Blind Stereo Quality Assessment Based on Learned Features From Binocular Combined Images

Maryam Karimi, Mansour Nejati, S. M. Reza Soroushmehr, Shadrokh Samavi, *Member, IEEE*, Nader Karimi, and Kayvan Najarian

***Abstract*—Quality assessment of stereo images confronts more challenges than its 2D counterparts. Direct use of 2D assessment methods is not sufficient to deal with the challenges of 3D perception. In this paper, an efficient general-purpose no-reference stereo image quality assessment, based on unsupervised feature learning, is presented. The proposed method extracts features without any prior knowledge about the types and levels of distortions. This property enables our method to be adaptable for different applications. The perceived contrast and phase of the binocular combination of original stereo images are utilized to learn individual dictionaries. For each distorted stereo image, two feature vectors are pooled, in a hierarchical manner, over all sparse representation vectors of phase and contrast blocks by their corresponding dictionaries. Performance results of learning a regression model by the features acknowledge the superiority of the proposed method to state-of-the-art algorithms.**

***Index Terms*—3D perception, binocular combination, no-reference (NR) image quality assessment (IQA), sparse representation, stereo image quality assessment (SIQA), unsupervised feature learning.**

I. Introduction

QUALITY of images degrades due to image processing applications such as compression, retargeting and super-resolution or by going through communication channels. This has created a significant need for image quality assessment (IQA) at the end user side. It is inconvenient, costly and time consuming to have all video and images subjectively assessed. Moreover, subjective results are vulnerable to circumstances and individual characteristics of viewers. Hence, lots of research has been done to design objective image quality assessment approaches [1]–[5].

Due to the increasing popularity of three-dimensional images and videos it is more likely that in near future a large portion of all transferred data would likely be 3D media. The solution of 3D IQA is beyond a simple combination of left and right 2D quality scores. It should involve measuring the impact of changes on the binocular rivalry, visual fatigue, visual discomfort, and depth perception of image pairs. Hence, efficient and fast 3D IQA methods which are highly correlated with subjective mean opinion scores (MOS) are desired for quality protection and improvement of 3D imaging applications.

Similar to 2D IQA methods, based on the availability of reference image pair, 3D methods are divided into three major categories: full reference (FR), reduced reference (RR) and no reference (NR). An original image pair is needed to assess the quality of distorted pair in FR methods while some partial information about the original pair is available for RR schemes. NR algorithms estimate quality of the distorted stereo images without any information of the reference pair.

In this paper, an efficient NR Stereo IQA (NR-SIQA) method is proposed. This method aims to generate a single view close to what is formed in human brain (cyclopean). Most of the synthesized cyclopean images are formed by weighted summation of the two stereo views that are matched using the disparity map. In our method, images are combined using a mechanism similar to what the brain uses to combine the two perceived stereo images. In this method depth estimation of the scene is not needed. Hence, two synthesized "phase" and "contrast" images are formed from each stereo pair. We evaluate each stereo pair by analyzing these synthesized images. We use sparse representation as an appropriate environment for feature extraction and evaluation of quality of images. An unsupervised method for learning sparse features is used without prior knowledge about distortions. In the proposed method, dictionary learning is performed on undistorted phase and contrast images. Hence, there is no need to learn large dictionaries to cover different distortions. The sparse coefficients, in each feature vector, have the contribution of all the dictionary atoms in approximating a local phase/contrast feature. The contribution of the dictionary atoms in representation of a distorted image is thus collectively presented by sparse coefficients associated with all local phase/contrast features of that image. More efficient and more general features, as compared to handcrafted ones, are gener-

Manuscript received July 10, 2016; revised December 6, 2016 and April 3, 2017; accepted April 8, 2017. Date of publication April 27, 2017; date of current version October 13, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Lingfen Sun. *(Corresponding author: S. M. Reza Soroushmehr.)*

M. Karimi, M. Nejati, and N. Karimi are with the Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran (e-mail: Maryam.karimi@ec.iut.ac.ir; mansour.nejati@ec.iut.ac.ir; nader.karimi@cc.iut.ac.ir).

S. M. R. Soroushmehr is with the Michigan Center for Integrative Research in Critical Care, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: ssoroush@umich.edu).

S. Samavi is with the Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran, and also with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada (e-mail: samavi@mcmaster.ca).

K. Najarian is with the Michigan Center for Integrative Research in Critical Care, Department of Computational Medicine and Bioinformatics, and the Department of Emergency Medicine, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: kayvan@umich.edu).

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.2017.2699082

1520-9210 © 2017 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

ated. This is done using a spatial pyramid average pooling on sparse coefficients of all patches of each image. This encoding not only encodes the spatial structure of the patches but it also considers their neighborhood to have a more global representation of distortions. Then, a regression model is learned using distorted images. This model is to distinguish between sparse representations of different distortions and estimate the quality of distorted images. The effective performance of our regression model demonstrates the discrimination power of this method for various image distortions with different severities.

In summary, our main contributions are: 1) Formation of two synthesized phase and contrast images from a pair of stereoscopic images. These synthesized images have maximum coverage of the visual discomfort characteristics. 2) Unsupervised feature learning based on dictionaries trained for sparse representation of undistorted phase and contrast images which is not restricted to any particular type of distortion. 3) Use of the sparse representation of distorted phase and contrast patches as local feature representations and aggregation of features over large neighborhoods for richer representation of distortions.

The rest of this paper is organized as follows. Section II presents a review of existing objective stereo image quality assessment methods. In Section III, details of the proposed method are provided. The experimental results are described in Section IV and conclusions are drawn in Section V.

II. Related Work

Based on the type of modeling, 3D IQA methods can be divided into two categories. The first category belongs to methods that use depth information in addition to 2D information of stereo-pairs. The second category of methods belongs to those that do not estimate the depth and only use 2D spatial information of stereo images.

A. SIQA Methods Using Depth Estimation

Since FR SIQA methods have access to the reference stereo image pair, the original depth map can be used as additional information. Also, in RR methods the original depth map may be available for the assessing unit.

An improved version of Structural Similarity Index Measure (SSIM [1]) with depth for compressed and blurred stereo images was proposed by Benoit *et al.* [6]. Another work in [7] utilized the quality of disparity in addition to 2D qualities of left and right images and it was shown that visual quality of 3D video is impressed by low disparity regions and the type of video content. An algorithm called cyclopean Multi Scale SSIM (MS-SSIM) [8] applies 2D MS-SSIM between the original and distorted Gabor cyclopean images. The Gabor cyclopean is a weighted summation of left image and the shifted right image by the disparity. Edge information of reference depth map has been used in an RR approach in [9]. In [10], sensitivity coefficients of cyclopean images and the coherence between disparity maps were combined to produce subjective qualities. Moreover, color plus depth was employed as the depth sensation in 3D video.

Although there is no access to the reference images in NR scenarios, some NR SIQA approaches study the depth perception by imprecise disparity estimation of distorted stereo images.

In [11], Gaussian distributions were fitted to the statistics of disparity map and the images in contourlet domain and the related parameters were used to produce the quality metric. The estimated depth was used in [12] to measure temporal outliers, temporal inconsistencies, and spatial outliers. The combination of these three parameters constitutes the final measure. Local information of encoded stereo images and their disparity map were used in [13] to predict 3D quality scores. In another work proposed by Akhter *et al.* [14] some features extracted from disparity map in combination with blockiness and blurriness degrees, form the final quality score of JPEG compressed stereo images. Another method proposed by Chen *et al.* [15], extracts 2D and 3D features by fitting Gaussian distributions to the histograms of a generated cyclopean image. The cyclopean image is a weighted summation of the left and shifted right image using the imprecise disparity map. The Gabor filter responses are extracted from both views and the cyclopean image is calculated as a weighted summation of the two images, where the weights are computed from the Gabor filter responses. The *Stereoscopic/3D BLind Image Naturalness Quality* (S3D-BLINQ) index in [16] after forming a convergent cyclopean image using disparity maps, extracts both spatial-domain and wavelet-domain univariate and bivariate natural scene statistics features to estimate the quality of stereoscopically viewed image pairs. In [17], a Bivariate Generalized Gaussian Density (BGGD) model was proposed for the joint statistics of luminance and disparity of natural stereo scenes. This model is used to design an NR SIQA algorithm called Stereo Quality Evaluator (StereoQUE).

Disparity estimation algorithms are not only time consuming, but also the quality of computed disparity map is not satisfying in distorted stereo pairs. Therefore, it is necessary to design accurate 3D IQA models with no need to depth information.

B. SIQA Methods Not Using Depth Estimation

Primal 3D FR SIQA metrics tried to evaluate the quality of left and right images using 2D image quality assessment metrics. In [18], SSIM, Universal Quality Index (UQI [19]) and the RR metric [20] are combined to estimate the quality of stereo images. Another method in [21] applied numerous 2D metrics in order to estimation of quality for color plus depth encoded video.

A number of methods have been proposed to improve the performance of SIQA metrics by studying the human binocular perception. In [22], the depth quality was estimated based on amplitude changes of binocular energy in the stereo pair. A binocular quality perception model in [23] measures the luminance masking and contrast sensitivity of similar blocks in 3D-DCT domain. A binocular perception based on SSIM was introduced in [24] that combines luminance similarity, structural similarity and contrast similarity measures. For each left/right reference and distorted image a BJND model in [25] was formed by assessing pixels independently, in different classes and their average makes the final score. Another FR approach in [26] uses the local amplitude of a bank of log-Gabor filters to weight the summation and difference between each reference and distorted image. Finally, a 2D quality metric between these two

synthesized images provides the 3D quality score. A consistency checking in [27] divides each image into three regions and each region is assessed independently on the basis of the amplitude and phase maps of the reference and distorted images. The final quality score is a combination of all region scores. An FR metric for compressed stereo images decomposes each reference and distorted image into some narrow spatial frequency bands. Finally, 2D quality scores in all bands are weighted in a binocular combination way and summed together [28]. A recent method in [29] learns a code book on the training reference images. In the test phase, the similarity index between sparse coefficient vectors of each reference image and its distorted version are computed and binocularly combined to get the final stereoscopic score. In [30] a measure called Information content and Distortion Weighted SSIM (IDW-SSIM) is developed to estimate the quality of single view images. Then, a multi-scale model inspired by binocular rivalry was proposed that predicts the quality of stereoscopic images from that of single view images. The predicted quality scores by this method on asymmetric distorted stereo images are superior to others.

Among NR SIQA methods, a few of them do not use depth estimation. Ryu and Sohn proposed a top-down NR method that measures the blurriness and blockiness of left and right images in wavelet domain and combines them using a binocular perception model [31]. In [32], two individual monocular scores were obtained for each left and right images and the weighted summation of them by the sparsity of related feature vectors, prepares the final quality score. The local statistical distributions from local magnitude pattern and local directional pattern of binocular responses are used for extracting features in [33].

A high correlation between MOS and the objective scores based on binocular combination has been reported and compared with other 2D information only SIQA schemes [28]. As these methods are faster than depth information based methods, they are suitable for real-time applications.

Recently, unsupervised learning techniques from raw data have attracted increasing attention of researchers as they provide more accurate results than hand-crafted features [34]–[36]. Sparse representation is naturally discriminative as it selects only those basis vectors (dictionary atoms) among many that most compactly represent a signal and therefore is useful for description of image structure.

Inspired by a binocular combination method, we propose a method to combine perceived phase and contrast in different spatial frequency layers, appropriate for quality assessment without using depth estimation. Moreover, utilizing a pyramidal sparse feature representation for both phase and contrast leads to superior results to existing methods.

III. Proposed Method

Fig. 1 shows major stages of the proposed method which consists of dictionary learning and train/test stages. In the dictionary learning stage, as shown in Fig. 1(a), a set of pristine images, apart from the main database, is used. Binocular combination is performed on the image pair to form phase and contrast images. Here, we use the luminance component of the images as in this component all structural distortions are visible while color components have little contribution in evaluation of image quality [1]. The color components also increase the computational overhead. Therefore, we first convert the left and right images to grayscale before employing them in the "binocular combination" unit. Two individual sparse dictionaries are learned on the phase and contrast of the undistorted images. Then, at the training stage, as shown in Fig. 1(b), the phase and contrast blocks of the distorted training set are sparse coded, using the dictionaries learned in the previous stage. The sparse representations of distorted phase and contrast images are used as local features for the distortions. To achieve a single feature vector representing the whole stereo image pair, a pooling scheme is applied on all of the local sparse representations of each phase and contrast image pair. The regression model is then trained on the concatenation of the two feature vectors. Similarly, binocular combination, sparse representation, and feature pooling operations, are performed on the test set. The trained regression model, obtained from the training set, is used to evaluate the quality of images of the test set.

A. Binocular Combination

The information received by the two eyes is combined in the visual nervous system. Behavior of human visual cortex, in performing binocular combination, is modeled by optometrists using simple sine-wave images [37]. The main goal of binocular combination is to generate a single view (i.e. cyclopean) from two stereo images. This single view should be close to what is formed in human brain. As the brain eventually combines stereoscopic images, it seems logical to employ the combination of left and right views to evaluate the quality of stereo images. The Ding, Sperling, Klein and Levi (DSKL) model [37] provides a phase dependent contrast combination method. This model considers the left and right views as two sine waves, which have different contrasts (amplitude) and positions (phase). After contrast modifications in a gain-control and gain-enhancement process, an initial cyclopean sine wave is estimated by arithmetic vector summation of the two views. Then, the monocular misaligned sine waves are driven towards the cyclopean phase to compensate the disparity between perceptions of the two eyes. The final cyclopean phase and contrast are formed after the second summation. The phases and contrasts of cyclopean images contain 3D information about the stereo images [37].

To form binocular combination of stereoscopic images, we decompose each image into phase and contrast parts in different spatial frequencies with different orientations and then employ DSKL method. In the DSKL method, each eye is assumed to apply gain-control on the total contrast energy received by the other eye, proportional to its received total contrast energy. The controlled energies are used for exerting gain-enhancement and gain-control on the contrast of the opposite view in each spatial frequency-orientation. Afterwards, an initial vector summation and a fusion stage that remaps the retinal coordinates, to compensate the disparity between two misaligned waves, provide final contrast and phase images. We combine all contrast images together as well as phase images to make them useful for feature extraction.

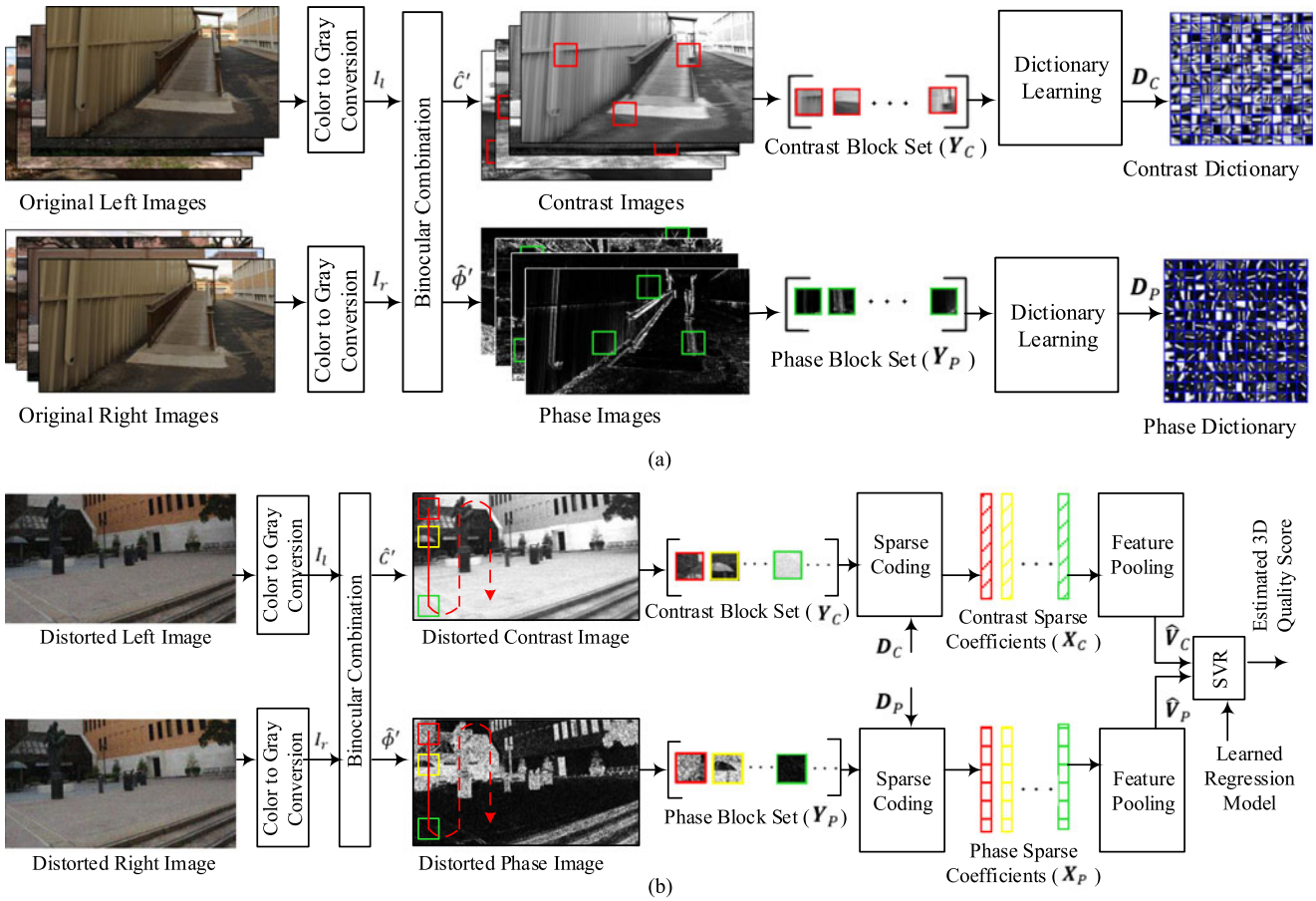

Fig. 1. Diagram of the proposed framework including (a) the dictionary learning stage and (b) the training/testing stage.

1) Extraction of Phase and Contrast for Each View: Regarding the binocular perception model in [38], we consider a spatial array of monocular linear neurons for each of the left and right views, $\nu \in \{l, r\}$, where their responses C_ν can be computed by convolving each image with a receptive field function such as a Gabor filter g with spatial frequency f_s and orientation θ

$$C_\nu(x, y; f_s, \theta) = \iint g(x-\xi, y-\eta; f_s, \theta) \cdot I_\nu(\xi, \eta)\, d\xi\, d\eta = |C_\nu(x, y; f_s, \theta)| \cdot e^{i\phi_\nu(x,y;f_s,\theta)} \quad (1)$$

where $I_\nu(\xi, \eta)$ is the intensity value of image received by view ν at each spatial position and (x, y) is the position of the receptive field center.

$$|C_\nu(x, y; f_s, \theta)| = \sqrt{Im[C_\nu(x, y; f_s, \theta)]^2 + Re[C_\nu(x, y; f_s, \theta)]^2} \quad (2)$$

$$\phi_\nu(x, y; f_s, \theta) = \tan^{-1}\left(\frac{Im[C_\nu(x, y; f_s, \theta)]}{Re[C_\nu(x, y; f_s, \theta)]}\right), \quad \nu \in \{l, r\} \quad (3)$$

The absolute value $|C_\nu(x, y; f_s, \theta)|$ of each response is defined as the contrast and ϕ_ν is the phase of response. $Re[.]$ and $Im[.]$ are the real and imaginary parts of the response respectively.

2) Total Contrast Energy (TCE): In this stage the total contrast energy E_ν for gain-control and total contrast energy E^*_υ for gain-enhancement across different orientations and frequency channels are computed for the view ν

$$E_\nu(x, y) = \sum_{f_s}\sum_{\theta} \frac{|C_\nu(x, y; f_s, \theta)|}{g_c}, \quad \nu \in \{l, r\} \quad (4)$$

$$E^*_\nu(x, y) = \sum_{f_s}\sum_{\theta} \frac{|C_\nu(x, y; f_s, \theta)|}{g_e}, \quad \nu \in \{l, r\} \quad (5)$$

where g_c is a gain-control threshold at which the contrast gain control becomes apparent and g_e is the gain-enhancement threshold.

3) Mutual Energy Gain-Control: The Gain-Control theory model explains the neural mechanism of binocular vision [28]. On the basis of this theory each eye, proportional to its total contrast energy, exerts divisive inhibition to the other eye's total energies with different gain-control efficiencies α and β. For the left view we have [37]

$$E'_l(x, y) = \frac{E_l(x, y)}{1 + \alpha E_r(x, y)} \quad (6)$$

$$E^{*'}_l(x, y) = \frac{E^*_l(x, y)}{1 + \beta E_r(x, y)}. \quad (7)$$

Similar equations are derived for the right eye.

4) Gain-Control and Gain-Enhancement on Signals: In this stage, each eye applies two gains corresponding to each of the controlled amount of total contrast energy to the other eye's signal [37]

$$|C_l\,(x,y;f_s,\theta)|_{ce} = \frac{|C_l\,(x,y;f_s,\theta)|\,.\,\left(1+E_r^{*'}\,(x,y)\right)}{1+E'_r\,(x,y)} \quad (8)$$

$$|C_r\,(x,y;f_s,\theta)|_{ce} = \frac{|C_r\,(x,y;f_s,\theta)|\,.\,\left(1+E_l^{*'}\,(x,y)\right)}{1+E'_l\,(x,y)}. \quad (9)$$

5) Vector Summation: After the interocular interaction, the monocular outputs can be combined using vector linear summation [37]. Let's show the monocular outputs as $\boldsymbol{I_v}\langle|C_v(x,y;f_s,\theta)|_{ce},\ \angle\phi_v(x,y;f_s,\theta)\,\rangle$ for $v\in\{l,r\}$ where $A\langle b,\angle\theta\rangle$ denotes the vector A with amplitude b and phase (angle) θ in polar coordinate system. Then, a vector arithmetic summation in each spatial position (x,y) produces an initial cyclopean image $\hat{\mathrm{I}}\langle\hat{C}(f_s,\theta),\ \angle\hat{\phi}(f_s,\theta)\rangle$ from which, phase and contrast are perceived [37].

$$\hat{I}\left\langle\hat{C},\angle\hat{\phi}\right\rangle = I_R\,\langle|C_r|_{ce},\ \angle\phi_r\rangle + I_L\,\langle|C_l|_{ce},\ \angle\phi_l\rangle \quad (10)$$

$$\hat{C} = \sqrt{|C_r|_{ce}^2+|C_l|_{ce}^2+2|C_r|_{ce}.\,|C_l|_{ce}.\cos\left(\phi_r-\phi_l\right)} \quad (11)$$

$$\hat{\phi} = \tan^{-1}\frac{|C_r|_{ce}.\,\sin\phi_r+\,|C_l|_{ce}.\,\sin\phi_l}{|C_r|_{ce}.\,\cos\phi_r+\,|C_l|_{ce}.\,\cos\phi_l} \quad (12)$$

Given that the spatial frequency f_s, the orientation θ and position (x,y) are common, we removed $(x,y;f_s,\theta)$ from the above equations for simplicity and brevity.

6) Fusion of Corresponding Points: Afterwards, the misaligned sine waves of the two eyes shift their phases in each spatial position (x,y) towards the cyclopean phase to be aligned with each other. Based on the fact that the phase difference in retinal coordinates becomes smaller than the physical coordinates, the motor/sensory fusion mechanism drives the two monocular misaligned vectors towards the cyclopean by an angle of the fraction "a" of the phase difference between that view and the initial cyclopean view

$$\begin{aligned}&\phi'_v\,(x,y;f_s,\theta) = \phi_v\,(x,y;f_s,\theta)+a\,(x,y;f_s,\theta)\\&\left(\hat{\phi}\,(x,y;f_s,\theta)-\phi_v\,(x,y;f_s,\theta)\right),\qquad \nu\in\{l,r\}.\end{aligned} \quad (13)$$

"a" is the output of gain-control on the disparity energy. The two right and left vectors are cross multiplied to calculate disparity energy E_d for fusion [37]

$$\begin{aligned}E_d\,(x,y;f_s,\theta) = |C_r\,(x,y;f_s,\theta)|_{ce}.|C_l\,(x,y;f_s,\theta)|_{ce}.\\ \sin\left(\phi_r\,(x,y;f_s,\theta)-\phi_l\,(x,y;f_s,\theta)\right)\end{aligned} \quad (14)$$

$$a\,(x,y;f_s,\theta) = \frac{E_d(x,y;f_s,\theta)^{\gamma_f}}{\left(g_f\right)^{2\gamma_f}+E_d(x,y;f_s,\theta)^{\gamma_f}} \quad (15)$$

where γ_f is the exponent value for gain-control in fusion step and g_f is a contrast threshold at which the fusion becomes apparent. At very low contrast $E_d \ll g_f^2$ the fusion does not occur and when $E_d \gg g_f^2$ the fusion mechanism causes a complete alignment. Finally, the arithmetic summation for each spatial position (x,y) is done again with the shifted phases. $(x,y;f_s,\theta)$ is removed to shorten (16) and (17)

$$\hat{C}' = \sqrt{|C_r|_{ce}^2+|C_l|_{ce}^2+2|C_r|_{ce}.\,|C_l|_{ce}.\,\cos\left(\phi'_r-\phi'_l\right)} \quad (16)$$

$$\hat{\phi}' = \tan^{-1}\frac{|C_r|_{ce}.\,\sin\phi'_r+\,|C_l|_{ce}.\,\sin\phi'_l}{|C_r|_{ce}.\,\cos\phi'_r+\,|C_l|_{ce}.\,\cos\phi'_l}. \quad (17)$$

Now, corresponding to each pair of f_s and θ we have a contrast image and a phase image. Both phase and contrast are affected by the disparity energy between the left and right waves. Therefore, having both the contrast and phase of the cyclopean image in all spatial frequencies and orientations contain all the needed 3D information. We can either extract features from each set resulted images or combine each set and then perform feature extraction. Since the latter is less complex, we combine each set and for doing that we use the root mean square (RMS) of contrast results and the maximum of phase over all spatial frequencies and orientations as the final contrast and phase images for extracting effective features for stereo quality assessment

$$\hat{C}'\,(x,y) = \sqrt{\frac{1}{n_{f_s}.n_\theta}\sum_{f_s}\sum_{\theta}\hat{C}'^2\,(x,y;f_s,\theta)} \quad (18)$$

$$\hat{\phi}'\,(x,y) = \max_{f_s,\,\theta}\hat{\phi}'\,(x,y;f_s,\theta) \quad (19)$$

where n_{f_s} and n_θ are the number of used spatial frequencies and orientations respectively. Eight orientations $\theta\in\{k\pi/8|k=0,\ldots,7\}$ with six different spatial frequencies $f_s\in\{1.5,\ 2.5,\ 3.5,\ 5,\ 7,\ 10\}(cycles/degree)$ are used in our experiments. Fig. 2 shows phase and contrast images for a pristine stereo image pair and its distorted versions.

B. Learning Sparsifying Dictionaries

This stage consists of constructing dictionaries used for local feature encoding. Learning sparsifying dictionaries from data is a recent approach to dictionary construction which has been strongly influenced by the latest advances in sparse representation field [39]. Given a set of N training examples, $\mathbf{Y}=[\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_N]\in\mathbb{R}^{n\times N}$, learning a dictionary $\mathbf{D}\in\mathbb{R}^{n\times K}$ $(K>n)$ with K atoms for sparse representation of $\mathbf{Y}$ is typically written as a joint optimization problem with respect to the dictionary $\mathbf{D}$ and sparse coefficients $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N]\in\mathbb{R}^{K\times N}$

$$\min_{\mathbf{D},\,\mathbf{X}}\ \sum_{i\,=\,1}^{N}\left\{\|\mathbf{y}_i-\mathbf{D}\mathbf{x}_i\|_2^2+\mu\psi\left(\mathbf{x}_i\right)\right\} \quad (20)$$

or equivalently in a matrix form

$$\min_{\mathbf{D},\,\mathbf{X}}\ \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2+\mu\Psi\left(\mathbf{X}\right) \quad (21)$$

where ψ is sparsity-inducing regularization function, $\Psi(\mathbf{X})=\sum_{i\,=\,1}^{N}\psi(\mathbf{x}_i)$, and μ the regularization parameter. Also, $\|\cdot\|_F$ denotes the Frobenius norm. The common choices for ψ are non-convex ℓ_0-pseudo-norm [40]–[42], that counts the number

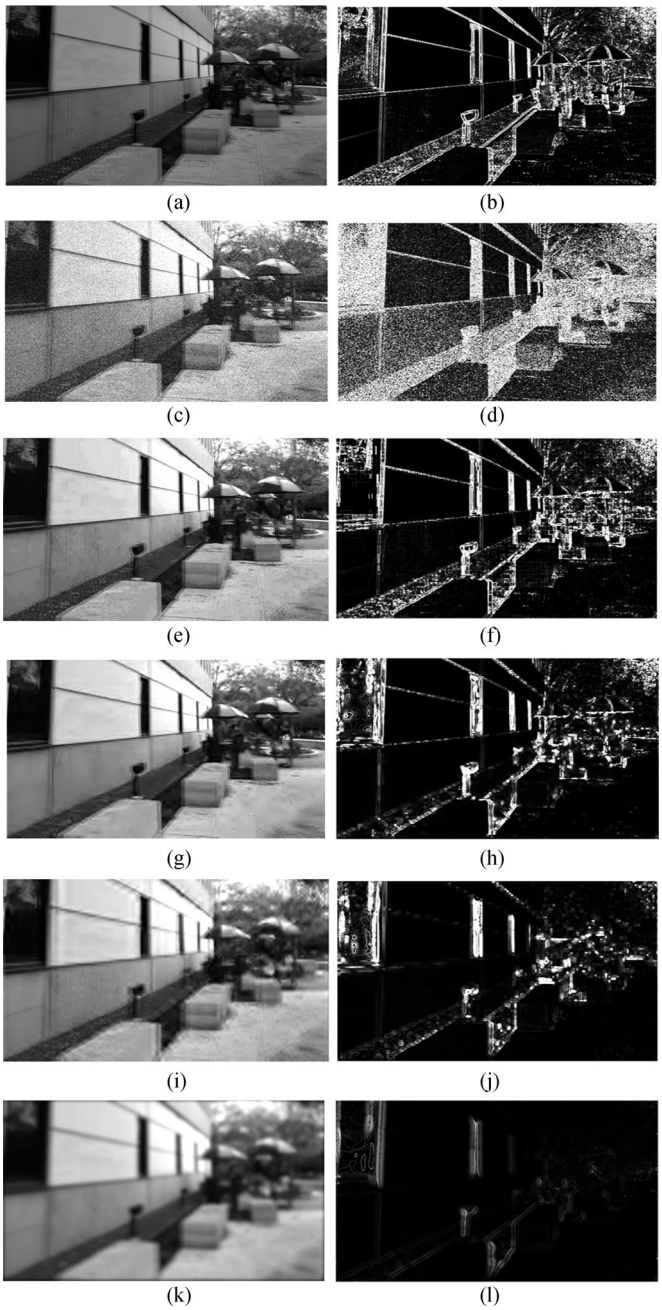

Fig. 2. (a) Perceived contrast and (b) phase of original stereo pair and corresponding distorted versions by (c), (d) WN, (e), (f) JPEG, (g), (h) JP2K, (i), (j) FF, and (k), (l) Blur.

of non-zero entries, and ℓ_1-norm [43], [44], as its convex counterpart.

In our algorithm, dictionary learning is applied on local features extracted from a set of undistorted stereoscopic image pairs. Dictionary learning on undistorted images makes it needless to learn a very large dictionary that includes all of the distortions. Also, it is possible to perform quality assessment with no knowledge about the type of distortion or its severity. The impact of different distortions on the sparse representation of undistorted blocks has enough discriminative characteristics that can lead to a well-trained regression.

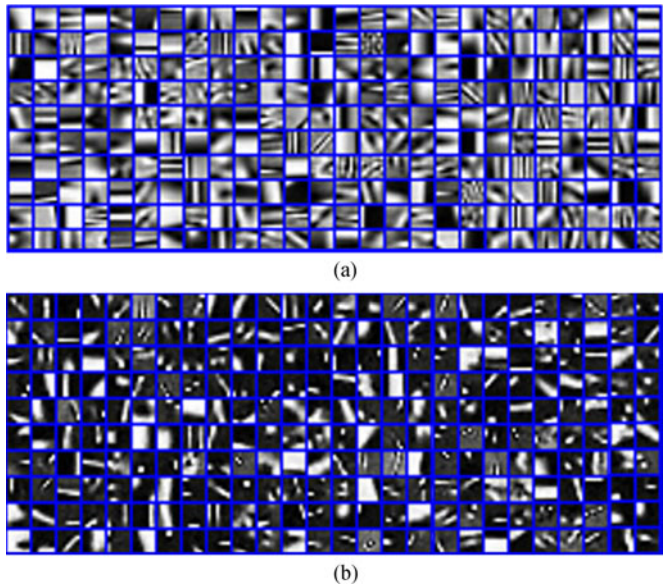

Fig. 3. Small parts of learned dictionaries on (a) contrast and (b) phase images.

Using the binocular combination presented in the previous section, we first produce pristine phase and contrast images for the training stereoscopic image pairs. After that, we need to create training samples for dictionary learning. To do this, small patches of size $\sqrt{n} \times \sqrt{n}$ pixels are randomly sampled as local features from the phase and contrast images. All patches are then rearranged into column vectors and the mean value of each patch is subtracted from it. Let $\mathbf{Y}_P = \{\mathbf{y}_{P,j} \in \mathbb{R}^n\}_{j=1}^N$ and $\mathbf{Y}_C = \{\mathbf{y}_{C,j} \in \mathbb{R}^n\}_{j=1}^N$ denote the set of training samples extracted from the phase and contrast images respectively. Given these two sets of local features, we train dictionaries $\mathbf{D}_P \in \mathbb{R}^{n \times K_P}$ and $\mathbf{D}_C \in \mathbb{R}^{n \times K_C}$ for sparse representation of $\mathbf{Y}_P$ and $\mathbf{Y}_C$ respectively. These dictionaries are then employed for encoding of local features. Two small parts of the learned dictionaries on contrast and phase images are displayed in Fig. 3.

C. Local Feature Sparse Coding

At the training and testing stages, the phase and contrast of distorted stereoscopic images are represented in terms of the corresponding learned dictionaries. Specifically, given a phase/contrast image, its non-overlapping patches as local features are sparsely represented over the learned phase/contrast dictionary. Let $\mathbf{Y}_P = \{\mathbf{y}_{P,i} \in \mathbb{R}^n\}_{i=1}^M$ denotes the set of all $\sqrt{n} \times \sqrt{n}$ non-overlapping patches of a given phase image. The sparse representation for the *i*-th patch, $\mathbf{y}_{P,i}$, is obtained by solving the following ℓ_1 -regularized sparse coding problem:

$$\hat{\mathbf{x}}_{P,i} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y}_{P,i} - \mathbf{D}_P \mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \qquad (22)$$

where $\hat{\mathbf{x}}_{P,i} \in \mathbb{R}^{K_P}$ is sparsely encoded representation of the local feature associated with the patch $\mathbf{y}_{P,i}$ using the phase dictionary $\mathbf{D}_P$. Also, λ is the sparsity regularization parameter empirically set to 0.05. To solve the above sparse coding problem, we utilize the popular method of Least Angle Regression

with Lasso modification (LARS) [45]. The sparse coding of non-overlapping patches $\mathbf{Y}_C = \{\mathbf{y}_{C,i} \in \mathbb{R}^n\}_{i=1}^{M}$ in a contrast image can be calculated in the same way.

Once the above sparse coding is performed for all non-overlapping patches of a phase-contrast pair computed for a given stereo image pair, we obtain two sets of sparse features $\mathbf{X}_P = \{\hat{\mathbf{x}}_{P,1}, \hat{\mathbf{x}}_{P,2}, \ldots, \hat{\mathbf{x}}_{P,M}\}$ and $\mathbf{X}_C = \{\hat{\mathbf{x}}_{C,1}, \hat{\mathbf{x}}_{C,2}, \ldots, \hat{\mathbf{x}}_{C,M}\}$. These feature sets are then aggregated using pooling operators to achieve richer representations for input stereo images.

D. Feature Pooling

Once the sparse vectors of local image features are obtained, we should combine them into a global stereo quality score. We have now sparse feature sets $\mathbf{X}_P$ and $\mathbf{X}_C$ for a given distorted stereo-pair. To achieve a single feature vector for each image pair, a pooling operator is employed. In our algorithm, we evaluate several pooling schemes including Max pooling (Max), Average pooling (AVG), Number of Non-Zeros pooling (NNZ), Hard Max Pooling (HMax) [32], the Spatial Pyramid Average pooling (SPA) and Spatial Pyramid Max pooling (SPM) [46] methods on the absolute value of sparse coefficients. The final feature vector is $\mathbf{v} = \begin{bmatrix}\hat{\mathbf{v}}_P \\ \hat{\mathbf{v}}_C\end{bmatrix}$ which is obtained by concatenating feature vectors $\hat{\mathbf{v}}_P$ and $\hat{\mathbf{v}}_C$.

1) Average Pooling (AVG): This feature pooling technique, applies averaging or element-wise summing of sparse vectors. We use this pooling on the absolute values of sparse coefficients

$$\hat{\mathbf{v}}_P(j) = \underset{1\le i\le M}{\text{AVG}}\left(|\hat{\mathbf{x}}_{P,i}(j)|\right), \quad \forall j = 1,\ldots,K_p \tag{23}$$

$$\hat{\mathbf{v}}_C(j) = \underset{1\le i\le M}{\text{AVG}}\left(|\hat{\mathbf{x}}_{C,i}(j)|\right), \quad \forall j = 1,\ldots,K_C \tag{24}$$

where $\hat{\mathbf{x}}_{P,i}(j)$ and $\hat{\mathbf{x}}_{C,i}(j)$ are the *j*-th element of *i*-th sparse representation in the feature sets $\mathbf{X}_P$ and $\mathbf{X}_C$ respectively. Also, $\hat{\mathbf{v}}_P(j)$ and $\hat{\mathbf{v}}_C(j)$ denote the *j*-th element of the phase and contrast pooled feature vector respectively.

2) Max Pooling (MAX): This method chose the maximum value in each row of $\mathbf{X}_P$ or $\mathbf{X}_C$

$$\hat{\mathbf{v}}_P(j) = \max_{1\le i\le M}\left(|\hat{\mathbf{x}}_{P,i}(j)|\right), \quad \forall j = 1,\ldots,K_p \tag{25}$$

$$\hat{\mathbf{v}}_C(j) = \max_{1\le i\le M}\left(|\hat{\mathbf{x}}_{C,i}(j)|\right), \quad \forall j = 1,\ldots,K_C \tag{26}$$

where $\hat{\mathbf{x}}_{P,i}(j)$ and $\hat{\mathbf{x}}_{C,i}(j)$ are the *j*-th element of *i*-th sparse representation in the feature sets $\mathbf{X}_P$ and $\mathbf{X}_C$ respectively. Also, $\hat{\mathbf{v}}_P(j)$ and $\hat{\mathbf{v}}_C(j)$ denote the *j*-th element of the phase and contrast pooled feature vectors respectively.

3) Hard Max Pooling (HMax): Based on the pooling method in [32], we set all entries in each sparse feature vector $\hat{\mathbf{x}}_{P,j}$ and $\hat{\mathbf{x}}_{C,j}$ to zero except the maximum sparse coefficient that is converted to 1. Specifically, we first apply hard-max function on the sparse feature vectors as

$$\tilde{\mathbf{x}}_{P,i}(j) = \begin{cases} 1, & \text{if } j = \underset{1\le j\le K_P}{\text{argmax}}\left(\hat{\mathbf{x}}_{P,i}(j)\right) \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

$$\tilde{\mathbf{x}}_{C,i}(j) = \begin{cases} 1, & \text{if } j = \underset{1\le j\le K_C}{\text{argmax}}\left(\hat{\mathbf{x}}_{C,i}(j)\right) \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

Then, the pooled feature vectors $\hat{\mathbf{v}}_P$ and $\hat{\mathbf{v}}_C$ are obtained by summing up all the modified sparse vectors in the corresponding sparse feature set as follows:

$$\hat{\mathbf{v}}_P(j) = \sum_{i=1}^{M} \tilde{\mathbf{x}}_{P,i}(j), \quad \forall j = 1,\ldots,K_p \tag{29}$$

$$\hat{\mathbf{v}}_C(j) = \sum_{i=1}^{M} \tilde{\mathbf{x}}_{C,i}(j), \quad \forall j = 1,\ldots,K_C. \tag{30}$$

4) Number of Non-zero Pooling (NNZ): This type of pooling scheme simply counts the number of non-zero elements in each row of sparse feature matrices $\mathbf{X}_P$ and $\mathbf{X}_C$ to produce the pooled features as

$$\hat{\mathbf{v}}_P(j) = \|\hat{\mathbf{x}}_{P,1}(j), \hat{\mathbf{x}}_{P,2}(j), \ldots, \hat{\mathbf{x}}_{P,M}(j)\|_0, \quad \forall j = 1,\ldots,K_p \tag{31}$$

$$\hat{\mathbf{v}}_C(j) = \|\hat{\mathbf{x}}_{C,1}(j), \hat{\mathbf{x}}_{C,2}(j), \ldots, \hat{\mathbf{x}}_{C,M}(j)\|_0, \quad \forall j = 1,\ldots,K_C \tag{32}$$

where $\|\cdot\|_0$ is the ℓ_0 pseudo-norm that counts the number of nonzero entries in a vector.

5) Spatial Pyramid Pooling: The Spatial Pyramid pooling [46] produces a global representation of an image by summarizing the distribution of the sparse codes in the bins of a spatial pyramid by a pooling step. For each phase and contrast image Spatial Pyramid Max/AVG pooling partitions each image into multiple level spatial bins. We apply Spatial Pyramid Max pooling (SPM) and Spatial Pyramid AVG pooling (SPA) on a three-level spatial pyramid with 1, 4 and 16 bins respectively. The features of each spatial bin B are the component-wise Max/AVG pooled over all sparse codes of the blocks within the Bin

$$\hat{\mathbf{v}}_P^B(j) = \underset{i\in B}{\text{OP}}\left(|\hat{\mathbf{x}}_{P,i}(j)|\right), \quad \forall j = 1,\ldots,K_p \tag{33}$$

$$\hat{\mathbf{v}}_C^B(j) = \underset{i\in B}{\text{OP}}\left(|\hat{\mathbf{x}}_{C,i}(j)|\right), \quad \forall j = 1,\ldots,K_C \tag{34}$$

where OP is the operator of Max/AVG, $\hat{\mathbf{x}}_{P,i}(j)$ and $\hat{\mathbf{x}}_{C,i}(j)$ are the *j*-th element of *i*-th sparse representation in the feature sets $\mathbf{X}_P$ and $\mathbf{X}_C$ respectively. Also, $\hat{\mathbf{v}}_P^B(j)$ and $\hat{\mathbf{v}}_C^B(j)$ denote the *j*-th element of the phase and contrast pooled feature related to bin B.

The final feature vectors $\hat{\mathbf{v}}_P$ and $\hat{\mathbf{v}}_C$ are the concatenation of aggregated sparse feature vectors in all spatial bins which are normalized by dividing with their ℓ_2-norm. The dimensionality of each pooled feature vector is $(16+4+1)M$, where M is

the size of dictionary.

$$\hat{\mathbf{v}}_P = \left[\frac{\hat{\mathbf{v}}_P^1}{\|\hat{\mathbf{v}}_P^1\|_2 + \varepsilon}, \frac{\hat{\mathbf{v}}_P^2}{\|\hat{\mathbf{v}}_P^2\|_2 + \varepsilon}, \ldots, \frac{\hat{\mathbf{v}}_P^{21}}{\|\hat{\mathbf{v}}_P^{21}\|_2 + \varepsilon} \right] \quad (35)$$

$$\hat{\mathbf{v}}_C = \left[\frac{\hat{\mathbf{v}}_C^1}{\|\hat{\mathbf{v}}_C^1\|_2 + \varepsilon}, \frac{\hat{\mathbf{v}}_C^2}{\|\hat{\mathbf{v}}_C^2\|_2 + \varepsilon}, \ldots, \frac{\hat{\mathbf{v}}_C^{21}}{\|\hat{\mathbf{v}}_C^{21}\|_2 + \varepsilon} \right] \quad (36)$$

We compare the performance of our method using all the above pooling methods in the section of experimental results. Among these methods, the one that has the best performances on both symmetric and asymmetric distorted stereo images, is SPA pooling. Our results show that not only it is robust against different parameters of our algorithm but also can result in very good performance even if we use small sizes of dictionaries.

E. Learning Regression Model for Quality Estimation

In addition to finding effective features, the method of feature fusion is important to produce an efficient quality estimator. We use support vector regression (SVR) to map our features to 3D subjective quality scores. If $\mathbf{v}_i$ is the feature vector of i -th stereo image and t_i is the corresponding subjective score, ε -SVR [47] tries to define a function $f(\mathbf{v})$ with the maximum deviation of ε from the subjective quality score for all the training image set

$$f(\mathbf{v}) = \sum_i \alpha_i t_i \varphi(\mathbf{v}_i)^T \varphi(\mathbf{v}) + \delta \quad (37)$$

where $\varphi(\mathbf{v})$ is a nonlinear function of feature vector $\mathbf{v}$ and δ is the bias. The goal is to find α and δ such that the error is less than ε. In the training stage, the training set is presented to the SVR system to estimate best values for α and δ. In the test phase the trained model is presented with test vectors to estimate the corresponding objective scores. The kernel Radial Basis Function (RBF) is used for the regression with the form of $K(\mathbf{v}_i, \mathbf{v}) = \varphi(\mathbf{v}_i)^T \varphi(\mathbf{v}) = \exp(-\gamma \|\mathbf{v}_i - \mathbf{v}\|^2)$ for each two samples $\mathbf{v}$ and $\mathbf{v}_i$, where the parameter γ is a positive parameter controlling the radius and it is estimated using cross-validation on the training set. The feature vectors of the training set and the corresponding 3D subjective scores are fed to SVR to generate a 3D score estimator model.

IV. Experimental Results and Analysis

A. Performance Measures

To benchmark our proposed method against state-of-the-art methods, we use three common performance measures: Pearson linear correlation coefficient (PLCC), Spearmen rank order correlation coefficient (SROCC), and root mean squared error (RMSE) between the predicted quality scores and subjective difference mean opinion scores (DMOS). Closer values to 1 for PLCC and SROCC and smaller amounts of RMSE are indicative of more accurate matching between objective and subjective scores.

A separate support vector regression (SVR) with an RBF kernel is trained for each dataset. The SVR estimates the quality of stereoscopic images. We train the SVR to produce scores in the range of DMOS values. Therefore, to evaluate the performance of this method on each dataset, we perform 1000 repetitions of the train-test process. This process is used by most of the learning based IQA methods [13]–[16], [32]. In each iteration, 80 percent randomly selected stereo images of the database are used for the training and the remaining 20 percent are used as the test set. Median of all 1000 obtained results, for each performance measure, is reported as the final result of the proposed model. This is done separately for the LIVE phase I and phase II databases. The obtained results for each dataset are compared with state-of-the-art SIQA methods.

B. Databases

We are presenting an objective method. To evaluate the performance of objective methods, they should be tested with datasets that have subjective scores. We use two popular subjective datasets of LIVE phase I [48] and LIVE phase II [25] to verify the performance of our proposed method. Most of the subjective IQA datasets provide subjective scores (DMOS) for each distorted stereo image pair. To create such databases subjective experiments are performed. Each image pair is shown in 3D displays to a number of human subjects who are wearing 3D glasses. Each person assigns a score to each image pair. For every image pair the average of difference opinion scores (DMOS) is reported. The stereo images of phase I dataset are symmetrically distorted while phase II dataset contains both symmetrically and asymmetrically distorted stereo-pairs.

1) LIVE 3D IQA Dataset Phase I: The database LIVE 3D IQA phase I consists of 365 distorted stereo image pairs generated from 20 pristine image pairs. Five types of distortions including JPEG, JPEG 2000 (JP2K), White Noise (WN), Gaussian Blur (Blur), and Fast-Fading (FF) model based on the Rayleigh fading channel are symmetrically applied to reference image pairs at different distortion levels. DMOS values in this database are in the range of -10 to 60.

2) LIVE 3D IQA Dataset Phase II: The LIVE 3D IQA phase II includes 8 reference stereo images distorted by five distortion types of JPEG, JP2K, WN, Blur and FF. Among 360 distorted stereo pairs in this dataset, 120 pairs are symmetrically distorted and the remaining 240 ones are distorted asymmetrically at various severities. The subjective quality scores of DMOS are in the range of 0 to 100.

C. Overall Performance Evaluation

In this section, the overall performance of the proposed no-reference SIQA algorithm is evaluated. The phase and contrast dictionaries are trained respectively on a set of 100,000 patches of size 8×8 randomly chosen from the phase and contrast of training stereoscopic image pairs. We take the original undistorted images from the LIVE II dataset as the training images to evaluate the proposed method on the LIVE I dataset and vice versa. The dictionary size is set to be $\mathrm{K} = 1000$ atoms for all dictionaries and the Recursive Least Squares Dictionary Learning (RLS-DL) [40] is used to train dictionaries. We utilize the method of LARS [45] to sparsely code the patches with the regularization parameter of $\lambda = 0.05$. The final feature vector for each stereo image is the concatenation of two vectors resulting from individually SPA pooling on sparse representation of

TABLE I
OVERALL PERFORMANCE OF THE PROPOSED METHOD AND OTHER STATE-OF-THE-ART METHODS IN TERMS OF PLCC, SROCC, AND RMSE ON LIVE PHASE I AND LIVE PHASE II DATABASES

Database		LIVE Phase I			LIVE Phase II		
Type	Method	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
FR	Benoit [6]	0.902	0.899	7.061	0.748	0.880	7.490
	FI-MS-SSIM [28]	0.695	0.902	–	0.729	0.712	–
	FI-WSNR [28]	0.853	0.901	–	0.705	0.684	–
	SDM-GSSIM [26]	0.933	0.925	7.857	–	–	–
	Shao [29]	0.935	0.903	5.816	0.863	0.849	5.706
	Bensalma [22]	0.887	0.875	**5.558**	0.770	0.751	7.204
	Shao [27]	0.925	0.922	6.252	0.759	0.745	7.355
	MS-SSIM [8]	0.917	0.916	6.533	0.900	0.889	4.987
	IDW-SSIM [30]	0.929	0.924	6.048	**0.915**	**0.918**	**4.549**
RR	Hewage [9]	0.830	0.814	9.139	0.891	0.501	9.365
	Wang [11]	0.892	0.889	7.408	–	–	–
NR	Akhter [14]	0.626	0.383	14.827	0.722	0.543	9.294
	Ryu [31]	0.800	0.860	7.930	–	–	–
	Chen [15]	0.895	0.891	7.247	0.895	0.880	5.102
	Shao [32]	**0.957**	**0.950**	–	–	–	–
	S3D-BLINQ [16]	–	–	–	0.913	0.905	4.657
	StereoQUE [17]	0.917	0.911	6.598	0.845	0.888	7.279
	Zhou [33]	0.928	0.887	6.025	0.861	0.823	5.779
	proposed	**0.963**	**0.958**	**4.566**	**0.959**	**0.951**	**3.289**

the phase and contrast image patches. Finally, a separate SVR model is learned for each dataset to estimate objective quality scores.

Best eighteen current 3D IQA methods have been selected to be compared with our method in terms of the above metrics. Among these methods the one proposed by Benoit [6], FI-MS-SSIM [28], FI-WSNR [28], SDM-GSSIM [26], [27], [29] by Shao, cyclopean MS-SSIM [8], IDW-SSIM [30] and Bensalma [22] are FR 3D IQA methods, Hewage [9] and Wang [11] are two RR methods and the NR 3D IQA methods are the ones proposed in [14], [31], [15], [32], [16], [17] and [33].

Among all these methods FI-MS-SSIM [28], FI-WSNR [28], SDM-GSSIM [26], [27], [29], Bensalma [22], IDW-SSIM [30], Ryu [31], [32] and [33] use only 2D information of stereo images while Benoit [6], cyclopean MS-SSIM [8], Hewage [9], Wang [11], Akhter [14], Chen [15], S3D-BLINQ [16] and StereoQUE [17] are 2D plus depth information based methods.

We compare the performance of our proposed approach with other 3D IQA methods on two selective databases LIVE Phase I and Phase II in terms of PLCC, SROCC and RMSE in Table I. The best two overall results are highlighted in bold.

Since FR methods have access to the reference stereopairs, they are expected to perform better than RR and NR 3D IQA approaches.

In LIVE Phase I dataset, our results not only surpass the best FR results but also are higher than those of RR and NR methods in terms of PLCC and SROCC as well as RMSE. The superiority of our method to other FR and RR methods on LIVE Phase II, is also acknowledged by the right part of Table I. The best NR reported results of PLCC and SROCC are improved about 5% by our proposed method. As can be seen, none of the NR methods is able to reach the FR method IDW-SSIM on LIVE Phase II. Table I shows that our results on this dataset are far better than those of IDW-SSIM for the first time.

D. Performance Evaluation on Individual Distortions

To evaluate the performance of our method more precisely, we compare the reported results of other methods with ours on each distortion type. Tables II, III and IV are containing per distortion results in terms of PLCC, SROCC and RMSE respectively. The left part of each table includes the results on LIVE Phase I and the results of LIVE Phase II are located in the right parts. The top two results in each column are highlighted in bold.

Similar to the overall results, most of the per-distortion values of our method on both datasets are still one of the top two in terms of PLCC, SROCC and RMSE.

In LIVE Phase I dataset and based on Table II, our PLCC results are superior to all other methods from 1% to 6% in all distortions except the Blur images that our results are about 1% lower than Shao's [32]. The relevant SROCC and RMSE values in Table III and Table IV have maintained their supremacy in all distortions. This excellence is very impressive about FF and JPEG compression so that our method has promoted their PLCC results by more than 3% and 6% respectively.

Since the LIVE phase II database includes both symmetrically and asymmetrically distorted stereo pairs, the performance results on this dataset is more meaningful to evaluate the methods in terms of modeling different cases of distortions. The right parts of the Tables II–IV infer that the performances are still the best on all distortions on LIVE Phase II except for the JPEG which is slightly lower than Chen's [15]. The improvement of performance results in all distortion groups by our proposed method is indicative of its power to mimic the binocular aspects of human quality perception.

E. Performance Evaluation on Symmetric and Asymmetric Distorted Images

To examine the capability of the state-of-the-art SIQA methods to deal with symmetrically and asymmetrically distorted images, we list the performance of the FR and NR algorithms with best overall performances, on related subsets of LIVE phase II dataset in Table V. It can be seen that most of the models can predict scores that correlate well with subjective evaluations when the stereo images are symmetrically distorted. However, their performances drop down for asymmetric distortions. Among those, our proposed NR SIQA method obtains the best performance on both symmetric and asymmetric distortions as well as overall achievements. Another important observation is that the performance gap between the results on symmetric and asymmetric distorted parts is filled somewhat by the proposed method.

F. Comparison With 2D IQA Methods

Even though 2D IQA methods are not expected to provide satisfying results on stereo images, we compare our results with three commonly used FR 2D IQA metrics: PSNR, SSIM [1] and

TABLE II
PLCC VALUES FROM OUR MODELS AND OTHER 3D IQA METHODS, PERFORMED ON LIVE PHASE I AND LIVE PHASE II DATABASES

Database		LIVE phase I					LIVE phase II				
Type	Method	WN	JP2K	JPEG	Blur	FF	WN	JP2K	JPEG	Blur	FF
FR	Benoit [6]	0.925	0.935	0.640	0.948	0.747	0.926	0.784	0.853	0.535	0.807
	FI-MS-SSIM [28]	0.814	0.840	0.579	0.817	0.568	0.933	0.867	0.874	0.707	0.745
	FI-WSNR [28]	0.930	0.911	0.611	0.863	0.693	0.961	0.908	0.827	0.771	0.702
	SDM-GSSIM [26]	0.935	0.940	0.671	0.952	0.865	–	–	–	–	–
	Shao [29]	0.945	0.921	0.520	0.959	0.859	0.946	0.782	0.747	0.958	0.905
	Bensalma [22]	0.915	0.839	0.380	0.937	0.734	0.944	0.667	0.858	0.908	0.909
	Shao [27]	0.941	0.924	0.656	0.951	0.840	0.850	0.838	0.750	0.827	0.881
	MS-SSIM [8]	0.942	0.912	0.603	0.942	0.776	0.957	0.834	0.862	0.960	0.901
	IDW-SSIM [30]	0.939	0.929	0.692	0.945	0.821	0.945	0.861	0.873	0.974	0.939
RR	Hewage [9]	0.895	0.904	0.530	0.798	0.669	0.891	0.664	0.734	0.450	0.746
	Wang [11]	0.913	0.916	0.570	0.957	0.783	–	–	–	–	–
NR	Akhter [14]	0.904	0.905	0.729	0.617	0.503	0.722	0.776	0.786	0.795	0.674
	Ryu [31]	0.940	0.860	0.630	0.960	0.780	**-**	**-**	**-**	**-**	**-**
	Chen [15]	0.917	0.907	0.695	0.917	0.735	0.947	**0.899**	**0.901**	0.941	0.932
	Shao [32]	0.938	**0.950**	0.796	**0.986**	0.837	–	–	–	–	–
	S3D-BLINQ [16]	–	–	**-**	–	–	**0.953**	0.847	**0.888**	0.968	**0.944**
	StereoQUE [17]	0.919	0.938	**0.806**	0.881	0.758	0.920	0.867	0.829	0.878	0.836
	Zhou [33]	**0.945**	0.915	0.695	0.973	**0.861**	0.936	0.781	0.757	**0.983**	0.900
	Proposed	**0.951**	**0.973**	**0.867**	**0.976**	**0.890**	**0.970**	**0.946**	0.869	**0.991**	**0.960**

TABLE III
SROCC VALUES FROM OUR MODELS AND OTHER 3D IQA METHODS, PERFORMED ON LIVE PHASE I AND LIVE PHASE II DATABASES

Database		LIVE phase I					LIVE phase II				
Type	Method	WN	JP2K	JPEG	Blur	FF	WN	JP2K	JPEG	Blur	FF
FR	Benoit [6]	0.930	0.910	0.603	0.931	0.699	0.923	0.751	**0.867**	0.900	0.933
	FI-MS-SSIM [28]	0.931	0.897	0.562	0.933	0.693	0.929	0.849	0.858	0.746	0.709
	FI-WSNR [28]	0.930	0.905	0.577	0.932	0.660	0.955	**0.901**	0.807	0.757	0.684
	SDM-GSSIM [26]	–	–	–	–	–	–	–	–	–	–
	Shao [29]	0.941	0.894	0.495	**0.940**	0.796	**0.965**	0.785	0.733	**0.920**	0.891
	Bensalma [22]	0.906	0.817	0.328	0.916	0.650	0.939	0.804	0.846	0.884	0.874
	Shao [27]	0.943	0.875	0.615	0.937	0.781	0.846	0.848	0.720	0.801	0.851
	MS-SSIM [8]	**0.948**	0.888	0.530	0.925	0.707	0.940	0.814	0.843	0.908	0.884
	IDW-SSIM [30]	0.928	0.891	0.629	0.924	0.732	0.944	0.848	0.861	0.911	**0.935**
RR	Hewage [9]	0.940	0.856	0.500	0.690	0.545	0.880	0.598	0.736	0.028	0.684
	Wang [11]	0.907	0.883	0.542	0.925	0.655	–	–	–	–	–
NR	Akhter [14]	0.914	0.866	0.675	0.555	0.640	0.714	0.724	0.649	0.682	0.559
	Ryu [31]	–	–	–	–	–	**-**	**-**	**-**	**-**	**-**
	Chen [15]	0.919	0.863	0.617	0.878	0.652	**0.950**	0.867	**0.867**	0.900	0.933
	Shao [32]	0.935	**0.936**	**0.818**	0.927	0.814	–	–	–	–	–
	S3D-BLINQ [16]	–	–	–	–	–	0.946	0.845	0.818	0.903	0.899
	StereoQUE [17]	0.910	0.917	0.782	0.865	0.666	0.932	0.864	0.839	0.846	0.860
	Zhou [33]	0.915	0.824	0.614	0.916	**0.867**	0.891	0.717	0.593	0.903	0.891
	Proposed	**0.943**	**0.950**	**0.835**	**0.943**	**0.843**	0.943	**0.919**	0.823	**0.952**	**0.940**

MS-SSIM [2]. Also three state-of-the-art NR 2D IQA methods including: BLIND/Referenceless Image Spatial Quality Evaluator (BRISQUE) [5], Distortion Identification-based Image Verity and Integrity Evaluation (DIIVINE) [4], and Blind Integrity Notator using DCT Statistics-II (BLIINDS-II) [3] are studying in the Table VI. The NR methods employ natural scene statistics changes in spatial domain, wavelet domain and DCT domain respectively. To utilize the 2D approaches in 3D case, we assess the quality of left and right images, individually and report the average of them as the 3D quality of the stereo image pair. The overall results of the 2D methods on LIVE Phase I and Phase II, in terms of PLCC and SROCC are listed in Table VI. The results in both datasets indicate the domination of our method over the top 2D IQA methods. The performance results by the 2D techniques show that 3D quality perception of stereo images is more than estimating the quality of left and right images especially about the LIVE Phase II that contains both symmetrically and asymmetrically distorted images. The results demonstrate that our proposed method is successful in modeling the human eye in the case of 3D quality perception.

TABLE IV
RMSE VALUES FROM OUR MODELS AND OTHER 3D IQA METHODS, PERFORMED ON LIVE PHASE I AND LIVE PHASE II DATABASES

Database		LIVE phase I					LIVE phase II				
Type	Method	WN	JP2K	JPEG	Blur	FF	WN	JP2K	JPEG	Blur	FF
FR	Benoit [6]	6.307	4.426	5.022	4.571	8.257	4.028	6.096	3.878	11.763	6.894
	FI-MS-SSIM [28]	–	–	–	–	–	–	–	–	–	–
	FI-WSNR [28]	–	–	–	–	–	–	–	–	–	–
	SDM-GSSIM [26]	7.853	5.909	6.465	5.919	8.312	–	–	–	–	–
	Shao [29]	–	–	–	–	–	–	–	–	–	–
	Bensalma [22]	–	–	–	–	–	–	–	–	–	–
	Shao [27]	–	–	–	–	–	–	–	–	–	–
	MS-SSIM [8]	5.581	5.320	5.216	4.822	7.837	**3.368**	5.562	**3.365**	3.747	4.966
	IDW-SSIM [30]	**4.596**	**3.814**	**4.040**	**3.409**	**5.607**	3.492	4.998	3.573	3.126	**3.944**
RR	Hewage [9]	7.405	5.530	5.543	8.748	9.226	10.13	7.343	4.976	12.436	7.667
	Wang [11]	6.777	5.189	5.374	4.178	7.725	–	–	–	–	–
NR	Akhter [14]	7.092	5.483	4.273	11.387	9.332	7.416	6.189	4.535	8.450	8.505
	Ryu [31]	**-**	–	–	–	–	**-**	**-**	**-**	**-**	**-**
	Chen [15]	6.433	5.402	4.523	5.898	8.322	3.513	**4.298**	**3.342**	4.725	4.180
	Shao [32]	–	–	–	–	–	–	–	–	–	–
	S3D-BLINQ [16]	–	–	–	–	–	3.547	5.482	4.169	4.453	4.199
	StereoQUE [17]	6.664	4.943	4.391	6.938	9.317	4.325	5.087	4.756	6.662	6.519
	Zhou [33]	**5.086**	4.999	4.286	**3.127**	5.750	3.575	5.802	4.502	**2.455**	4.375
	Proposed	5.564	**3.284**	**3.343**	3.565	**5.622**	**2.892**	**3.528**	3.783	**2.130**	**3.347**

TABLE V
COMPARISON OF STATE-OF-THE-ART SIQA METHODS ON SYMMETRICALLY AND ASYMMETRICALLY DISTORTED STIMULI ON LIVE PHASE II DATABASE IN TERMS OF PLCC AND SROCC

		PLCC			SROCC		
Type	Method	Sym	Asym	All	Sym	Asym	All
FR	Benoit [6]	0.734	0.770	0.762	0.696	0.747	0.744
	MS-SSIM [25]	0.938	0.875	0.907	0.925	0.854	0.901
	IDW-SSIM [30]	0.937	0.898	0.916	0.923	0.902	0.919
NR	Chen [15]	–	–	0.895	0.918	0.834	0.880
	S3D-BLINQ [16]	–	–	0.913	0.937	0.849	0.905
	StereoQUE [17]	–	–	0.845	0.857	0.872	0.888
	Proposed	**0.963**	**0.952**	**0.958**	**0.946**	**0.935**	**0.950**

TABLE VI
PLCC AND SROCC VALUES OF OUR PROPOSED METHOD VERSUS STATE-OF-THE-ART 2D IQA METHODS, PERFORMED ON LIVE PHASE I AND LIVE PHASE II DATABASES

Database		LIVE phase I		LIVE phase II	
Type	Method	PLCC	SROCC	PLCC	SROCC
FR	PSNR	0.834	0.834	0.665	0.665
	SSIM [1]	0.872	0.876	0.792	0.792
	MS-SSIM [2]	0.926	0.926	0.777	0.776
NR	BRISQUE [5]	0.910	0.901	0.749	0.701
	DIIVINE [4]	0.939	0.929	0.697	0.669
	BLINDS-II [3]	0.917	0.910	0.736	0.700
	Proposed	**0.963**	**0.958**	**0.959**	**0.951**

TABLE VII
PLCC COMPARISON FOR EACH COMPONENT OF THE PROPOSED SCHEME ON LIVE PHASE I AND LIVE PHASE II DATASETS

	LIVE I					
Components	**WN**	**JP2K**	**JPEG**	**Blur**	**FF**	**All**
Phase	0.898	0.953	0.800	0.953	0.834	0.925
Contrast	**0.948**	0.973	**0.868**	**0.978**	**0.903**	**0.963**
Phase-Contrast	0.946	**0.973**	0.866	0.976	0.899	0.961
	LIVE II					
Phase	0.953	0.925	0.839	0.979	0.943	0.942
Contrast	0.970	0.938	**0.879**	0.983	0.935	0.945
Phase-Contrast	**0.970**	**0.946**	0.870	**0.990**	**0.959**	**0.958**

G. Effect of Parameters

The proposed framework includes a number of parameters such as components of the feature vectors, dictionary learning algorithm, the number of code words in the dictionary, the sparsity level and the feature pooling method. We investigate different conditions of the parameters individually in this section.

1) Phase and Contrast Components: To demonstrate the impact of phase and contrast components in the proposed approach, we retried our 1000 train-test iterations on the two separate sets of features. The results of PLCC for phase only, contrast only and the final model which uses both of them are illustrated in Table VII. Even though using features extracted from contrast images is more efficient than the features extracted from phase on LIVE phase I dataset, the performance of the main model is superior to both of the single models on LIVE phase II. It can also be argued that both phase and contrast components are effective in order to assessing the quality of stereo images.

2) Pooling Schemes: The effectiveness of SPA pooling can be verified in Table VIII, and Fig. 4 where the PLCC of subjective and objective scores using six different pooling strategies are illustrated. While the performance indices of the proposed approach with NNZ, SPA and AVG poolings closely are better than other three ones on LIVE Phase I dataset, on

TABLE VIII
PERFORMANCE COMPARISON OF THE PROPOSED METHOD USING DIFFERENT POOLING METHODS IN TERMS OF PLCC ON LIVE PHASE I AND LIVE PHASE II DATASETS

Pooling	WN	JP2K	JPEG	Blur	FF	All
			LIVE I			
HMax	0.580	0.607	0.256	0.761	0.653	0.707
Max	0.949	0.950	0.783	0.963	0.874	0.951
AVG	**0.976**	0.965	0.849	0.972	0.850	0.960
NNZ	0.957	0.973	0.844	0.974	**0.903**	**0.964**
SPM	0.947	0.967	0.833	0.970	0.872	0.955
SPA	0.948	**0.973**	**0.867**	**0.976**	0.885	0.961
			LIVE II			
HMax	0.815	0.402	0.649	0.951	0.671	0.721
Max	0.968	0.941	0.835	0.980	0.909	0.934
AVG	0.970	0.941	**0.877**	0.980	0.844	0.947
NNZ	**0.973**	0.911	0.868	0.990	**0.964**	0.951
SPM	0.956	0.939	0.845	0.989	0.953	0.947
SPA	0.970	**0.945**	0.873	**0.991**	0.960	**0.958**

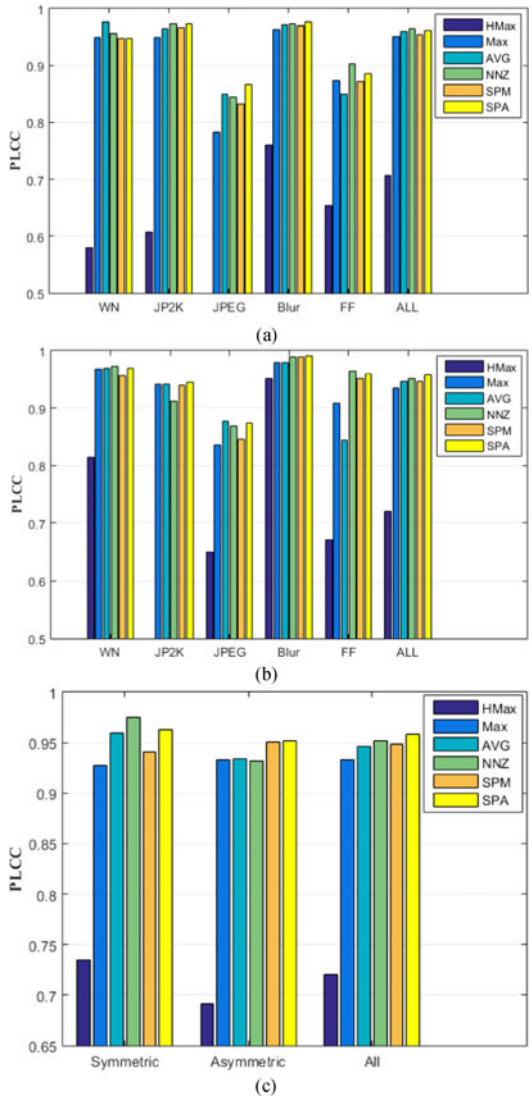

Fig. 4. Performance comparison of the proposed method using different pooling methods in terms of PLCC on (a) LIVE Phase I, (b) LIVE Phase II, and (c) on symmetrically and asymmetrically distorted parts of LIVE Phase II database.

TABLE IX
COMPARISON OF CORRELATION RESULTS OF THE PROPOSED METHOD USING DIFFERENT POOLING METHODS ON SYMMETRICALLY AND ASYMMETRICALLY DISTORTED STIMULI ON LIVE PHASE II DATABASE

Method	PLCC			SROCC		
	Sym	Asym	All	Sym	Asym	All
HMax	0.735	0.691	0.720	0.728	0.610	0.675
Max	0.927	0.933	0.933	0.917	0.910	0.923
AVG	0.960	0.934	0.946	0.935	0.917	0.941
NNZ	**0.975**	0.932	0.952	**0.957**	0.910	0.945
SPM	0.941	0.951	0.948	0.925	0.933	0.942
SPA	0.963	**0.952**	**0.958**	0.946	**0.935**	**0.950**

TABLE X
PERFORMANCE COMPARISON OF THE PROPOSED METHOD USING DIFFERENT DICTIONARY LEARNING ALGORITHMS IN TERMS OF PLCC ON LIVE PHASE I AND LIVE PHASE II

Dictionary Learning	WN	JP2K	JPEG	Blur	FF	All
			LIVE I			
RLS-DL [40]	0.946	0.973	0.866	0.976	0.899	0.961
KSVD [41]	0.947	0.973	0.868	0.974	0.890	0.959
ODL [43]	0.947	0.971	0.866	0.971	0.887	0.960
			LIVE II			
RLS-DL [40]	0.970	0.946	0.870	0.990	0.959	0.958
KSVD [41]	0.968	0.951	0.872	0.991	0.957	0.958
ODL [43]	0.966	0.950	0.870	0.991	0.962	0.958

database LIVE Phase II, SPA performs better than NNZ and AVG. The SPM, Max and HMax results are respectively lower than the other three pooling methods.

We repeated the experiment for symmetric and asymmetric parts of database LIVE phase II to evaluate different pooled features on asymmetrically distorted stereo image pairs. In Table IX and Fig. 4(c) it can be seen that although the performance results of SPA on symmetric distorted stereo images are not as good as NNZ, its overall performances and those of asymmetric distorted pairs has defeated all the existing methods.

3) Dictionary Learning Algorithm: We test three different well known dictionary learning methods in our feature learning framework including K-SVD [41], RLS-DL [40], and online dictionary learning (ODL) [43] algorithms. The K-SVD and RLS-DL methods make use of ℓ_0-pseudo-norm as the sparsity-inducing regularization function, while an ℓ_1 -regularized dictionary learning problem is the objective in ODL. Using these methods, the dictionaries $\mathbf{D}_P \in \mathbb{R}^{n\times K}$ and $\mathbf{D}_C \in \mathbb{R}^{n\times K}$ are trained with $K = 1000$ from 8×8 patches of training phase and contrast images respectively. Refer to Table X it seems there is not much difference between the quality assessment results achieved by different dictionary learning methods. We select RLS-DL dictionary learning algorithm just because of its slight lead in overall correlation results over the LIVE phase I dataset.

4) Dictionary Size: The number of dictionary atoms is an important parameter of a dictionary which is mainly set by

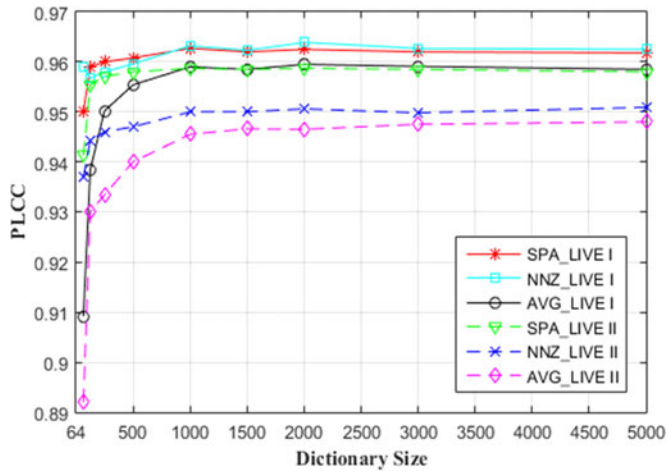

Fig. 5. Accuracy of the proposed method in terms of PLCC with various sizes of dictionaries on LIVE Phase I and LIVE Phase II.

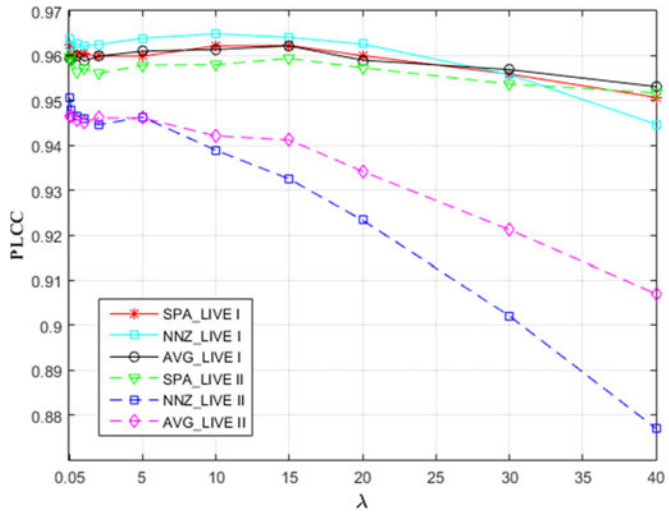

Fig. 6. Accuracy of the proposed method in terms of PLCC with different values of sparsity regularization parameter λ on LIVE Phase I and LIVE Phase II.

experience or based on empirical evaluations. We carry out quantitative experiments to study the influence of this parameter on the performance of our quality assessment method. We test different sizes of dictionaries for best three feature pooling schemes in our algorithm. As shown in Fig. 5 as the number of dictionary atoms grows more than 1000, the evaluation accuracy of the proposed method does not improve significantly, in terms of PLCC on both of the LIVE Phase I and LIVE Phase II. This indicates that the proposed method does not highly depend on huge sizes of dictionary. Another important observation is that for both of the LIVE phase I and LIVE phase II datasets using SPA can result in the best correlation results and early convergence to its final performance with small sizes of dictionaries. Therefore, using SPA, we can select very small sizes of dictionaries while the performance results do not drop yet.

5) Sparsity Regularization Parameter: As it is mentioned in Section III-C, an ℓ_1 -regularized sparse coding is adopted in our method for encoding of local image features. This sparse coding is performed by solving the convex optimization problem in (22) in which the ℓ_1-norm encourages the solution to be sparse and the parameter $\lambda \geq 0$ is used to control the trade-off between data fitting and the sparsity of solution. In general, increasing λ leads to sparser solution. To illustrate the effect of changing the value of this parameter on the final performance of our method, PLCC values between subjective and objective scores on datasets LIVE Phase I and LIVE Phase II for different λ values are plotted in Fig. 6. It is observed that the performance of our approach is almost consistent with $\lambda < 5$. It can be seen that comparing to NNZ and AVG, the SPA pooling method not only provides better results, but also is more robust against different values of parameter λ on both of the datasets.

V. Conclusion

In this paper we employed the physiological discoveries in 3D perception of human vision to propose a 3D quality assessment method for stereo images. The perceived phase and contrast of the cyclopean wave were produced in a binocular combination manner. An efficient general-purpose algorithm for NR SIQA problem was presented that outwent the state-of-the-art. We used the sparse representation of phase and contrast patches as local descriptors. A spatial pyramidal pooling of the patch descriptors also provided a representation of images. The unsupervised sparse feature representation framework was adaptable to all types of distortions and strengths. The proposed algorithm outperformed the current 2D and 3D IQA methods on both LIVE Phase I and LIVE Phase II datasets.

References

[1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[2] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Conf. Rec. 37th Asilomar Conf. Signals, Syst. Comput.*, 2004, pp. 1398–1402.

[3] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 583–586, Jun. 2010.

[4] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[5] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

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

[7] J. Yang, C. Hou, R. Xu, and J. Lei, "New metric for stereo image quality assessment based on HVS," *Int. J. Imag. Syst. Technol.*, vol. 20, no. 4, pp. 301–307, 2010.

[8] 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.

[9] C. T. Hewage and M. G. Martini, "Reduced-reference quality metric for 3D depth map transmission," in *Proc. 2010 Conf. True Vis.-Capture, Transm. Display 3D Video*, 2010, pp. 1–4.

[10] 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.

[11] X. Wang, Q. Liu, R. Wang, and Z. Chen, "Natural image statistics based 3D reduced reference image quality assessment in contourlet domain," *Neurocomputing*, vol. 151, no. 2, pp. 683–691, 2015.

[12] M. Solh and G. AlRegib, "A no-reference quality measure for DIBR-based 3D videos," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2011, pp. 1–6.
[13] Z. P. Sazzad, S. Yamanaka, and Y. Horita, "Spatio-temporal segmentation based continuous no-reference stereoscopic video quality prediction," in *Proc. 2nd Int. Workshop Qual. Multimedia Experience*, 2010, pp. 106–111.
[14] R. Akhter, Z. P. Sazzad, Y. Horita, and J. Baltes, "No-reference stereoscopic image quality assessment," *Proc. SPIE*, vol. 7524, pp. 75240T-1–75240T-12, 2010.
[15] M.-J. Chen, L. K. 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.
[16] C.-C. Su, L. K. Cormack, and A. C. Bovik, "Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1685–1699, May 2015.
[17] B. Appina, S. Khan, and S. S. Channappayya, "No-reference stereoscopic image quality assessment using natural scene statistics," *Signal Process., Image Commun.*, vol. 43, pp. 1–14, 2016.
[18] P. Campisi, P. Le Callet, and E. Marini, "Stereoscopic images quality assessment," in *Proc. 15th Eur. Signal Process. Conf.*, 2007, pp. 2110–2114.
[19] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[20] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *Proc. SPIE*, vol. 5666, pp. 149–159, 2005.
[21] C. T. Hewage, S. T. Worrall, S. Dogan, S. Villette, and A. M. Kondoz, "Quality evaluation of color plus depth map-based stereoscopic video," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 304–318, Apr. 2009.
[22] R. Bensalma and M.-C. Larabi, "Towards a perceptual quality metric for color stereo images," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 4037–4040.
[23] L. Jin, A. Boev, A. Gotchev, and K. Egiazarian, "3D-DCT based perceptual quality assessment of stereo video," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2521–2524.
[24] S. Ryu, D. H. Kim, and K. Sohn, "Stereoscopic image quality metric based on binocular perception model," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep./Oct. 2012, pp. 609–612.
[25] W. Hachicha, A. Beghdadi, and F. A. Cheikh, "Stereo image quality assessment using a binocular just noticeable difference model," in *Proc. 20th IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 113–117.
[26] J. Yang, Y. Liu, Z. Gao, R. Chu, and Z. Song, "A perceptual stereoscopic image quality assessment model accounting for binocular combination behavior," *J. Vis. Commun. Image Represent.*, vol. 31, pp. 138–145, 2015.
[27] 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.
[28] Y.-H. Lin and J.-L. Wu, "Quality assessment of stereoscopic 3D image compression by binocular integration behaviors," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1527–1542, Apr. 2014.
[29] F. Shao *et al.*, "Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 2971–2983, Oct. 2015.
[30] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3400–3414, Nov. 2015.
[31] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 24, no. 4, pp. 591–602, Apr. 2014.
[32] F. Shao, K. Li, W. Lin, G. Jiang, and M. Yu, "Using binocular feature combination for blind quality assessment of stereoscopic images," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1548–1551, Oct. 2015.
[33] W. Zhou and L. Yu, "Binocular responses for no-reference 3D image quality assessment," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1077–1084, Jun. 2016.
[34] X. Ren and D. Ramanan, "Histograms of sparse codes for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 3246–3253.
[35] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 1098–1105.
[36] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2009, pp. 1794–1801.
[37] J. Ding, S. A. Klein, and D. M. Levi, "Binocular combination of phase and contrast explained by a gain-control and gain-enhancement model," *J. Vis.*, vol. 13, no. 2, pp. 1–37, 2013.
[38] D. J. Fleet, H. Wagner, and D. J. Heeger, "Neural encoding of binocular disparity: Energy models, position shifts and phase shifts," *Vis. Res.*, vol. 36, no. 12, pp. 1839–1857, 1996.
[39] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.
[40] K. Skretting and K. Engan, "Recursive least squares dictionary learning algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2121–2130, Apr. 2010.
[41] M. Aharon, M. Elad, and A. Bruckstein, "KSVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[42] K. Engan, S. O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 1999, vol. 5, pp. 2443–2446.
[43] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 689–696.
[44] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, 2010.
[45] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Stat.*, vol. 32, no. 2, pp. 407–499, 2004.
[46] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, Jun. 2006, vol. 2, pp. 2169–2178.
[47] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.
[48] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Process., Image Commun.*, vol. 28, no. 8, pp. 870–883, 2013.

Maryam Karimi received the B.Sc. degree in computer engineering from the Amirkabir University of Technology, Tehran, Iran, in 2006, the M.Sc. degree in computer engineering from the Sharif University of Technology, Tehran, in 2009, and is currently working toward the Ph.D. degree electrical and computer engineering at the Isfahan University of Technology, Isfahan, Iran.

She is also a Visiting Researcher with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. Her research interests include subjective and objective image quality assessment and perceptual modeling.

Mansour Nejati received the B.S. degree (Hons.) in electrical engineering from Azad University, Najafabad Branch, Iran, in 2006, and the M.S. and Ph.D. degrees in electrical engineering from the Isfahan University of Technology, Isfahan, Iran, in 2010 and 2016, respectively.

His research interests include image processing, sparse representations, and dictionary learning.

S. M. Reza Soroushmehr received the B.Sc., M.Sc., and Ph.D. degrees from the Isfahan University of Technology, Isfahan, Iran, in 2000, 2004 and 2013, respectively.

He was previously a Postdoctoral Fellow with the Electrical and Computer Engineering Department, McMaster University, Hamilton, ON, Canada. He is currently a Postdoctoral Research Fellow with the Emergency Medicine Department, University of Michigan, Ann Arbor, MI, USA, where he is conducting research on medical image processing. His main research interests include image processing, video compression, algorithm design, and optimization.

Shadrokh Samavi (M'08) received the B.S. degree in industrial technology from California State University, Los Angeles, CA, USA, in 1980, the B.S. degree in electrical engineering from California State University, Bakersfield, CA, USA, in 1983, the M.S. degree in computer engineering from the University of Memphis, Memphis, TN, USA, in 1985, and the Ph.D. degree in electrical engineering from Mississippi State University, Starkville, MS, USA, in 1989.

He is currently a Professor of computer engineering with the Isfahan University of Technology, Isfahan, Iran. He is also an Adjunct Professor with the ECE Department, McMaster University, Hamilton, ON, Canada, where he is a member of the Multimedia Signal Processing Laboratory. He is also a Research Affiliate with the Biomedical and Clinical Informatics Laboratory, Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. His research interests include image processing and hardware implementation and optimization of image processing algorithms. He is also interested in compression and processing of biomedical images, and VLSI design and computer arithmetic.

Prof. Samavi is a Registered Professional Engineer, USA. He is a Member of the Eta Kappa Nu and Tau Beta Pi honor societies.

Nader Karimi received the B.S. degree (*summa cum laude*) in computer engineering from Azad University, Arak Branch, Iran, in 2002, and the M.Sc. and Ph.D. degrees (Hons.) in computer engineering and electrical engineering from the Isfahan University of Technology, Isfahan, Iran, in 2004 and 2012, respectively.

He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Isfahan University of Technology. His research interests include image compression, hardware implementation and optimization of image processing algorithms, and watermarking.

Kayvan Najarian received the B.Sc. degree in electrical engineering from Sharif University, Tehran, Iran, in 1990, the M.Sc. degree in biomedical engineering from Amirkabir University, Tehran, Iran, in 1994, and the Ph.D. degree in electrical and computer engineering from the University of British Columbia, Vancouver, BC, Canada, in 2000.

He is currently an Associate Professor with the Department of Computational Medicine and Bioinformatics and the Department of Emergency Medicine, University of Michigan, Ann Arbor, MI, USA. He also serves as the Director of the Michigan Center for Integrative Research in Critical Care's Biosignal, Image and Computational Core program. His research interests include design of signal/image processing and machine learning methods to create computer assisted clinical decision support systems that improve patient care.

Prof. Najarian serves as the Editor-in-Chief of a journal in the field of biomedical engineering and the Associate Editor of two journals in the field of biomedical informatics. He is also a member of many editorial boards and has served as a Guest Editor of special issues for several journals.