USING DISPARITY FOR QUALITY ASSESSMENT OF STEREOSCOPIC IMAGES

Alexandre Benoit[+], *Patrick Le Callet*[+], *Patrizio Campisi*[*], *Romain Cousseau*[+]
[+]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: alexandre.benoit@univ-nantes.fr, Patrick.Lecallet@univ-nantes.fr
*Dipartimento di Elettronica Applicata, Università degli Studi di Roma "Roma Tre",
Via della Vasca Navale 84, 00146 Roma, Italy
phone: +39.06.55177064, fax: +39.06.55177026, e-mail: campisi@uniroma3.it

ABSTRACT

3DTV has been widely studied these last years from a technical point of view but the related quality evaluations does not follow this enthusiasm.
This article reviews the quality assessment for 3DTV. Compared to 2D quality measure, the third dimension adds several new problems and quality assessment becomes a complex issue. Nevertheless, efforts made for 2D content quality estimation can be used for an extension to 3D. In this paper we propose a first attempt to adapt such 2D metrics to 3D content and add the contribution of a measure of the distortion on the disparity map for stereoscopic image pairs. This 3D metric performances has been evaluated with subjective tests.

Index Terms *quality assessment, 3D, 2D metrics, disparity distortion*

1. INTRODUCTION

The aim of the paper is to review the quality assessment issue for 3D visualisation and present an attempt to assess 3D image quality. 3D imaging is a wide research area in which enthusiastic research efforts have been recently revealed in [1]. From John Logie Baird who introduced the first version of stereo TV, many approaches have been developed [2]: stereoscopic vision with glasses, auto stereoscopic displays, holographic systems. In parallel, methods for 3D scene representation [3] and data content broadcasting [4] have been widely studied.
Nevertheless, even if technology solutions are now available, the issue of stereo images quality evaluation has to be investigated in depth. Indeed, contrary to the 2D imaging community, no 3D quality metric has already been developed. Several new problems appear when dealing with 3D which are also linked to human factor such as accommodation problems and 3D artefacts perception [5]. However, requirements for perceptual quality evaluations have already been identified [6] opening a way to investigate 3D quality metrics.
Up to now, few objective quality metrics have been proposed, the task is difficult and the idea should be to involve 2D metrics. For example in [7] a method which uses 2D quality metric for the quality assessment of stereo images is proposed. We present in this work an improvement of this method by taking into account the disparity map between stereoscopic image pairs.
The paper is presented as the following: section 2 presents a review of quality issue in 3D. Section 3 presents the subjective tests on which this work is based. Section 4 and 5 present respectively the objective quality metric we propose and the related results.

2. QUALITY ISSUE IN 3D

3D quality measure constitutes a real challenge for several reasons. First, from a technical point of view, many approaches have been developed. From stereo vision using glasses which requires stereo image pairs to multivue autostereoscopic displays which use multiple views of a visual scene or a 2D image plus related pixel depth, it is difficult to built an objective metric able to cope with all these formats.
Second, from a visual point of view, 3D perception involves new critical points which have to be taken into account. We can draw a short list of the psychovisual measures that this new scene representation should include:
. Image quality from an imagine coding point of view: artefacts such as blockiness, blurring, jerkiness and ghosting are caused by the algorithm involved in the image coding process (spatial filtering, temporal low pass filtering and resolution reduction).
. Depth precision: as the depth of the image can be rendered using different technologies and data formats, depth representation and coding impact on the visual perception.
. Field of view: can the observer move his head freely and perceive behind the boundaries of the objects without perceiving artefacts.
. Human factors: in addition to the image quality evaluation from a technology point of view, we can address another problem which is linked to human perception. We can distinguish several factors, general problems such as depth perception, accommodation problems but also physiological differences between people which induces different perception (age, inter-pupillary distance etc.). All these factors are difficult to integrate but they yield to visual discomfort, perceived 3D distortions, etc. For example, when evaluating image quality for stereo images, observers can give good scores to 3D scenes representation even if one of the images perceived

by one eye is highly disturbed. The visual system has indeed a tendency to compensate the lower quality perceived by an eye by the quality of the other eye perception. Nevertheless eye fatigue phenomena can be observed in such case. As a consequence, image quality assessors must take such factors into account in order to really evaluate the user experience.
This brief overview shows the wide research area for quality metric design that still waits to be investigated. Nevertheless, in a first step, all the parameters can not be taken into account. It is necessary to first, choose a technology and focus on a limited set of factors. Such evaluation task started recently for example in [8] where the impact of the compression of the depth information for 2D plus depth visual scene coding on auto stereoscopic displays has been studied. The preference of the observers in regard of the compression method and related bit rates are investigated. Also, in [7] is presented the beginning of a metric design based on subjective measures that allows 3D image quality assessment. The aim of this work is to analyse the relevancy of 2D quality metrics applied to stereo content. Different metrics where evaluated for stereoscopic vision with glasses. Nevertheless, introducing 2D quality metrics in a 3D context is a real challenge when attempting to integrate depth information. We propose an attempt for such factor integration by involving a quality metric on disparity maps between views. Then several questions appears from the choice of the quality measure operator to the fusion if its result in the original metric. This first attempt is still limited to stereo vision with glasses, an extension to 2D plus depth for auto stereoscopic displays is expected in future research.

3. SUBJECTIVE STEREO QUALITY ASSESSMENT

In [7] we proposed a methodology for subjective assessment of stereo images by using well established 2D image quality objective metrics. Specifically, we focused on the quality assessment when either compression (JPEG and JPEG2000) or blurring were applied symmetrically to the stereo pair.

We have applied these distortions on six different contents. The test was performed in a controlled environment as recommended in ITU BT 500-11, 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.

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. OBJECTIVE STEREO QUALITY ASSESSMENT

In this Section a new metric for objective quality assessment is proposed. Specifically, in Section 4.1, we revise the approach proposed in [7] for the quality assessment of 3D images, where 2D metrics are applied separately to the left and right image of the stereo pair and then properly fused. In Section 4.2, we present a new metric which makes use of the depth information of the stereo pair.

4.1 Using 2D objective metrics

In [7] is presented a metric dedicated to stereo vision using glasses. In this work, four 2D quality metrics have been evaluated in 3D context including Structural SIMilarity (SSIM) and C4. All these metrics were applied separately on each eye and fusion methods were investigated. The correlation between *DMOS* and each of the objective metrics for each of considered distortions are given after a "mapping" operation in order to evaluate the performances of the metrics. Specifically, mapping refers to the application of non linear function as recommended by VQEG [10] in order to map metrics scores into subjective score space. For each condition, parameters of the mapping function have been optimized. It appeared that the most effective fusion was the average of both left and right eye measure.

4.2 3D quality metrics using depth information and 2D objective metrics

In this Section, we propose more refined metrics for the quality assessment of 3D images which stem from the ones used in [7]. However, since the metrics used in [7] suffer from the drawback that no information related to the 3D nature of the images are taken into account, we resort to take into account also the disparity map and to fuse it with the scores coming from the metrics employed in the previous method. Indeed, as well know [1, 11], the sense of stereovision can be artificially induced by presenting two different images of the same scene, shifted one with respect to the other, thus mimicking two different viewpoints, namely the left and the right image of the stereo pair, to the left and the right eye. The difference in the viewpoints generates *disparity* in the images. More in details, given two corresponding points in the left and the right image of a stereo pair the vector between the two points is called disparity. In general disparity can be used to reproduce one of the two images of the stereo pairs having the other one. In this paper, we do not intend to study the impact of the choice of the disparity computation algorithm and chose the one described in [12].
Figure 1 shows an overview of such a system.

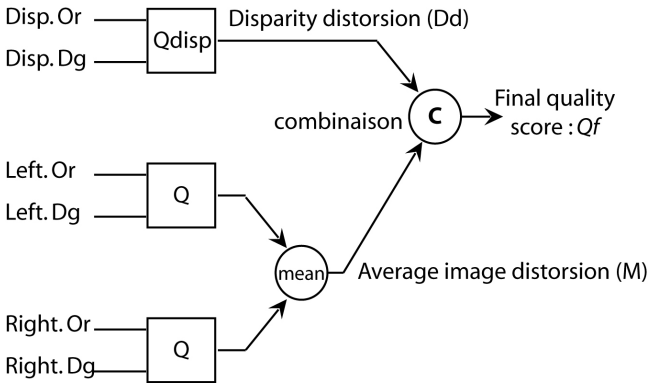


Fig. 1: quality estimation of stereo pairs using original Left and Right views (Left.Or, Right.Or) compared with the degraded versions (Left.Dg, Right.Dg) and the related original disparity map compared to the degraded disparity map (Disp.Or and Disp.Dg).

On this figure, the distortion is measured independently between both eyes with metric *Q*, this metric being either SSIM or C4. These two measures are averaged in order to get the global 2D image distortion measure *M*. In parallel, the disparity distortion measure *Dd* is computed using correlation coefficient between the original disparity maps and the corresponding disparity maps processed after image degradation (cf. Fig. 2). As disparity maps are not natural images, using psychovisual based distortion metrics is non-sens, however simpler distortion methods such as correlation coefficient are better appropriate.
The final objective quality measure *Qf* is obtained after the fusion of the disparity and the averaged left and right image distortion measures. Two linear combinations have been tested and we add a third method which only includes the disparity distortion measure *Dd* as it was done for *M* in [7]:

$$Qf_1 = M.D^{1/2} \; ; \; Qf_2 = M.(1+D) \; ; \; Qf_3 = D$$

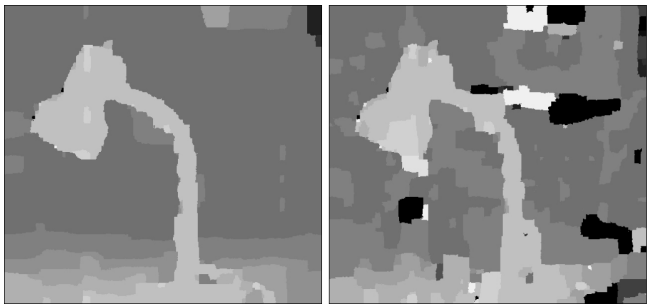


Fig. 2: original disparity map (left) and disparity map after JP-G2000 compression

5. RESULTS

We computed these quality metrics on stereo pairs when JPEG and JPEG2000 compression are performed as described in [7], using the same database. Nevertheless, contrary to [7] where the metrics were evaluated independently on JPG and JPG2000 compressions and finally fused to get the performance on both compressions, here, we evaluate the performance of the metrics on JPG and JPEG200 compressions at the same time. As a consequence, we consider simultaneously a larger spectrum of possible distortions. Results are presented on table I. We compare the correlation coefficient *CC* obtained with the original SSIM and C4 method applied to *M* [7] with the proposed method and related combination's. After mapping, more indicators become available such as Mean Square Errors (*RMSE* on a scale of 100) and the number of out layers (*OR*) which allow in-depth performance analysis.
Table I: metrics performances synthesis before mapping

	SSIM	SSIM-Qf_1	SSIM-Qf_2	C4	C4 Qf_1	C4 Qf_2	SSIM-Qf_3
CC	0,869	0,902	0,899	0,872	0,932	0,943	0,823
RMSE	10,43	10,43	10,55	11,91	8,82	8,1	13,8
OR	2	2	2	2	1	1	3

Significant performance increases can be observed with the proposed new metrics. Combinations Qf_1, Qf_2 always give better results than the original metric. More precisely, Qf_1 is more efficient with SSIM method and Qf_2 is more accurate with C4 metric. In parallel, Qf_3 shows that, as the original method, using only either *M* or *Dd* is not sufficient and gives lower results.
Moreover, with the proposed combinations, the number of outlayers remains constant or is reduced with C4 metric. We can also see that MSE remains stable with SSIM but is significantly reduced with C4 metric. This can be observed on

figure 3 where couples of points (DMOS, Mapped objective score) for C4 metric are shown for the original method and the new C4 with combination. On this figure, we can see the correlation coefficient increase and the RMSE reduction with the proposed new method.

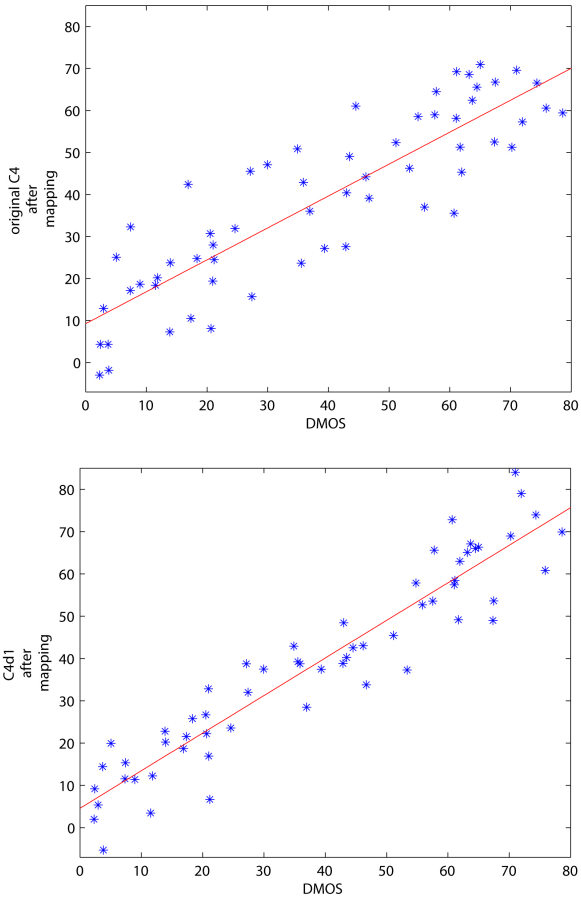


Figure 3 - Couples of points (DMOS, Mapped MOS) for JPEG + JPEG2000/C4 with the original method (top) and the proposed C4/ Qf_2 metric (bottom)

6. CONCLUSION

In this paper was emphasised the need of metrics for quality assessment in stereovision. An extension of the metric presented in [7] which involves the measure of the disparity map distortion was proposed. A significant metric performance enhancement was observed when using linear combinations of the disparity map distortion and the measure of the 2D image quality on both eyes. Future work will attempt to evaluate the influence of the choice of the disparity map computation algorithm and the extension of the metric to other display technologies.

7. ACKNOWLEDGMENT

This work was supported by FuturIm@ge project within the Media and Networks competitive pole fromthe French Pays de la Loire region.

8. REFERENCES

[1] IEEE Signal Processing Magazine, November 2007

[2] Benzie, P.; Watson, J.; Surman, P.; Rakkolainen, I.; Hopf, K.; Urey, H.; Sainov, V.; von Kopylow, C.; A Survey of 3DTV Displays: Techniques and Technologies , IEEE Transactions on Circuits and Systems for Video Technology, Volume 17, Issue 11, Nov. 2007 Page(s):1647 - 1658, DOI 10.1109/TCSVT.2007.905377

[3] Alatan, A.A.; Yemez, Y.; Gudukbay, U.; Zabulis, X.; Muller, K.; Erdem, C.E.; Weigel, C.; Smolic, A.; Scene Representation Technologies for 3DTV A Survey , IEEE Transactions on Circuits and Systems for Video Technology, Volume 17, Issue 11, Nov. 2007 Page(s):1587 - 1605, DOI 10.1109/TCSVT.2007.909974

[4] Smolic, A.; Mueller, K.; Stefanoski, N.; Ostermann, J.; Gotchev, A.; Akar, G.B.; Triantafyllidis, G.; Koz, A.; Coding Algorithms for 3DTV A Survey , IEEE Transactions on Circuits and Systems for Video Technology, Volume 17, Issue 11, Nov. 2007 Page(s):1606 - 1621, DOI 10.1109/TCSVT.2007.909972

[5] Lambooij M.T.M.; IJsselsteijn W.A.; Heynderickx I.; Visual Discomfort in stereoscopic Displays: A Review , SPIE-IS&T, Volume 6490, 64900I-1, January 2007

[6] Meesters, L.M.J.; IJsselsteijn, W.A.; Seuntiens, P.J.H., A survey of perceptual evaluations and requirements of three-dimensional TV , IEEE Transactions on Circuits and Systems for Video Technology, Volume 14, Issue 3, March 2004 Page(s): 381 - 391, DOI 10.1109/TCSVT.2004.823398

[7] P. Campisi, P. Le Callet, and E. Marini, Stereoscopic images quality assessment, European Signal Processing Conference (EUSIPCO) 2007, September 2007, Poznan, Poland.

[8] Fehn C.; 3D TV Broadcasting , chapter from 3D Video communication, pages: 23-38, Jan. 2006, DOI 10.1002/0470022736.ch2

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

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

[11] 3D Videocommunication. Algorithms, concepts and real-time systems in human centered communication , Edited by O. Schreer, P. Kauff, t. Sikora, WILEY 2005.

[12] Felzenszwalb P.F.; Huttenlocher D.P.; Efficient Belief Propagation for Early Vision , International Journal of Computer Vision, Vol. 70, No. 1, October 2006.