Signal Processing: Image Communication 58 (2017) 287–299

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment

Lixiong Liu [a], Bao Liu [a], Che-Chun Su [b], Hua Huang [a,*], Alan Conrad Bovik [b]

[a] Beijing Laboratory of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China
[b] Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Keywords:
Stereopair quality assessment
No-reference
Binocular rivalry
Spatial activity
Reverse saliency

ABSTRACT

We develop a new model for no-reference 3D stereopair quality assessment that considers the impact of binocular fusion, rivalry, suppression, and a reverse saliency effect on the perception of distortion. The resulting framework, dubbed the S3D INtegrated Quality (SINQ) Predictor, first fuses the left and right views of a stereopair into a single synthesized cyclopean image using a novel modification of an existing binocular perceptual model. Specifically, the left and right views of a stereopair are fused using a measure of "cyclopean" spatial activity. A simple product estimate is also calculated as the correlation between left and right disparity-corrected corresponding binocular pixels. Univariate and bivariate statistical features are extracted from the four available image sources: the left view, the right view, the synthesized "cyclopean" spatial activity image, and the binocular product image. Based on recent evidence regarding the placement of 3D fixation by subjects viewing stereoscopic 3D (S3D) content, we also deploy a reverse saliency weighting on the normalized "cyclopean" spatial activity image. Both one- and two-stage frameworks are then used to map the feature vectors to predicted quality scores. SINQ is thoroughly evaluated on the LIVE 3D image quality database (Phase I and Phase II). The experimental results show that SINQ delivers better performance than state of the art 2D and 3D quality assessment methods on six public databases, especially on asymmetric distortions.

1. Introduction

In the past two decades there has been tremendous progress on both the theoretical and practical aspects of visual quality assessment (QA) [1]. The problem of 2D picture quality prediction has been intensively studied [2–7]; however, the field of 3D perceptual quality assessment is relatively nascent [8]. Although 3D capture and display technologies have greatly advanced and become widely affordable, progress still needs to be made on developing better tools that can effectively evaluate, maintain and improve the quality of 3D visual signals [8]. With the rapid growth of available 3D content, the need for accurate and efficient methods for assessing the quality of 3D visual signals is increasing.

Here we consider the problem of No-Reference Stereoscopic 3D (S3D) picture quality prediction, hence no reference signal is assumed present. Several studies have been presented to address the problem of NR 3D visual quality assessment. Akhter et al. proposed a NR 3D QA algorithm which extracts features from stereopairs and an estimated disparity map [9]. Ryu et al. discussed the relationship between stereo quality and distortion types [10] and Chen et al. proposed an NR method that extracts features from synthesized "cyclopean" images, computed disparity maps and uncertainty maps [11]. Su et al. modeled stereopairs with bivariate statistical models [12,13] and proposed a 3D blind image naturalness quality index based on both univariate and bivariate natural scene statistics (NSS) models [14]. Su et al. proposed an improved cyclopean image synthesis model, and extracted perceptually relevant NSS features using bivariate statistical models from spatially adjacent bandpass coefficients of the synthesized cyclopean images [14]. Natural scene statistics features measured on 3D visual coordinates propel the NR 3D IQA model in [15]. Lopez et al. [16] proposed an S3D video quality assessment method based on depth maps and video motion. While these early advances on the NR S3D IQA problem are promising, further improvements are possible by better accounting for human stereo perception mechanisms. Algorithms for conducting S3D IQA have been developed by modeling a variety of binocular vision characteristics, e.g., binocular energy responses [17],

* Corresponding author.
E-mail addresses: lxliu@bit.edu.cn (L. Liu), liubao@bit.edu.cn (B. Liu), ccsu@utexas.edu (C.-C. Su), huahuang@bit.edu.cn (H. Huang), bovik@ece.utexas.edu (A.C. Bovik).

http://dx.doi.org/10.1016/j.image.2017.08.011
Received 17 December 2016; Received in revised form 19 August 2017; Accepted 20 August 2017
Available online 30 August 2017
0923-5965/

gain-control mechanisms [18], models of visual information extraction [10] and of Gabor filter responses [19]. However, there is still room for improvement beyond available S3D IQA models with regards to correlation with human subjective judgments. Here we develop a way to account for both binocular perception and the differences in low-level content between the two views in a new NR S3D picture quality assessment model. The proposed model and resulting algorithm contains three main innovations. First, we use a simple measure of spatial activity to model the binocular perception process by calculating the differences between the two views arising from distortion, and use it to modify an existing method of synthesizing a cyclopean image. Second, based on recent evidence regarding the placement of 3D fixations by humans viewing S3D content, we deploy a model of visual reverse saliency to optimize the process of feature extraction. Lastly, we compute a novel disparity-corrected correlation image obtained by computing the products of corresponding pixels between the two stereopair images. The experimental results demonstrate the effectiveness of the proposed method in improving the prediction of the perceived quality of S3D picture content.

The rest of the paper is organized as follows. Section 2 reviews related S3D image quality assessment models. Section 3 describes the binocular perception model we use. We introduce the concept of the S3D product image or between-view empirical correlation in Section 4. The various features are integrated into a new NR stereopair quality assessment model, called SINQ. We evaluate the performance of SINQ against the state-of-the-art in Section 5 and conclude the paper in Section 6.

2. Related work

S3D image quality prediction models can be classified according to the types of 2D and 3D information that is computed from the stereoscopic pairs. The simplest 3D IQA algorithms do not incorporate depth/disparity information, instead using features drawn directly from the left and right view images. Yasakethu et al. [20] used various 2D IQA models to predict the quality of left and right view images independently, then synthesized both scores to obtain predictions of S3D quality. Most 2D-based 3D IQA algorithms belong to this category.

The deployment of models of stereoscopic perception and disparity computations can be used to improve on 2D-based models. You et al. applied a variety of 2D IQA algorithms on stereopairs as well as on disparity maps to predict perceptual quality [21]. Sazzad et al. [22] took disparity information into account to assess the quality of symmetrical and asymmetrical JPEG compression stereopairs. Zhao et al. proposed a binocular JND (BJND) model, which uses basic properties of binocular vision in response to asymmetric noise [23]. Lopez et al. utilized depth maps and video motion to measure the quality of a 3D video [16]. The authors of [24] proposed a theory that binocular perception is not uniformly distributed and can be classified into different types of perception patterns. Based on this theory, Shao et al. classified S3D images into non-corresponding, binocular fusion, and binocular suppression regions, and evaluated the quality of each region independently [24]. Lin et al. [25] assessed the quality of compression distortion considering binocular integration, and Silva et al. [19] proposed a novel video quality metric based on binocular suppression theory.

In addition to using depth/disparity information, accounting for the fused perception of a given S3D scene by computing an estimated cyclopean image is a useful way to improve S3D quality predictions. A growing number of algorithms have utilized a cyclopean image to construct better S3D IQA models. Maalouf et al. [26] utilized disparity information to compute a cyclopean image by averaging the left view and compensated right view. In [18], a new 3D image quality model specifically tailored for mobile 3D video was proposed, which considers a cyclopean view, binocular rivalry, and the scene geometry. More recently, Chen et al. developed a framework for assessing the quality of both symmetrically and asymmetrically distorted stereopairs [27]. In their framework, given the left and right view images, an estimated disparity map was computed using a stereo algorithm, while the perceptual responses are generated on the stereo images using a Gabor filter bank. Finally, a "cyclopean image" was synthesized from the stereo image pair, the estimated disparity map and the Gabor filter responses. Chen et al. further developed this framework by proposing an NR S3D IQA algorithm in [11]. Su et al. improved the cyclopean image synthesis model and proposed a 3D blind image naturalness quality index [14], which uses a parallel-viewing geometry to generate a convergent cyclopean image from the left and right views images, then extracts perceptually relevant NSS features from spatially adjacent bandpass coefficients of the cyclopean images. Yang et al. [28] proposed a saliency map to define an S3D model that processes a "cyclopean image". A "cyclopean image" synthesis framework is also adopted here in the proposed SINQ model, but with a different fusion strategy based on a measure of relative spatial activity.

3. Binocular spatial activity model

In order to analyze the mechanisms that drive binocular vision in an accessible way, we thoroughly consider the phenomena of binocular fusion, rivalry and suppression. The relationship between the amount of spatial activity, which is a measure of normalized image variance, contained in an S3D image pair [10] and perceived image quality is also studied. Levelt [29] demonstrated that low-level sensory factors can strongly influence binocular rivalry and suppression. He pointed out that strong stimuli, such as areas of high contrast or sharp edges tend to dominate rivalry between the two views. Following this idea, we construct a cyclopean fusion model based on the hypothesis that an image in a (grayscale) stereopair containing the greater amount of spatial activity will dominate perception.

3.1. Binocular fusion, rivalry and suppression

The human visual system (HVS) receives visual stimuli from the two eyes and combines them into a single combined percept [30]. When the two views share similar characteristics, binocular fusion can happen easily, driven by processes of vergence, accommodation, and other visual adaptations when viewing the real, natural world. However, binocular rivalry can occur when unusual dissimilar (dichoptic) monocular stimuli are presented to the corresponding retinal locations of the two eyes. Then, rather than perceiving a single, stable fused image, one either experiences alternations in perceptual awareness over time as the two stimuli compete for perceptual dominance, or diplopia, which is what usually occurs when viewing stereopairs containing asymmetric distortions. If one of the stimuli dominates the other, then binocular suppression will occur. A related challenging issue is how to model the process of binocular combination, when evaluating asymmetrically distorted stimuli [30]. Since the binocular perception of degrees or characteristics of distortion can greatly influence the perceived quality of S3D images [31], determining how to synthesize the two views to simulate the perception of a 3D scene is a key ingredient of a successful 3D images quality assessment model.

Normal viewing of stereo images in daily life usually results in binocular fusion rather than binocular rivalry, since the stereopairs share similar content. However, when viewing electronically displayed S3D content, the introduction of distortions may destroy this similarity, leading to binocular rivalry. This rivalry can modify the perception of the distortions [11,14,27,32], and can even lead to feelings of physiological discomfort [33]. Clearly, binocular rivalry is an important factor in S3D quality perception [18,19,24,25,27].

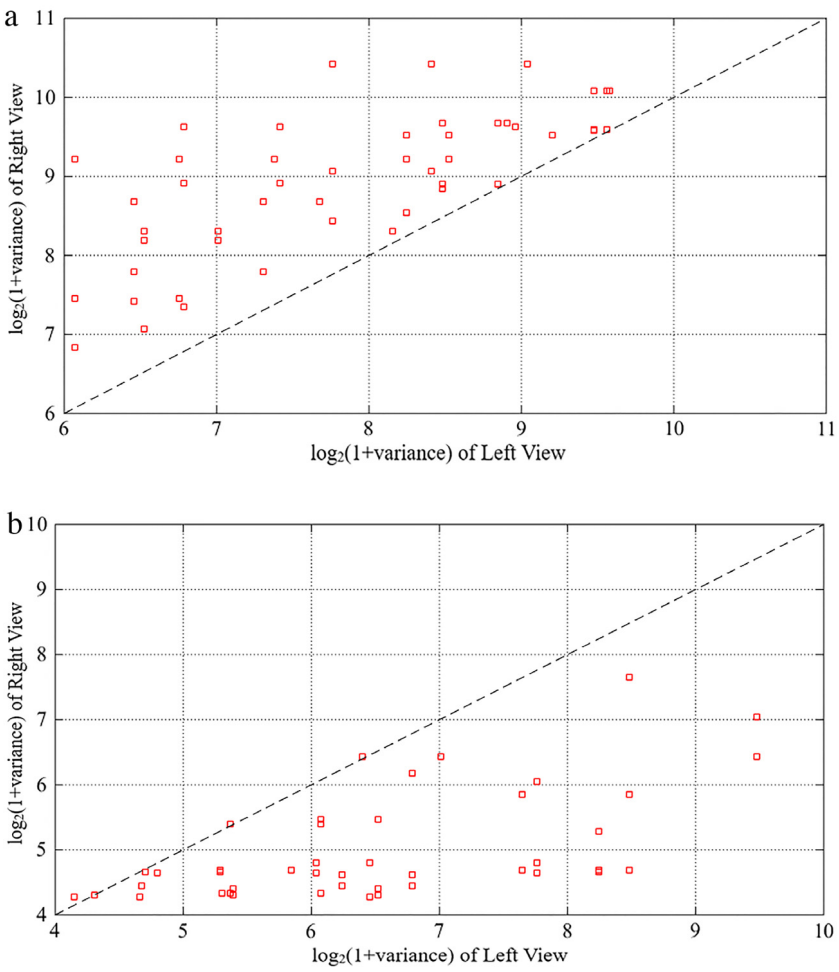

Fig. 1. Comparison of variances of corresponding left and right views of asymmetrically distorted stereopairs. (a) Noise distorted. (b) Blur distorted.

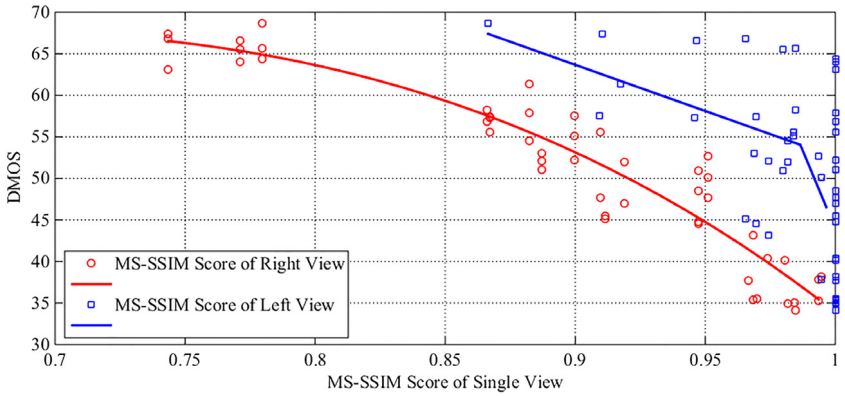

Fig. 2. Scatter plot of DMOS versus the corresponding MS-SSIM scores of asymmetrically noisy left/right views. The fitted curves show that DMOS correlates with the MS-SSIM scores of the right view images better than with the left view images. The SROCC between DMOS and MS-SSIM scores of right views was 0.9369, while the SROCC between DMOS and MS-SSIM scores of left views was 0.5180.

3.2. Spatial information and binocular perception

We conducted studies on blur and noise distorted images to further clarify the connection between binocular vision and these distortions. If one of the views is blurry and the other is sharp, the visual system may reduce the loss of high frequency information of the blurred view, resulting in a fused stereoscopic perception highly similar to the sharper view [34,35]. The perception of stereopairs tends to be dominated by the sharper, higher-quality image when one or both are distorted by blur [10,36]. However, when noise is introduced, the perception of stereopairs may be dominated by the low-quality component, i.e., a high-frequency distortion [10,33]. Blur tends to decrease the amount of spatial activity in images, while noise tends to increase the amount of spatial activity, i.e. by comparing the amount of spatial activity in the left and right views, we can determine the dominant factors affecting S3D quality prediction. Thus, our approach to S3D quality prediction operates under the hypothesis that the perception of S3D quality tends to be dominated by the image containing the greater

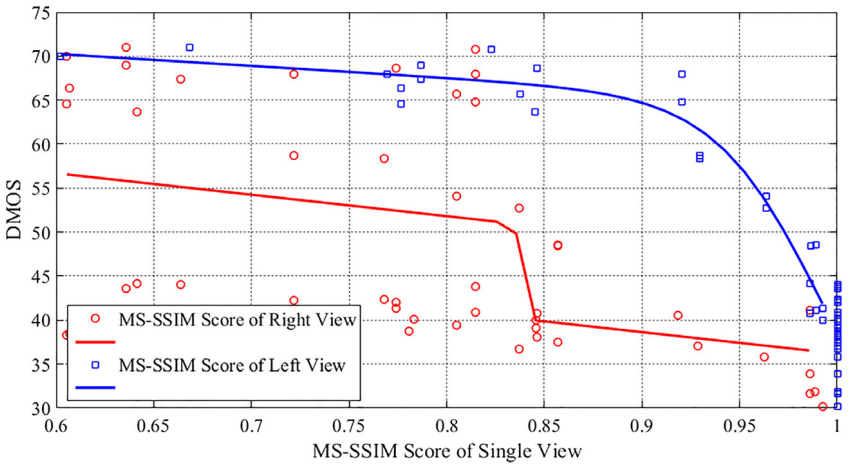

Fig. 3. Scatter plot of DMOS versus the corresponding MS-SSIM scores of asymmetrical blurred left/right views. The fitted curves show that DMOS correlates with the MS-SSIM scores of the left view images better than with the right view images. The SROCC between DMOS and MS-SSIM scores of right views was 0.5564, while the SROCC between DMOS and MS-SSIM scores of left views was 0.8729.

amount of spatial activity. To validate this hypothesis, we conducted a series of experiments on a subset of the LIVE 3D Image Quality Database (Phase II) [37]. Specifically, we used the set of asymmetric stereopairs distorted by blur and noise. By design, all of the selected left views have less distortion than the corresponding right views in the LIVE database. We used the MS-SSIM [38] index as a proxy to measure the quality of each of the left and right images.

We first analyzed the relationship between image distortion (noise and blur) and spatial activity, measured using variance [39], (see Fig. 1). When the left and right views of images from this database are asymmetrically distorted, the right view is more severely distorted and hence, the scatter points gather on one side of the diagonal. Clearly, image noise increased spatial activity, while blur decreased spatial activity. We used the individual (left and right) image of the reference and distorted stereopairs as inputs to compute MS-SSIM scores, then analyzed the correlation between the MS-SSIM scores and the DMOS provided by the database (see Figs. 2 and 3). For the case of noise distortion, the results show that the right view (which contains more spatial information than the left view) MS-SSIM scores better correlated with human judgments (DMOS) than did the left-view scores. For blur distortion, the left view (with more spatial information than the right view) MS-SSIM scores better correlated with human quality perception. These experimental results support our hypothesis, suggesting that the (left or right) view containing greater spatial activity will dominate S3D quality perception.

3.3. Binocular spatial activity model

Based on the experiments we described above, we have observed that spatial activity is not only useful for determining whether there will be binocular rivalry (suppression) between the left and right view images, but it can also be used to further evaluate the relative degrees of influence of the two views when binocular rivalry occurs. We next describe the quantitative binocular spatial activity model expressive of our hypothesis. Stereopsis is largely a local process [40], and the relative amounts of information contained in stereopairs may vary locally over space, hence we define a local spatial binocular activity measure as follows. Let $I_L(i)$ denote the left view (luminance) image, where $i = (x, y)$ are spatial image coordinates. Then, let $\mathbf{S}_L(i)$ be a neighborhood centered at i, e.g. defined by a moving $N \times N$ window. We have observed that the performance of the model is not very sensitive to the window size, and we fixed it to $N = 17$. Then express the spatial activity within $\mathbf{S}_L(i)$ as:

$$\varepsilon\left[S_L(i)\right] = \log_2\left[\sigma_L^2(i) + 1\right], \tag{1}$$

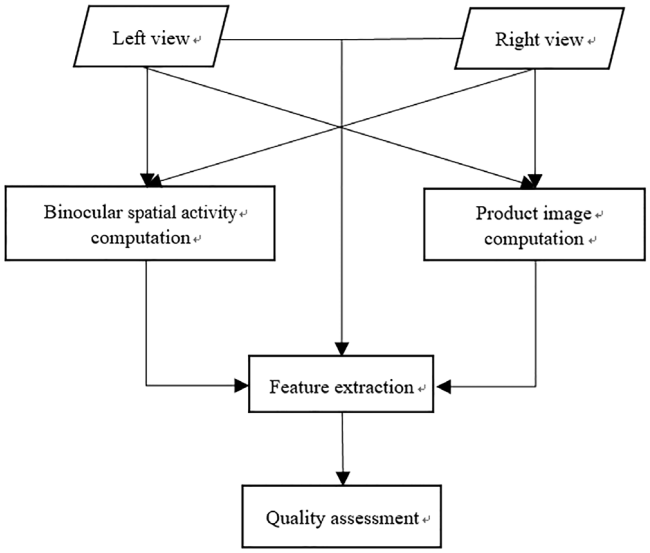

Fig. 4. Flow diagram of proposed NR S3D quality assessment model.

where the unit additive constant guarantees a positive spatial information, and $\sigma_L^2(i)$ is the variance of $\mathbf{S}_L(i)$. The monotonic logarithm serves to compress the activity values to a smaller range. Similarly defining quantities $\mathbf{S}_R(i)$, $\sigma_R^2(i)$, and $\varepsilon\left[\mathbf{S}_R(i)\right]$ on the right image $I_R(i)$, a synthesized "cyclopean" spatial activity image $O(i)$ can be modeled as:

$$O(i) = \frac{\left\{\varepsilon\left[S_L(i)\right] + C\right\} \cdot I_L(i) + \left\{\varepsilon\left[\mathbf{S}_R\left(i + d_i\right)\right] + C\right\} \cdot I_R\left(i + d_i\right)}{\left\{\varepsilon\left[\mathbf{S}_L(i)\right] + C\right\} + \left\{\varepsilon\left[\mathbf{S}_R\left(i + d_i\right)\right] + C\right\}}, \tag{2}$$

where d_i is the horizontal disparity at coordinate i, and $C = 0.01$ is a small positive number to guarantee stability. Thus, the calculation of the binocular spatial activity measure $O(i)$ follows three steps: (1) Generate a disparity map between the left and right view images using a simple stereo disparity estimation algorithm (we deploy a simple stereo algorithm which utilizes the SSIM index as a matching criterion, as motivated in [27]); (2) Generate spatial activity maps on the left and right view images using (1); (3) Synthesize a final "cyclopean" spatial activity image using (2).

4. No-reference stereopair quality assessment

Next we describe our proposed model of S3D picture quality prediction. Fig. 4 shows a flow diagram of the model.

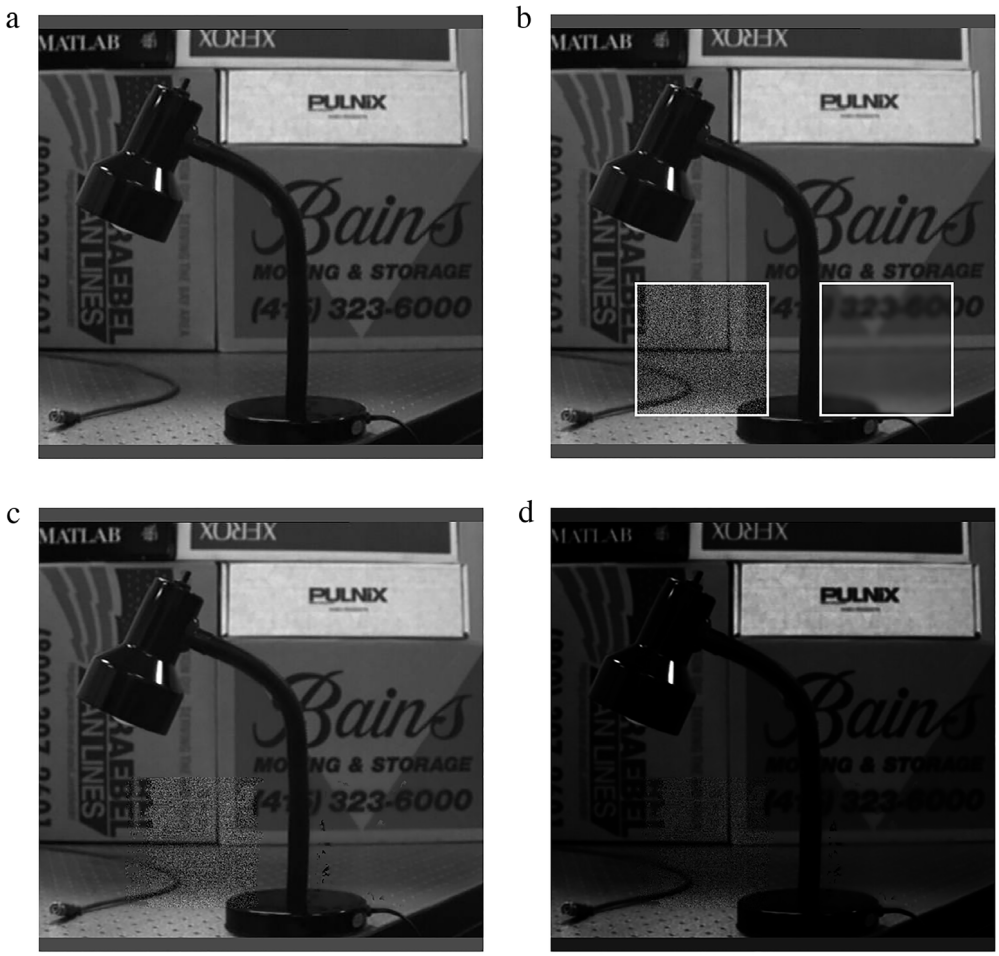

Fig. 5. Four inputs. (a) Left view image. (b) Right view image with two distorted patches: white Gaussian noise and Gaussian blur, respectively. (c) "cyclopean" spatial activity image. (d) Product image.

Along one path that accepts both images as inputs, a "cyclopean" measure of binocular spatial activity is generated using (2). Along a parallel path, an empirical correlation image is obtained by computing the products of corresponding disparity-shifted pixels of the two stereopair images. We tried two methods to analyze the degree of correlation between the left and right views. One uses a multivariate generalized gaussian distribution (MGGD) to model the relationship between the different views, the other uses a bivariate product. It has been shown that using the bivariate product has lower time complexity and still delivers good performance. This element-by-element product image is a simple way of estimating the degree of correlation between the corresponding pixels [6]. We employ it to capture the disparity compensated empirical correlation between the left and right view images. These two matrices ("cyclopean" spatial activity and correlation) form an S3D quality feature vector.

The product image $P(i)$ is computed as:

$$P(i) = I_L(i) \cdot I_R\left(i + d_i\right), \tag{3}$$

which is a simple measure of disparity compensated correlation. This product image can be very efficiently computed. The combined effectiveness of the synthesized "cyclopean" spatial activity image and the binocular product image will be studied in Section 4.

As described above, the process of feature extraction at this stage has produced four feature maps: the left view image, the right view image, the synthesized "cyclopean" spatial activity image and the product image. Fig. 5 shows an example of each of these four inputs. The left view image is undistorted, while the right view image contains two distorted patches containing additive noise and blur, respectively. The "cyclopean" spatial activity image responds strongly to noise, but is insensitive to the blur. This result correlates well with human visual perception. Finally, the product image is shown.

These four feature maps represent a large amount of data. Towards reducing the feature maps into a small number of "quality-aware" descriptors we first apply a process of normalization. The normalization process is reminiscent of the one used by the natural scene statistics (NSS) motivated BRISQUE 2D IQA model [6], which has achieved success as a blind 2D image quality predictor. When viewing a given scene, humans may cast their attention and gaze at different places. Spatial acuity peaks at the foveola and rapidly declines away from fixation. Naturally, locating the points of fixation, then weighting an IQA model accordingly, could provide better correlation with perceived quality. Liu et al. [41] conducted a series of eye tracking experiments which showed that human viewers tend to visually fixate on smooth depth regions, while avoiding large disparity gradients associated with depth discontinuities. This observation, named reverse saliency, showed that depth information can be used in the prediction of visual saliency. It has been successfully deployed in a 3D saliency model and as part of the S3D visual discomfort model [42,43]. Here we also deploy this interesting and powerful concept to compute a weighting map from the disparity gradient image computed on each stereopair, then use this map to "reverse-saliency-weight" the normalized "cyclopean" spatial activity image. As with BRISQUE, we extract parametric features from a model fit to define a small set of quality-predictive features by using the generalized Gaussian distribution (GGD) and asymmetric GGD.

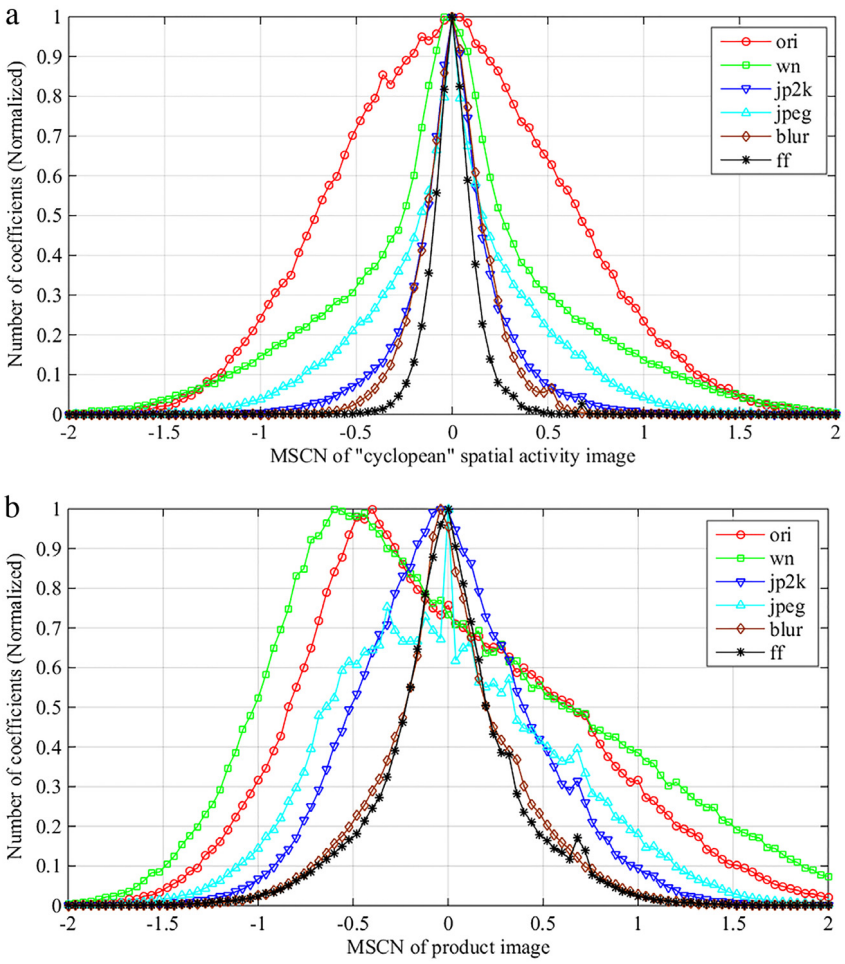

Fig. 6. Histograms of MSCN coefficients of (a) the "cyclopean" spatial activity image and (b) the product image computed on a natural undistorted stereopair and several distorted versions of it. ori: original image. wn: white noise. jp2k: JPEG2000. jpeg: JPEG compression. blur: Gaussian blur. ff: fast fading.

4.1. Univariate parametric features

On each of the four feature maps, mean subtracted contrast normalized (MSCN) coefficients [6] are computed, as follows. Given a normalized map $F(i)$ (of size $M \times N$),

$$MSCN_F(i) = \frac{F(i) - \mu_F(i)}{\sigma_F(i) + C_1}, \tag{4}$$

where μ_F and σ_F are the local mean and standard deviation of F at location i, and C_1 is a stabilizing constant. The local mean and standard deviation values are computed over Gaussian-weighted windows as in [6], and the size of window is also set at 7×7. Then, the reverse-saliency weighted MSCN coefficients of the "cyclopean" spatial activity image are computed as follows:

$$MSCN_O(i) = W_i \cdot MSCN_F(i), \tag{5}$$

$$W_i = \frac{1}{1 + |\nabla d_i|}, \tag{6}$$

and $|\nabla d_i|$ is the gradient magnitude of disparity d_i. The MSCN histograms of the "cyclopean" spatial activity image and the product image are plotted in Fig. 6 for a number of distorted versions of an original image. It may be seen that the histograms of the MSCN coefficients of the "cyclopean" spatial activity image computed on the original image and on the product image have Gaussian-like appearances. However, the introduction of various distortions changes the shape of the histogram in characteristic ways.

We utilize a generalized Gaussian distribution (GGD) to fit the MSCN coefficient distributions of the four maps (the left view image, the right view image, the synthesized "cyclopean" spatial activity image and the product image). The GGD with zero mean can be expressed as:

$$f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left[-\left(\frac{|x|}{\beta}\right)^{\alpha}\right], \tag{7}$$

where

$$\beta = \sigma\sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}, \tag{8}$$

and $\Gamma(\cdot)$ is the standard gamma function. The fitting procedure is the same as that used in [6]. The resulting best-fit shape parameter α and scale parameter σ^2 from each of the four feature maps (i.e. the left view, the right view, the synthesized "cyclopean" spatial activity image and the binocular product image) are both adopted as quality features.

4.2. Bivariate Parametric Features

We also model the local spatial bivariate product statistics of spatially adjacent MSCN coefficients to capture the local structure information. The asymmetric GGD is used to model pairwise products of neighboring MSCN coefficients along four orientations-horizontal (H), vertical (V), main-diagonal ($D1$) and secondary-diagonal ($D2$) [6]. We adopt the same strategy on the four feature maps (similarly defining quantities $MSCN_O(i)$ on the MSCN histograms of the "cyclopean" spatial activity image $MSCN_F(i)$). Pairwise products of MSCN coefficients

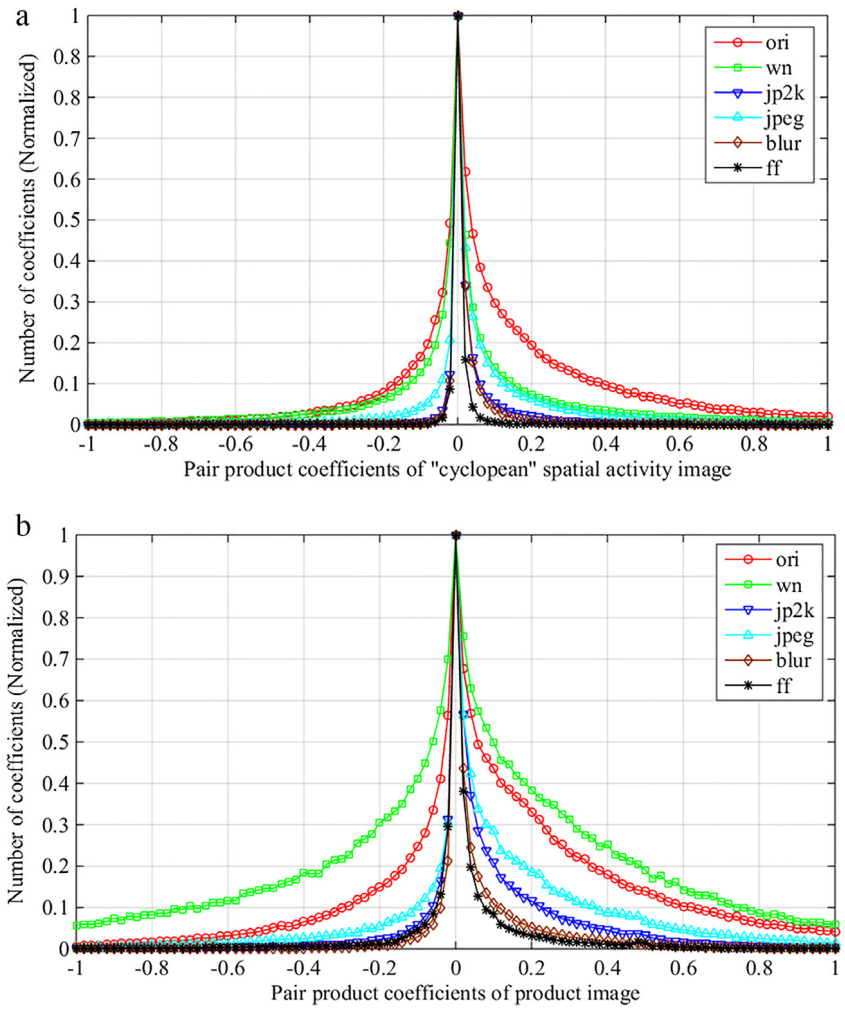

Fig. 7. Histograms of paired products of horizontally adjacent MSCN coefficient feature maps of an undistorted stereopair and corresponding distorted versions of the stereopair. (a) "cyclopean" spatial activity image. (b) binocular product image. ori: original image. wn: white noise. jp2k: JPEG2000. jpeg: JPEG compression. blur: Gaussian blur. ff: fast fading.

along four orientations are computed as follows.

$$H(m,n) = MSCN(m,n)\,MSCN(m,n+1), \tag{9}$$

$$V(m,n) = MSCN(m,n)\,MSCN(m+1,n), \tag{10}$$

$$D1(m,n) = MSCN(m,n)\,MSCN(m+1,n+1), \tag{11}$$

$$D2(m,n) = MSCN(m,n)\,MSCN(m+1,n-1), \tag{12}$$

where m and n index rows and columns in the MSCN map. Fig. 7 plots the pairwise product MSCN histograms of the "cyclopean" spatial activity image and the binocular product image (using horizontally-adjacent MSCN coefficients). The MSCN product histograms of feature maps computed on the original image tend towards a slightly asymmetrical Laplacian-like appearance with a longer right tail than left, while the various distortions include different, narrower shapes and scales in characteristic ways.

We use an asymmetric GGD model to fit the empirical distributions of paired products of neighboring MSCN coefficients along four orientations: horizontal, vertical, main diagonal and counter diagonal. These are computed on each of the four feature maps. The asymmetric GGD model with zero mean is given by:

$$f\left(x;\nu,\sigma_l^2,\sigma_r^2\right) = \begin{cases} \dfrac{\nu}{(\beta_l+\beta_r)\,\Gamma(1/\nu)} \exp\left[-\left(\dfrac{-x}{\beta_l}\right)^{\nu}\right], & x<0 \\ \dfrac{\nu}{(\beta_l+\beta_r)\,\Gamma(1/\nu)} \exp\left[-\left(\dfrac{x}{\beta_l}\right)^{\nu}\right], & x\geq 0, \end{cases} \tag{13}$$

where

$$\beta_l = \sigma_l \sqrt{\frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}}, \tag{14}$$

$$\beta_r = \sigma_r \sqrt{\frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}}. \tag{15}$$

The shape parameter ν and left and right scale parameters β_l and β_r are used as quality-predictive features.

Some discussion of the normalization of the four types of feature maps is worthwhile. The univariate and bivariate parametric features found by the GGD and AGGD fits form an effective summarization of the spatial behavior of each of the feature maps. In the case of the left and right view images, the parametric features have perceptual significance, since the MSCN coefficients have an interpretation both as a model of retinal processing [44] and as a canonical natural scene statistic model [45]. The MSCN parameters also form the core of a successful no-reference image quality predictor [6]. On the other two feature maps, we simply apply the MSCN process to remove redundancies from the data and extract the same parameters, although we do not ascribe a neural mechanism to these constructs.

Images are naturally multiscale, since they contain objects of different depths and sizes, and so is low-level retinal and cortical processing in primate vision systems. Moreover, multiscale processing has been shown to improve the efficacy of leading image quality models [6,38]. Therefore, all of the above features are computed at two scales: the

Table 1
Comparison of median SROCC of compared S3D IQA models against DMOS on LIVE Phase I database.

Method		WN	JP2K	JPEG	Blur	FF	All
2D	PSNR	0.932	0.799	0.121	0.902	0.588	0.834
	SSIM [3]	0.938	0.858	0.436	0.879	0.586	0.877
	MS-SSIM [38]	0.942	0.893	0.611	0.926	0.734	0.924
	VIF [52]	0.931	**0.901**	0.581	**0.934**	0.804	0.920
	VSI [53]	0.932	0.874	0.602	0.905	0.718	0.872
	BRISQUE [6]	0.940	0.812	0.569	0.860	0.784	0.901
	SSEQ [54]	0.909	0.827	0.438	0.879	0.732	0.903
	GWH-GLBP [55]	0.903	0.791	0.465	0.891	0.641	0.886
3D	Chen FR [27]	0.948	0.888	0.530	0.925	0.707	0.916
	Chen NR [11]	0.919	0.863	0.617	0.878	0.652	0.891
	Su [14]	0.906	0.838	0.603	0.791	0.679	0.903
	SINQ	**0.954**	0.897	**0.656**	0.883	**0.841**	**0.936**

original image scale and a coarser scale downsampled by a factor of 2, as in [6].

Next we deploy a two-stage framework first introduced in [46] to map the feature vectors to perceptual quality scores. The two-stage framework is (1) distortion classification followed by (2) distortion-specific quality assessment, using a support vector machine (SVM) to classify by distortion in the first stage, and a support vector regression (SVR) to form a mapping of features to quality predictions in the second stage [47,48]. The Libsvm-3.16 toolbox [49] was utilized to implement the SVM and SVR using the radial basis function (RBF) kernel.

5. Experimental results

We tested the performance of our proposed method on the LIVE 3D image quality database. There are two phases contained in this publicly available database: Phase I [50], in which all of the stereopairs are symmetrically distorted, and Phase II [37], which consists of both symmetrical and asymmetrical distortions, i.e., where the left and right view images are impaired by different degrees of the same distortion. In addition, we further evaluated the performance of SINQ on four other 3D IQA databases. The performance was evaluated via three metrics: Spearman's Rank Ordered Correlation Coefficient (SROCC), Pearson's Correlation Coefficient (LCC) and the Root Mean Squared Error (RMSE) between the predicted quality scores and DMOS. The predicted scores were mapped to DMOS using the 5-parameter logistic nonlinearity function described in [51] before computing LCC and RMSE. The lower RMSE and higher LCC and SROCC values indicate better correlation with human perception. To achieve a more comprehensive comparison, several 2D and 3D IQA models are also included in our comparison experiments. Note that NR IQA methods are italicized, while all others are FR and RR IQA methods.

5.1. Performance on Phase I database

The LIVE Phase I Stereo Image Quality database consists of 20 reference images and 365 corresponding distorted images (80 each for JPEG, JP2K, WN and FF and 45 for Blur) with co-registered human scores in the form of DMOS. Smaller values of DMOS correspond to better image quality. All distortions are symmetric. For the purpose of comparison, several top-performing 2D IQA models were selected, including the FR algorithms PSNR, SSIM [3], MS-SSIM [38], VIF [52], and VSI [53] and the NR models BRISQUE [6], SSEQ [54], and GWH-GLBP [55]. These 2D IQA algorithms were applied to both views of each stereo pair, resulting in two quality scores. The mean value of the two scores was used as the predicted score. Several available 3D IQA models were also tested. These included a cyclopean natural scene statistic based method proposed by Chen et al. [27], which also targets binocular rivalry, and the NR IQA methods proposed by Chen et al. [11] and Su et al. [14]. The software releases of all of the selected IQA models are publicly available. For the FR methods, all of the reference and distorted

Table 2
Comparison of median LCC of compared S3D IQA models against DMOS on LIVE Phase I database.

Method		WN	JP2K	JPEG	Blur	FF	All
2D	PSNR	0.935	0.788	0.228	0.916	0.701	0.836
	SSIM	0.942	0.875	0.489	0.918	0.670	0.877
	MS-SSIM	0.948	0.931	0.684	0.944	0.807	0.930
	VIF	0.931	0.937	0.681	0.965	0.862	0.925
	VSI	0.958	0.928	0.674	**0.973**	0.810	0.871
	BRISQUE	0.941	0.847	0.615	0.926	0.853	0.910
	SSEQ	0.932	0.906	0.626	0.956	0.787	0.919
	GWH-GLBP	0.932	0.863	0.582	0.961	0.733	0.912
3D	Chen FR	0.942	0.912	0.603	0.942	0.776	0.917
	Chen NR	0.917	0.907	0.695	0.917	0.735	0.895
	Su	0.961	0.917	**0.746**	0.953	0.721	0.922
	SINQ	**0.970**	**0.948**	0.734	0.967	**0.898**	**0.956**

Table 3
Comparison of median RMSE of compared S3D IQA models against DMOS on LIVE Phase I database.

Method		WN	JP2K	JPEG	Blur	FF	All
2D	PSNR	5.897	7.976	6.369	5.807	8.865	9.010
	SSIM	5.578	6.262	5.703	5.740	9.226	7.891
	MS-SSIM	5.296	4.738	4.769	4.774	7.336	6.028
	VIF	6.076	4.510	4.787	3.776	6.302	6.232
	VSI	4.594	4.562	4.467	**3.200**	6.511	7.969
	BRISQUE	5.640	6.894	5.158	5.473	6.491	6.793
	SSEQ	5.748	5.337	4.639	3.798	7.068	6.381
	GWH-GLBP	5.835	6.195	4.827	3.567	7.675	6.657
3D	Chen FR	5.581	5.320	5.216	4.822	7.837	6.533
	Chen NR	6.433	5.402	4.523	5.898	8.322	7.247
	Su	3.931	4.873	**3.959**	4.326	8.613	6.258
	SINQ	**3.834**	**3.985**	4.049	3.554	**5.018**	**4.781**

stereopairs were used for testing, while for the NR methods, which are generally learning-based, we randomly divided the reference stereopairs and corresponding distorted counterparts into 80% training and 20% test subsets (the same as [6,11] and [14]). There was no overlap of content between the training and test subsets. In order to avoid any bias introduced by image content on training, the random procedure of 80% training and 20% test was repeated over 1000 trials. The median SROCC, LCC and RMSE results across the 1000 random trials are tabulated in Tables 1–3. The best performing method for distortion type and on the entire database is marked in bold type.

From Tables 1–3, it may be observed that, in most cases, SINQ performed better than all of the compared 2D and 3D IQA models. In particular, the comparison with [6] shows that our model effectively captures the effects on quality perception introduced by stereopsis, since the non-stereo features used in SINQ are identical to those that define [6]. For every type of distortion SINQ delivered competitive performance. Since there are only symmetric distortions in the Phase I database, we may infer that SINQ is able to predict the perceptual severity of symmetric distortions quite effectively.

Table 4
Comparison of median SROCC of compared S3D IQA models against DMOS on LIVE Phase II database.

Method		WN	JP2K	JPEG	Blur	FF	All
2D	PSNR	0.665	0.641	0.492	0.868	0.747	0.701
	SSIM	0.921	0.703	0.679	0.836	0.835	0.793
	MS-SSIM	0.947	0.800	0.854	0.800	0.827	0.777
	VIF	0.838	0.822	0.846	**0.951**	**0.933**	0.819
	VSI	0.948	0.787	0.804	0.845	0.854	0.799
	BRISQUE	0.846	0.593	0.769	0.862	0.935	0.770
	SSEQ	0.938	0.769	0.841	0.922	0.843	0.802
	GWH-GLBP	0.903	0.905	0.849	0.950	0.928	0.804
3D	Chen FR	0.940	0.814	0.843	0.908	0.884	0.889
	Chen NR	0.950	0.867	**0.867**	0.900	0.933	0.880
	Su	0.946	0.845	0.818	0.903	0.899	0.905
	SINQ	**0.957**	**0.909**	0.839	0.909	0.924	**0.931**

Table 5
Comparison of median LCC of compared S3D IQA models against DMOS on LIVE Phase II database.

Method		WN	JP2K	JPEG	Blur	FF	All
2D	PSNR	0.782	0.653	0.467	0.915	0.788	0.721
	SSIM	0.930	0.725	0.669	0.842	0.867	0.802
	MS-SSIM	0.951	0.759	0.864	0.799	0.868	0.793
	VIF	0.835	0.831	0.847	**0.987**	0.931	0.837
	VSI	**0.975**	0.817	0.834	0.916	0.909	0.804
	BRISQUE	0.845	0.681	0.795	0.951	0.931	0.782
	SSEQ	0.960	0.739	0.840	0.973	0.873	0.826
	GWH-GLBP	0.918	**0.923**	0.881	0.981	**0.945**	0.825
3D	Chen FR	0.957	0.834	0.862	0.963	0.901	0.900
	Chen NR	0.947	0.899	**0.901**	0.941	0.932	0.895
	Su	0.953	0.847	0.888	0.968	0.944	0.913
	SINQ	0.970	0.920	0.871	0.984	0.938	**0.936**

Table 6
Comparison of median RMSE of compared S3D IQA models against DMOS on LIVE Phase II database.

Method		WN	JP2K	JPEG	Blur	FF	All
2D	PSNR	6.673	7.439	6.481	5.604	7.081	7.824
	SSIM	3.934	6.758	5.451	7.510	5.736	6.737
	MS-SSIM	3.306	6.390	3.695	8.379	5.720	6.874
	VIF	5.907	5.464	4.087	**2.221**	4.197	6.184
	VSI	**2.089**	4.723	3.780	5.049	4.720	6.790
	BRISQUE	5.731	7.193	4.448	4.323	4.206	7.038
	SSEQ	2.809	5.346	3.494	3.575	5.738	6.519
	GWH-GLBP	3.974	3.563	**3.338**	2.423	**3.652**	6.282
3D	Chen FR	3.368	5.562	3.865	3.747	4.966	4.987
	Chen NR	3.513	4.298	3.342	4.725	4.180	5.102
	Su	3.547	5.482	4.169	4.453	4.199	4.657
	SINQ	2.519	**3.463**	3.476	2.481	3.803	**3.959**

Table 7
Median classification accuracy and standard deviation of stage 1 of SINQ across 1000 trials on LIVE Phase II database.

	WN	JP2K	JPEG	Blur	FF	All
Mean acc. (%)	95.9	89.7	99.4	94.9	88.1	93.6
Standard deviation	0.079	0.146	0.035	0.061	0.107	0.041

Table 8
Performance of single-stage realization of SINQ on LIVE Phase II and Phase II databases.

		WN	JP2K	JPEG	Blur	FF	All
Phase I	SROCC	0.959	0.869	0.671	0.886	0.809	0.933
	LCC	0.974	0.939	0.734	0.967	0.876	0.953
	RMSE	3.590	4.363	4.079	3.433	5.364	4.982
Phase II	SROCC	0.940	0.922	0.837	0.920	0.940	0.926
	LCC	0.963	0.931	0.866	0.980	0.951	0.929
	RMSE	2.962	3.316	3.575	2.818	3.326	4.090

5.2. Performance on Phase II database

The LIVE Phase II Stereo Image Quality database consists of 8 reference images and 360 distorted images (72 each for JP2K, JPEG, WN, FF and Blur) with co-registered human scores in the form of DMOS. For each type of distortion, every reference stereopair was processed to create three symmetric distorted counterparts and six asymmetric distorted counterparts. Hence the database contains a mixture of symmetric and asymmetric distortions. Asymmetric distortions are an important consideration, as they can occur in a variety of ways, as for example proposed asymmetric 3D picture compression algorithms. The experiments were conducted in the same way as on the Phase I dataset, using SROCC, LCC and RMSE as performance metrics, and compared with the same 2D and 3D IQA models. The median results across 1000 random trials are given in Tables 4–6. We can see that when the test images contain asymmetric distortions, the prediction accuracy of each 2D IQA method severely deteriorated as compared with symmetric distortions. However, SINQ was able to predict the quality of asymmetric distortions quite well, and again outperformed all of the other tested 2D and 3D IQA models on the Phase II database. These results further confirm the efficacy of the SINQ S3D quality prediction model.

5.3. Statistical significance

Although the results showed apparent differences between the models, we conducted statistical tests to determine whether the differences were statistically significant. Because the Phase II database contains a mixture of symmetric and asymmetric distortions, and is more complex than the Phase I database, we only conducted experiments on the Phase II database. We determined significance using the t-test on SROCC of the compared algorithms across 1000 train-test trials. Figs. 8 and 9 show a box plot of SROCC and the results of the t-test, respectively. The null hypothesis we applied is that the value of the mean correlation in the row is equal to the value of the column with a confidence of 95%. The alternate hypotheses are that the mean correlation value of the row is lesser or greater than the value of the column. Fig. 9 indicates whether an algorithm in a row is statistically inferior ('-1'), statistically equivalent ('0') or statistically superior ('1') to the algorithm in a column. From Fig. 8 it may be seen that SINQ delivered the most stable results among all methods. SINQ also was superior as measured by the t-test (see Fig. 9). Thus we may conclude that the performance of SINQ is quite stable, and statistically better than all the other compared FR and NR algorithms.

5.4. Classification accuracy

As mentioned earlier, SINQ embodies a two-stage framework [46] to map the feature vectors to the distortion categories and then to predicted quality scores. The probability of the tested image belonging to each distortion type is predicted using a SVM multi-class classification engine. Then, five separate SVR models are trained to predict the perceptual quality of the tested image, one for each distortion type. Classification efficacy was tested over 1000 80% training and 20% testing trials. The mean classification accuracy and standard deviation across 1000 trials were used as the performance metrics, as tabulated in Table 7. The mean classification accuracy of SINQ across the entire database was 93.6% with a relatively small standard deviation 0.041. SINQ was also able to distinguish the individual distortion types with high accuracy, especially on the JPEG and White Noise (WN) distortions.

5.5. Single-stage performance

In order to show that the features defining SINQ are effective without a first classification stage, we also created a single-stage SINQ model where features are directly mapped to quality without distortion identification. We conducted the performance of this model on the LIVE 3D Image Quality Database (Phase I and Phase II). The median

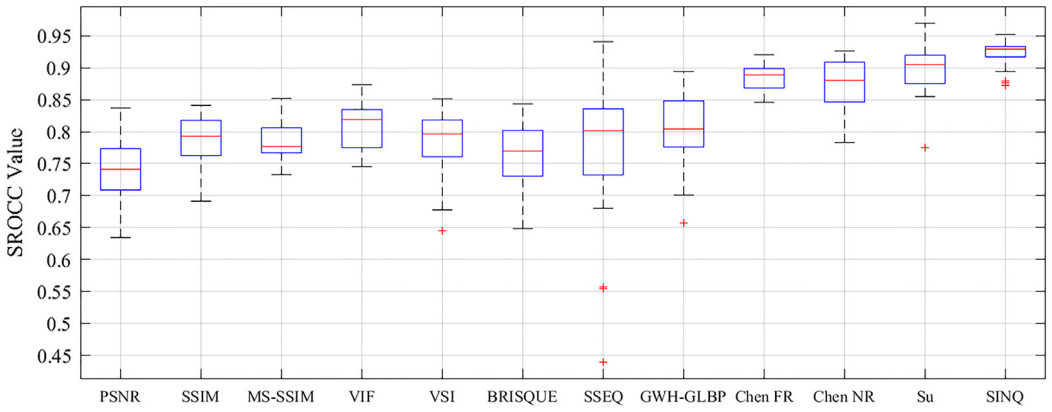

Fig. 8. Box plot of SROCC of algorithms across 1000 train-test trials on the Phase II database.

	2D PSNR	2D SSIM	2D MS-SSIM	2D VIF	2D VSI	2D BRISQUE	2D SSEQ	2D GWH-GLBP	Chen FR	Chen NR	Su	SINQ
2D PSNR	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2D SSIM	1	0	1	-1	-1	1	-1	-1	-1	-1	-1	-1
2D MS-SSIM	1	-1	0	-1	-1	1	-1	-1	-1	-1	-1	-1
2D VIF	1	1	1	0	1	1	1	1	-1	-1	-1	-1
2D VSI	1	1	1	-1	0	1	-1	-1	-1	-1	-1	-1
2D BRISQUE	1	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1
2D SSEQ	1	1	1	-1	1	1	0	-1	-1	-1	-1	-1
2D GWH-GLBP	1	1	1	-1	1	1	1	0	-1	-1	-1	-1
Chen FR	1	1	1	1	1	1	1	1	0	1	-1	-1
Chen NR	1	1	1	1	1	1	1	1	-1	0	-1	-1
Su	1	1	1	1	1	1	1	1	1	1	0	-1
SINQ	1	1	1	1	1	1	1	1	1	1	1	0

Fig. 9. Results of T-test between SROCC values of various IQA algorithms.

SROCC, LCC and RMSE values across 1000 random trials are tabulated in Table 8. When compared with the two-stage performance in Tables 1–6, the performances are nearly indistinguishable suggesting that two stages are only needed if the distortion categorization is required for some other process, such as image correction [56].

5.6. Individual feature map contributions

The proposed SINQ model uses summary statistical parameters drawn from four different feature maps: original left and right view images, the synthesized "cyclopean" spatial activity image and the binocular product image. It is of interest to study the contributions of each, as well as the efficacy of the reverse saliency model. We divided the features into five groups corresponding to feature maps and tested the prediction power of each group. The median SROCC, LCC and RMSE scores across 1000 80% training and 20% testing trials were again tabulated, in Table 9. We found that features drawn from the binocular product image performed about as well those from the "cyclopean" spatial activity image and much better than those from the original left and right views. The reverse saliency model delivers an obvious improvement in performance. When compared with the experimental results in Tables 4–6, SINQ is still able to reach a relatively high level of performance if using only features drawn from "cyclopean" spatial activity images or binocular product images. By combining features from all four sources, even better performance is achieved, as shown in Table 9. Thus, each type of feature map makes an important contribution to the performance of SINQ.

5.7. Performances on symmetric and asymmetric distortions

We further compared the same set of 2D and 3D IQA models on the separate Phase I and Phase II databases, and found that almost all of the 2D and 3D methods are able to reasonably predict the perceptual severity of symmetric distortions, but generally perform worse on asymmetric distortions. By contrast, SINQ performs well on both symmetric and asymmetric distortions. Since the Phase II database is composed of both symmetric and asymmetric distortions, we conducted a further comparative test to study the performance of SINQ. We divided the Phase II database into two separate subsets of only symmetric and asymmetric distortions and tested the performance of several 2D and 3D IQA methods on both subsets. The results are listed in Table 10.

On only symmetric distortions, SINQ outperformed all the other tested IQA methods except Su's NR model [14], while the 2D MS-SSIM, 2D VIF, Chen FR and Chen NR models all obtained good performance. On asymmetric distortions, SINQ outperformed all the others by a significant margin. We found that all of the 2D and 3D IQA models performed significantly worse on asymmetric distortions than on symmetric ones, except SINQ, which had only a modest drop in performance.

Table 9
Performance of feature map on LIVE Phase II database.

	SROCC	LCC	RMSE
Left view	0.723	0.781	7.140
Right view	0.675	0.702	7.954
"Cyclopean" spatial activity image without reverse saliency	0.873	0.889	5.157
"Cyclopean" spatial activity image with reverse saliency	0.922	0.923	4.232
Binocular product image	0.860	0.878	5.308
SINQ	0.931	0.936	3.959

Table 10
Performance on symmetric and asymmetric distortions on LIVE Phase II database.

Method		SROCC	
		Symmetric	Asymmetric
2D	PSNR	0.776	0.587
	SSIM	0.828	0.733
	MS-SSIM	0.912	0.684
	VIF	0.916	0.793
	VSI	0.836	0.742
	BRISQUE	0.849	0.667
	SSEQ	0.822	0.662
	GWH-GLBP	0.853	0.678
3D	Chen FR	0.923	0.842
	Chen NR	0.918	0.834
	Su	**0.937**	0.849
	SINQ	0.933	**0.905**

Table 11
Performance comparison between 2D baseline IQA models and binocular enhanced IQA models on LIVE Phase II database.

	SROCC	
	2D Baseline	Binocular model
PSNR	0.701	0.722
SSIM	0.793	0.836
MS-SSIM	0.777	**0.879**
VIF	**0.806**	0.859

Table 12
Comparison of time complexity of eight IQA methods.

Method		Time (s)
2D	SSIM	0.082
	VIF	1.595
	BRISQUE	0.141
	GWH-GLBP	0.200
3D	Chen FR	23.110
	Chen NR	11.667
	Su	170.396
	SINQ	3.232

5.8. Generality of cyclopean spatial activity map

To further study the usefulness of the "cyclopean" spatial activity image, we conducted a verification experiment on several 2D IQA algorithms (PSNR, SSIM, MS-SSIM and VIF) by applying them on the "cyclopean" image. The results of the 2D baseline were obtained by averaging the left and right algorithm scores, and the results of the binocular model were obtained by computing the 2D algorithms on the synthesized "cyclopean" spatial activity images. The comparison results on the Phase II database are tabulated in Table 11.

It may be observed that the median SROCC values of the 2D IQA methods increased sharply, especially for MS-SSIM and VIF. These results further confirm that the "cyclopean" spatial activity model is an effective descriptor of the perceptual quality of stereopairs.

5.9. Computational complexity

The computations expended by our model are mainly consumed in the process of synthesizing a cyclopean image, computing a product image, and subsequent feature extraction. Since all these steps are processed in the spatial domain, the complexity of SINQ is quite reasonable. We selected seven IQA models and compared their complexities. The original MATLAB code of each algorithm was run on the chosen 100 3D images from the LIVE Phase II database, and the mean computation time of each model over these images was used as the complexity metric. The results in Table 12 show that SINQ is faster than the compared 3D methods, but slower than the 2D methods, using a PC with a quad-core 3.4 GHz and 8 GB RAM.

5.10. Performance on four other databases

We further tested the performance of SINQ on four other 3D IQA databases. These databases are the MLC 3D Database [57], the IRRCyN IVC Database [58], and the Waterloo IVC 3D IQA Database [59], which includes two datasets, called Phase I and Phase II. We summarize the relevant details of these databases in Table 13, including the numbers of reference images, symmetric distorted images, asymmetric distorted images and distortion types. The experiments were conducted in the same way as on the LIVE database, using SROCC, LCC and RMSE as performance metrics. The median results across 1000 random trials on the four image databases are given in Table 14. SINQ also achieved highly competitive prediction accuracy and outperformed all of the other compared 2D and 3D IQA models, especially on the MCL and IRRCyN IVC databases. Also, we divided the Waterloo IVC database into two parts: symmetric and asymmetric distorted images, and evaluated the performance of SINQ on each subset separately. The results are showed in Table 15. Although other tested methods' results fluctuated severely, SINQ predicted the quality uniformly quite well on the Waterloo IVC database. These results again confirm the efficacy of the proposed SINQ model.

6. Conclusion

We have developed a new NR S3D IQA model called SINQ. A novel modification of an existing binocular perceptual model is utilized to form synthesized "cyclopean" spatial activity images. A binocular product image is also generated to express correlations between the left and right disparity-corrected corresponding binocular pixels. A powerful reverse saliency weighting strategy was also deployed that boosts performance further. SINQ was extensively validated on several S3D image databases, and shown to deliver standout performance. We

Table 13
Information of the four databases.

Database	Reference image No.	Symmetric distorted image No.	Asymmetric distorted image No.	Distortion type No.
Waterloo IVC phase I	6	78	252	3
Waterloo IVC phase II	10	130	330	3
MCL	9	648	0	6
IRRCyN IVC	6	96	0	5

Table 14
Comparison of compared S3D IQA models on four databases.

Method		Waterloo IVC Phase I			Waterloo IVC Phase II			MCL			IRRCyN IVC		
		SROCC	LCC	RMSE	SROCC	LCC	RMSE	SROCC	LCC	RMSE	SROCC	LCC	RMSE
2D	SSIM	0.631	0.743	10.526	0.577	0.671	14.192	0.783	0.765	1.677	0.709	0.743	15.668
	VIF	0.682	0.761	10.213	0.628	0.693	13.794	**0.865**	**0.862**	1.317	0.741	0.847	12.454
	BRISQUE	**0.866**	**0.870**	**7.547**	**0.866**	**0.855**	**10.000**	0.786	0.810	1.500	0.565	0.658	16.912
	GWH-GLBP	0.831	0.800	9.687	0.830	0.811	11.202	0.843	0.858	**1.307**	**0.818**	**0.855**	**10.949**
3D	Chen FR	0.569	0.674	11.623	0.444	0.569	15.740	0.839	0.762	1.686	0.676	0.683	17.100
	Chen NR	0.911	0.926	6.232	0.884	0.882	8.961	0.799	0.813	1.497	0.835	0.851	12.088
	SINQ	**0.937**	**0.956**	**4.960**	**0.911**	**0.922**	**7.377**	**0.928**	**0.951**	**0.806**	**0.900**	**0.929**	**8.155**

Table 15
SROCC of compared models on symmetric and asymmetric distortions of Waterloo IVC database.

Method		Waterloo IVC phase I		Waterloo IVC phase II	
		Symmetric	Asymmetric	Symmetric	Asymmetric
2D	SSIM	0.608	0.620	0.549	0.551
	VIF	0.720	0.654	0.641	0.588
	BRISQUE	**0.956**	**0.836**	**0.941**	**0.825**
	GWH-GLBP	**0.956**	0.743	0.924	0.724
3D	Chen FR	0.863	0.491	0.642	0.390
	Chen NR	0.934	0.907	0.897	0.867
	SINQ	**0.967**	**0.934**	**0.917**	**0.904**

found several aspects of the problem to be of particular interest which might merit further research. For example, since reverse saliency is based only on depth information, perhaps a more general saliency model could achieve better performance. In the future we plan to extend these ideas towards developing S3D video quality prediction models that incorporate space–time measures of binocular spatial activity as well as space–time natural scene statistic models.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under grant 61672095 and grant 61425013.

References

[1] A.C. Bovik, Automatic prediction of perceptual image and video quality, Proc. IEEE 101 (9) (2013) 2008–2029.
[2] A.C. Bovik, Z. Wang, Modern Image Quality Assessment, Morgan & Claypool, Morgan, CA, USA, 2006.
[3] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[4] M. Saad, A.C. Bovik, C. Charrier, Blind image quality assessment: A natural scene statistics approach in the DCT domain, IEEE Trans. Image Process. 21 (8) (2012) 3339–3352.
[5] L. Liu, H. Dong, H. Huang, A.C. Bovik, No-reference image quality assessment in curvelet domain, Signal Process. Image Commun. 29 (4) (2014) 494–505.
[6] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[7] L. Liu, Y. Hua, Q. Zhao, H. Huang, A.C. Bovik, Blind image quality assessment by relative gradient statistics and adaboosting neural network, Signal Process. Image Commun. 40 (2016) 1–15.
[8] A.K. Moorthy, C.-C. Su, A. Mittal, A.C. Bovik, Subjective evaluation of stereoscopic image quality, Signal Process. Image Commun. 28 (8) (2013) 870–883.
[9] R. Akhter, J. Baltes, Z.M. Parvez Sazzad, Y. Horita, No reference stereoscopic image quality assessment, in: Proc. SPIE, vol. 7524, Feb. 2010, p. 75240T.
[10] S. Ryu, K. Sohn, No-reference quality assessment for stereoscopic images based on binocular quality perception, IEEE Trans. Circuits Syst. Video Technol. 24 (4) (2014) 591–602.
[11] M. Chen, L.K. Cormack, A.C. Bovik, No-Reference quality assessment of natural stereopairs, IEEE Trans. Image Process. 22 (9) (2013) 3379–3391.
[12] C.-C. Su, L.K. Cormack, A.C. Bovik, Bivariate statistical modeling of color and range in natural scenes, in: Proceedings of SPIE, Human Vision and Electronic Imaging XIX, vol. 9014, Feb. 2014.
[13] C.-C. Su, L.K. Cormack, A.C. Bovik, Closed-Form correlation model of oriented bandpass natural images, IEEE Signal Process. Lett. 22 (1) (2015) 21–25.
[14] C.-C. Su, L.K. Cormack, A.C. Bovik, Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation, IEEE Trans. Image Process. 24 (5) (2015) 1685–1699.
[15] K. Lee, A.K. Moorthy, S. Lee, A.C. Bovik, 3D visual activity assessment based on natural scene statistics, IEEE Trans. Image Process. 23 (1) (2014) 450–465.
[16] J.P. Lopez, J.A. Antonio, D. Jimenez, J.M. Menendez, Stereoscopic 3D video quality assessment based on depth maps and video motion, EURASIP J. Image Video Process. (2013).
[17] F. Shao, G. Jiang, M. Yu, Binocular energy response based quality assessment of stereoscopic images, Digit. Signal Process. 29 (2014) 45–53.
[18] L. Jin, A. Boev, K. Egiazarian, A. Gotchev, Quantifying the importance of cyclopean view and binocular rivalry-related features for objective quality assessment of mobile 3D video, EURASIP J. Image Video Process. (2014).
[19] V.D. Silva, H.K. Arachchi, E. Ekmekcioglu, A. Kondoz, Toward an impairment metric for stereoscopic video: A full-reference video quality metric to assess compressed stereoscopic video, IEEE Trans. Image Process. 22 (9) (2013) 3392–3404.
[20] S.L.P. Yasakethu, C.T.E.R. Hewage, W.A.C. Fernando, A.M. Kondoz, Quality analysis for 3D video using 2D video quality models, IEEE Trans. Consum. Electron. 54 (4) (2008) 1969–1976.
[21] J. You, L. Xing, A. Perkis, X. Wang, Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis, in: Proc. Int. Workshop Video Process. Qaulity Metrics Consum. Elect. 2010, pp. 1–6.
[22] Z.M.P. Sazzad, R. Akhter, J. Baltes, Y. Horita, Objective noreference stereoscopic image quality prediction based on 2D image features and relative disparity, Adv. Multimedia 2012 (2012) Art. ID 8.
[23] Y. Zhao, Z. Chen, C. Zhu, Y. Tan, L. Yu, Binocular just-noticeable-difference model for stereoscopic images, IEEE Signal Process. Lett. 18 (1) (2011) 19–22.
[24] F. Shao, W. Lin, S. Gu, G. Jiang, T. Srikanthan, Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics, IEEE Trans. Image Process. 22 (5) (2013) 1940–1953.
[25] Y. Lin, J. Wu, Quality assessment of stereoscopic 3D image compression by binocular integration behaviors, IEEE Trans. Image Process. 23 (4) (2014) 1527–1542.
[26] A. Maalouf, 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.
[27] M. Chen, C. Su, D.K. Kwon, L.K. Cormack, A.C. Bovik, Full-reference quality assessment of stereopairs accounting for rivalry, Signal Process. Image Commun. 28 (9) (2013) 1143–1155.
[28] J. Yang, Y. Wang, B. Li, W. Lu, Q. Meng, Z. Lv, D. Zhao, Z. Gao, Quality assessment metric of stereo images considering cyclopean integration and visual saliency, Inform. Sci. 373 (2016) 251–268.
[29] W.J.M. Levelt, On Binocular Rivalry, Mouton, Paris, France, 1968.
[30] R. Blake, D.H. Westendorf, R. Overton, What is suppressed during binocular rivalry?, Perception 9 (2) (1980) 223–231.
[31] F. Shao, W. Tian, W. Lin, G. Jiang, Q. Dai, Toward a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions, IEEE Trans. Image Process. 25 (5) (2016) 2059–2074.
[32] M.J. Chen, L.K. Cormack, A.C. Bovik, Distortion conspicuity on stereoscopically fused 3D images may correlate to scene content and distortion type, J. Soc. Inf. Disp. 21 (11) (2013) 491–503.
[33] M.T.M. Lambooij, W.A. Ijsselsteijn, I. Heynderickx, Visual discomfort in stereoscopic displays: a review, Proc. SPIE 6490 (2007) 649001–1–13.
[34] G. Saygili, C.G. Gurler, A.M. Tekalp, Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming, IEEE Trans. Broadcast. 57 (2) (2011) 593–601.
[35] A.K. Jain, A.E. Robinson, T.Q. Nguyen, Comparing perceived quality and fatigue for two methods of mixed resolution stereoscopic coding, IEEE Trans. Circuits Syst. Video Technol. 24 (3) (2014) 418–429.
[36] B. Julesz, Foundations of Cyclopean Perception, University of Chicago Press, 1971.
[37] A. Moorthy, C.-C. Su, M.-J. Chen, A. Mittal, L.K. Cormack, A.C. Bovik, LIVE 3D Image Quality Database Phase II. [Online]. Available: http://live.ece.utexas.edu/research/quality/live_3dimage_phase2.html.
[38] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment, in: Proc. 37th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR), vol. 2, Nov. 2003, pp. 1398–1402.

[39] S. Aja-Fernandez, R. San-José-Estépar, C. Alberola-Lopez, C. Westin, Image quality assessment based on local variance, in: Proc. 28th Annu. IEEE Int. Conf. Engineering in Medicine and Biology Society (EMBS 2006), Aug. 2006, pp. 4815-4818.
[40] M.K. Kapadia, M. Ito, C.D. Gilbert, G. Westheimer, Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys, Neuron 15 (4) (1995) 843–856.
[41] Y. Liu, L.K. Cormack, A.C. Bovik, Dichotomy between luminance and disparity features at binocular fixations, J. Vis. 10 (12) (2010) 1–17.
[42] H. Kim, S. Lee, A.C. Bovik, Saliency prediction on stereoscopic videos, IEEE Trans. Image Process. 23 (4) (2014) 1476–1490.
[43] T. Kim, S. Lee, A.C. Bovik, Transfer function model of physiological mechanisms underlying temporal visual discomfort experienced when viewing stereoscopic 3D images, IEEE Trans. Image Process. 24 (11) (2015) 4335–4347.
[44] M. Carandini, D.J. Heeger, Normalization as a canonical neural computation, Nat. Rev. Neurosci. 13 (1) (2011) 51–62.
[45] D.L. Ruderman, The statistics of natural images, Netw., Comput. Neural Syst. 5 (4) (1994) 517–548.
[46] A.K. Moorthy, A.C. Bovik, A two-stage framework for blind image quality assessment, in: Proc. IEEE Int. Conf. Image Process. Sep. 2010, pp. 2481–2484.
[47] B. Scholkopf, A.J. Smola, R.C. Williamson, P.L. Bartlett, New support vector algorithms, Neural Comput. 12 (5) (2000) 1207–1245.
[48] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discov. 2 (2) (1998) 121–167.
[49] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) Art. ID 27. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[50] A. Moorthy, C.-C. Su, M.-J. Chen, A. Mittal, L.K. Cormack, A.C. Bovik, LIVE 3D Image Quality Database Phase I. [Online]. Available: http://live.ece.utexas.edu/research/quality/live_3dimage_phase1.html.
[51] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15 (11) (2006) 3440–3451.
[52] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2) (2006) 430–444.
[53] L. Zhang, Y. Shen, H. Li, VSI: a visual saliency-induced index for perceptual image quality assessment, IEEE Trans. Image Process. 23 (10) (2014) 4270–4281.
[54] L. Liu, B. Liu, H. Huang, A.C. Bovik, No-reference image quality assessment based on spatial and spectral entropies, Signal Process. Image Commun. 29 (8) (2014) 856–863.
[55] Q. Li, W. Lin, Y. Fang, No-reference quality assessment for multiply-distorted images in gradient domain, IEEE Signal Process. Lett. 23 (4) (2016) 541–545.
[56] A.K. Moorthy, A. Mittal, A.C. Bovik, Perceptually optimized blind repair of natural images, Signal Process. Image Commun. 28 (10) (2013) 1478–1493.
[57] R. Song, H. Ko, C.C. Kuo, MCL 3D Database. [Online]. Available: http://mcl.usc.edu/mcl-3d-database/.
[58] A. Benoit, P. Callet, P. Campisi, R. Cousseau, IRCCyN IVC Quality Assessment of Stereoscopic Images Database. [Online]. Available: http://www.irccyn.ec-nantes.fr/~lecallet/platforms.htm.
[59] J. Wang, A. Rehman, K. Zeng, S. Wang, Z. Wang, Waterloo IVC 3D Image Quality Database. [Online]. Available: http://ivc.uwaterloo.ca/database/3DIQA/3DIQA.php.