Binocular Combination and Fractional Differential Based 3D Image Quality Assessment

Haiyong Xu1, Gangyi Jiang1, Mei Yu1
1. Faculty of Information Science and Engineering
Ningbo University
Ningbo, 315211, China

Yun Zhang2, Ting Luo1, Zongju Peng1, Feng Shao1
2. Shenzhen Institutes of Advanced Technology
Chinese Academy of Sciences
Shenzhen, 518055, China

Abstract—Three dimensional image quality assessment (3DIQA) is an important issue for three-dimensional video technologies and faces more challenges than its two-dimensional counterparts. In this paper, a novel 3DIQA method is proposed based on human visual system considering the influence of binocular combination between the left and right views of stereo image. To be more specific, difference-of-Gaussian response is defined and used to construct cyclopean images of stereo images. Also, fractional differential masks are defined instead of the gradient operator to improve image local feature. Then, a fractional differential-based feature similarity metric is used to derive a single quality score. Experimental results demonstrate that the proposed method achieves better consistency with subjective quality assessment compared with other representative methods.

Keywords—three dimensional image quality assessment; cyclopean image; difference-of-Gaussian response; fractional differential mask

I. INTRODUCTION

With the rapid development of three-dimensional (3D) movies and display devices, 3D videos and images are becoming one of the most popular types of multimedia in our daily life. Since the limits on bandwidth and the physical properties of devices, the output 3D videos and images may be unsatisfactory. Thus, image quality assessment (IQA) plays a key role for 3D multimedia applications [1].

Similar to two-dimensional IQA (2DIQA), 3DIQA can be divided into two classes of subjective and objective quality assessments. Since human is the final receiver of visual information, subjective IQA [2, 3] is the best approach to percept the effect of distortions. However, subjective IQA is time-consuming, expensive, complexity and difficult to be used for real-time systems. Thus, objective 3DIQA has been extensively conducted for decades. Although various 2DIQA models, such as mean-squared-errors (MSE), peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [4], visual information fidelity (VIF) [5], feature similarity (FSIM) [6], etc., had been proposed, 3DIQA is still a problem to be widely researched. Compared with 2DIQA, 3DIQA involves not only evaluating two images’ (left and right views) qualities, but also depth perception quality. In [7], 2DIQA models were used to evaluate qualities of the left and right views independently, then calculated the mean of the two views’ qualities as stereo image quality score. In [8], color plus depth image was evaluated by using 2DIQA metrics after being rendered to 3D images. In addition, in [9], 2D image quality measurement and depth maps were used to evaluate 3D video coding quality. However, human binocular perception is not sufficiently considered in these IQA methods. In fact, the quality of 3D image is not a simple combination of the qualities of its two views.

To consider this problem, some 3DIQA methods have been proposed by evaluating the quality of two images and display. Lambooij et al. [10] presented a 3DIQA model with a higher level evaluation concept as a weighted sum of perceived image quality and perceived depth. Jin et al. [11] proposed a stereo image quality metric based on 3D digital cosine transform, contrast sensitive function (CSF) and luminance marking. Maalouf et al. [12] computed the cyclopean images from the left and right views of stereo image and used CSF to measure quality of the cyclopean image. Chen et al. [13] used a linear model of binocular rivalry to construct the cyclopean image from the left and right views. Shao et al. [14] proposed a stereo image quality assessment by considering binocular perception property of different image region. Lin et al. [15] proposed a frequency integrated metric for stereo image.

In this paper, stereo image, as one kind of 3D representation, is researched, we propose a new 3DIQA method based on a gain-control theory model of binocular combination and fractional differential. Experimental results show that the proposed 3DIQA method outperforms other methods and can predict human visual perception of stereo image more accurately.

II. RELATED WORK

A. Gain-Control Theory of Binocular Combination

In the human vision system, cyclopean perception means that a single combined image is perceived when different images are presented to the left and right eyes. Let \( I_L \) and \( I_R \) be the left and right views of stereo images. A single perceived cyclopean image \( I_C \) can be described as a binocular combination functional, \( \Gamma \), of two views \( I_L \) and \( I_R \). i.e.

\[
I_C = \Gamma(I_L, I_R)
\] (1)
A Gain-Control model in [16] was proposed to describe an early stage of binocular combination

\[ \Gamma(I_L, I_R) = \frac{1}{1 + \varepsilon_L(I_R)\varepsilon_R(I_L)} I_L + \frac{1}{1 + \varepsilon_R(I_R)\varepsilon_L(I_L)} I_R \]  

(2)

where \( \varepsilon_L(I_L) \) and \( \varepsilon_R(I_R) \) are the sums of the total visually weighted contrast energies for gain control of the left-eye and right-eye, respectively.

B. Feature Similarity (FSIM)

Based on the fact that human visual system (HVS) understands an image mainly according to its low-level features, FSIM was proposed with phase congruency (PC) and gradient magnitude (GM). PC of image is used as the primary feature in FSIM, and GM of image is computed as the secondary feature to encode contrast information. FSIM model can be expressed as

\[ FSIM = \frac{\sum_{(x,y)\in \Omega} S_{PC}(x,y) S_{GM}(x,y) P_{CM}(x,y)}{\sum_{(x,y)\in \Omega} P_{CM}(x,y)} \]  

(3)

where \( S_{PC} \) and \( S_{GM} \) denote the similarities in image’s PC and GM, respectively, and \( P_{CM} \) denotes the maximum PC between the distorted and reference images.

In this paper, the proposed 3DIQA method aims to study gain-control theory of binocular combination and FSIM model to achieve comparable performance.

III. THE PROPOSED 3DIQA METHOD

A framework of the proposed 3DIQA method with stereo image as a 3D image representation is shown in Fig. 1. In the framework, cyclopean image is synthesized from stereo image pair, and disparity map and difference-of-Gaussian (DOG) responses are estimated, respectively. Finally, the fractional differential-based feature similarity metric is proposed to predict 3D image quality scores between the distorted cyclopean image and reference cyclopean image.

Here, the disparity map between the left and right views is first estimated to obtain depth information. The disparity map is computed by using a segment-based stereo matching method [17].

With the disparity map, Eq. (2) is rewritten by

\[ \Gamma(I_L, I_R) = E_L(x,y)I_L(x,y) + E_R(x+d,y)I_R(x+d,y) \]  

(4)

where \( d \) denotes disparity value, and \( E_L \) and \( E_R \) denote gain weights for left and right views, which can be calculated by

\[ E_L(x,y) = \frac{1 + \varepsilon_L(x,y)}{1 + \varepsilon_R(x,y) + \varepsilon_L(x+d,y)} \]  

(5)

and

\[ E_R(x+d,y) = \frac{1 + \varepsilon_R(x+d,y)}{1 + \varepsilon_R(x,y) + \varepsilon_L(x+d,y)} \]  

(6)

To construct a cyclopean image of the distorted or reference stereo image, we use DOG responses on their left and right views to compute \( \varepsilon_L(I_L) \) and \( \varepsilon_R(I_R) \) in Eqs. (5) and (6) over all the frequency channels. DOG decomposition \( I_{DOG} \) of an image \( I \) is defined by

\[ I_{DOG} = (G(k\sigma) - G(\sigma)) * I \]  

(7)

where \( G(\cdot) \) is the Gaussian kernel function with standard deviation \( \sigma \), and \( k \) is a constant. Actually, \( I_{DOG} \) in Eq. (7) represents an image \( I \) convoluted to the difference of two Gaussian kernels.

Let \( D(I) \) be the vector which represents different spatial frequency band components of the image \( I \), that is,

\[ D(I) = (D_0, D_1, \ldots, D_{n-1}, D_n) \]  

(8)

where

\[ D_0 = G(\sigma) * I - I \]  

(9)

and

\[ D_i = (G(k^{i}\sigma) - G(k^{i-1}\sigma)) * I, i = 1, \ldots, n \]  

(10)

Following Eq. (8), the reference left image \( I_L^R \) and right image \( I_R^R \) can respectively be represented by

\[ D(I_L^R) = (D_0^L, D_1^L, \ldots, D_{n-1}^L, D_n^L) \]  

(11)

and

\[ D(I_R^R) = (D_0^R, D_1^R, \ldots, D_{n-1}^R, D_n^R) \]  

(12)

Using Eqs. (11) and (12), \( \varepsilon_L(I_L^R) \) and \( \varepsilon_R(I_R^R) \) can be computed as

\[ \varepsilon_L(I_L^R) = \sum_{i=0}^{n} (D_i^L)^2 \]  

(13)

and

\[ \varepsilon_R(I_R^R) = \sum_{i=0}^{n} (D_i^R)^2 \]  

(14)

Then, using the Gain-Control theory of binocular combination given in Eq. (4), a single cyclopean image \( I_{Cyc} \) of the reference image can be computed. Similarly, a cyclopean image \( I_{Cyc}^D \) of the distorted image is obtained.

In IQA methods, integral differential-based algorithms [6, 18], such as gradient operators, have been widely applied. However, images often exhibit many features including smooth area of an image, texture, edges and noise. Therefore, texture detail of smooth area will have substantially linear attenuation if traditional operators are used, such as Sobel
operator and Laplace operator. It verified that integer order differential-based methods cannot well detect texture details in image smooth regions. In contrast, fractional differential-based method can be used to linearly preserve low frequency of signal. Here, fractional differential masks are defined and used instead of the image gradient operator in FSIM.

As the generalization of the integer-order derivative, the fractional-order differential has long history but until now it has no uniform definition. The commonly used definitions of the fractional-order differential are Gr"{u}mwald-Letnikov definition and Riemann-Liouville definition. Here, $\nu$-order Gr"{u}mwald-Letnikov based fractional differential is used and it is defined as

$$D_{G-L}^{\nu} s(x) = \lim_{N \to \infty} \left( \frac{x-a}{N} \right)^{\nu-1} \Gamma(\nu) \sum_{k=0}^{N-1} \Gamma(1-k) \times s(x-k \frac{x-a}{N}),$$

(15)

Where $\nu$ denotes a real number (including fraction), the range of the signal $s(x) = [a, x]$, $s(x-k \frac{x-a}{N})$ is the discrete sampling, and $\Gamma(\cdot)$ is the Gamma function.

Based on the above definition, fractional differential masks [19] can be constructed. When $N$ is big enough, the limits symbol can be removed and the numerical algorithm of fractional differential is obtained. For 2D image $I(x, y)$, fractional differential can be defined as follows

$$\frac{\partial^\nu I(x, y)}{\partial x^\nu} \cong I(x, y) + (-\nu) I(x - 1, y) + \frac{(-\nu)(-\nu+1)}{2} I(x - 2, y) + \cdots
$$

$$+ \frac{\Gamma(n - \nu - 1)}{(n-1)! \Gamma(-\nu)} I(x - n + 1, y)$$

(16)

Here, $n$ is empirically set as 3. Similarly, $\frac{\partial^\nu I(x, y)}{\partial y^\nu}$ can be obtained.

Here, fractional differential masks on eight symmetric directions are defined as Fig. 2, which are eight $3 \times 3$ fractional differential masks and denoted by $M_1$, $M_2$, $M_3$, $M_4$, $M_5$, $M_6$, $M_7$ and $M_8$. Thus, fractional partial derivative $W_i$ of the image $I(x, y)$ along eight directions is defined as follows

$$W_i = M_i \ast I(x, y)$$

(17)

Fractional differential magnitude (FDM) of $I(x, y)$ is obtained by

$$F = \frac{1}{8 - 12\nu + 4\nu^2} \sum_{i=1}^{8} W_i$$

(18)

Then, FDMs $F_1^d$ and $F_v^v$ between the distorted cyclopean image $I_1^d$ and reference cyclopean image $I_2^c$ are compared, and the similarity measure is defined as follows

$$S_F = \frac{2F_1^d \cdot F_v^v + C}{(F_1^d)^2 + (F_v^v)^2 + C}$$

(19)

Finally, $S_F$ is used instead of the term $S_G$ in Eq. (3), and accordingly, the fractional differential-based feature similarity metric (FDFSM) is defined by

$$FDFSM = \frac{\sum_{(x,y) \in \Omega} S_{PC}(x,y) \cdot S_F(x,y) \cdot PC_m(x,y)}{\sum_{(x,y) \in \Omega} PC_m(x,y)}$$

(20)

### IV. EXPERIMENTAL RESULTS

To verify the performance of the proposed 3DIQA method, the LIVE 3D Image Database phase I [20] is used, which includes 20 reference stereo images and 365 symmetric distorted stereo images spanning five different distortion categories: JPEG 2000 (JP2K) and the JPEG compression standards, additive white Gaussian noise (WN), Gaussian blur (Blur) and a fast-fading (FF) model based on the Rayleigh fading channel. Each distorted stereo image is associated with a difference mean opinion scores (DMOS). Smaller DMOS value implies higher subjective image quality.

In the experiments, three indexes that measure the consistency between the results of the proposed method and DMOS are used: the Spearman rank order correlation coefficient (SROCC) and the Pearson linear correlation coefficient (PLCC), which measures the prediction monotonicity; the root mean squared error (RMSE), which measures the prediction accuracy. A perfect matching between the objective and subjective scores will give $SROCC = PLCC = 1$, and $RMSE = 0$. The proposed method will be compared with a 2DIQA metric (MS-SSIM [21]) and four 3DIQA metrics (SSIM [4] combining the cyclopean image, FSIM [6] combining the cyclopean image, Chen’s method [13] and Shao’s method [14]). For the
nonlinear regression, the 4-parameter logistic function is defined as follows:

\[
DMOS_P = \frac{\beta_1 - \beta_2}{1 + \exp \left( -\frac{x - \beta_3}{\beta_4} \right)} + \beta_2
\]

where \(\beta_1, \beta_2, \beta_3\) and \(\beta_4\) are parameters of the regression model.

Fig. 3. Scatter plot of objective scores versus subjective scores for the proposed 3DIQA method.

Fig. 3 shows the scatter plot for the proposed method. The horizontal axis represents the predicted DMOS value \((DMOS_P)\) and the vertical axis represents the subjective ratings of the perceived distortions. The values of SROCC, PLCC and RMSE of each distortion type with the database are list in Table I. The best results across the six 3DIQA methods for each distortion type are highlighted in boldface. From Table I, the best overall performance is conducted by the proposed method with 0.9336 PLCC value, 0.9322 SROCC value and 5.8754 RMSE value. For JP2K distortion, the proposed method and “cyclopean image + FSIM” perform better than the other IQA methods. For JPEG, WN and FF distortion, the proposed method is comparable in performance to Shao’s method. Meanwhile, the proposed method performs better than the other IQA methods except Shao’s method on blur distortion.

V. CONCLUSIONS

In this paper, we have proposed a novel method for objective stereo image quality assessment. First, we constructed the cyclopean image, which is based on the Gain-Control theory of binocular combination considering the disparity information and difference-of-Gaussian responses between the left image and right image. Furthermore, we use the fractional difference mask to construct the image quality metric. Moreover, experimental results of the propose method has good consistency with subjective quality assessment. In the future, we will extend the proposed method to measure the quality assessment of stereoscopic video sequence.

ACKNOWLEDGMENT

This work was supported by Natural Science Foundation of China (U1301257, 61171163, 61271270, 61271021, 61311140261 and 61471348).

### TABLE I. PERFORMANCE COMPARISON AMONG FIVE SCHEMES ON PHASE I DATABASE (CASES IN BOLD DENOTE BEST PERFORMANCE)

<table>
<thead>
<tr>
<th>Distortion</th>
<th>Criteria</th>
<th>Metrics</th>
<th>MS-SSIM</th>
<th>Cyclopean image + SSIM</th>
<th>Cyclopean image + FSIM</th>
<th>Chen’s method</th>
<th>Shao’s method</th>
<th>The proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>JP2K</td>
<td>PLCC</td>
<td>0.9191</td>
<td>0.8547</td>
<td>0.9205</td>
<td>0.9006</td>
<td>0.9114</td>
<td>0.9300</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SROCC</td>
<td>0.8932</td>
<td>0.8525</td>
<td>0.9010</td>
<td>0.8649</td>
<td>0.8619</td>
<td>0.8987</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>5.1032</td>
<td>6.7237</td>
<td>5.0605</td>
<td>5.6296</td>
<td>5.3305</td>
<td>4.7592</td>
<td></td>
</tr>
<tr>
<td>JPEG</td>
<td>PLCC</td>
<td>0.6794</td>
<td>0.4293</td>
<td>0.6136</td>
<td>0.5192</td>
<td>0.6623</td>
<td>0.6806</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SROCC</td>
<td>0.6106</td>
<td>0.3660</td>
<td>0.5210</td>
<td>0.4400</td>
<td>0.6148</td>
<td>0.5841</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>4.7979</td>
<td>5.9060</td>
<td>5.1648</td>
<td>5.5886</td>
<td>4.8992</td>
<td>4.7908</td>
<td></td>
</tr>
<tr>
<td>WN</td>
<td>PLCC</td>
<td>0.9385</td>
<td>0.9400</td>
<td>0.9328</td>
<td>0.9289</td>
<td>0.9409</td>
<td>0.9385</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SROCC</td>
<td>0.9421</td>
<td>0.9438</td>
<td>0.9339</td>
<td>0.9365</td>
<td>0.9431</td>
<td>0.9450</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>5.7439</td>
<td>5.6760</td>
<td>5.9962</td>
<td>6.1599</td>
<td>5.6328</td>
<td>5.7449</td>
<td></td>
</tr>
<tr>
<td>Blur</td>
<td>PLCC</td>
<td>0.9438</td>
<td>0.9166</td>
<td>0.9446</td>
<td>0.9387</td>
<td>0.9512</td>
<td>0.9494</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SROCC</td>
<td>0.9261</td>
<td>0.8842</td>
<td>0.9249</td>
<td>0.9240</td>
<td>0.9375</td>
<td>0.9200</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>4.7823</td>
<td>5.7857</td>
<td>4.7510</td>
<td>4.9900</td>
<td>4.4668</td>
<td>4.5457</td>
<td></td>
</tr>
<tr>
<td>FF</td>
<td>PLCC</td>
<td>0.8056</td>
<td>0.7075</td>
<td>0.8186</td>
<td>0.7994</td>
<td>0.8473</td>
<td>0.8438</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SROCC</td>
<td>0.7341</td>
<td>0.5885</td>
<td>0.7394</td>
<td>0.7578</td>
<td>0.7915</td>
<td>0.8049</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>7.3626</td>
<td>8.7819</td>
<td>7.1375</td>
<td>7.4659</td>
<td>6.5999</td>
<td>6.6688</td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>PLCC</td>
<td>0.9258</td>
<td>0.8868</td>
<td>0.9273</td>
<td>0.9180</td>
<td>0.9231</td>
<td>0.9336</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SROCC</td>
<td>0.9240</td>
<td>0.8724</td>
<td>0.9225</td>
<td>0.9090</td>
<td>0.9200</td>
<td>0.9322</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


