Contents lists available at ScienceDirect





journal homepage: www.elsevier.com/locate/neucom

Natural image statistics based 3D reduced reference image quality assessment in contourlet domain



Xu Wang^a, Qiong Liu^{b,*}, Ran Wang^{a,c}, Zhuo Chen^d

^a Department of Computer Science, City University of Hong Kong, Hong Kong

^b Department of Electronics & Information Engineering, Huazhong University of Science & Technology, Wuhan 430074, China

^c Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518067, China

^d National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China

ARTICLE INFO

Article history: Received 13 November 2013 Received in revised form 26 May 2014 Accepted 26 May 2014 Available online 4 November 2014

Keywords: 3D image Reduced reference image quality assessment Contourlet transform Natural image statistics Feature similarity

ABSTRACT

Designing a reliable and generic perceptual quality metric is a challenging issue in three-dimensional (3D) visual signal processing. In many practical 3D application scenarios, the original stereoscopic images (i.e., the reference image) cannot be fully accessed, leaving difficulties to quality assessment. To handle this problem, 3D reduced-reference image quality assessment (RRIQA) metrics have been investigated, which only extract small amount of information from the original stereoscopic images. In this paper, we propose a novel 3D RRIQA metric based on 3D natural image statistics in contourlet domain. In this metric, the Gaussian scale mixtures (GSM) model is employed to normalize the coefficients in the contourlet subband of luminance image and disparity map of the 3D images. After divisive normalization transform, we find that the marginal distribution of the coefficients is approximately Gaussian distributed. Based on these investigations, the standard derivations of the fitted Gaussian distribution are determined as the feature parameters in our metric for each contourlet subband. Then, the feature similarity index is employed to measure the 3D visual quality at the receiver side without accessing the reference stereoscopic images. Experimental results demonstrate that the proposed metric has good consistency with 3D subjective perception of human, and can be implemented in many end-to-end 3D video systems.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

With the rapid development of content generation and display technology, three-dimensional (3D) applications and services [1] become common in nearly every facet of people's daily life. For example, people can go to 3D movies for immersive viewing experiences, and play 3D video games for attractive scenarios. With the help of 3D display, doctors can accomplish complicated diagnose and surgeries. For these applications, the quality of 3D content is the most important part to guarantee the viewing experiences. However, in the 3D processing chain including capturing, processing, coding, transmitting, reconstruction [2], retrieving [3], etc., artifacts are inevitable due to the resource shortage in processing [4,5]. Compared to the conventional 2D image, the artifacts of 3D image have more effects on human visual system (HVS) [6–8]. Therefore, how to design a reliable and generic perceptual quality metric becomes a challenging issue in 3D visual signal processing.

* Corresponding author. Tel.: +86 2787542667.

E-mail addresses: xuwang6-c@my.cityu.edu.hk (X. Wang),

q.liu@hust.edu.cn (Q. Liu), ranwang3-c@my.cityu.edu.hk (R. Wang), chenzhuo.zoom@gmail.com (Z. Chen).

http://dx.doi.org/10.1016/j.neucom.2014.05.090 0925-2312/© 2014 Elsevier B.V. All rights reserved. The goal of 3D IQA metric is to automatically assess the quality of 3D images or videos in agreement with human quality judgments, and then the assessment outcome can be feedback to optimize the 3D image or video systems [9]. The quality of 3D content [10] contains several aspects, such as perceived image quality, depth perception and visual fatigue. In this work, we focus on the metric of perceived image quality. Considering whether the original stereoscopic references are available, the perceptual quality metrics for 3D image can be classified into three categories:

- *Full reference (FR) IQA*: The original stereoscopic images are fully available. This kind of metrics have been widely investigated [11,12], which can achieve the best correlation between predictive quality and subjective perceptual quality.
- No reference (NR) IQA: The original stereoscopic images are totally inaccessible. To measure the visual quality of distorted images, natural image statistic features and learning based methods [13,14] are widely used in designing a high performance NR IQA metric. However, due to the lack of knowledge on 3D perception of HVS, there is still a long way to improve performance of NR metrics.
- Reduced reference (RR) IQA: Compared to the FR metrics and NR metrics, RR metrics can achieve a trade-off on performance and

amount of information, where the original stereoscopic images are partial accessible. The features that can reflect the intrinsic quality characteristics are extracted from the original stereoscopic images and stored as auxiliary information [15,16]. After receiving the auxiliary information, the RR metrics evaluate the perceptual quality of the distorted stereoscopic images by incorporating the corresponding extracted features.

There are two main research directions for 3D RRIQA metrics. One is image feature description based metrics. The other direction is information theory based metrics. The image features [17] include distortion driven features such as binary edge mask [18,19], and HVS features such as contrast sensitivity index [20]. Due to the lack of knowledge in binocular vision, the performances of feature description based metrics are usually unsatisfactory. Thus, we only focus on information theory based metrics. For the 2D image, Wang et al. [21] computed the relative entropy between the probability distributions of the reference and distorted images based on the generalized Gaussian distribution (GGD). To improve the performance and reduce the number of features, the probability distribution was represented by the Gaussian scale mixture (GSM) model in wavelet domain [22] and contourlet domain [23]. Soundararajan et al. [24] computed the differences of entropies of the reference and distorted images by the GSM model. However, all the abovementioned works were based on the 2D natural images statistical modeling, which cannot exploit the 3D natural image statistical characteristics.

Recent works on 3D natural images statistical modeling have shown their power on 3D image application. For example, the marginal distributions of disparity [25] subband coefficients can be well fitted by the GGD model, which is similar to the luminance images. This property was employed as prior information to improve the performance of Bayesian disparity estimation [26]. Motivated by these works, we propose a novel 3D RRIQA metric based on 3D natural image statistics in contourlet domain in this paper. Compared with the state-of-the-art works, the main contributions of our algorithm are listed as follows:

- The relationship between 3D natural images statistical information (including luminance image and disparity map) degradation and 3D perceptual quality is first investigated in contourlet domain, which is useful in predicting the subjective perceptual quality of 3D images.
- The number of features is reduced where each subband only needs one feature for the luminance image/disparity map of 3D images. In other words, our metric just need a little auxiliary information. This is convenient for practical environments, and is potential for future 3D applications.

The rest of this paper is organized as follows. Section 2 overviews the related works. In Section 3, the proposed metric is described in detail. Section 4 introduces the subjective image database used for evaluating the metric and provides experimental results for performance analysis. Finally, conclusions are given in Section 5.

2. Related works

For 3D images, the ultimate optimization criterion is the subjective perceptual quality. However, subjective quality assessment is time-consuming. Thus, designing an effective objective 3D IQA metric is very important. Existing perceptual quality metrics for 3D image applications can be divided into two categories, named 2DIQA extension model and 3DIQA model.

The first category extends the 2DIQA models directly to analyze the quality degradation of stereoscopic image. For example, Hewage et al. [27] investigated the correlation between subjective quality scores and three quality metrics, including Peak Signal-to-Noise Ratio (PSNR), Video Quality Model (VQM) [28], and Structural Similarity Model (SSIM) [29] for the 3D video content. The simulation results demonstrated that the VQM metric is better than the other two metrics for predicting the overall perceptual quality of 3D content. Similar work has also been done in [30]. However, this kind of metrics does not consider the binocular vision properties, thus lead to unsatisfactory performance.

The other category exploited the HVS properties such as binocular vision and depth perception to improve the performance of IOA metrics. For example, Benoit et al. [31] used the fusion of the depth (or disparity) information and 2D guality metrics to analyze 3D visual quality. The integration of disparity information into quality assessment was also fully investigated in [32]. Boev et al. [33] combined the monoscopic quality component and the stereoscopic quality component for developing a stereo-video quality metric. The cyclopean image concept was first introduced for fusing the left and right views. To further improve the performance of 3DIQA, the binocular fusion and rivalry properties are widely investigated. For example, Wang et al. [34] proposed a binocular spatial sensitivity (BSS) weighted metric based on the binocular just noticeable difference model [35]. Chen et al. [36] developed a framework for assessing the quality of stereoscopic images that have been afflicted by possibly asymmetric distortions. In [37], the linear rivalry model was employed in the metrics to exploit the binocular rivalry property. However, the features employed in the binocular vision properties based metric are local, which may not work well when the original 3D images are inaccessible. As we know, the natural image statistics can be used to extract the global feature to reflect the quality degradation of image. Therefore, in order to design a high performance 3D RRIQA metric, it is necessary to investigate the global features based on 3D natural image statistics.

3. Proposed metric

The framework of our proposed 3D RRIQA metric is described in Fig. 1. At the sender side, the original stereoscopic images and the corresponding disparity map are decomposed to subbands with different scales and directions by contourlet transform. Each subband is first modeled by GSM, and then processed by divisive normalization transform. The statistical features are extracted at the sender side and then sent to the receiver side as the auxiliary information. At the receiver side, the distorted stereoscopic images and their corresponding disparity map are processed by the same procedure at the sender side for the statistical features. Finally, the quality of distorted images can be measured by the feature similarity index.

3.1. 3D natural images statistics

3.1.1. Contourlet transform

The contourlet transform has the features of multi-resolution representation, localized analysis and direction-sensitivity, which has been proved be efficient for computational image representation [38]. Comparing to the wavelet transform, the contourlet transform can provide a much richer set of direction and shape basis. In other words, contourlet transform is specialized in capturing smooth contours and geometric structures in images. From the knowledge of HVS, the characteristics of the receptive fields in the visual cortex are also localized, oriented and bandpass. Therefore, the features in the contourlet domain can be employed to reflect the features of visual perception. As shown in Fig. 2, majority of pixels in the subband are in black, indicating that the non-zero coefficients in contourlet domain are sparse, which reveals the features of visual perception.



Fig. 1. The framework of the proposed 3D RRIQA metric.



Fig. 2. Example of the contourlet transform on the left view of 3D images. The image is decomposed into two pyramidal levels, each level is decomposed into four directional subbands. Small coefficients are shown in black while large coefficients are shown in white. (a) Original luminance image and (b) contourlet decomposition.

3.1.2. 3D natural images statistics in contourlet domain

For convenience, the coefficient relationship in the contourlet subband is defined in Fig. 3. For each contourlet coefficient X (the pixel in red), its neighbors (denoted as NX) are defined as the adjacent coefficients in the same subband. The coefficient in the same spatial location of the coarse scale is defined as its parent (denoted as PX), and those coefficients in the same spatial location of the finer scale are its children. The cousins (denoted as CX) of X are defined as the coefficients with the same scale and spatial location but different directions. The generalized neighborhood of coefficients X (denoted as GX) is finally defined as the set of its parent (PX), neighbors (NX) and cousins (CX).

The subband marginal distributions of natural images in the contourlet domain are highly non-Gaussian. For example, the distribution of coefficients in contourlet domain for 3D images and disparity map are shown in Fig. 4. They are with sharp peaks at zero amplitude and heavy tails on both sides of the peak. This phenomenon indicates that the majority of coefficients are close to zero. The kurtosis of the distributions is much higher than the kurtosis of 3 for Gaussian distributions. Therefore, using Gaussian distribution to model the coefficient distribution may result in large fitting errors. However, the conditional Gaussian distribution holds for any linear combination of the magnitudes of the generalized neighborhood [39].



Fig. 3. Contourlet coefficients relationships. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)



Fig. 4. Marginal statistics of four contourlet subbands of 3D images. X-axis represents the coefficient amplitude and Y-axis represents the probability density. (a) Luminance images: the kurtosis of the four distributions are 17.75, 6.53, 6.16 and 9.59; (b) disparity map: the kurtosis of the four distributions are 8.70, 7.47, 7.95 and 8.71.

3.1.3. 3D natural images statistics in divisive normalization transform domain

Divisive normalization is a simple nonlinear efficient coding transform that can be used to reduce the statistical dependencies of coefficients in contourlet domain for natural images. The transformed coefficients are approximately Gaussian distributed [22]. To normalize the output coefficients, each coefficient is divided by the energy of its neighborhood coefficients. For example, for the coefficient *X* as shown in Fig. 3, the normalized coefficient \tilde{X} can be computed as $\tilde{X} = X/p$, where the divisive normalization factor *p* is a positive constant derived from the generalized neighborhood coefficients *GX*.

GSM model is an effective local statistical image model for computing the normalization factor *p*. For the natural images, the

GSM model [40] can accurately model the marginal and joint distributions of the coefficients in wavelet transform domain. As an extension of wavelet transform, the statistical distribution of coefficients in contourlet domain can also be characterized by GSM model. Based on the definition of GSM [40], a length-N random vector Y is a GSM if Y=zU, where $z \ge 0$ is a scalar random. U is a zero-mean Gaussian random vector with covariance Q, z and U are independent.

Suppose that the probability density of the mixing density *z* is $\Phi_z(z)$, then the density of *Y* is written as

$$p_{Y}(Y) = \int_{-\infty}^{\infty} p_{Y}(Y|z) \Phi_{z}(z) dz$$

=
$$\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{N/2} |z^{2}Q|^{1/2}} \exp\left(-\frac{Y^{T}Q^{-1}Y}{2z^{2}}\right) \Phi_{z}(z) dz$$
(1)

In this paper, the vector *Y* is formed by clustering a set of generalized neighborhoods of coefficient *X*. For each location of coefficient *X*, the value of *z* is fixed to simplify the model. Then *Y* is a zero-mean Gaussian vector whose covariance is z^2Q . The normalized representation is defined as dividing the original coefficient vector *Y* by an estimate of *z* computed from its neighboring coefficients, where *z* is the divisive normalization factor. The coefficient cluster *Y* is applied as a moving window across a subband. At each step, only the center coefficient y_c of the vector *Y* is normalized and the new coefficient under the divisive normalization representation becomes y_c/\hat{z} , where \hat{z} is the estimate of *z*. A convenient method to obtain \hat{z} is the maximum likelihood estimation [40] as shown in the following equation:

$$\hat{z} = \arg \max_{z} \log p_{Y}(Y|z)$$

$$= \arg \max_{z} (N \log z + Y^{T}Q^{-1}Y/2z^{2})$$

$$= \sqrt{Y^{T}Q^{-1}Y/N}$$
(2)

The covariance matrix $Q = E(UU^T)$ is estimated from the entire contourlet subband. As shown in Fig. 5, the kurtosis of the distributions for the coefficients for 3D images and disparity map after divisive normalization transform are close to 3. It indicates that the distribution of normalized coefficients in each subband can be modeled approximately by Gaussian distribution with small fitting error.

3.2. Influence of distortion on 3D natural image statistics properties

In this section, experiments are conducted to investigate the influence of distortion on the 3D natural image statistics properties. We impose different types of distortion on the original 3D images and investigate the marginal distribution. Experimental result shows that the distortions on the 3D images can significantly change the original near-Gaussian distribution of the coefficients. Meanwhile, the change varies with the distortion type. For easy illustration, the marginal distribution effect of the distorted 3D images is provided in Fig. 6. For example, the marginal distribution of the distorted luminance image has smaller standard deviation than that of the original luminance image in Fig. 6(a). The marginal distribution of distorted disparity map has larger standard deviation than that of the original disparity map in Fig. 6(g). Based on these observations, it can be concluded that the standard deviation of the marginal distribution can be used to measure the visual quality of distorted 3D images.

3.3. Feature extraction

For the original stereoscopic images and corresponding disparity map, the probability density function $p_i(x_i)$ of the coefficients in each subband *i* can be well fitted with a zero-mean Gaussian model as follows:

$$p_i^m(x) = \frac{1}{\sqrt{2\pi\sigma_{i,o}^2}} \exp\left(-\frac{x_i^2}{2\sigma_{i,o}^2}\right),\tag{3}$$

which just includes one parameter, i.e., standard derivation $\sigma_{i,o}$, to describe the whole histogram distribution. Therefore, 3N features (left/right luminance image, disparity map) will be sent to the receiver side as auxiliary information where N is the number of subbands. At the receiver side, the feature extraction operation is the same as the sender side. The standard derivation $\sigma_{i,d}$ is obtained from the subband i of the distorted stereoscopic images and the corresponding disparity map.

3.4. Quality pooling

At the quality pooling stage, the final quality score is determined by the feature similarity index as follows:

$$Q = \frac{2\sum_{i=1}^{3N} (\sigma_{i,o} \cdot \sigma_{i,d}) + c}{\sum_{i=1}^{3N} \sigma_{i,o}^2 + \sum_{i=1}^{3N} \sigma_{i,d}^2 + c'},$$
(4)

where c is a small positive offset to prevent division by zero. The quality score is in the range [0, 1].

4. Experimental results

To evaluate the performance of our proposed 3D RRIQA metric, the public LIVE 3D IQA dataset [41] is tested. The disparity estimation described in [42] is employed to obtain the disparity map for the original and distorted 3D images. The filters for the multi-scale decomposition and multi-directional decomposition stage in the contourlet transform are set as 9–7 biorthogonal filter and CD filter, respectively. The number of scale is three. For each scale, the corresponding high-pass band is decomposed into subbands with four directions. Therefore, the total number of subbands is 12. For each subband, 13 neighboring coefficients are used in the algorithm as the input of divisive normalization transform. As shown in Fig. 3, there are nine coefficients from the same subband, 1 from the parent band, and 3 from the same spatial location in the bands with other orientations at the same spatial scale. For the original 3D images, 36 features are extracted in total.

The LIVE 3D IQA dataset contains five datasets of 365 subjectrated 3D images with five types of distortions at different distortion levels. The distortion types include JPEG compression (denoted as JPEG), JPEG2000 compression (denoted as JP2K), white noise contamination (denoted as WN), Gaussian blur (denoted as GBLUR), and fast fading channel distortion of JPEG2000 compressed bitstream (denoted as FF). To validate the robustness of the metrics, the performance on the entire dataset (denoted as ALL) is also evaluated.

For fair comparison, both 2DIQA extension model and 3DIQA model including FR and RR metrics are evaluated in the experiment. As shown in Table 1, the name of 2DIQA extension model is italics. To fairly remove nonlinearity introduced by the subjective rating process and facilitate empirical comparison of different IQA metrics, we first map the objective quality score to subjective quality score by a five parameters logistic function. Afterwards, we choose three criteria to evaluate the mapping performance: (1) correlation coefficient (CC): accuracy of objective metrics; (2) Spearman's rank order correlation coefficient (SROCC): monotonicity of objective metrics; and (3) root mean-squared-error (RMSE).

Experimental results are provided in Tables 1–3. The nonlinear scatter plots of subjective DMOS vs. our proposed 3D RRIQA metric are provided in Fig. 7. For easy illustration, the metric that achieves the best performance is labeled with the ‡. Regarding the RR metrics, the one that achieves the best performance is in bold face.



Fig. 5. Marginal statistics of four contourlet subbands of 3D images after divisive normalization transforming. *X*-axis represents the coefficient amplitude and *Y*-axis represents the probability density. (a) Luminance images: the kurtosis of the four distributions are 3.15, 2.70, 2.75 and 2.92. (b) Disparity map: the kurtosis of the four distributions are 3.01, 2.84, 2.86 and 3.09.



Fig. 6. Marginal distribution of contourlet subbands after divisive normalization transforming. *X*-axis represents the coefficient amplitude and *Y*-axis represents the probability density. The blue and red curves are the distribution of original and distorted images, respectively. (a) Luminance images: JPEG2000 compression; (b) luminance images: JPEG compression; (c) luminance images: white Gaussian noise; (d) luminance images: Gaussian blur; (e) disparity map: JPEG2000 compression; (f) disparity map: JPEG compression; (g) disparity map: white Gaussian noise; (h) disparity map: Gaussian blur. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

As shown in Tables 1–3, our proposed metric can achieve the best performance among the RR metrics for all distortion types (except the distortion WN). It has been discovered that the receptive-field

of human eyes are sensitive to the tuning of individual voxels for space, orientation and spatial frequency, and this characteristic can be estimated by subband coding (e.g. contourlet based) directly

Table 1CC results summarization for IQA metrics.

Metric		JP2K	JPEG	WN	GBLUR	FF	ALL
FR	PSNR	0.7757	0.1167	0.9335	0.9161	0.6663	0.8239
	SSIM [29]	0.8540	0.4699	0.9416	0.9170	0.7293	0.8681
	MSSIM[43]	0.8814	0.6694 [‡]	0.9219	0.9441	0.7818	0.9188 [‡]
	VSNR[44]	0.8885	0.4107	0.9099	0.7599	0.7664	0.7575
	3DFR [36]	0.8340	0.2699	0.9129	0.9084	0.6766	0.8787
	You [32]	0.8770	0.4870	0.9410	0.9190	0.7300	0.8810
RR	RRED [24]	0.8051	0.3367	0.9425 [‡]	0.9272	0.6959	0.8818
	Hewage [18]	0.9040	0.5300	0.8950	0.7980	0.6690	0.8300
	Proposed	0.9162 [‡]	0.5697	0.9133	0.9574 [‡]	0.7833 ‡	0.8921

 ‡ The corresponding metric achieves the best performance.

Table 2

SROCC results summarization for IQA metrics.

Metric		JP2K	JPEG	WN	GBLUR	FF	ALL
FR	PSNR	0.7957	0.1186	0.9319	0.9001	0.5686	0.8236
	SSIM	0.8535	0.4317	0.9372	0.8813	0.5908	0.8700
	MSSIM	0.8962 [‡]	0.5524 [‡]	0.9419 [‡]	0.9282 [‡]	0.6937 [‡]	0.9137 [‡]
	VSNR	0.8300	0.4065	0.9038	0.8316	0.6931	0.8775
	You	0.8600	0.4390	0.9400	0.8820	0.5880	0.8780
	3DFR	0.8189	0.2390	0.9324	0.9000	0.6446	0.8599
RR	RRED	0.7628	0.3166	0.9307	0.9116	0.6610	0.8423
	Hewage	0.8560	0.5000	0.9400	0.6900	0.5450	0.8140
	Proposed	0.8832	0.5420	0.9066	0.9246	0.6548	0.8890

[‡] The corresponding metric achieves the best performance.

Table 3

RMSE results summarization for IQA metrics.

Metric		JP2K	JPEG	WN	GBLUR	FF	ALL
FR	PSNR	8.1732	6.4945	5.9668	5.8024	9.2653	9.2923
	SSIM	6.7385	5.7723	5.6016	5.7747	8.5013	8.1403
	MSSIM	6.1166	4.8578 [‡]	6.4437	4.7726	7.7472	6.4723 [‡]
	VSNR	5.9435	5.9622	6.9016	9.4076	7.9822	10.7058
	You	6.2060	5.7090	5.6210	5.6790	8.4920	7.7460
	3DFR	7.1456	6.2965	6.7890	6.0524	9.1498	7.8276
RR	<i>RRED</i>	7.6823	6.1575	5 . 5607 [‡]	5.4216	8.9237	7.7332
	Hewage	5.5300	5.5430	7.4050	8.7480	9.2260	9.1390
	Proposed	5.1890 [‡]	5.3741	6.7772	4.1777 [‡]	7.7245 [‡]	7.4081

[‡] The corresponding metric achieves the best performance.

689

from responses evoked by natural images [45]. This discovery suggests that the structural distortions may have more influences on human perception. As for image processing on natural images, the process of JPEG, JP2K, GBLUR, FF and WN always results in structural distortions, but WN is different from the others. For the WN distortion, the variance of white noise will be reduced after divisive normalization transform, thus the variance of the original and distorted 3D images are very similar. Therefore, the proposed metric shows better performances on JPEG, JP2K, GBLUR and FF than other RR metrics. Compared with the FR metrics such as PSNR. SSIM, MSSIM, VSNR, You and 3DFR, our proposed metric can also achieve a better performance for almost all the distortion types. except for the distortion IPEG and ALL. It is noteworthy that among these FR metrics, PSNR, SSIM, MSSIM and VSNR are proposed for traditional 2D natural images, while You and 3DFR are for the 3D images. The performance of PSNR, SSIM, MSSIM and VSNR are limited since the binocular vision properties are not considered. For You, the depth information is employed for predicting the quality score, where the performance is mainly depends on the fitting parameters. While for 3DFR, the binocular rivalry characteristic is modeled. Currently, the mechanism of binocular summation is still an open issue, thus the computation model of the rivalry property may not be accurate enough for assessing the perceptual quality of 3D images. Based on the observations above, we can make a conclusion that our proposed metric is powerful for predicting the 3D visual quality.

5. Conclusion

In this paper, we propose a novel 3D RRIQA metric based on 3D natural image statistics in contourlet domain. For the coefficient in the contourlet subband of luminance image and disparity map of the 3D images, the Gaussian scale mixtures' model is employed to normalize the coefficients. After the divisive normalization transform, the marginal distribution of the coefficients is approximately Gaussian distributed. For each contourlet subband of the luminance image and disparity map of the 3D images, the standard derivations of the fitted Gaussian distribution are extracted as feature parameters. At the receiver side, the feature similarity index is employed to measure the 3D visual quality. Experiments show that the proposed metric has good consistency with 3D subjective perception of human.



Fig. 7. Nonlinear scatter plots of subjective DMOS vs. proposed 3D RRIQA metric. (a) JP2K, (b) JPEG, (c) WN, (d) GBLUR, (e) FF, and (f) ALL.

Acknowledgments

This work was partly supported by National Natural Sciences Foundation of China under the Grants 61202301 and 61202242, and Shenzhen New Industry Development Fund under the Grant no. JCYJ20120617120716224.

References

- R. Ji, L.-Y. Duan, J. Chen, T. Huang, W. Gao, Mining compact 3D patterns for low bit rate mobile visual search, IEEE Trans. Image Process 23 (7) (2014) 3099–3113.
- [2] Y. Yang, Q. Liu, R. Ji, Y. Gao, Remote dynamic three-dimensional scene reconstruction, PloS One 8 (5) (2013) e55586.
- [3] Y. Gao, J. Tang, R. Hong, S. Yan, Q. Dai, N. Zhang, T.-S. Chua, Camera constraintfree view-based 3D object retrieval, IEEE Trans. Image Process. 21 (4) (2012) 2269–2281.
- [4] C. Hewage, M. Martini, Quality of experience for 3D video streaming, IEEE Commun. Mag. 51 (5) (2013) 101–107.
- [5] Y. Gao, M. Wang, Z. Zha, Q. Tian, Q. Dai, N. Zhang, Less is more: efficient 3d object retrieval with query view selection, IEEE Trans. Multimed. 11 (5) (2011) 1007–1018.
- [6] S. Daly, R. Held, D. Hoffman, Perceptual issues in stereoscopic signal processing, IEEE Trans. Broadcast. 57 (2 PART 2) (2011) 347–361.
- [7] X. Wang, M. Yu, Y. Yang, G. Jiang, Research on subjective stereoscopic image quality assessment, in: Proceedings of the SPIE, vol. 7255, 2009.
- [8] Y. Yang, X. Wang, T. Guan, J. Shen, L. Yu, A multi-dimensional image quality prediction model for user-generated images in social networks, Inf. Sci. 281 (2014) 601–610.
- [9] Y. Yang, Q. Dai, Contourlet-based image quality assessment for synthesised virtual image, Electron. Lett. 46 (7) (2010) 492–494.
- [10] M. Lambooij, W. Ijsselsteijn, D. Bouwhuis, I. Heynderickx, Evaluation of stereoscopic images: beyond 2D quality, IEEE Trans. Broadcast. 57 (2) (2011) 432–444.
- [11] V. De Silva, H. Arachchi, E. Ekmekcioglu, A. Kondoz, Toward an impairment metric for stereoscopic video: a full-reference video quality metric to assess compressed stereoscopic video, IEEE Trans. Image Process. 22 (9) (2013) 3392–3404.
- [12] R. Bensalma, M.-C. Larabi, A perceptual metric for stereoscopic image quality assessment based on the binocular energy, Multidimens. Syst. Signal Process. 24 (2) (2013) 281–316.
- [13] R. Ji, H. Yao, W. Liu, X. Sun, Q. Tian, Task dependent visual codebook compression, IEEE Trans. Image Process. 21 (4) (2012) 2282–2293.
- [14] Y. Gao, M. Wang, R. Ji, X. Wu, Q. Dai, 3d object retrieval with Hausdorff distance learning, IEEE Trans. Ind. Electron. 61 (4) (2014) 2088–2098.
- [15] Q. Liu, H. Xiao, Semi-global depth estimation algorithm for mobile 3-d video applications, Tsinghua Sci. Technol. 17 (2) (2012) 128–135.
- [16] Q. Liu, Y. Yang, R. Ji, Y. Gao, L. Yu, Cross-view down/upsampling method for multiview depth video coding, IEEE Signal Process. Lett. 19 (5) (2012) 295–298.
- [17] Y. Gao, M. Wang, D. Tao, R. Ji, Q. Dai, 3D object retrieval and recognition with hypergraph analysis, IEEE Trans. Image Process. 21 (9) (2012) 4290–4303.
- [18] C. Hewage, M. Martini, Reduced-reference quality assessment for 3D video compression and transmission, IEEE Trans. Consum. Electron. 57 (3) (2011) 1185–1193.
- [19] G. Nur, G. Akar, An abstraction based reduced reference depth perception metric for 3D video, in: Proceedings of the International Conference on Image Processing (ICIP), 2012, pp. 625–628.
 [20] A. Maalouf, M.-C. Larabi, CYCLOP: a stereo color image quality assessment
- [20] A. Maalouf, M.-C. Larabi, CYCLOP: a stereo color image quality assessment metric, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2011, pp. 1161–1164.
- [21] Z. Wang, G. Wu, H. Sheikh, E. Simoncelli, E.-H. Yang, A. Bovik, Quality-aware images, IEEE Trans. Image Process. 15 (6) (2006) 1680–1689.
- [22] Q. Li, Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, IEEE J. Select. Top. Signal Process. 3 (2) (2009) 202–211.
- [23] X. Wang, G. Jiang, M. Yu, Reduced reference image quality assessment based on contourlet domain and natural image statistics, in: Fifth International Conference on Image and Graphics, 2009, pp. 45–50.
- [24] R. Soundararajan, A. Bovik, RRED indices: reduced reference entropic differencing for image quality assessment, IEEE Trans. Image Process. 21 (2) (2012) 517–526.
- [25] Y. Liu, A.C. Bovik, L.K. Cormack, Disparity statistics in natural scenes, J. Vis. 8 (11) (2008) 1–14.
- [26] Y. Liu, L. Cormack, A. Bovik, Statistical modeling of 3-D natural scenes with application to Bayesian stereopsis, IEEE Trans. Image Process. 20 (9) (2011) 2515–2530.
- [27] C. Hewage, S. Worrall, S. Dogan, A. Kondoz, Prediction of stereoscopic video quality using objective quality models of 2-D video, Electron. Lett. 44 (2008) 963–965.
- [28] M. Pinson, S. Wolf, A new standardized method for objectively measuring video quality, IEEE Trans. Broadcast. 50 (3) (2004) 312–322.

- [29] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
- [30] P. Campisi, P. Le Callet, E. Marini, Stereoscopic images quality assessment, in: European Signal Processing Conference, 2007, pp. 2110–2114.
- [31] A. Benoit, P. Callet, P. Campisi, R. Cousseau, Using disparity for quality assessment of stereoscopic images, in: Proceedings of the International Conference on Image Processing (ICIP) 2008, pp. 389–392.
- [32] J. You, L. Xing, A. Perkis, X. Wang, Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis, in: Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics, January 2010.
- [33] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, G. Bozdagi Akar, Towards compound stereo-video quality metric: a specific encoder-based framework, in: Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation, vol. 2006, 2006, pp. 218–222.
- [34] X. Wang, S. Kwong, Y. Zhang, Considering binocular spatial sensitivity in stereoscopic image quality assessment, in: 2011 IEEE Visual Communications and Image Processing (VCIP) 2011, 2011.
- [35] Y. Zhao, Z. Chen, C. Zhu, Y.-P. Tan, L. Yu, Binocular just-noticeable-difference model for stereoscopic images, IEEE Signal Process. Lett. 18 (1) (2011) 19–22.
- [36] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. Cormack, A. Bovik, Full-reference quality assessment of stereopairs accounting for rivalry, Signal Process.: Image Commun. 28 (9) (2013) 1143–1155.
- [37] M.-J. Chen, L. Cormack, A. Bovik, No-reference quality assessment of natural stereopairs, IEEE Trans. Image Process. 22 (9) (2013) 3379–3391.
- [38] M.N. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (12) (2005) 2091–2106.
- [39] D.D.-Y.M.N. Po, M.N. Do, Directional multiscale modeling of images using the contourlet transform, IEEE Trans. Image Process. 15 (6) (2006) 1610–1620.
- [40] M.J. Wainwright, E.P. Simoncelli, Scale mixtures of Gaussians and the statistics of natural images, in: Advances in Neural Information Processing Systems, 2000, pp. 855–861.
- [41] A.K. Moorthy, C.-C. Su, A. Bovik, Subjective evaluation of stereoscopic image quality, Signal Process.: Image Commun. 28 (8) (2012) 870–883.
- [42] A. Ogale, Y. Aloimonos, A roadmap to the integration of early visual modules, Int. J. Comput. Vis. 72 (1) (2007) 9–25.
- [43] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multi-scale structural similarity for image quality assessment, in: in Proceedings of the IEEE Asilomar Conference on Signals, Systems, and Computers, 2003, pp. 1398–1402.
 [44] D. Chandler, S. Hemami, VSNR: a wavelet-based visual signal-to-noise-ratio
- [44] D. Chandler, S. Hemami, VSNR: a wavelet-based visual signal-to-noise-ratio for natural images, IEEE Trans. Image Process. 16 (9) (2007) 2284–2298.
- [45] K.N. Kay, T. Naselaris, R.J. Prenger, J.L. Gallant, Identifying natural images from human brain activity, Nature 452 (7185) (2008) 352–355.



Xu Wang received the B.S. degree from South China Normal University, Guangzhou, China, in 2007, and M.S. degree from Ningbo University, Ningbo, China, in 2010. He received his Ph.D. degree from the Department of Computer Science, City University of Hong Kong, Hong Kong in 2014. He is currently a senior research associate at the Department of Computer Science/Shenzhen Research Institute, City University of Hong Kong. His research Interests are 3D video coding and 3D image/ video quality assessment.



Qiong Liu received her B.S. degree in Computer Science and Technology from Central China Normal University and Ph.D. degree in National Engineering Research Center for Multimedia Software from Wuhan University in 2008. She is now the faculty of Division of Communication and Intelligent Network, WNLO. Her research interests include multi-view video coding, depth estimation for multi-view video, low complexity video coding, region-of-interest video compression, scalable video coding and video surveillance.



Ran Wang received her B.S. degree and Ph.D. degree from the College of Information Science and Technology, Beijing Forestry University, China, and Department of Computer Science, City University of Hong Kong, Hong Kong, in 2009 and 2013, respectively. She is currently a research associate at the Department of Computer Science, City University of Hong Kong. Her current research interests include support vector machines, extreme learning machines, decision tree induction, active learning, multiclass classification, and the related applications of machine learning.



Zhuo Chen received the B.S. degree from the School of Electronic and Information Engineering, Beijing Jiaotong University, China in 2014. He is currently a research assistant at the National Astronomical Observatories, Chinese Academy of Sciences. His research interests include 3D image quality assessment and micro-electro-mechanic system design.