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: +39.06.57337064, fax: +39.06.57337026, e-mail: campisi@uniroma3.it

ABSTRACT

3DTV has been widely studied these last years from a technical point of view but the related perceived quality evaluations do not follow this hype. This paper firstly reviews quality assessment issues for 3DTV. Compared to 2D quality measure, depth information adds several new problems considering quality assessment. Nevertheless, efforts made for 2D content quality estimation can be used for an extension to 3D. In a second part, this paper proposes an adaptation of 2D metrics to 3D in the context of coding artifacts and stereoscopic images. Distortion on disparity is introduced to improve conventional 2D metrics. Performances have been evaluated using subjective tests.

Index Terms— image quality assessment, stereo image quality metric, disparity distortion.

1. INTRODUCTION

3D imaging is a wide research area in which several successful efforts have been recently revealed in the special section of IEEE Signal Processing Magazine November 2007 [1]. From John Logie Baird who introduced the first version of stereo TV, many approaches have been developed [2]: stereoscopic vision with polarizing glasses, auto stereoscopic displays for free viewpoint TV, or sophisticated holographic systems. In parallel, methods for 3D scene representation [3] and data content broadcasting [4] have been widely studied.

Since technology solutions are now available, it is of paramount importance to take care about 3D images quality assessment. 2D image quality assessment has concentrated many efforts for almost three decades, whereas 3D quality assessment is still an immature topic. From a visual point of view, 3D perception involves new critical points which have to be taken into account. First, subjective experiments [5,6,7] have to be performed in order to identify the main new issues. We can draw a short list of some psychological effects that this new scene representation implies:

-Depth perception and quality constancy regarding field of view: as the depth of the image can be rendered using different technologies and data formats, depth representation impacts on the visual perception.

-Depending on the technology, visual effects such as key-stone distortion, depth-plane curvature, puppet theatre effect, cross-talk, cardboard effect, shear distortion, picket-fence effect and image flipping can appear. Also blockiness, blurring, jerkiness and ghosting appear due to compression.

-Visual perception issues: as the visual system integrates 3D from the 2D images captured by the eyes, general problems such as accommodation issues, inter-eye masking effect can appear but also physiological differences between people impact on individual perception.

All these general cues yield to observer fatigue and impact on visual comfort. This brief overview shows the wide research area for, on one hand physiological issues investigation and on the other hand objective quality metric design. This last goal has not been widely explored yet, as the number of factors is important. In a first step, it is wiser to preliminary choose a technology and focus on a limited set of parameters. Regarding efforts done for 2D images, one could think to apply this knowledge to 3D, especially in the case of stereo images that can be seen as two pairs of 2D images. In [8], such approach is proposed simply combining 2D quality score of each image of the stereo pair. In this paper, we propose an improvement of the method described in [8] by taking into account the disparity information.

The paper is organized as the following: section 2 presents the subjective tests on which this work is based. Section 3 and 4 present respectively the objective quality metric we propose and the related results.

2. SUBJECTIVE STEREO QUALITY ASSESSMENT

In [8], 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 was applied symmetrically to the stereo pair. Figure 1 shows a capture of the experiment.

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. 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. . 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 1: capture of one experiment, the user is facing the screen with crystal shutter glasses.

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

$$DMOS=MOS_{hr} - MOS_i$$

Subjective experiments lead to ninety *DMOS* values.

3. OBJECTIVE STEREO QUALITY ASSESSMENT

In [8], we have 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) [13] and C4 [14]. All these metrics were applied separately on each image (left and right eye) and fusion methods between "per image" quality score 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 (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.

Figure 2: original disparity map (left) and disparity map after JPG2000 compression

In this paper, we propose more refined metrics for the quality assessment of 3D images which stem from the ones used in [8]. Previous study suffers 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 stereo vision 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]. Applying coding, distortions can lead to modify this disparity between right and left views (see Figure 2). Therefore, it could be useful to introduce this modification in order to assess the quality.

When designing a quality metric for stereo data content, the disparity information can be taken into account in different ways. We propose the method sketched in Figure 3. Specifically, two types of distortions are considered:

- one coming from the difference between original (left or right) images and the corresponding (left or right) distorted version. For this purpose, one can use usual 2D perceptual quality metric such as SSIM or C4. From [8], the two measures per pair are averaged in order to get the global 2D image distortion measure *M*.
- one coming from the comparison between original disparity and disparity computed on distorted stereo pair. As disparity maps are not natural images, using perceptual based distortion metrics is non-sense. In this paper disparity distortion measure D_d is computed using global correlation coefficient between the original disparity maps and the corresponding disparity maps processed after image degradation.

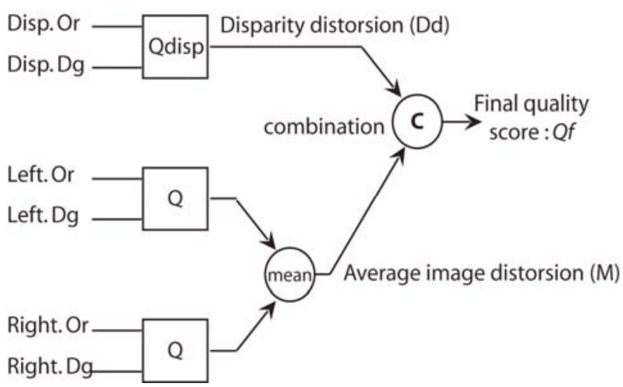

Figure 3: quality estimation of stereo pairs using original Left and Right views (Left.Or, Right.Or) compared with the degraded versions (Left.Dg, Right.Dg) and the related original disparity map compared to the degraded disparity map (Disp.Or and Disp.Dg).

The final quality measure *d* is obtained after the fusion of the disparity and the averaged left and right image distortion measures. In this paper, we focus on linear combinations of the 2D and disparity distortion measures and we propose two linear combinations d_1 and d_2. We add a third method which only includes the disparity distortion measure D_d in order to evaluate the relevance of this measure when used alone:

$$d_1=MD_d^{\frac{1}{2}}; \qquad d_2=M(1+D_d); \qquad d_3=D_d$$

By using C4 and SSIM metrics we obtain seven different metrics to perform quality assessment: SSIM (no disparity), C4 (no disparity), d_3 (disparity only), SSIM using d_1, C4 using d_1, SSIM using d_2 and C4 using d_2.

4. RESULTS

We computed these quality metrics on stereo pairs when applying JPEG and JPEG2000 compression. Nevertheless, contrary to [8] where the metrics were evaluated independently on JPEG and JPEG2000 compressions, we evaluate here the performance of the metrics on JPEG and JPEG2000 compressions at the same time (e.g. mapping is applied on the overall database). As a consequence, we consider simultaneously a larger set of possible distortions. Results before mapping are presented in table I. We compare the correlation coefficient *CC* obtained with the original SSIM and C4 methods (only applied to the right and left views) with the proposed new metrics which include a global disparity distortion measure (D_d).

	SSIM	SSIM d_1	SSIM d_2	C4	C4 d_1	C4 d_2	d_3
CC	0.826	0.900	0.896	0.935	0.930	0.938	0.815

Table I: metrics performances synthesis before mapping

Significant performance increases can be observed with the proposed new SSIM based metrics while no significant increase is observed with C4 based metrics. This shows that SSIM metric performances are increased with the added value of the disparity distortion measure. As a result its performance reach values close to the one of a perceptually based metric like C4. In parallel, C4 metric is sufficient on its own in this test setup so that the depth information do not add relevant information. In parallel, d_3 shows that, as the original method, using only either *M* or D_d is not sufficient and gives lower performances. In comparison, table II presents more complete results obtained after data mapping. After this operation, more indicators become available such as Root Mean Squared Error (*RMSE* on a scale of 100) and the number of outliers (*OR*) which allow in-depth performance analysis.

	SSIM	SSIM d_1	SSIM d_2	C4	C4 d_1	C4 d_2	d_3
CC	0.869	0.902	0.899	0.939	0.932	0.943	0.823
RMSE	11.94	10.43	10.55	11.91	8.82	8.1	13.8
OR	2	2	2	2	1	1	3

Table II: metrics performances synthesis after mapping

Compared to the original method (C4, SSIM without disparity), the increase of the correlation coefficient due to mapping is almost insignificant for the new metrics. We obtain a correlation coefficient increase of 0.002 with SSIM after mapping using metric d_1 metric and 0.005 for C4 using d_2 while the original SSIM without disparity increased by 0.04. Indeed, since the new SSIM based metrics are already highly correlated with the *DMOS* before mapping, the mapping do not increase significantly performances. Also, with the proposed combinations, the number of outliers remains constant or is reduced with C4 metric. In addition we can see that RMSE remains stable with SSIM but is significantly reduced with C4 metric. This can be observed in Figure 4 where couples of points (DMOS, Mapped objective score) for C4 metric are shown for the original method and the new C4 with d_2 combination. Also it is evident from figure 4 that the correlation coefficient increase and the RMSE reduces when the proposed metric is applied.

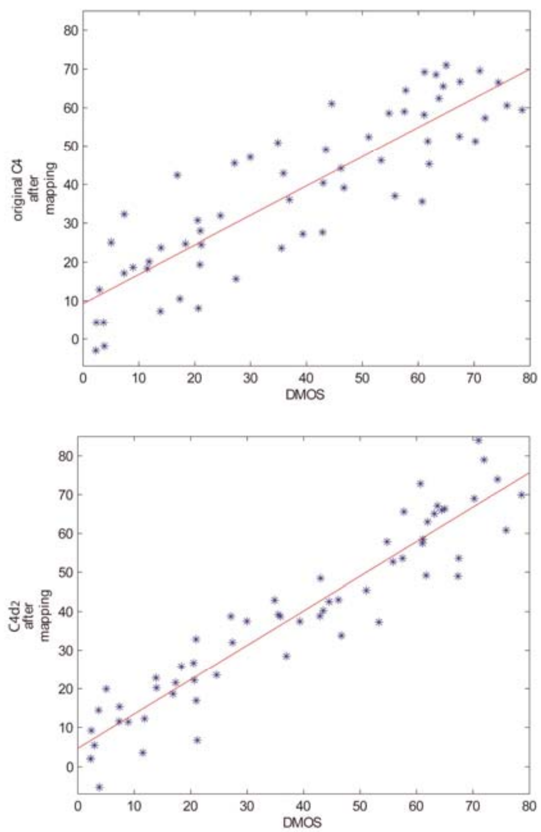

Figure 4: Couples of points (DMOS, Mapped MOS) for JPEG + JPEG2000/C4 with the original method (top) and the proposed C4/ d_2 metric (bottom)

5. CONCLUSIONS

In this paper was emphasized the need of metrics for quality assessment in stereo vision. An extension of the metric presented in [8] which involves the measure of the disparity map distortion was proposed. A significant metric performance enhancement was observed with SSIM metric 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.

6. ACKNOWLEDGMENT

This work was supported by FuturIm@ge project within the “Media and Networks” French cluster.

7. REFERENCES

[1] IEEE Signal Processing Magazine, special section-multiview imaging and 3DTV, 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 Trans. on Circuits and Systems for Video Technology, Volume 17, Issue 11, Nov. 2007 Page(s):1647 - 1658

[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

[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

[5] Lambooij M.T.M.; IJsselsteijn W.A.; Heynderickx I.; “Visual Discomfort in stereoscopic Displays: A Review”, SPIE-IS&T, Volume 6490, 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 Trans. on Circuits and Sys. for Video Techn., Vol. 14, No.3, March 2004 Page(s): 381- 391

[7] Fehn C.;“3D TV Broadcasting”, chapter from 3D Videocommunication, pages: 23-38, Jan. 2006

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

[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.

[13] 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.

[14] 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.