Stereoscopic images quality assessment

Patrizio Campisi
Università Degli Studi Roma Tre

Patrick Le Callet
University of Nantes

STEREOSCOPIC IMAGES QUALITY ASSESSMENT

Patrizio Campisi, Patrick Le Callet+, and Enrico Marini*+*

*Dipartimento di Elettronica Applicata, Università degli Studi Roma "Roma Tre",
Via della Vasca Navale 84, 00146 Roma, Italy
phone: +39.06.55177064, fax: +39.06.55177026, e-mail: campisi@uniroma3.it
+IRCCyN, Ecole polytechnique de l'Université de Nantes,
Rue Ch. Pauc, la chantrerie, 44306 Nantes Cedex 3, Nantes, France
phone: +33.2.40.68.30.47, fax: +33. 2.40.68.32.32 e-mail: Patrick.Lecallet@univ-nantes.fr

ABSTRACT

Although several metrics have been proposed in literature to assess the perceptual quality of bidimensional images, no similar effort has been devoted to quality assessment of stereoscopic images.

Therefore, in this paper, we propose a methodology for subjective assessment of stereo images. Moreover, in the process of defining an objective metric specifically designed for stereoscopic images, we evaluate whether 2-D image quality objective metrics are also suited for quality assessment of stereo images. Specifically, distortions deriving from both coding and blurring are taken into account and the quality degradation of the stereo pair is estimated.

1. INTRODUCTION

With the always increasing interest in stereoscopic technology [1], driven both by the entertainment industry and by scientific applications, several signal processing operations [2], [3] have been specifically designed for stereoscopic images. Therefore, it is evident the necessity to define standardized protocols to assess the perceived quality of the processed stereo images.

Quality assessment is achievable either through subjective tests or through objective metrics. The best way to assess image and video quality would surely be to run subjective tests according to standardized protocols, which are defined in order to obtain correct, universal, and reliable quality evaluations. However, the use of subjective tests is a time consuming approach. Furthermore, the analysis of the obtained results is not straightforward. Therefore, the definition of objective metrics reliably predicting the perceived quality of images would be a great improvement in the quality assessment field.

A great effort has been devoted by both the academic and the industrial community to develop objective metrics able to quantitatively evaluate the amount of degradation undergone by a signal, an image, or a video sequence. In fact objective metrics can be used in order to accomplish different tasks. Among the multitude of possible applications it is worth pointing out they can be used for benchmarking purposes in order to choose among several processing systems which can be used for the same purpose on a digital media; the system providing the best metric value will be used. Moreover, when image and video delivery takes place in an error prone scenario, the use of objective quality metrics can be used as side information for the image and video server to take the necessary actions to improve the quality of the received data, like prefiltering, optimal bit assignment algorithms, error concealment methods, and so on.

Objective image and video quality metrics can be classified according to the availability of the distortion free image and video signal, which may be used as a reference to compare a distorted image or video signal against the distorted counterpart. The metrics which assume that the original data are available are referred to as *full-reference* (FR) image and video quality metrics. When the reference images or video sequences are not accessible we refer to *no-reference* (NR) image and video quality assessment. In some scenarios we can assume that although the original image or video signal are not fully available, some features are available. Therefore, they can be used to support the quality assessment process. This is referred to as *reduced-reference* (RR) image and video quality assessment.

However, although several subjective and objective quality assessment methods have been proposed in literature for images and videos, no comparable effort has been devoted to the quality assessment of stereoscopic images. With the widespread of 3-D technology applied to different fields such as entertainment, CAD, medical applications, to cite only a few, 3-D images and videos needs to be processed. Therefore, the necessity to define both subjective procedures and objective metrics to assess the quality of the processed stereo images is becoming an issue of paramount importance.

In ITU-R BT.1438 [4] subjective quality assessment of stereoscopic television pictures is described. In [5] the effect of low-pass filtering one channel of a stereo sequence was explored in terms of perceived quality, depth, and sharpness. A comprehensive analysis of the perceptual requirements for 3-D TV is made in [6] along with a description of the main artefacts which may arise when dealing with stereo TV. Roughly speaking, stereo images perceived quality depends on several factors such as the rendered perception of depth, stereoscopic impairments [6] (key-stone distortion, depth-plane curvature, puppet theatre effect, cross talk, cardboard effect, shear distortion, picket-fence effect and image flipping), and visual (dis)comfort [7].

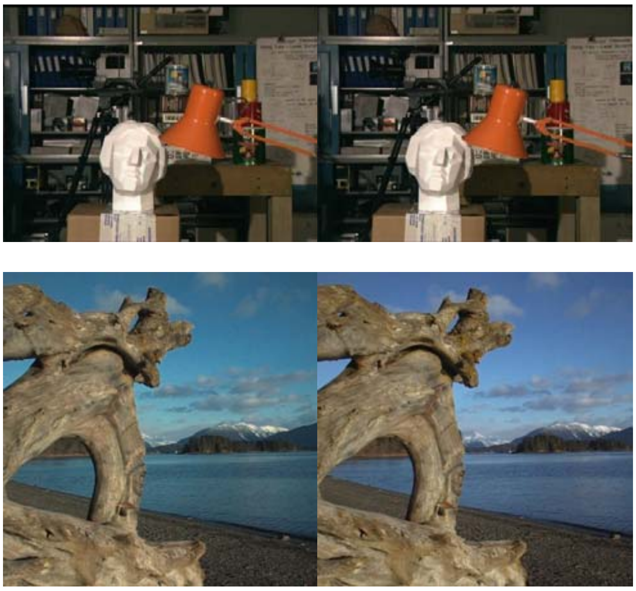

Figure 1 – Examples of some of the studied stereo pairs.

In this paper we first propose a methodology for subjective assessment of stereo images. Then, in the process of defining an objective metric specifically designed for stereoscopic images, we evaluate whether 2-D image quality objective metrics are also suited for quality assessment of stereo images. Specifically, we focus on the quality assessment when either compression or blurring are applied to the stereo pair.

The paper is organized as follows. In Section 2 a brief introduction on stereo images is given. Quality assessment of stereo images by means of subjective tests is detailed in Section 3. The performance analysis of several objective metrics for quality evaluation of stereoscopic images and conclusions are drawn in Section 4 and 5 respectively.

2. STEREO IMAGES

Stereo images and videos have been widely studied in the recent literature because stereoscopic viewing is one basic and popular way to perceive a scene in 3-D, that is by rendering the perception of depth. In general, 3-D perception is based on various depth cues such as illumination, relative size, motion, occlusion, texture gradient, geometric perspective, disparity, and many others. However, a very effective depth perception sensation is obtained by viewing a scene from slightly different viewing positions. From a physiological point of view, given a scene in the real world, 2-D slightly different scenes are projected on the retina of each eye. This implies that the 3-D depth information is lost at this stage. Then, the primary visual cortex in the brain fuses the stereo pair by means of a stereopsis and a prior knowledge on the 3-D world. Therefore, humans can perceive the depth starting from the bidimensional images on the retina of each eye. When 3-D imaging systems try to mimic the behaviour of the human visual system, the role of the eyes is taken over by stereo cameras that capture a scene from slightly different positions. The depth information can be obtained using stereo vision techniques by means of the disparity, the relative displacement of the stereo camera as well as its geometry (see Figure 1).

Roughly speaking the systems used to display stereo images present alternatively to the left and right eye two slightly different images in such a way that the human visual system gets a perception of depth. More in details the 3-D rendering systems can be classified as either *autostereoscopic* or *stereoscopic* displays. Autostereoscopic displays do not need any special viewing glasses, but the viewing angle is not very wide. On the other side, stereoscopic displays require viewing glasses such as red-and-blue lenses or polarized glasses, but they are more affordable than autostereoscopic displays and they can be used in commercial theatre as well as in a home environment. These systems allow the left and right images to be projected onto a screen with different polarization or colours. Among the stereoscopic systems it is worth citing the *active* systems where liquid crystal shutter glasses, which are synchronized with a display, are used.

3. QUALITY ASSESSMENT OF STEREO IMAGES THROUGH SUBJECTIVE TESTS

In this paper we perform quality assessment on stereo pairs when blurring, JPEG, and JPEG2000 compression are performed. We have applied these distortions on six different contents (see Figure 1 for some examples).

After generating several distorted images a pre-test was ran to evaluate the quality of the produced images, selecting fifteen distortions for each content: five JPEG compression ratios, five JPEG2000 compression ratios, and five levels of blur were considered. The selection criterion was not to choose the same distortions for different contents, but to have the possibly widest range of impairments so that subjective scores could be uniformly distributed for each class of distortion on the entire range going from very poor to very high quality (bad, poor, fair, good, and excellent)

Finally the test was performed in a controlled environment as recommended in ITU BT 500-11 [8], following SAMVIQ [9] protocol by using displays with active liquid crystal shutter glasses. SAMVIQ is a methodology for subjective test of multimedia applications using computer displays, whose application can be extended to embrace the full format television environment as well. The method proposed by SAMVIQ specification makes possible to combine quality evaluation capabilities and ability to discriminate similar levels of quality, using an implicit comparison process. The proposed approach is based on a random access process to play sequence files. Observers can start and stop the evaluation process as they wish and can follow their own paces in rating, modifying grades, repeating play out when needed. Therefore, SAMVIQ can be defined as a multi stimuli continuous quality scale method using explicit and hidden references. It provides an absolute measure of the subjective quality of distorted sequences which can be compared directly with the reference. As the assessors can directly compare the impaired sequences among themselves and against the reference, they can grade them accordingly. This feature permits a high degree of resolution in the grades given to the systems. Further, there is no continuous sequential presentation of items as in double-stimulus-continuous-quality-scale (DSCQS) method: this characteristic reduces possible errors due to a lack of concentration, thus offering higher reliability. Nevertheless, since each sequence can be played and assessed as many times as the observer wants, it is time consuming and less conditions can be tested during a session.

Figure 2 - SEOVQ User Interface.

The used protocol was implemented by the SEOVQ (see Figure 2) software developed by France Telecom R&D. SEOVQ is expressly designed for testing video images so that three players (WindowsMedia, RealVideo and QuickTime) are implemented, and two more (Div-x and Envivio) are defined as compatible. Unfortunately no one of these players offers the possibility of displaying stereoscopic images. OpenGL was then used to program a completely new player.

Seventeen observers originally took part to the test. Only three of them were discarded because the correlations between their individual scores and the mean opinion score were lower than a threshold (0.85). All the other observers had correlation values higher than the threshold.

The Difference Mean Opinion Score (*DMOS*) for the *i*-th image is computed as the difference between the *MOS* for the hidden reference MOS_{hr} and the one relative to the image *i*, MOS_i

$$DMOS_i = MOS_{hr} - MOS_i \, .$$

Subjective experiments lead to ninety *DMOS* values.

4. QUALITY ASSESSMENT OF STEREO IMAGES THROUGH OBJECTIVE METRICS

Four still image quality metrics, namely SSIM [10], UQI [11], C4 [12], and RRIQA [13], have been selected among the plethora of metrics proposed in the literature in order to design a quality metric tailored to stereo images.

A brief description of the aforementioned metrics is given here.

- **Structural SIMilarity (SSIM)** [10] is an objective metric for assessing perceptual image quality, working under the assumption that human visual perception is highly adapted for extracting structural information from a scene. Quality evaluation is thus based on the degradation of this structural information assuming that error visibility should not be equated with loss of quality as some distortions may be clearly visible but not so annoying. Finally SSIM does not attempt to predict image quality by accumulating the errors associated with psychophysically understood simple patterns, proposing to directly evaluate the structural changes between two complex-structured signals.
- **Universal Quality Index (UQI)** [11] is a universal objective image quality index, designed to model any image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. UQI is a mathematically defined metric such as the widely used Root Mean Squared Error (RMSE) or the PSNR.
- **C4** [12] is a metric based on the comparison between the structural information extracted from the distorted and the original images. What makes this metric interesting is that it uses reduced references containing perceptual structural information and exploiting an implementation of a rather elaborated model of the Human Visual System. The full process can be decomposed into two phases. During the first step, perceptual representation is built for the original and the distorted images, then, during the second stage, representations are compared in order to compute a quality score.
- **Reduced Reference Image Quality Assessment (RRIQA)** [13] is a reduced reference metric, based on a natural image statistic model in the wavelet transform domain. The Kullback-Leibler distance between the wavelet coefficients marginal probability distributions of the reference and the distorted image is used as a measure of the image distortion. A generalized Gaussian model is employed to summarize the marginal distribution of the reference image wavelet coefficients, so that only a relatively small number of images' features are needed for quality evaluation. The basic assumption behind this approach is that most image distortions modify image statistics and make the distorted version "unnatural". The measured unnaturalness can then be used to quantify image quality distortion.

Referring to stereo pairs, we have considered a limited set of distortions in the attempt to better evaluate the degradations perceived by the observer in the fruition of the stereo image.

Specifically, we have applied the following distortions:

- blurring,
- JPEG compression,
- JPEG2000 compression,
- JPEG & JPEG2000 compression,

with the same strength to both images of the stereo pair.

	Global	*Blur*	*JPEG*	*JPEG 2000*	*JPEG & JPEG2000*
SSIM	0.58	0.42	0.81	0.83	0.82
UQI	0.65	0.49	0.79	0.84	0.81
C4	0.73	0.40	**0.93**	**0.94**	**0.94**
RRIQA	0.59	**0.69**	0.65	0.45	0.55

Table 1 - Correlation values for the "Average" approach, for different metrics grouped per distortion before the mapping process.

	Global	*Blur*	*JPEG*	*JPEG 2000*	*JPEG & JPEG2000*
SSIM	0.74	0.51	0.85	0.88	0.86
UQI	0.71	0.49	0.79	0.85	0.82
C4	0.79	0.48	**0.94**	**0.97**	**0.95**
RRIQA	0.59	**0.72**	0.66	0.58	0.55

Table 2 - Correlation values for the "Average" approach, for different metrics grouped per distortion after the mapping process.

Basically, quality scores on both images (right and left) of the stereo content are evaluated by means of the four aforementioned metrics, thus obtaining two scores for the right and the left image of the stereo pair. Then, in order to obtain a single metric for the quality assessment of the stereo image, the two so obtained quality scores must be combined. Three different combination approaches have been taken into account:

- "**average**" approach,
- "**main eye**" approach,
- "**visual acuity**" approach.

Results were grouped according to the class of distortion, obtaining for each metric five correlation values: a global one for all the images, one for blurred images, one for JPEG compression, one for JPEG2000 compression, and finally one for JPEG & JPEG2000 compression.

In the **"average"** approach the scores coming from the metrics applied to the right and the left image separately are averaged. Correlation between *DMOS* and the results deriving from the tested objective metrics were thus used for comparing the performances of the employed metrics.

In Table 1 the correlation between *DMOS* and each of the four objective metrics for each of considered distortions are given. In Table 2 the same values obtained after the so called "mapping" are shown. Specifically, mapping refers to the application of non linear function as recommended by VQEG [14] in order to map metrics scores into subjective score space. For each condition, parameters of the mapping function have been optimized.

It can be easily noticed that among the tested metrics, the RRIQA is the metric that better represents the perceived degradation of the stereo pair because of blurring. The C4 metric is the metric giving the best performance for evaluating the perceived distortion on the stereo pair. No one metric performs acceptably for all kinds of distortions.

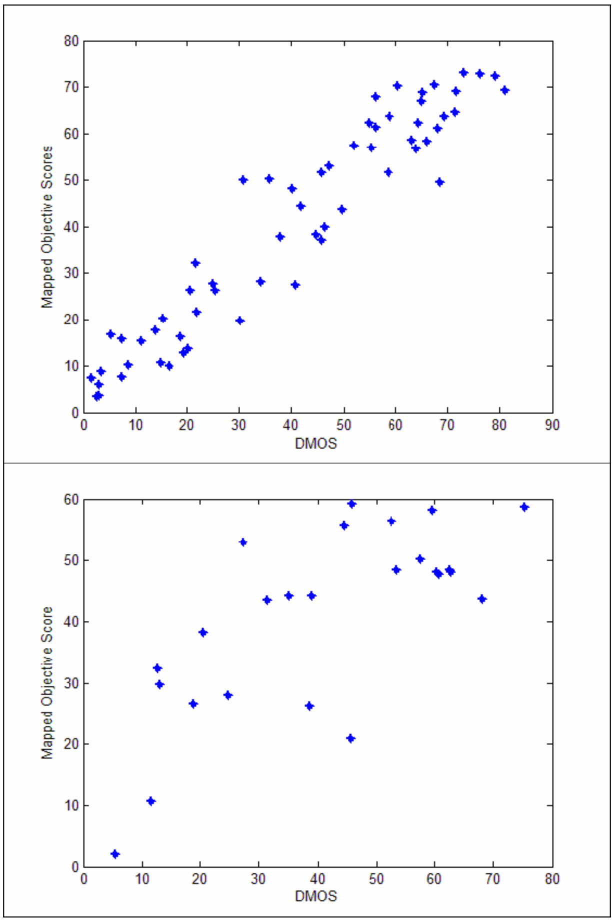

Figure 3 - Couples of points (DMOS, Mapped MOS) for JPEG & JPEG2000/C4 (top) and blur/RRIQA (bottom).

In Figure 3, couples of points (DMOS, Mapped objective score) are shown to compare performances of different metrics in the case of blurred images and JPEG plus JPEG2000 distortions in the case of the "average" approach.

The objective metrics evaluated separately on the right and left image of the stereo pair have also been combined by means of what we have called the **"main eye"** approach, where the main eye of each observer has been taken into account. Specifically, the "main eye" objective score has been defined as follows:

$$Objectivescore = \frac{1}{N_{obs}}(N_{left}Score_{left} + N_{rigth}Score_{right})$$

where N_{obs} is the total number of observers, N_{left} and N_{right} is the number of observers whose main eye is respectively the left or the right one, and $Score_{left}$ and $Score_{right}$ are objective scores for the left and the right images. This combination tries to weight different objective scores giving more importance to the evaluation of the image relative to the main eye of the observer. However, the quantitative evaluation of this approach has revealed that no significant performance improvement has been obtained.

The objective metrics evaluated independently on the right and left image of the stereo pair have also been combined by means of the **"visual acuity"** approach. Specifically the scores obtained for the left and right image have been weighted by using the visual acuity of the observers. The rational behind this approach is that the perceived quality could decrease along with the visual acuity of the observer. Specifically, the employed combining rule is:

$$Objectivescore = \frac{1}{N_{obs}}\left(\frac{\sum_i Acuity_{left,i}}{Acuity_{\max}} Score_{left} + \right.$$
$$\left. + \frac{\sum_i Acuity_{right,i}}{Acuity_{\max}} Score_{right} \right)$$

where $Acuity_{left,i}$ and $Acuity_{right,i}$ represent the visual acuity of the left and right eye of the *i*-th observer respectively, and $Acuity_{\max}$ represents the maximum value for the visual acuity that has been set equal to ten. The weights for $Score_{left}$ and $Score_{right}$ have been observed to be equal to 0.43 and 0.57 respectively, which are not significantly different from the weights used in the "average" approach (0.5). Therefore, also the use of this metric has led to no performance improvement with respect to the "average" approach.

5. CONCLUSIONS

The goal of this work is twofold. First, we propose a methodology for subjective assessment of stereo images. Then, we test some well known bidimensional objective metrics to verify the possibility to use them for quality assessment of stereoscopic images. The path followed to accomplish this task was to compare results of tested metrics with those of subjective tests. We found out that none of the employed metrics succeeded in assessing quality of blurred images. Specifically, RRIQA metric performed better than the other employed metrics in assessing the quality of blurred stereo images, but it did not give satisfactory results when either JPEG or JPEG2000 compression were applied. More in detail, when JPEG & JPEG2000 compression is taken into account the C4 metric achieves optimal performances.

These results were obtained by simply averaging objective scores for left and right images. Several attempts to combine the objective left and right scores taking into account different aspects of visual characteristics of the observers have been taken into account, although they did not improve the results. However, future works is oriented toward the test of different combinations for left and right scores in order to obtain more appealing results.

It is worth pointing out that this research was done applying a distortion of the same entity to both the images of the stereoscopic pair. Future research would embrace the characterization, by means of objective metrics, of the distortions perceived on the stereo pair when the same distortion but with different intensity is applied to the images composing the stereo pair. The preliminary study we have conducted so far points out that a real masking effect occurs between the two eyes. Moreover, a very annoying sensation, almost leading to sickness, occurs if the difference between the left and right images is significant. This behaviour affects quality assessment in a way that is not captured by bidimensional image quality metrics. Therefore, this is clearly a wide open challenging issue.

ACKNOWLEDGMENT

Authors would like to acknowledge Arnaud Tirel for providing support during subjective quality assessment experiments.

REFERENCES

[1] L. Lipton. *Foundations of the stereoscopic cinema: a study in depth*. Van Nostrand Reinhold, New York, 1982.

[2] M.Z. Brown, D. Burschka, and G.D. Hager, "Advances in computational stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, pp. 993–1008, 2003.

[3] A. Puri, R. V. Kollarits, B. G. Haskell, "Basic of stereoscopic video, new compression results with MPEG-2 and a proposal for MPEG-4," *Signal Processing: Image Communication,* 10 pp. 210–234, 1997.

[4] Rec ITU-R BT.1438, Subjective Assessment of Stereoscopic Television Pictures.

[5] L. Stelmach, W. J. Tam, D. Meegan, and A. Vincent, "Stereo Image Quality: Effects of Mixed Spatio-Temporal Resolution," *IEEE Transaction on Circuits Systems and Video Technology*, vol 10, no. 2, pp. 188–193, 2000.

[6] L.M.J. Meesters, W.A. IJsselsteijn, P.J.H. Seuntiens, "A survey of perceptual evaluations and requirements of three-dimensional TV", *IEEE Transaction on Circuits Systems and Video Technology,* vol. 14, no. 3, pp. 381–391, Mar. 2004.

[7] M.T.M. Lambooij, W.A. IJsselsteijn, I. Heynderickx, "Visual discomfort in stereoscopic displays: a review", *SPIE Electronic Imaging, Stereoscopic Displays and Virtual Reality Systems XIV,* vol. 6490, 2007.

[8] ITU-R BT.500-11. Methodology for the Subjective Assessment of the Quality of Television Pictures.

[9] EBU Project Group B/VIM Video In Multimedia. SAMVIQ: Subjective Assessment Methodology for Video Quality. May 2003.

[10] Z. Wang, A.C. Bovik A, H.R. Sheikh, E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing,* Volume 13, No. 4, April 2004.

[11] Z. Wang and A. C. Bovik, "A Universal Image Quality Index," *IEEE Signal Processing Letters,* Vol. 9, No. 3, March 2002.

[12] M. Carnec, P. Le callet, and D. Barba, "An Image Quality Assessment Method Based on Perception of Structural Information," *IEEE Int. Conf. on Image Processing ICIP 2003*, Barcelona, Spain, September 2003.

[13] Z. Wang, and E. P. Simoncelli, "Reduced-Reference Image Quality Assessment Using a Wavelet-Domain Natural Image Statistic Model," *Human Vision and Electronic Imaging X, Proceedings of the SPIE,* Volume 5666, March 2005.

[14] VQEG, Video Quality Expert Group, www.its.bldrdoc.gov/vqeg