Correspondences

Reduced Reference Stereoscopic Image Quality Assessment Based on Binocular Perceptual Information

Feng Qi, *Member, IEEE*, Debin Zhao, *Member, IEEE*, and Wen Gao, *Fellow, IEEE*

***Abstract*—In this paper, we propose a novel reduced reference stereoscopic image quality assessment (RR-SIQA) metric by using binocular perceptual information (BPI). BPI is represented by the distribution statistics of visual primitives in left and right views' images, which are extracted by sparse coding and representation. Specifically, entropy of the left view's image and entropy of the right view's image are used to represent monocular cue. Their mutual information is used to represent binocular cue. Constructively, we represent BPI as three numerical indicators. The difference of the original and distorted images' BPIs is taken as perceptual loss vector. The perceptual loss vector is used to compute the quality score for a stereoscopic image by a prediction function which is trained using support vector regression (SVR). Experimental results show that the proposed metric achieves significantly higher prediction accuracy than the state-of-the-art reduced reference SIQA methods and better than several state-of-the-art full reference SIQA methods on the LIVE phase II asymmetric databases.**

***Index Terms*—Binocular perceptual information (BPI), mutual information, sparse representation, stereoscopic image quality assessment (SIQA).**

I. Introduction

STEREOSCOPIC image quality assessment (SIQA) is one of the most fundamental yet challenging issues in 3D image processing technology. Because of the remarkable distinction between human monocular and binocular vision, SIQA can't be replaced by a simple combination of two individual views' quality assessment [1]. Apart from some novel visual experience issues (e.g.,unnatural depth experience, visual discomfort, visual fatigue, crosstalk), as a traditional distortion issue, asymmetric distortion assessment is a new question in image quality assessment (IQA). This paper focuses on asymmetric quality assessment for stereoscopic image in which two views' images have different visual quality levels.

Driven by the rapid development of 3D image applications, several works have been done in the last decade to design objective SIQA metrics which can automatically predict the perceived distortion in the distorted stereoscopic image. Since a 3D image consists of two views' 2D images, Gorley *et al.* [2], Benoit *et al.* [3], Campisi *et al.* [4] and You *et al.* [5] used a straightforward way to apply the state-of-the-art 2D-IQA methods in their SIQA metrics. They evaluated the qualities of the left and right view's images by 2D IQA metrics separately and then combined the two qualities into one quality score. Obviously, the simple combination of two views' images quality is not consistent with human binocular visual perception. Based on the study of binocular visual properties, Bensalma *et al.* proposed a binocular energy based SIQA metric [6]. Chen *et al.* proposed a SIQA metric accounting for rivalry [7]. Shao *et al.* proposed a SIQA metric by using binocular visual characteristics [8]. Zhou *et al.* proposed a perceptual modulated feature similarity metric for SIQA [9]. The aforementioned metrics depend on the entire original stereoscopic image to predict the quality of the distorted image, they are termed as full-reference (FR) SIQA. Without any information about the reference image, Chen *et al.* employed the natural scene statistics features to train a support vector machine model for their no-reference (NR) SIQA metric [10]. Akhter *et al.* proposed a NR metric to predict 3D image quality by perceptual features extracted from stereopairs and an estimated disparity map [11]. Sazzad *et al.* proposed two NR metrics for stereoscopic images and videos based on the segmented local features of artifacts, disparity and spatio-temporal segmentation, respectively [12][13]. Ryu *et al.* explored the relationship between the perceptual quality of stereoscopic images and visual information, and proposed a NR quality metric by a binocular quality perception model [14]. Since the NR-SIQA metrics are designed for specific and limited types of distortion, they may not be mature enough to some 3D applications.

Reduced reference (RR) metrics achieve a good tradeoff between FR and NR metrics [15], [16], as they can achieve higher prediction accuracy in terms of a limited transmitting data extracted from the reference image. Several studies have been conducted for RR-SIQA in recent years. Hewage *et al.* evaluated color plus depth 3D video by using the extracted edge information of depth maps and extracted information from the corresponding color image in the areas in the proximity of edges

Manuscript received December 15, 2014; revised May 25, 2015 and July 30, 2015; accepted October 08, 2015. Date of publication October 26, 2015; date of current version November 13, 2015. This work was supported in part by the Major State Basic Research Development Program of China 973 Program under Grant 2015CB351804 and in part by the National Science Foundation of China under Grant 61272386. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Christian Timmerer.

F. Qi and D. Zhao are with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: fqi@jdl.ac.cn).

W. Gao is with the School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2015.2493781

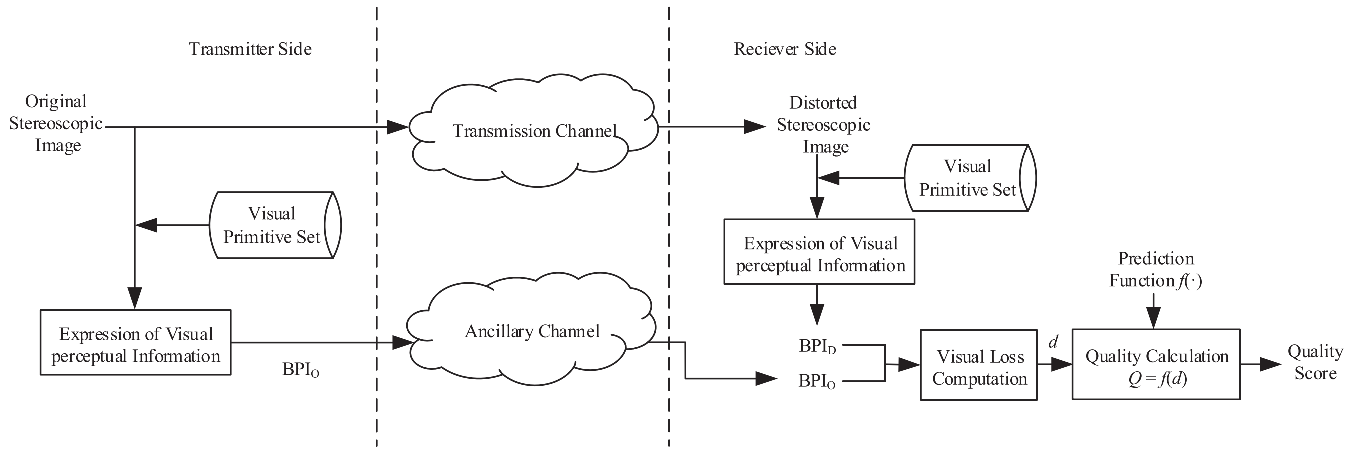

Fig. 1. Framework of the proposed SIQA metric.

[17]. Maalouf *et al.*'s metric is based on the discrepancy in the disparity maps of the reference and the distorted pair of images and the perceptual difference between the reference and the distorted cyclopean images [18]. Considering the binocular fusion and binocular rivalry, Zheng *et al.* proposed a RR-SIQA model based on binocular perceptual properties of HVS [19]. Xu *et al.* proposed a RR-SIQA model through measuring structural degradation and saliency based parallax compensation model [20].

Human eyes are the front-end binocular system. It has been discovered that cells in the retina of each eye individually encode their received visual signal, and then the coded information, later merged in lateral geniculate nucleus (LGN), formulate the ultimate stereoscopic image in the brain [20]. Human visual perception depends on both monocular and binocular cues. Image artifacts may make the unnatural perception of HVS. When the artifacts are asymmetric, the balance of monocular perception is broken and the binocular perception is also changed. Therefore, both the monocular and binocular cues should be taken into account in SIQA metric.

Neurobiological studies have suggested that sparse coding is one of the most important properties of receptive field in response to natural images [21]. Sparseland model is a powerful method to describe signals based on the sparsity and redundancy of signal representations [22] and it is efficient in dealing with visual information contained in natural scenes [23]. Since sparse representation is closely related to the cognitive behavior of HVS, it has been an efficient approach in image quality assessment [24], [25]. In this paper, inspired by the concept of sparse representation and visual primitive [26], we propose a RR-SIQA metric based on binocular perceptual information. In the proposed RR-SIQA, both monocular and binocular cues are taken into account. To measure monocular loss, we first exploit sparse coding to calculate the visual perceptual information for each view's image, which is termed as entropy. Then, monocular loss is expressed as the difference of visual perceptual information between the original and distorted two views' images. To measure binocular loss, the mutual information of both view's images is calculated. Similar with the monocular loss, binocular loss is expressed as the difference of mutual information between the original and distorted two views' images. Finally, both monocular loss and binocular loss are combined into one quality score by a perdition function.

The rest of this paper is organized as follows. Section II elaborates the proposed SIQA metric. Section III provides the experimental results and Section IV concludes the paper.

II. The Proposed SIQA Metric

The framework of the proposed SIQA metric is shown in Fig. 1. Taking the instance of a typical 3D visual communication system, at the transmitter side, binocular perceptual information (BPI) which consists of each view's entropy and both views' mutual information are calculated. An off-line visual primitive set (VPS) is used in BPI calculation. Note that VPS is constructed by the sparse representation model and is independent of the testing stereoscopic image. The BPI of original stereoscopic image is transmitted to the receiver side by an ancillary channel. At the receiver side, by loading the same off-line VPS, the distorted stereoscopic image's BPI is calculated. A perceptual loss vector d is constructed by the difference between the original and distorted stereoscopic images' BPIs. Finally, the quality score Q is computed using the perceptual loss vector d by a prediction function $f(\cdot)$. More details will be described in the following subsections.

A. Visual Primitive Set

To obtain compact representation from the observed signal, sparse representation can adaptively account for all or most of the information of a signal with the linear combination of a small number of elementary signals [24]. As for an image, its visual information can be efficiently described by its visual primitives [26]. Specifically, for a given image, the basic units of sparse representation are the patches. The vector representation of the image and image patches of size $\sqrt{B_s} \times \sqrt{B_s}$ at location $[i, i = 1, 2, \ldots, n,]$ are mathematically denoted by $X \in \mathbb{R}^N$ and $x_i \in \mathbb{R}^{B_s}$, where N is the number of the image vectors, B_s is the size of each patch vector, and n is the number of patches in an image [27], [28]. Then, we have

$$x_i = R_i X \tag{1}$$

where $R_i \in \mathbb{R}^{B_s \times N}$ is a matrix operator that extracts patch x_i from X. Note that patches are usually overlapped, and such

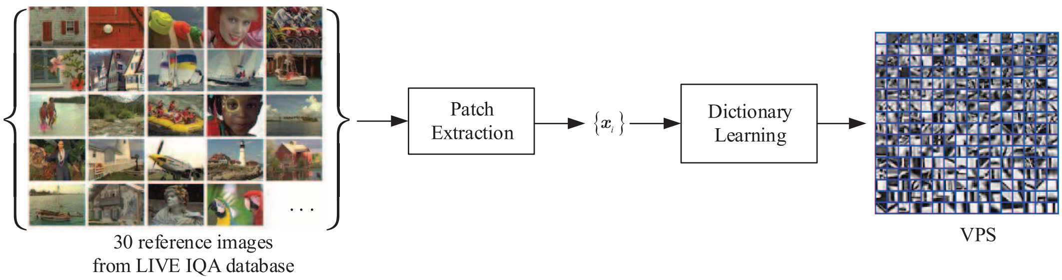

Fig. 2. Illustration of VPS construction.

patch-based representation is highly redundant. Therefore, the recovery of X from x_i becomes an over-determined system, from which it is straightforward to obtain by the least-square solution [29] as follows:

$$X = \left(\sum_{i=1}^{n} R_i^T R_i\right)^{-1} \sum_{i=1}^{n} \left(R_i^T x_i\right) \tag{2}$$

which is nothing but an abstraction strategy of averaging all the overlapped patches.

Given patches $\{x_i\}$, the purpose of dictionary learning is to search the best possible dictionary for the sparse representation of $\{x_i\}$.

In our case, since the given patches $\{x_i\}$ are extracted from a number of images, the optimal dictionary is able to represent these images in the sparsest way. If image samples are large enough, the dictionary can represent any image efficiently. We regard the dictionary as visual primitive set (VPS) which is the basis to express visual information for any image. In sparse representation, the dictionary learning process is formulated as

$$D, \{a_i\} = \underset{D,\{a_i\}}{\arg\min} \sum_i \|x_i - Da_i\|_2^2 \quad \text{s.t.} \quad \|a_i\|_0 < T \tag{3}$$

where $D \in \mathbb{R}^{B_s \times M}$ is an over-complete dictionary matrix that contains M visual primitives. The vector $a_i \in \mathbb{R}^M$ is coefficient vector for patch $\{x_i\}$. $\|a_i\|_0$ is the ℓ_0 norm, which counts the non-zero entries of the vector a_i, and T is the constraint of the non-zero number of a_i. In the experiment, M is set to 256 and T is set to 3.

Generally, (3) can be efficiently solved by the K-SVD algorithm [27]. How to construct VPS is shown in Fig. 2. In our implementation, 30 reference images in LIVE IQA database[1] are chosen as image samples. Each image is decomposed into patches. All the patches are concatenated to form the patches $\{x_i\}$. VPS is trained from the patches $\{x_i\}$ by the dictionary learning algorithm.

B. Binocular Perceptual Information Expression

The visual primitives in VPS are the basic visual perceptual elements to represent an image. How to use visual primitives to express visual perceptual information of an image is still a problem that needs to be solved. In this paper, we attempt to use the distribution statistics of visual primitives in left and right views' images to quantify visual perceptual information of a stereoscopic image.

For a stereoscopic image I, the visual primitive set (D) approximates I as

$$I = \left(\sum_{i=1}^{n} R_i^T R_i\right)^{-1} \sum_{i=1}^{n} \left(R_i^T D a_i\right) \tag{4}$$

where a_i denotes the ith column of A, A is the coefficient matrix of D, which is computed by the orthogonal matching pursuit (OMP) algorithm [30]. n is the number of patches in I. Assume d_k is the kth visual primitive of D, where M is the number of visual primitives in D. The sum of coefficients that d_k is used to reconstruct both the ith patch in left view and the jth patch in the right view can be expressed as

$$a_{d_k}(i,j) = \begin{cases} |a_i[k]| + |a_j[k]|, & \text{if } a_i[k] \neq 0 \text{ and } a_j[k] \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $a_i[k], a_j[k], (i, j = 1, \cdots, n/2)$ are the coefficients of d_k in the ith patch of left view's image and the jth patch of right view's image, respectively. Then, the sum of coefficients that d_k is used to reconstruct the patches in both the left view and the right view is calculated by

$$a^k = \sum_{i=1}^{n/2} \sum_{j=(n/2+1)}^{n} a_{d_k}(i,j). \tag{6}$$

Then, the joint probability density of visual primitive d_k for the left view's image I_L and the right view's image I_R is calculated by

$$p_k = \frac{a^k}{\sum\limits_{k=1}^{M} a^k}. \tag{7}$$

The probability density of visual primitive d_k for I_L is calculated by

$$p_k^L = \frac{\|a_L^k\|_1}{\sum\limits_{k=1}^{M} \|a_L^k\|_1} \tag{8}$$

where $\|a_L^k\|_1$ is the sum of coefficients that are used by d_k to reconstruct I_L. The probability density of visual primitive d_k to reconstruct I_R can be calculated in the same way.

[1]"LIVE image quality assessment database release 2," [Online]. Available: http://live.ece.utexas.edu/research/quality

According to Shannon theory, entropy is the uncertainty of a single random variable, which is used to represent the information quantities. Mutual information is a measure of the dependence between the two random variables. In this paper, entropy is used to represent the monocular visual perceptual information of one view's image. Mutual information of two views' images is used to represent the binocular visual perceptual information. The entropy of I_L can be calculated by

$$H(I_L) = -\sum_{k=1}^{M} p_k^L \log\left(p_k^L\right). \tag{9}$$

In the same way, we can obtain the entropy of I_R.

The mutual information of two views' images can be calculated by

$$MI(I_L; I_R) = \sum_{k\in\Omega}^{M} p_k \cdot \log\left(\frac{p_k}{p_k^L \cdot p_k^R}\right) \tag{10}$$

where $\Omega = \{k | p_k^L \times p_k^R \neq 0, k = 1, 2, \ldots, M\}$.

Therefore, the binocular perceptual information (BPI) of a stereoscopic image can be represented by $\{H(I_L), H(I_R), MI(I_L; I_R)\}$.

C. Quality Calculation

As aforementioned, when observer watches an asymmetric distorted stereoscopic image, both monocular and binocular visual perceptual information have some loss. The value of visual perceptual information loss reflects the distortion level between the original and distorted stereoscopic images. The loss of binocular perceptual information can be represented by $(BPI_O - BPI_D)$. $BPI_O : \{H(I_L), H(I_R), MI(I_L; I_R)\}$ and $BPI_D : \{H(I_L{}'), H(I_R{}'), MI(I_L{}'; I_R{}')\}$ are binocular perceptual information of original and distorted stereoscopic images, respectively. Specifically, for the left view's distortion, the loss of visual perceptual information is computed by

$$d(H_L) = H(I_L) - H(I_L{}') \tag{11}$$

where I_L and $I_L{}'$ are the original and distorted left view images, respectively. Similarly, the loss of right view's visual perceptual information is computed by

$$d(H_R) = H(I_R) - H(I_R{}') \tag{12}$$

where I_R and $I_R{}'$ are the original and distorted right view images, respectively. The loss of binocular perception information is computed by

$$d(MI) = MI(I_L; I_R) - MI(I_L'; I_R'). \tag{13}$$

The BPI difference between the original and distorted stereoscopic images can be expressed as a perceptual loss vector d : $\{d(H_L), d(H_R), d(MI)\}$. The quality score of a stereoscopic image is computed using the perceptual loss vector d by a prediction function $f(\cdot)$. That is, the final quality score is given by

$$Q = f(d) \tag{14}$$

where $f(\cdot)$ is trained in advance using support vector regression (ε-SVR) [31]. Here, $f(\cdot) : R^2 \to R$ takes d as input and produces output as a corresponding quality score.

In the ε-SVR, the unknown function $f(\cdot)$ is constructed by linearly combining the results of a nonlinear transformation of the input samples

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{15}$$

where α_i and α_i^* are the Lagrange multipliers, and $K(x_i, x)$ is the kernel function to perform nonlinear transformation. Here, we choose the radial basis function (RBF) kernel as follows:

$$K(x_i, x) = e^{-\gamma \|x_i - x_j\|^2} \tag{16}$$

where γ is a positive number, which represents the variance of the kernel function. In solving the SVR, there are two parameters should be determined. The ε-insensitive loss function is used to ignore errors that are smaller than a certain threshold, and the penalty parameter C is used to control the complexity of the prediction function $f(\cdot)$. In the experiment, the regression of the prediction function is performed using the LIBSVM.[2] Three parameters are finally fixed by a grid search, e.g.$\varepsilon = 0.5$, $C = 32$, $\gamma = 1$.

III. Experimental Results

A. Experiment Setting

To the best of our knowledge, there are only one public database for asymmetric distortion assessment of SIQA. The LIVE PhaseII 3D IQA database consists of 8 reference stereoscopic images and 360 distorted stereoscopic images with difference mean opinion scores (DMOS) [7], [10]. The database provides 5 distortion types, including White Noise (WN), JP2 K, JPEG, Gaussian Blur (GB), and Fast Fading (FF). For each distortion type, every reference stereopair is processed to create three symmetric distorted stereopairs and six asymmetric distorted stereopairs. Therefore, each distortion type has 72 distorted stereopairs.

To better divide the training and testing set for each distortion type stereopairs, we use 100-times 9-fold cross validation in the performance evaluation. Eight-ninths of all distorted stereopairs are randomly chosen for training, the rest stereopairs are used for testing. Three popular evaluation criteria are chosen to compare the predicted quality score after nonlinear regression with DMOS, e.g. Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC) and root mean square error (RMSE). A good objective method should have high PLCC and SROCC values but low RMSE value.

B. Performance Evaluation

To evaluate the efficiency of the proposed SIQA metric, we choose seven state-of-the-art objective metrics for comparison, i.e., Benoit *et al.*'s metric [3], You *et al.*'s metric [5], Chen *et al.*'s metric [7], Zhou *et al.*'s metric [9], Shao *et al.*'s metric [8], Hewage *et al.*'s metric [15], and Xu *et al.*'s metric [20].

[2] "LIBSVM: A library for support vector machines," [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

TABLE I
PERFORMANCE COMPARISON OF PLCC ON LIVE PHASEII DATABASE

	FR-SIQA					RR-SIQA		
Metrics	Benoit *et al.*	You *et al.*	Chen *et al.*	Zhou *et al.*	Shao *et al.*	Hewage *et al.*	Xu *et al.*	Proposed
WN	0.926	0.912	0.957	**0.958**	0.854	0.891	0.918	0.891
JP2K	0.784	**0.905**	0.834	0.806	0.782	0.664	0.752	0.858
JPEG	0.853	0.830	0.862	0.829	0.787	0.734	0.788	**0.871**
GB	0.535	0.784	0.963	0.971	0.934	0.450	0.938	**0.981**
FF	0.807	0.915	0.901	0.912	**0.933**	0.746	0.914	0.925
ALL	0.748	0.800	0.900	0.902	0.840	0.558	0.824	**0.915**

TABLE II
PERFORMANCE COMPARISON OF SROCC ON LIVE PHASEII DATABASE

	FR-SIQA					RR-SIQA		
Metrics	Benoit *et al.*	You *et al.*	Chen *et al.*	Zhou *et al.*	Shao *et al.*	Hewage *et al.*	Xu *et al.*	Proposed
WN	0.923	0.909	0.940	**0.945**	0.843	0.880	0.940	0.904
JP2K	0.751	**0.894**	0.814	0.788	0.781	0.598	0.751	0.776
JPEG	**0.867**	0.795	0.843	0.805	0.760	0.736	0.768	0.736
GB	0.455	0.813	0.908	0.909	**0.917**	0.028	0.900	0.871
FF	0.773	0.891	0.884	0.886	**0.923**	0.684	0.920	0.854
ALL	0.728	0.786	0.889	**0.892**	0.825	0.501	0.820	0.867

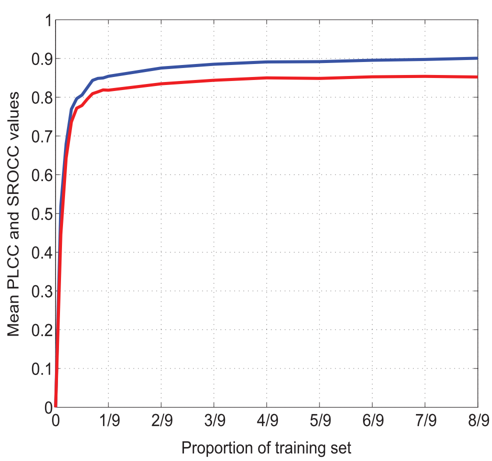

Fig. 3. PLCC and SROCC with different fold cross validation on LIVE phaseII database.

Note that Hewage *et al.*'s and Xu *et al.*'s metrics are RR-SIQA methods, while other metrics are FR-SIQA methods. The performance comparisons of PLCC and SROCC values for each distortion type on the LIVE phaseII database are listed in Table I and Table II, where the indicator that gives the best performance is highlighted in bold.

As shown in Table I, from the values of PLCC in the last line, the proposed metric achieves the best performance. Among individual distortion types, the proposed metric performs best in JPEG and GB distortion types. It should be noted that only Zhou *et al.*'s metric and the proposed metric adopt the same 9-fold cross validation method.

TABLE III
PERFORMANCE OF ENTROPY AND MI ON LIVE PHASEII DATABASE

Method	PLCC	SROCC	RMSE
MI	0.850	0.809	5.745
Entropy	0.896	0.838	4.964
Combination	**0.915**	**0.867**	**4.409**

From the values of SROCC in Table II, the proposed metric ranks three for all distortion types. In sum, Chen *et al.*'s metric, Zhou *et al.*'s metric and the proposed metric have close perdition accuracy. Since Chen *et al.*'s, Zhou *et al.*'s, Shao *et al.*'s and Xu *et al.*'s metrics all take the monocular and binocular perception properties into account, they achieve better prediction performance than the other SIQA metrics.

When compared to RR metrics: Hewage *et al.*'s metric and Xu *et al.*'s metric, the proposed metric achieves better performance both on PLCC and SROCC.

In addition, we also used 100-times cross validation to test the proposed metric under different proportion of training set and testing set. The same SVR parameters are applied. The experimental result is shown in Fig. 3. The blue line is the PLCC result, while the red line is the SROCC result. It can be seen that when the proportion is larger than 4/9, the PLCC and SROCC values change slightly.

Table III compares the performances only using MI factor and only using both views' entropy factors and their combination. It can be seen that both views' entropy factor contributes more prediction accuracy than MI factor. The combination achieves the best performance in the overall evaluation. It can be easily understood from the utilization of monocular and binocular cues point of view. The MATLAB source code of RR-SIQA and the experimental results are publicly available online at http://www.escience.cn/people/qifeng/index.html.

IV. Conclusion

This paper proposes a RR-SIQA metric based upon the binocular perceptual information (BPI). In the proposed SIQA metric, a visual primitive set is firstly constructed by offline training on the LIVE IQA 2D database. Then, according to the distribution statistics of visual primitives, entropy of each view's image and their mutual information are calculated by sparse coding and representation model. These three numerical indicators consist of BPI, which is used to describe monocular perception and binocular perception of a stereoscopic image. The difference of the original and distorted stereoscopic image's BPIs is taken as a perceptual loss vector. The perceptual loss vector is used to compute the quality score for the stereoscopic image by a prediction function which is trained using support vector regression.

Acknowledgment

The authors would like to thank the anonymous reviewers and the Associate Editor for their comments and suggestions that resulted in the improvement of the manuscript. They would also thank the authors of all the objective SIQA metrics used in this work, especially Dr. F. Shao and Dr. G. Zhai.

References

[1] M. Lambooij, W. IJsselsteijn, D. G. Bouwhuis, and I. Heynderickx, "Evaluation of stereoscopic images: Beyond 2D quality," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 432–444, Jun. 2011.

[2] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," in *Proc. Stereoscopic Displays Appl. XIX, SPIE*, 2008, vol. 6803, pp. 680305-1–12.

[3] A. Benoit, P. L. Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process*, vol. 659024, pp. 1–13, 2008.

[4] P. Campisi, P. L. Callet, and E. Marini, "Stereoscopic images quality assessment," in *Proc. Eur. Signal Process. Conf.*, 2007, pp. 2110–2113.

[5] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in *Proc. Int. Workshop Video Process. Quality Metrics Consum. Electron.*, 2010, vol. 9, pp. 1–6.

[6] R. Bensalma and M. C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimedia Syst. Signal Process.*, vol. 24, no. 2, pp. 281–316, 2012.

[7] M. J. Chen, C. C. Su, D. K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Process.: Image Commun.*, vol. 28, no. 9, pp. 1143–1155, 2013.

[8] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1940–1953, May 2013.

[9] W. Zhou, G. Jiang, M. Yu, F. Shao, and Z. Peng, "PMFS: A perceptual modulated feature similarity metric for stereoscopic image quality assessment," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 1003–1006, Aug. 2014.

[10] M. J. Chen, L. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3379–3391, Sep. 2013.

[11] R. Akhter, Z. M. P. Sazzad, Y. Horita, and J. Baltes, "No-reference stereoscopic image quality assessment," in *Proc. Stereoscopic Displays Appl. XXI, SPIE*, 2010, vol. 7524, pp. 1–12.

[12] Z. M. P. Sazzad, S. Yamanaka, Y. Kawayoke, and Y. Horita, "Stereoscopic image quality prediction," in *Proc. Int. Workshop QoMEX*, 2009, pp. 180–185.

[13] Z. M. P. Sazzad, S. Yamanaka, and Y. Horita, "Spatio-temporal segmentation based continuous no-reference stereoscopic video quality prediction," in *Proc. Int. Workshop QoMEX*, 2010, pp. 106–111.

[14] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 591–602, Apr. 2014.

[15] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[16] S. Wang, X. Zhang, S. Ma, and W. Gao, "Reduced reference image quality assessment using entropy of primitives," in *Proc. Picture Coding Symp.*, 2013, pp. 193–196.

[17] C. Hewage and M. G. Martini, "Reduced-reference quality assessment for 3D video compression and transmission," *IEEE Trans. Consum. Electron.*, vol. 57, no. 3, pp. 1185–1193, Aug. 2011.

[18] A. Maalouf and M. C. Larabi, "CYCLOP: A stereo color image quality assessment metric," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2011, pp. 1161–1164.

[19] K. Zheng *et al.*, "New reduced-reference objective stereo image quality assessment model based on human visual system," in *Proc. IEEE 3DTV-Conf.*, Jul. 2014, pp. 1–4.

[20] Q. Xu, G. Zhai, M. Liu, and K. Gu, "Using structural degradation and parallax for reduced-reference quality assessment of 3D images," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.*, Jun. 2014, pp. 1–6.

[21] B. Olshausen and D. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1," *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.

[22] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.

[23] A. Hyvarinen, J. Hurri, and P. O. Hoyer, *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision*. New York, NY, USA: Springer, 2009.

[24] L. He, D. Tao, X. Li, and X. Gao, "Sparse representation for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 1146–1153.

[25] T. Guha, E. Nezhadarya, and R. K. Ward, "Sparse representation-based image quality assessment," *Signal Process.: Image Commun.*, vol. 29, no. 10, pp. 1138–1148, 2014.

[26] J. Zhang, S. Ma, R. Xiong, D. Zhao, and W. Gao, "Image primitive coding and visual quality assessment," in *Proc. Pacific-Rim Conf. Multimedia*, 2012, vol. 7674, pp. 674–685.

[27] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[28] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[29] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[30] J. A. Tropp and A. A. Gilber, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[31] V. N. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.

Feng Qi (S'13–M'15) received the B.S. degree in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 2002, the M.S. degree in computer science from Northeast Petroleum University (NEPU), Daqing, China, in 2008, and is currently working towards the Ph.D. degree in computer science and technology, HIT.

Since 2010, he has been with the National Engineering Lab for Video Technology, Peking University, Beijing, China, as a Research Assistant. His research interests include image/video quality assessment and 3D visual perception.

Debin Zhao (M'11) received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 1985, 1988, and 1998, respectively.

He is currently a Professor with the Department of Computer Science, HIT. He has authored or coauthored over 200 technical articles in refereed journals and conference proceedings. His research interests include the areas of image and video coding, video processing, video streaming and transmission, and pattern recognition.

Wen Gao (S'87–M'88–SM'05–F'09) received the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He is currently a Professor of Computer Science with Peking University, Beijing, China. Before joining Peking University, he was a Professor of Computer Science with the Harbin Institute of Technology (HIT), Harbin, China, from 1991 to 1995, and a Professor at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He has authored or coauthored five books and over 600 technical articles in refereed journals and conference proceedings. His research interests include the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bio-informatics.

Dr. Gao has served or currently serves on the editorial board for several journals, such as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, the *EURASIP Journal of Image Communications*, and the *Journal of Visual Communication and Image Representation*. He chaired a number of prestigious international conferences on multimedia and video signal processing, such as the IEEE ICME and ACM Multimedia, and also served on the advisory and technical committees of numerous professional organizations