetrical' images and reversely. The smallest differences can be noticed for the CQM metric.

For the Phase I images the increase of the PCC values to over 0.93 can be obtained for each of combined metrics with only 0.915 for the FSIM whereas the SROCC values remain similar. For the Phase II images, characterized by different perception of distortions due to their asymmetry, the increase of the Pearson's correlation to 0.8953 can be observed for the EHIS metric with 0.8464 for the Gradient SSIM. Spearman correlation increases as well, from 0.833 to over 0.89 respectively.

Table 1. Obtained results of the PCC and SROCC values for various single metrics and the combined metrics with default weights

Subset	Phase I		Phase II	
Metric	PCC	SROCC	PCC	SROCC
SSIM	0.86127	0.87724	0.78463	0.79320
MS-SSIM	**0.87387**	0.89747	0.72748	0.73131
GSSIM	**0.89366**	0.91174	**0.84639**	**0.83305**
QILV	0.64446	0.91385	0.42955	0.67806
IW-SSIM	**0.89300**	**0.93353**	0.63632	0.74955
R-SVD	0.54962	0.52507	0.63548	0.61600
IFC	0.46750	0.91981	0.52732	0.76531
VIF	**0.86985**	**0.92054**	0.83407	0.81641
VIFp	0.84215	0.91357	0.80501	0.80557
VSNR	0.77026	0.88202	0.66007	0.73344
RFSIM	0.75834	0.82678	0.79863	0.78687
WFSIM	**0.89228**	**0.93033**	0.68225	0.74970
FSIM	**0.91504**	**0.92769**	0.76984	0.78638
FSIMc	**0.91389**	**0.92698**	0.78245	0.79771
CQM	0.74246	**0.92079**	0.46043	0.74973
CISI	**0.86599**	**0.92718**	**0.86137**	**0.84771**
EHIS	0.13234	0.43470	0.39562	0.52314

Table 2. Obtained results of the PCC and SROCC values for the optimized combined metrics

Subset	Phase I		Phase II	
Metric	PCC	SROCC	PCC	SROCC
CQM_{opt1}	**0.93358**	0.92813	**0.85183**	**0.86375**
CQM_{opt2}	0.88153	0.91301	**0.88968**	**0.88251**
$CISI_{opt1}$	**0.93444**	**0.93000**	0.81918	0.81587
$CISI_{opt2}$	0.88281	0.90219	**0.87239**	**0.86717**
$EHIS_{opt1}$	**0.93618**	**0.93286**	0.82138	0.81956
$EHIS_{opt2}$	0.89774	0.91867	**0.89528**	**0.89090**

5 Concluding Remarks

Application of the combined metrics for 3D image quality assessment leads to meaningful increase of quality prediction accuracy and monotonicity, expressed as the Pearson's and Spearman's correlation coefficients respectively. The results provided in this paper confirm both the usefulness of the hybrid metrics also for 3D images and the different perception of similar distortions present in 2D and 3D images, especially applied in an asymmetric way for the left and right images from the consecutive stereoscopic pairs.

It is worth to notice that the direct use of well-known metrics may lead to relatively good performance for symmetrically distorted images whereas the presence of asymmetric distortions causes a significant decrease of the PCC and SROCC values for most single metrics. Nevertheless, optimization of the weighting coefficients for the combined metrics allows to obtain much better performance than for any of single metrics used in the experiments.

Since in the conducted experiments no information related to depth has been utilized, a natural direction of further research is the combination of the developed metrics with depth or disparity information provided in most of available databases, including the LIVE 3D Image Database used in the paper.

References

1. Okarma, K.: Colour image quality assessment using Structural Similarity index and Singular Value Decomposition. In: Bolc, L., Kulikowski, J.L., Wojciechowski, K. (eds.) ICCVG 2008. LNCS, vol. 5337, pp. 55–65. Springer, Heidelberg (2009)
2. Okarma, K.: Video quality assessment using the combined full-reference approach. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 2. AISC, vol. 84, pp. 51–58. Springer, Heidelberg (2010)
3. Jayaraman, D., Mittal, A., Moorthy, A.K., Bovik, A.C.: Objective image quality assessment of multiply distorted images. In: Conf. Rec. 46th Asilomar Conf. Signals, Systems and Computers, pp. 1693–1697 (2012)
4. Benoit, A., LeCallet, P., Campisi, P., Cousseau, R.: Quality assessment of stereoscopic images. EURASIP Journal on Image and Video Processing Article ID 659024, 13 (2008)
5. Chen, M.-J., Su, C.-C., Kwon, D.-K., Cormack, L.K., Bovik, A.C.: Full-reference quality assessment of stereopairs accounting for rivalry. Signal Processing: Image Communication 28(9), 1143–1155 (2013)
6. Chen, M.-J., Cormack, L.K., Bovik, A.C.: No-reference quality assessment of natural stereopairs. IEEE Trans. Image Process. 22(9), 3379–3391 (2013)
7. You, J., Xing, L., Perkis, A., Wang, X.: Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis. In: Proc. 5th Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), pp. 61–66 (2010)
8. Gorley, P., Holliman, N.: Stereoscopic image quality metrics and compression. In: Proc. SPIE, vol. 6803, p. 5 (2008)
9. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: From error measurement to Structural Similarity. IEEE Trans. Image Proc. 13(4), 600–612 (2004)
10. Hewage, C.: Reduced-reference quality metric for 3D depth map transmission. In: Proc. 4th 3DTV Conf.: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), pp. 1–4 (2010)
11. Yang, J., Hou, C., Zhou, Y., Zhang, Z., Guo, J.: Objective quality assessment method of stereo images. In: Proc. 3rd 3DTV Conf.: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), pp. 1–4 (2009)
12. Zhu, Z., Wang, Y.: Perceptual distortion metric for stereo video quality evaluation. WSEAS Trans. Signal Process. 5(7), 241–250 (2009)

13. Shen, L., Yang, J., Zhang, Z.: Stereo picture quality estimation based on a multiple channel HVS model. In: Proc. 2nd IEEE Int. Congress on Image and Signal Processing (CISP), pp. 1–4 (2009)
14. Akhter, R., Parvez Sazzad, Z., Horita, Y., Baltes, J.: No-Reference stereoscopic image quality assessment. In: Proc. SPIE. Stereoscopic Displays and Applications XXI, vol. 7524, p. 75240T (2010)
15. Okarma, K.: Combined full-reference image quality metric linearly correlated with subjective assessment. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2010, Part I. LNCS (LNAI), vol. 6113, pp. 539–546. Springer, Heidelberg (2010)
16. Wang, Z., Simoncelli, E., Bovik, A.C.: Multi-Scale Structural Similarity for image quality assessment. In: Proc. 37th IEEE Asilomar Conf. on Signals, Systems and Computers, pp. 1398–1402 (2003)
17. Sheikh, H., Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process. 15(2), 430–444 (2006)
18. Mansouri, A., Mahmoudi-Aznaveh, A., Torkamani-Azar, F., Jahanshahi, J.: Image quality assessment using the Singular Value Decomposition theorem. Opt. Rev. 16(2), 49–53 (2009)
19. Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: A Feature Similarity index for image quality assessment. IEEE Trans. Image Proc. 20(8), 2378–2386 (2011)
20. Okarma, K.: Combined image similarity index. Opt. Rev. 19(5), 249–254 (2012)
21. Liu, T.-J., Lin, W., Kuo, C.-C.J.: Image quality assessment using multi-method fusion. IEEE Trans. Image Process. 22(5), 1793–1807 (2013)
22. Okarma, K.: Extended Hybrid Image Similarity - combined full-reference image quality metric linearly correlated with subjective scores. Elektronika ir Elektrotechnika 19(10), 129–132 (2013)
23. Zhang, L., Zhang, L., Mou, X.: RFSIM: A feature based image quality assessment metric using Riesz transforms. In: Proc. 17th IEEE Int. Conf. Image Processing, pp. 321–324 (2010)
24. International Telecommunication Union: Recommendation BT.601-7 - Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios (2011)