

CYCLOP: A Stereo Color Image Quality Assessment Metric

Aldo Maalouf[†]*, Mohamed-Chaker Larabi[†]

[†] XLIM Laboratory, UMR CNRS 6172,

University of Poitiers, 2 Bd Marie et Pierre Curie, PO Box 30179,

86962 Futuroscope Chasseneuil, France

larabi@sic.univ-poitiers.fr

* Geenov lab, Groupe Glaizer, 43 Rue Pierre Valette, 92240 Malakoff, France

aldo.maalouf@geenov.com

Abstract—In this work, a reduced reference (RR) perceptual quality metric for color stereoscopic images is presented. Given a reference stereo pair of images and their "distorted" version, we first compute the disparity map of both the reference and the distorted stereoscopic images. To this end, we define a method for color image disparity estimation based on the structure tensors properties and eigenvalues/eigenvectors analysis. Then, we compute the cyclopean images of both the reference and the distorted pairs. Thereafter, we apply a multispectral wavelet decomposition to the two cyclopean color images in order to describe the different channels in the human visual system (HVS). Then, contrast sensitivity function (CSF) filtering is performed to obtain the same visual sensitivity information within the original and the distorted cyclopean images. Thereafter, based on the properties of the human visual system (HVS), rational sensitivity thresholding is performed to obtain the sensitivity coefficients of the cyclopean images. Finally, RR stereo color image quality assessment (SCIQA) is performed by comparing the sensitivity coefficients of the cyclopean images and studying the coherence between the disparity maps of the reference and the distorted pairs. Experiments performed on color stereoscopic images indicate that the objective scores obtained by the proposed metric agree well with the subjective assessment scores.

Keywords—Stereo quality assessment, cyclopean image, 2D multi-valued wavelet, disparity field.

I. INTRODUCTION

With the increasing advances in 3D TV and 3D cinema and the increasing interests in enhancing the viewing displays by incorporating multiview/stereo imaging, more and more digital stereo images are being compressed, transmitted, and used in computer systems. Therefore, stereo color image quality assessment (SCIQA) methods are needed to evaluate the quality of stereo images. For SCIQA, objective or subjective methods are used.

The subjective measurement Mean Opinion Score (MOS) is a widely used method for the assessment of image or video quality, but it has several obvious disadvantages. It is very tedious, expensive and impossible to be executed automatically. Instead, an objective image or video quality metric can provide a quality value for a given image or video automatically in a relatively short time. This is very important factor for real time

applications.

Generally, quality metrics can be classified into three categories, including full-reference (FR), reduced reference (RR) and no-reference (NR) according to the availability of the original signal [1] [14].

On the other hand, FR metrics [13] need full information of the original images and demand ideal images as references which can be hardly achieved in practice for some applications (such as broadcasting...). The traditional methods of FR (such as peak signal-to-noise-ratio PSNR) are based on pixel-wise error and have not always been in agreement with perceived quality measurement.

On the other hand, NR metrics aim to evaluate distorted images without any cue from their original ones. However, most of the proposed NR quality metrics are designed for one or a set of predefined specific distortion types and are unlikely to be generalized for evaluating images degraded with other types of distortions.

While RR metrics are between FR and NR, they make use of a part of the information from the original images in order to evaluate the visual perception quality of the distorted ones. As the transformed and stored data are reduced, RR metrics have great potential in image quality assessment (IQA).

Several IQA metrics have been proposed in the literature integrating the properties of the HVS. However, few efforts have been devoted to the quality assessment of stereoscopic images and no efforts, to our knowledge, to the SCIQA.

In [4], recommendations on subjective quality assessment of stereoscopic television pictures are given. In [8], the factors affecting the quality of stereo images are described. These factors include the perception depth, stereoscopic impairments and the visual discomfort. In [5], the authors performed a series of subjective tests to find whether 2D evaluation metrics can be applied for the evaluation of stereoscopic images. They have also proposed a methodology for subjective evaluation of stereoscopic IQA and they concluded after a series of experiments that 2D IQA methods are not suitable for stereoscopic IQA (SIQA). In [2] and [3], Benoit et al. proposed to combine depth information and 2D quality metrics to develop an objective SIQA metric. For that, they compared the disparity maps, to measure the disparity distortion, and the corresponding stereo images to obtain an average measure

This work was supported by the French National Research Agency (ANR) under the QuIAVU project

of the image distortion. Finally, they combined both measures into one objective SIQA.

In [9], Gorley et al. proposed a SIQA metric based on the point matches regions of high spatial frequency between the left and right views of the stereo pair. Their metric accounts for HVS sensitivity to contrast and luminance changes in regions of high spatial frequency. Recently, in [11], Sazzad et al. proposed an objective SIQA based on the fact that perceived distortion and depth of any stereoscopic images are strongly dependent of local features such as textures and edges. Their method is intended for the evaluation of JPEG compressed images and takes into account the blockiness of the images. All the above mentioned approaches do not take into consideration the properties of color images (i.e. multispectral images, multispectral disparity...). To cope this problem, we propose a method for the estimation of the disparity of color images using structure tensors. Then, the cyclopean images of the reference and distorted pairs are obtained and the multispectral wavelet transform proposed in [12] is used to describe the different channels in the HVS. In order to achieve a uniform visual sensitivity to different scales and bands for human perception, each wavelet band is masked with the CSF. For that, we have used the CSF model proposed by Daly [7]. Thereafter, contrast sensitivity and rational thresholding are performed to obtain the sensitivity coefficients. Finally, error pooling is performed by studying the coherence between the disparity maps and comparing the sensitivity coefficients of the cyclopean images.

The rest of the paper is organized as follows: in section 2, the procedure for obtaining color disparity maps is described. In section 3, the procedure for computing the cyclopean images and obtaining the sensitivity coefficients is described. In section 4 we describe our SCIQA metric. Section 5 is devoted to the experimental evaluation and section 6 concludes.

II. COLOR DISPARITY MAP

Color image geometry can be described by using the structure tensor proposed by Di Zenzo in [10]. In fact, Di Zenzo pointed out that the correct method to combine the first order derivative structure is by using a local tensor. Analysis of the shape of the tensor leads to an orientation and a gradient norm estimate. For a multichannel image $I = (I^1, I^2, \dots, I^m)^T$ ($m = 3$ for color images) the structure tensor is given by

$$M = \begin{pmatrix} I_x^T I_x & I_x^T I_y \\ I_y^T I_x & I_y^T I_y \end{pmatrix} \quad (1)$$

The multi-channel structure tensor describes the 2D first order differential structure at a certain point in the image.

In order to describe the disparity between two color images, we first extend the tensor defined in (1) to account for the difference between the two stereo images I_L (the left image) and I_R (the right image). The extended tensor is defined by:

$$S = \begin{pmatrix} \sum_{i=1}^m I_x^2 & \sum_{i=1}^m I_x I_y & \sum_{i=1}^m I_x I_z \\ \sum_{i=1}^m I_x I_y & \sum_{i=1}^m I_y^2 & \sum_{i=1}^m I_y I_z \\ \sum_{i=1}^m I_x I_z & \sum_{i=1}^m I_y I_z & \sum_{i=1}^m I_z^2 \end{pmatrix} \quad (2)$$

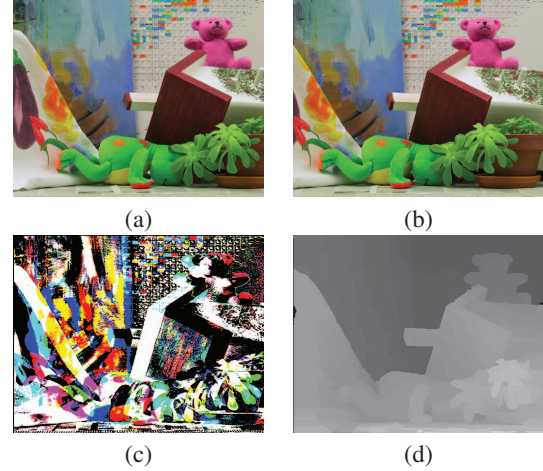


Fig. 1. (a) and (b) "teddy" stereo pair, (c) the residual image and (d) the disparity map

where I_x and I_y are the horizontal and the vertical derivatives, respectively, computed on I_L and I_R , and I_z is defined by: $I_z = I_L - I_R$. The spectral elements of S are for the eigenvalues:

$$\beta_1 = I_x^2 + I_y^2 + I_z^2, \beta_2 = 0 \text{ and } \beta_3 = 0.$$

The corresponding eigenvectors are defined by:

$$\theta^+ = \begin{pmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \\ \frac{\partial I}{\partial z} \end{pmatrix}, \theta^- = \begin{pmatrix} \frac{\partial I}{\partial y} \\ -\frac{\partial I}{\partial x} \\ 0 \end{pmatrix}, \theta^z = \begin{pmatrix} \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \\ \frac{\partial I}{\partial y} \frac{\partial I}{\partial z} \\ -\left(\frac{\partial I^2}{\partial x^2} + \frac{\partial I^2}{\partial y^2}\right) \end{pmatrix} \quad (3)$$

θ^+ is the gradient vector, and the subspace generated by θ^- and θ^z is orthogonal to the gradient ∇I . In fact, θ^- and θ^z are chosen so that $(\theta^+, \theta^-, \theta^z)$ form an orthogonal basis. The disparity tensor can therefore be defined by:

$$F = \theta^+ \theta^{+T} + \theta^- \theta^{-T} + \theta^z \theta^{zT} \quad (4)$$

Clearly, F can be approximated using the largest eigenvector, therefore we let:

$$F = \theta^z \theta^{zT} \quad (5)$$

The disparity map is finally defined by the largest eigenvector ν of F which is computed at each pixel. Figure 1 shows an example of a residual image and the disparity map computed using the disparity tensor F . Once we have obtained the disparity map, we proceed to compute the cyclopean image. This step is described in the next section.

III. CYCLOPEAN IMAGE AND PERCEPTUAL PROCESSING

The ability of the brain to perceive a single image by combining the images from left and right eye, so-called cyclopean image, is a result of combination of locally matched stereo regions in a single global image. To obtain the cyclopean image, we proceed as in [6]. That is, for each pixel, we take a local window from the left view and averaging it with the disparity-matched window from the right view. Figure 2(a) shows an example of a cyclopean image for the *teddy* pair shown in Fig. 1. Once we obtained the cyclopean images for



Fig. 2. (a) Cyclopean image for the "teddy" pair and (b) its color wavelet transform

the reference stereo pair and the distorted one, we proceed to perceptually compare them in order to assess the quality.

To do that, we have first to perform a wavelet decomposition on the two cyclopean images to describe the different channels in the HVS. For that, we have used the color wavelet transform proposed in [12]. This wavelet transform makes use of structure tensors to accumulate color edge information. Therefore, all detail information of the different bands is accumulated into one single representation. The representation allows for a multiscale edge description of color images. Figure 2(b) shows the color wavelet transform of the cyclopean image shown in figure 2(a).

After computing the color wavelet transform, we have to find the sensitivity coefficients of the two cyclopean images, i.e., the most attractive wavelet coefficients as seen by an observer, so that we can measure the perceptual similarity by using few coefficients. We begin by performing contrast sensitivity function (CSF) filtering.

The CSF describes the variations in visual sensitivity as a function of spatial frequency and orientation. In fact, it has been reported that human visual system has different sensitivity to signals with different frequencies. So, to make full use of the wavelet coefficients and orientations in different wavelet scales, we used the CSF to render the sensitivity of human visual system the same to different frequencies. Particularly, we used the model proposed by Daly in [7]. The CSF is applied for all wavelet bands at a given scale j as follows:

$$\hat{W}_j = W_j \cdot N^{CSF} \quad (6)$$

where W is the wavelet coefficient and N^{CSF} is the value of the 2D CSF defined by Daly [7].

After the normalization of the frequencies at different bands and scales, we apply sensitivity thresholding to find the sensitivity coefficients. For that purpose we used the thresholds proposed by Daly [7]. In this model, the visibility threshold elevation $T_j(x, y)$ at site (x, y) and scale j is given by:

$$T_j(x, y) = \left(1 + \left(k_1 \cdot \left(k_2 \cdot |\hat{W}_j(x, y)| \right)^f \right)^b \right)^{\frac{1}{b}} \quad (7)$$

where k_1 and k_2 determine the pivot point of the curve, and the parameter b determines how closely the curve follows the asymptote in the transition region. In the initial work of Daly, a value for the learning slope is chosen depending on the cortex subband. Ideally, this value should depend on the uncertainty of the signal masking. f is the slope of the curve of the

masking contrast.

The value of f is between 0.65 and 1.0, $k_1 = 6^{-7/3}$, $k_2 = 6^{10/3}$ and $b = 4$ [7].

After performing the thresholding, SCIQA is performed by comparing the sensitivity coefficients of both cyclopean images and by studying the coherence of the disparity maps of the reference and distorted pair of images. This step is described in the following section.

IV. CYCLOP: THE PROPOSED METRIC

Our SCIQA metric is based on the fact that any distortion in the received image pair leads to a distortion in disparity map and a perceptual difference between the sensitivity coefficient of the reference cyclopean image and the distorted one. We first define a coherence metric to measure the discrepancy in the disparity maps of the reference and the distorted pair of images, then, we define a perceptual measure to quantify the difference between the two corresponding cyclopean images. Our SCIQA metric will therefore be the sum of two measures. The coherence measure is defined as follows:

Let ξ_r and ξ_t be the largest eigenvalues of equation (5) for the reference and the transmitted image pairs respectively. Any change in the image content between the reference and the transmitted pairs will lead to a change in the disparity direction characterized by the largest eigenvectors ν and consequently, to a change in the values of the eigenvalues. Based on that, we define the coherence metric that measures whether the disparity of the reference pair is coherent (or the same) with the transmitted. The coherence measure is defined by:

$$coh = 1 - \frac{\sum_{H,P} \left(\frac{\xi_r - \xi_t}{\xi_r + \xi_t} \right)^2}{H \times P} \quad (8)$$

where H and P are the height and the width of the image respectively. It is to be noted that, if there is no difference between the two disparities (i.e. the same eigenvalues and eigenvectors) the coherence measure (8) is maximum and consequently, coh tends to 0.

For the perceptual difference measure between the reference and the distorted cyclopean images, we have adopted the normalized absolute difference

$$d_{perceptual} = \frac{1}{R} \sum_{i=1}^R \left(\hat{W}_{j,r} - \hat{W}_{j,t} \right)^2 \quad (9)$$

where R is the total number of the remaining coefficients after applying the sensitivity thresholding, $\hat{W}_{j,r}$ and $\hat{W}_{j,t}$ are the sensitivity coefficients of the reference cyclopean image and the distorted one.

Cyclop the SCIQA metric is finally defined by:

$$D = \frac{coh + d_{perceptual}}{2} \quad (10)$$

V. EXPERIMENTAL RESULTS

We evaluate our metric named Cyclop on the stereo image database used in [11]. The evaluation is performed by comparing the objective quality scores obtained by our SCIQA metric with the subjective Mean Opinion Scores (MOS) provided by

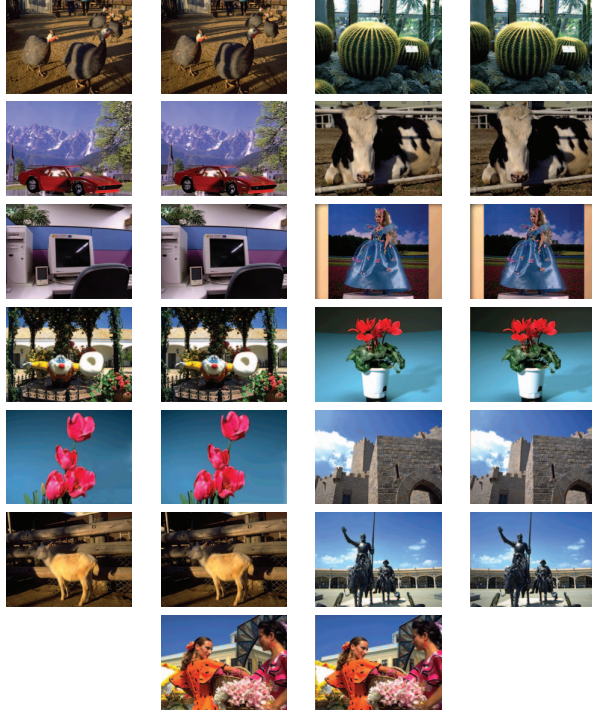


Fig. 3. Stereoscopic Images

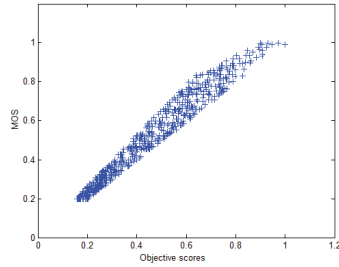


Fig. 4. Proposed SCIQA metric performance, MOS vs Objective scores

the authors of [11] for the images shown in figure 3. These images were JPEG coded with seven quality scales (QS: 10, 15, 27, 37, 55, 79 and 100) resulting in 637 stereoscopic symmetric (same bitrate for the left and right images) and asymmetric (different bitrates for the left and right images) coded image pairs in the database. The MOS vs the objective scores obtained by Cyclop are shown in figure 4.

We evaluate our SCIQA metric by computing the three measurements recommended by VQEG [1] to test the consistency of our Cyclop and the subjective perception, namely, the correlation coefficients (CC), the rank order correlation coefficient (ROCC) and the outlier ratio (OR). We also compared our results with others obtained by the NR metric proposed in [11]. The results are shown in Table 1 for both metrics. From these results, we can see that our metric has a high prediction accuracy (larger CC), stricter monotonicity (larger ROCC) and better stability (smaller OR).

TABLE I
OBJECTIVE ASSESSMENT

Model	CC	ROCC	OR
Cyclop	0.981	0.950	0.050
Metric [11]	0.96	0.920	0.069

VI. CONCLUSION

In this paper, we presented a new stereoscopic color image quality assessment metric. The proposed SCIQA metric makes use of a color disparity tensor to estimate the disparity map of color stereo images. The properties of the HVS are integrated in the proposed metric by performing CSF filtering and rational thresholding. The SCIQA is performed by computing the coherence of the disparities and the perceptual difference between the reference image pair and the distorted one. The experimental results showed that the proposed SCIQA metric performs well and has good consistency with subjective quality assessment.

For future work, we are extending Cyclop to measure the quality of stereoscopic video sequences.

REFERENCES

- [1] Vqeg: The video quality experts group, <http://www.vqeg.org>.
- [2] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, *Using disparity for quality assessment of stereoscopic images*, International conference on Image processing ICIP (2008).
- [3] Alexandre Benoit, Patrick Le Callet, Patrizio Campisi, and Romain Cousseau, *Quality assessment of stereoscopic images*, EURASIP Journal on Image and Video Processing, special issue on 3D Image and Video Processing **2008** (2008).
- [4] ITU-R BT.1438, *Subjective assessment of stereoscopic television pictures*.
- [5] Patrizio Campisi, Patrick Le Callet, and Enrico Marini, *Stereoscopic images quality assessment*, Proceedings of the 15th European Signal Processing Conference, Eurasp EUSIPCO 2007, Poznan, Poland (2007).
- [6] HENKEL R. D., *Fast stereovision by coherence detection*, Lecture notes in computer science, Springer **1296/1997** (1997), 297–304.
- [7] S. Daly, *The visible difference predictor: an algorithm for the assessment of image fidelity*, Proc. of SPIE **1666** (1992), 2–15.
- [8] Lambooi Marc, T. M. IJsselstein, Wijnand A., and Heynderickx I., *Visual discomfort in stereoscopic displays: a review*, SPIE electronic imaging, stereoscopic displays and virtual reality systems XIV **6490** (2007).
- [9] GORLEY Paul and HOLLIMAN Nick, *Stereoscopic image quality metrics and compression*, Proceedings of SPIE, the International Society for Optical Engineering (2008).
- [10] DiZeno S., *A note on the gradient of multi images*, Computer Vision Graphics and Image processing **33** (1986), no. 1, 116–125.
- [11] Z.M. Sazzad, S. Yamanaka, Y. Kawayoke, and Y. Horita, *Stereoscopic image quality prediction*, First International Workshop on Quality of Multimedia Experience QOMEX, San Diego, California, U.S.A. (2009).
- [12] Paul Scheunders, *An orthogonal wavelet representation of multivalued images*, IEEE Transactions on Image Processing **12** (2003), no. 6, 718–725.
- [13] H.R. Sheikh, A.C. Bovik, and G. de Veciana, *An information fidelity criterion for image quality assessment using natural scenes statistics*, IEEE Trans. Image Process. **14** (2005), 2117–2128.
- [14] Z. Wang, H. R. Sheikh, and A. C. Bovik, *Objective video quality assessment*, In The Handbook of Video Databases: Design and Applications (2003), 1041–1078.