IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 8, 2020, accepted November 26, 2020, date of publication December 1, 2020,
date of current version December 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3041612

Stereoscopic Video Quality Assessment Using Oriented Local Gravitational Force Statistics

YUJIAN HOU[1], LIXIONG LIU[1], (Member, IEEE), YONGMEI ZHANG[2], AND QINGBING SANG[3]
[1]Beijing Laboratory of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China
[2]School of Information Science and Technology, North China University of Technology, Beijing 100144, China
[3]School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China

Corresponding author: Lixiong Liu (lxliu@bit.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61672095 and Grant 61371143, and in part by the Jiangsu Natural Science Foundation of Jiangsu Province of China under Grant BK20171142.

ABSTRACT We develop a new no-reference (NR) stereoscopic video quality assessment model that adopts oriented local gravitational force (OLGF) statistics in the space-time domain. The OLGF is a novel extension of an existing local gravitational force descriptor and includes two new components: relative local gravitational force magnitude and relative local gravitational force orientation. The resulting algorithm, called Stereoscopic Video Integrity Predictor using OLGF Statistics (SVIPOS), first uses our previous work to generate a cyclopean image and a product image in spatial domain from a stereoscopic video sequence pair, and calculates a frame difference image in temporal domain from a left video sequence. Then the OLGF model is deployed to compute various local gravitational force responses on the cyclopean image, the product image and the frame difference image. Finally, a support vector regression is used to map statistical features extracted from the computed gravitational force responses to stereoscopic video quality predications. SVIPOS is thoroughly tested on three publicly stereoscopic video quality databases. The experimental results show that SVIPOS outperforms state-of-the-art stereoscopic video quality methods.

INDEX TERMS Stereoscopic video quality assessment, oriented local gravitational force, binocular perception, natural scene statistics.

I. INTRODUCTION

In recent years, stereoscopic videos have been widely used in movies, games, education and other industries due to their good immersion experience and broad application prospects [1], [2]. However, inevitable distortions are generated during stereoscopic video processing [3] and distortions adversely affect human judgements of video content [4]. Therefore, it is still important to design effective algorithms for measuring stereoscopic video quality [5].

Here we focus on addressing the problem of no-reference (NR) stereoscopic video quality assessment (SVQA). Recently, some researchers have proposed NR SVQA algorithms that deliver good prediction performance. Yang *et al*. [6] used sparse representation and spatio-temporal saliency model to train a dictionary, and predicted a quality score of stereoscopic videos by using stacked auto-encoder and support vector machine. Chen *et al*. [7] proposed an NR SVQA method that combines natural scene statistics with disparity entropy measurement based on autoregressive prediction. Qi *et al.* [8] considered just-noticeable difference model to evaluate the quality of stereoscopic videos. Jiang *et al*. [9] considered motion information extracted in tensor decomposition domain to evaluate the stereoscopic video quality. Fang *et al.* [10] designed a 2D-TO-3D video quality assessment model, which uses edge and frame difference information in stereoscopic video to weight 2D image quality assessment (IQA)/video quality assessment (VQA) scores. Yang *et al*. [11] proposed an NR SVQA algorithm by using curvelet transform and an optical flow method to extract spatial and temporal features. However, just using left and right view summation or difference map to simulate the properties of binocular vision is not sufficient for representing stereoscopic video content. A way to fully consider the fused perception of a stereoscopic scene contributes to video quality predictions. Galkandage *et al*. [12] considered binocular suppression and used an IQA method to evaluate the quality of each frame, and proposed an optimized pooling method

The associate editor coordinating the review of this manuscript and approving it for publication was Gangyi Jiang.

to evaluate stereoscopic video quality. Appina *et al*. [13] proposed an NR SVQA algorithm using binocular disparity and motion component joint statistical modeling. Wang *et al*. [14] considered the binocular spatio-temporal internal mechanism and the free-energy theory to compute binocular difference maps, and use the multi-channel natural statistics to evalute 3D video quality. Although these algorithms have achieved some success, due to the complication of stereoscopic video content, binocular perception models and effective spatial and temporal feature representation are required to comprehensively integrate various aspects of stereoscopic videos [15].

The image gradient carries rich information of image structures and edges, and has been widely used as a promising image feature in the field of IQA/ VQA [16], [17]. Xue *et al.* [18] calculated gradient similarity to evaluate image quality. Lu *et al*. [19] extended the gradient descriptors to the spatio-temporal domain and synthesized local spatial and temporal information to obtain the quality of a stereoscopic video. Li *et al*. [20] generated new local descriptors for evaluating image quality by using local binary pattern to capture image feature contained in gradient maps. Yang *et al*. [21] trained gradient and phase dictionaries to learn the quality-influencing features in stereo images and used a support vector regression to obtain the final quality score. Ji *et al*. [22] used the gradient map, the intensity map, and orientation selectivity based visual patterns to jointly predict the image quality. However, these gradient-based methods only used the difference of two adjacent pixels in horizontal and vertical directions to measure image gradient magnitude and orientation, while ignoring the pixel information in other orientations. Very recently, Bhattacharjee *et al*. [23] proposed a local gravitational force (LGF) descriptor that considers relationships between all of the surrounding image pixels. Thus this descriptor can better capture local information than conventional gradient methods. There are two components in the descriptor: local gravitational force magnitude, which is to measure image interest areas (edges and textures), and local gravitational force orientation that conveys local geometry structure. LGF has been successfully used for image recognition, this makes us believe that the descriptor has good potential and can be further applied to other fields, especially in the field of visual quality assessment. Inspired by this, we study the principle of this descriptor and further explore the discriminative power of the local gravitational force magnitude and orientation feature components. Here we conducted a novel gravitational force based descriptor by considering the masking effect and intensity changes on local structures.

Binocular perception model, as an important form of stereo vision, can effectively measure the perceived quality of stereo views. Shao *et al*. [24] constructed a binocular perception model based on binocular and monocular energy responses to evaluate the quality of stereo images. Since a cyclopean image can reflect the comprehensive stimulus and binocular perception of left and right views, it is often used to calculate the perceived quality of three-dimensional scene [15]. Shen *et al*. [25] proposed an NR 3D IQA method, which uses a new disparity search algorithm to compute a cyclopean image. Liu *et al*. [26] computed a cyclopean image by utilizing a 3D saliency map to weight the left and right views. Yue *et al*. [27] use the 2D log-Gabor filters to extract energy and homogeneity features from the cyclopean image. In our previous work [28], we proposed a binocular spatial activity model that thoroughly considers the impact of binocular rivalry and suppression. Here, we use this model to synthesize a cyclopean image due to its good performance. A product image [28] is a simple way of measuring empirical correlation between binocular pixels and also suitable for representing spatial information in stereoscopic videos. Therefore, the cyclopean image and product image are used to characterize spatial stereoscopic video content.

Video contains rich temporal information, which has a significant impact on visual perception [5]. Several researchers have made good use of temporal information in video quality predication. Chen *et al*. [29] proposed an novel Recurrent-In-Recurrent Network, which can express the spatio-temporal distortion in the video more effective by encoding the motion information.Wu *et al*. [30] used the optical flow to calculate the motion trajectory as the temporal information of the video, and then computed the similarity of motion trajectories to evaluate the video quality. Huang *et al.* [31] represented the temporal domain information by comprehensively analyzing the ratio of freezing duration, the entropy value of an optical flow map, and the change of light intensity. Zhang *et al*. [32] utilized 3D-DCT coefficients to extract the spatio-temporal features of distorted videos, and then used the convolutional neural network to obtain a quality score. Although these methods well represent the temporal information of the videos, due to their own model complexity, they generally cost a lot of time in a process of characterizing the temporal information. Alternatively, the temporal frame difference, as a simple and efficient feature representation, has been widely used in the VQA field. Saad *et al*. [33] extracted DCT coefficients of frame difference and the motion vector between frames, and then used a regression model to obtain a video quality score. Soundararajan *et al*. [34] studied the statistical model of wavelet coefficients between video frames and designed entropic differencing indices to evaluate video quality. Liu *et al*. [35] proposed a video quality assessment model using space-time slice (STS) mappings, which uses frame difference to characterize the STS temporal representation. These studies have proved that the video frame difference has a certain distribution law, which can represent the moving edge in the video. Here, we extract motion information by computing a frame difference image. Although the above-mentioned cyclopean image, product image and frame difference image have been used in the IQA/VQA field, we collectively compute them to more effectively characterize spatial and temporal stereoscopic content due to their good performance in binocular perception. Further,

the proposed gravitational force based descriptor is used to extract quality-predictive information embedded in these feature images.

Here we develop a no-reference stereoscopic video quality assessment model, called SVIPOS. The resulting model first computes a cyclopean image, a product image and a frame difference image from the left and right videos, then uses a novel local gravitational force descriptor to generate four gravitational force maps on the computed images. Lastly, statistical features extracted from these gravitational force responses are used to predict quality scores. The contributions of this work are as follows. (1) We propose a novel Oriented Local Gravitational Force descriptor that considers the contributions of relative information of the local gravitational force magnitude and orientation feature components. This gravitational force based descriptor is deployed, for the first time, in the field of visual quality assessment. (2) We collectively use the cyclopean image, the product image and the frame difference image to represent spatio-temporal information in stereoscopic video. (3) The OLGF statistics are deployed to efficiently represent the computed spatio-temporal information, and integrate them into the proposed algorithm. (4) SVIPOS delivers highly competitive performance relative to state-of-the-art stereoscopic video quality methods.

The rest of this work is organized as follows. Section II describes the proposed OLGF descriptor. In Section III, the proposed SVIPOS model is described in detail. We evaluate the performance of our proposed model in Section IV, and conclude our paper in Section V.

II. ORIENTED LOCAL GRAVITATIONAL FORCE DESCRIPTOR

The local gravitational force descriptor is motived by law of universal gravitation, and can better capture local information [36]. Following this descriptor, we propose a novel oriented local gravitational force descriptor that incorporates the advantages of the LGF descriptor and additionally consider the contributions of relative information of its feature components.

A. LOCAL GRAVITATIONAL FORCE DESCRIPTOR

Law of universal gravity points out that any two objects in nature attract each other with a force and this force is inversely proportional to the square of their distance and proportional to the product of their masses [36]. For two objects having masses m_1 $\mathrm{m_1m_1m_1}$ and $m_2\mathrm{m_2}$, the gravitational force is

$$F = C_1 \frac{m_1 m_2}{r^2} \tag{1}$$

where r is the distance of the two objects, and C_1 is a gravitational constant. An image pixel is quite analogous to an object. Then express the horizontal direction component F_x and vertical direction component F_y of the force as:

$$F_x = F \times \cos\theta \tag{2}$$

$$F_y = F \times \sin\theta \tag{3}$$

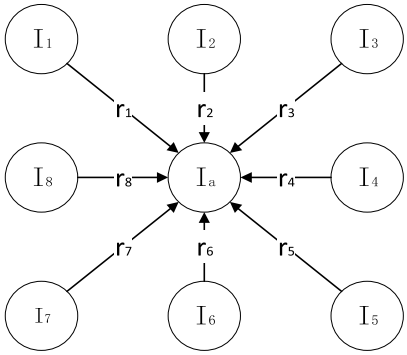


FIGURE 1. Gravitational forces exerted by eight adjacent pixels on the center pixel.

where θ is an angle between two pixels with respect to horizontal direction. Fig. 1 shows the gravitational force exerted by eight neighboring pixels on the central pixel in a 3 x 3 window. The F_x *and* F_y of the resultant force acting on the central pixel are computed as:

$$F_x = \sum_{i=1}^{p} \left(C_1 \times \frac{I_a \times I_i}{r_i^2} \times \cos\theta_{ai} \right) \tag{4}$$

$$F_y = \sum_{i=1}^{p} \left(C_1 \times \frac{I_a \times I_i}{r_i^2} \times \sin\theta_{ai} \right) \tag{5}$$

where I_a and I_i are the intensity of the central pixel and ith adjacent pixel respectively, θ_{ai} is the angle between these two pixels with respect to horizontal direction, p is the number of adjacent pixels and r_i is the Euclidean distance between these two pixels.

The force has gravitational force magnitude and orientation components. More specifically, the local gravitational force descriptor is composed of local gravitational force magnitude (LGFM), and local gravitational force orientation (LGFO). The gravitational force magnitude exerted on the central pixel is

$$F_M = \sqrt{(F_x)^2 + (F_y)^2} = I_a \times C_1 \sqrt{\sum_{i=1}^{p} \left(\frac{I_i}{r_i^2} \times \cos\theta_{ai} \right)^2 + \sum_{i=1}^{p} \left(\frac{I_i}{r_i^2} \times \sin\theta_{ai} \right)^2}, \tag{6}$$

and the gravitational force orientation is

$$F_O = \arctan\left(\frac{F_y}{F_x} \right) = \arctan\left(\frac{\sum_{i=1}^{p} \left(\frac{I_i}{r_i^2} \times \sin\theta_{ai} \right)}{\sum_{i=1}^{p} \left(\frac{I_i}{r_i^2} \times \cos\theta_{ai} \right)} \right). \tag{7}$$

The LGFM feature is to measure image interest areas (edges and textures), and the LGFO represents local geometry structure. In the following subsection, we will expand on the LGF descriptor and conduct a novel gravitational force based

descriptor by considering relative information of its feature components.

B. ORIENTED LOCAL GRAVITATIONAL FORCE DESCRIPTOR

Since visual cortical neurons are highly sensitive to local information [37], distortion may change the anisotropy of local areas and affect people's perception quality. Accounting for local gravitation force magnitude and orientation responses and their relevant potential may be a useful way to enhance the performance of the LGF descriptor. The relative gradient orientation and magnitude have been widely used in the field of IQA due to conveying information concerning changes in local structure and area [16]. They have good ability to capture textural changes and structure shifts arising from distortion. Inspired by this and the advantages of the LGF descriptor [23], we propose a novel gravitational force descriptor, called the Oriented Local Gravitational Force (OLGF), to explore the relative contributions of local gravitational force magnitude and orientation. Specifically, OLGF is composed of LGFM, LGFO, and two novel components: relative local gravitational force magnitude (RLGFM) and relative local gravitational force orientation (RLGFO). The RLGFO feature captures the deviation caused by local structure degradations, and the RLGFM feature measures the intensity changes in local structure. The relative magnitude and orientation in the local neighborhood capture the changes of the local video content. Thus, the relative gravitational force magnitude and orientation responses complement the LGF descriptor in a relative manner.

Here, we calculate the relative gravitational force orientation F_{RO} by subtracting the average gravitational force orientation exerted on an image pixel.

$$F_{RO} = F_O - F_{\hat{O}} \tag{8}$$

where $F_{\hat{O}}$ represents the local average orientation and is defined to be

$$F_{\hat{O}} = \arctan\left(\frac{F_{\hat{y}}}{F_{\hat{x}}}\right) \tag{9}$$

using the average gravitational force magnitudes of F_x and F_y

$$F_{\hat{y}} = \frac{\sum_{i=1}^{n} F_y}{n} \tag{10}$$

and

$$F_{\hat{x}} = \frac{\sum_{i=1}^{n} F_x}{n} \tag{11}$$

where n is the number of pixels in the neighborhood. To reduce descriptor complexity, the neighborhood size is set at 3 x 3.

Likewise, the relative gravitational force magnitude F_{RM} is defined as

$$F_{RM} = \sqrt{\left(F_y - F_{\hat{y}}\right)^2 + \left(F_x - F_{\hat{x}}\right)^2} \tag{12}$$

Fig. 2 shows various gravitational force representations of a video frame from the NAMA3DS1-COSPAD1 database [38].

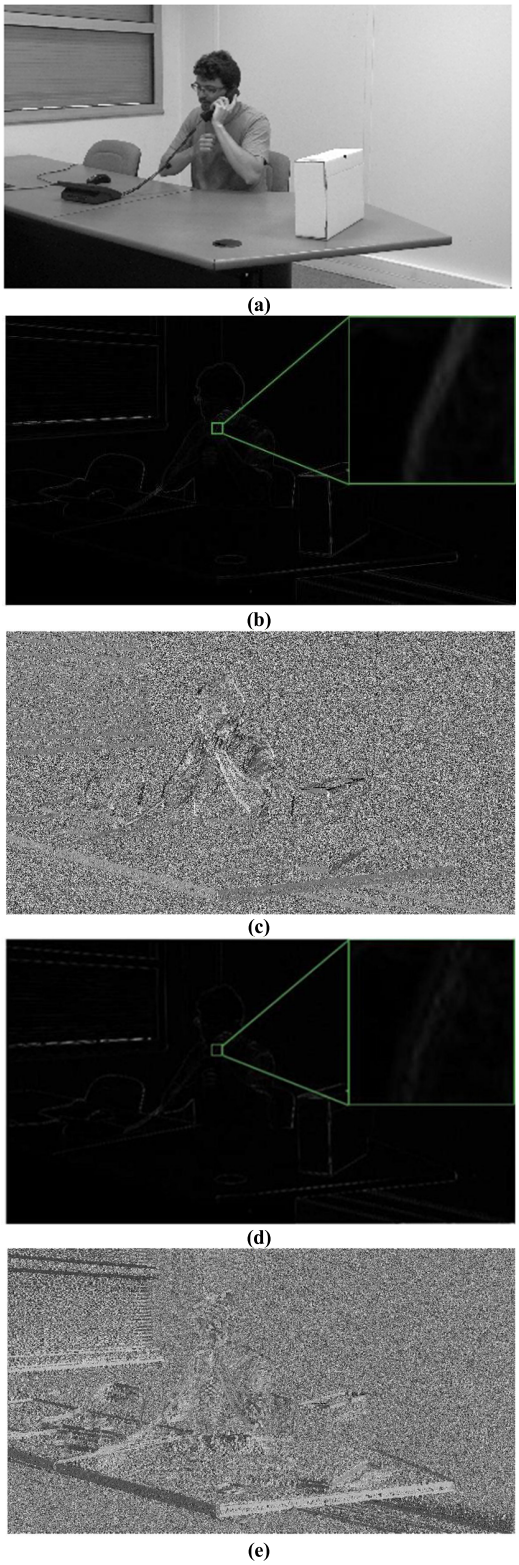


FIGURE 2. Gravitational force maps of an (a) original left video frame from the NAMA3DS1-COSPAD1 database. (b) Gravitational force magnitude and its zoomed image patch, (c) gravitational force orientation, (d) relative gravitational force magnitude and its zoomed image patch, and (e) relative gravitational force orientation.

It can be seen that there is apparent difference in local structure between the gravitational force orientation and the

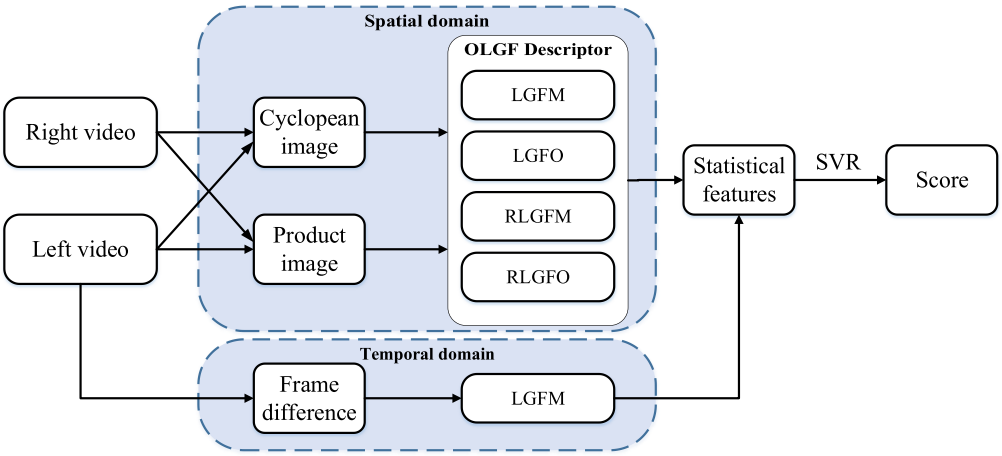


FIGURE 3. Framework of proposed SVIPOS model.

relative gravitational force orientation maps, and that the gravitational force magnitude and relative gravitational force magnitude are slightly different in edge and texture (see Figs. 2b and d). Although the differences between these four maps are not always apparent, each of them still represents obvious characters to a certain extent. Therefore, the combination of these feature components may be more suitable for different applications (such as texture recognition). Images and videos contain rich local structures and textures. We believe that our OLGF descriptor which represents the changes of the local structures and degradations may perform well in VQA. In the following we use the proposed OLGF descriptor to effectively represent spatial and temporal stereoscopic video content.

III. STEREOSCOPIC VIDEO INTEGRITY PREDICTOR

Next we describe the proposed SVIPOS method in detail. Fig. 3 shows the framework of the method. Along a top path that uses the left and right videos as inputs, a cyclopean image is computed using an existing binocular spatial activity model, and a product image is generated by measuring empirical correlation between binocular pixels. Both computed images are used to characterize spatial stereoscopic video content. All four components of the proposed OLGF descriptor are used to extract quality-predictive information embedded in the computed images, yielding eight gravitational force maps. Along the bottom path, we extract temporal information from the left video by computing a frame difference image and then only use the LGFM of the OLGF descriptor to capture the motion information on the computed frame difference image, yielding one gravitational force map. Since all the generated gravitational force responses show obvious statistical characteristics, we extract statistical features from all nine generated responses. Finally, a support vector regression is used to map generated statistical features to stereoscopic video quality predictions.

A. BINOCULAR PERCEPTION MODEL AND SPATIAL INFORMATION

Stereoscopic video is usually composed of left and right channels of video, so it is very important to consider the interaction between two monocular videos in the reconstruction of stereoscopic vision [39]. Generally, synthesizing cyclopean image using binocular perception model is an effective way to predict the perceived quality of stereo video content. Since the amount of spatial activity contained in a stereo pair is closely related to its perceived quality, we here use a binocular perception model [28] based on a measure of spatial activity to simulate the stereoscopic binocular perception response.

The cyclopean image (CI) is computed as (13), as shown at the bottom of the page, where $S_{L(k)}$ is a moving 17 x 17 window centered at coordinate k, $\varepsilon\left[S_{L(k)}\right]$ and $\varepsilon\left[S_{R(k+d)}\right]$ represent the spatial activity within neighborhood $S_{L(k)}$ of the left view $I_{L(k)}$ and that of the disparity-corrected right view $I_{R(k+d)}$ respectively, d represents the disparity at k, and C_2 is a constant to ensure stability. The disparity map is generated using an existing stereo matching algorithm [40], in which the SSIM metric is used to choose the best matching. The spatial activity $\varepsilon(\cdot)$ is given by:

$$\varepsilon\left(S_{L(k)}\right) = log_2[1+\sigma_L^2(k)] \tag{14}$$

where $\sigma_L^2(k)$ is the variance of $S_{L(k)}$. The measure of spatial activity is able to determine whether there exists binocular rivalry or suppression between the left view and right view, and to further gauge the relative degree of influence of both views while existing binocular rivalry.

$$CI = \frac{\left\{\varepsilon\left[S_{L(k)}\right]+C_2\right\}\times I_{L(k)} + \left\{\varepsilon\left[S_{R(k+d)}\right]+C_2\right\}\times I_{R(k+d)}}{\left\{\varepsilon\left[I_{L(k)}\right]+C_2\right\} + \left\{\varepsilon\left[I_{R(k+d)}\right]+C_2\right\}} \tag{13}$$

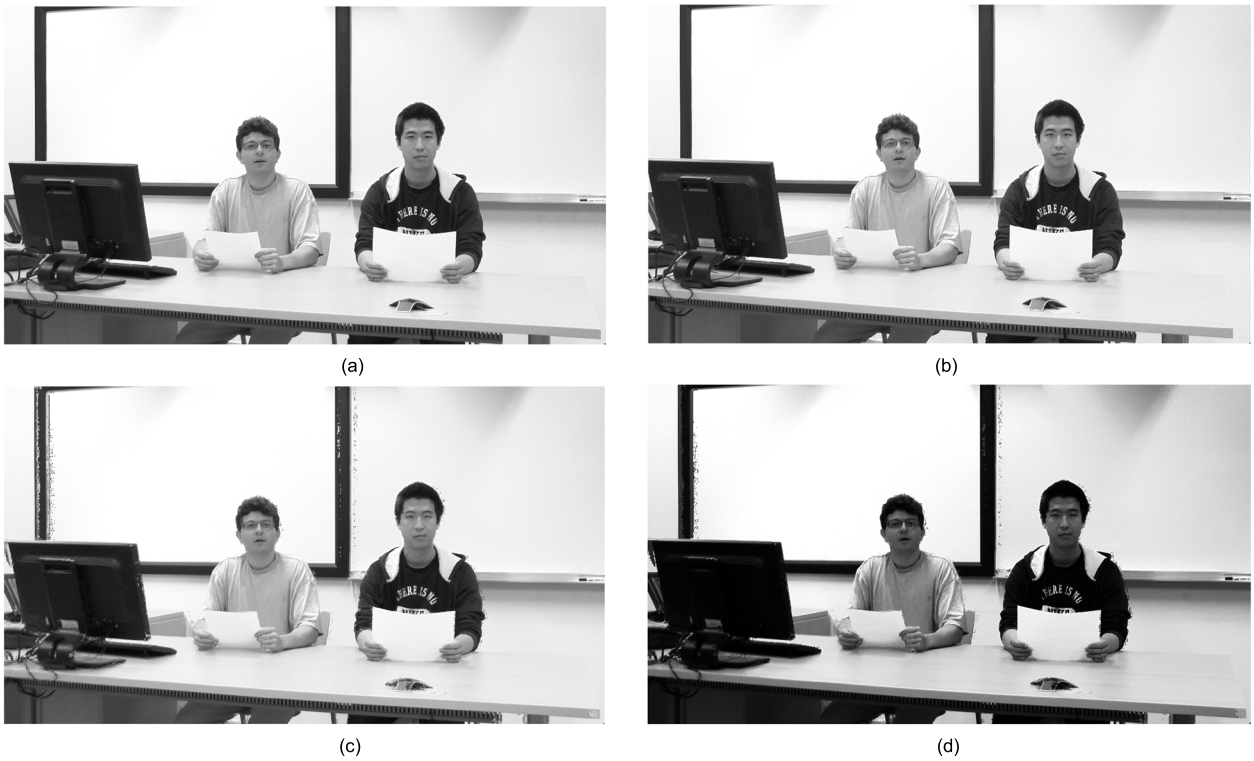


FIGURE 4. Four images. (a) Left video frame and (b) right video frame from the NAMA3DS1-COSPAD1 database, (c) cyclopean image, and (d) product image.

The product image is an estimate of disparity compensated correlation. Using it may achieve good performance and still has low time complexity [28]. We also consider this simple product estimate to express empirical correlation between binocular pixels. The product image is given by:

$$PI = I_{L(k)} \times I_{R(k+d)} \tag{15}$$

The computed cyclopean and product images represent considerable amount of data. Using spatial activity model or bivariate product to generate a feature map has lower time complexity, which is useful for stereoscopic video applications with limited computing resources. Fig. 4 shows the left view, the right view, the cyclopean image and the product image.

B. SPATIO-TEMPORAL GRAVITATIONAL FORCE RESPONSES

As mentioned above, the synthesized cyclopean and product images can effectively simulate the human stereoscopic visual perception system. Therefore, we deploy the binocular perception models on each frame of the stereoscopic videos and use these synthesized images as the spatial representation of the stereoscopic videos.

The temporal information in stereoscopic video also has a great impact on human perception. Considering temporal information contributes to the quality prediction of stereoscopic videos. Frame difference, as a widely used temporal representation, can be used to express motion information in videos. The neural responses of the visual cortex are almost separated in time domain [41]. Thus we also extract the frame difference image in time domain to represent motion information in stereoscopic videos. The frame difference is given by:

$$DF_t = I_{t+1} - I_t, \quad t \in [1, T-1] \tag{16}$$

where I_t represents the tth frame, and T is the number of frames in a video.

The frame difference and the above-mentioned cyclopean and product images collectively represent spatial and temporal stereoscopic video content. The proposed OLGF descriptor is then used to extract quality-predictive information embed in these three feature images, yielding gravitational force based feature responses. Specifically, all four components of the OLGF are deployed to capture discriminative features from the cyclopean and product images, while we try to use the frame difference as the complementary quality information and only use the local gravitational force magnitude to capture features from it in order to reduce computational complexity. All the four generated gravitational force responses to the cyclopean and product images and the generated local gravitational force magnitude response to the frame difference are then used for feature extraction.

C. STATISTICAL FEATURES

The natural scene statistics (NSS) features achieved good quality predictive power. Zhou *et al.* [42] used NSS features

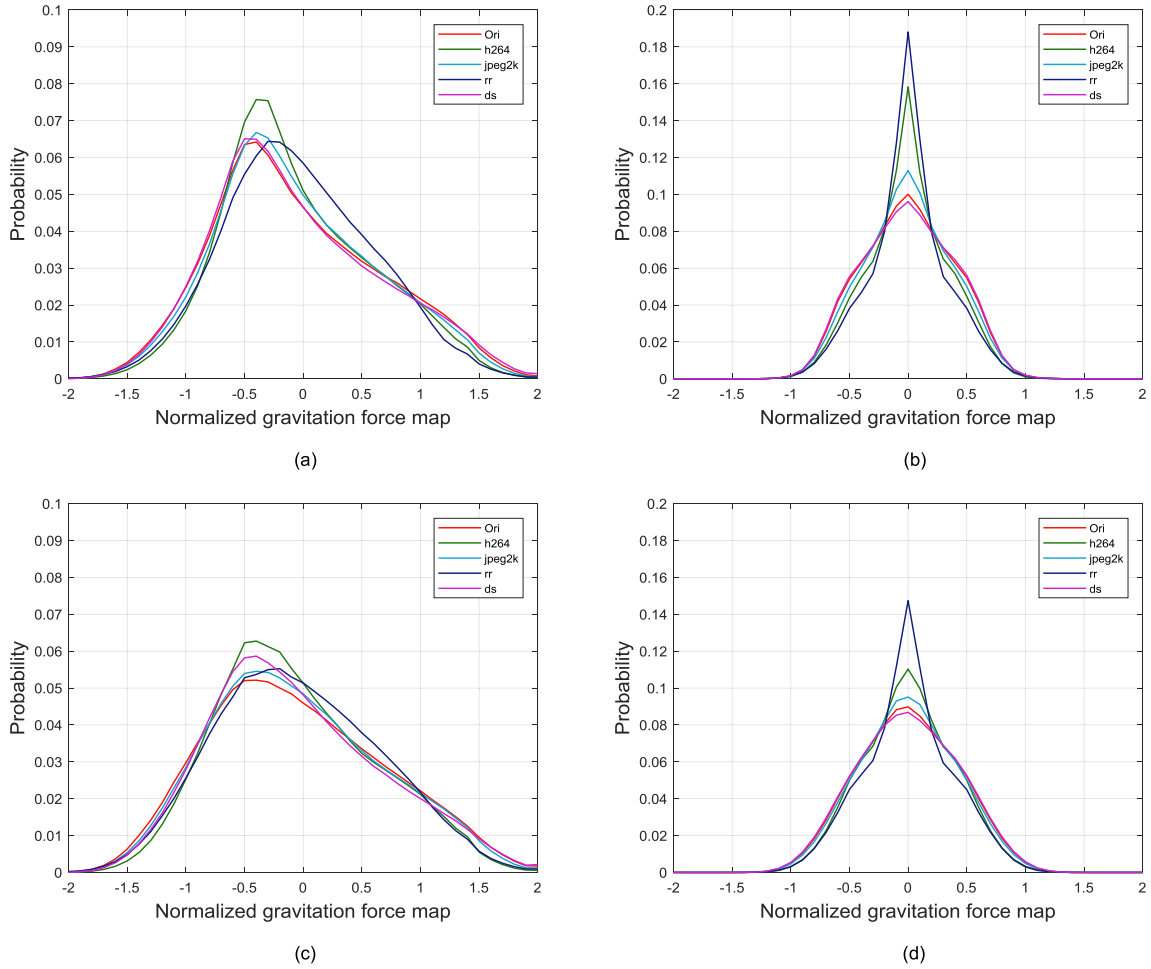


FIGURE 5. Histograms of normalized gravitational force based maps of the cyclopean image computed on an original stereoscopic video and its different distorted versions. (a) Local gravitational force magnitude, (b) local gravitational force orientation, (c) relative local gravitational force magnitude, and (d) relative local gravitational force orientation. ori: original stereoscopic video. h264: H.264. jpeg2k: JPEG2000. rr: reduction of resolution. ds: downsampling with sharpening.

with a multivariate Gaussian model to construct a blind stereoscopic image quality metric. Gu *et al.* [43] proposed an NR IQA method by multiscale NSS analysis. Here we also extract NSS features from the computed gravitational force based maps. A normalization process [44] is first applied to reduce spatial redundancy on a gravitation force based map F_j as follows:

$$\hat{F}_j = \frac{F_j - \mu_j}{\sigma_j + C_3}, \quad j \in [M, O, RM, RO] \tag{17}$$

where σ_j and μ_j are the standard deviation and local mean value of F_j respectively, and C_3 is a constant to ensure stability. The histograms of the four normalized gravitational force based maps of the cyclopean image computed on an original stereoscopic video and different distorted versions of it from the NAMA3DS1-COSPAD1 database [38] are shown in Fig. 5. We may see that the histograms of the normalized coefficients of various gravitational force based maps are similar to a slightly asymmetrical Gaussian appearance, while the distribution curves of various distortions have obvious differences. We also show the histograms of a normalized local gravitational force orientation map of the cyclopean image computed on the H.264 and JPEG2000 distorted versions of an original stereoscopic video at different distortion degrees in Fig. 6. The results again show that our gravitational force feature components are highly sensitive to the distortions.

We use an asymmetric generalized Gaussian distribution (AGGD) [44] to quantify the empirical distribution of normalized coefficients of a gravitation force based map $\hat{F}_j$. The AGGD model with zero mean is

$$f\left(z;\gamma,\sigma_l^2,\sigma_r^2\right) = \begin{cases} \dfrac{\gamma}{(\beta_l+\beta_r)\,\Gamma\left(\frac{1}{\gamma}\right)} e^{-\left(\frac{-z}{\beta_l}\right)^{\gamma}}, & z < 0 \\ \dfrac{\gamma}{(\beta_l+\beta_r)\,\Gamma\left(\frac{1}{\gamma}\right)} e^{-\left(\frac{z}{\beta_r}\right)^{\gamma}}, & z \geq 0 \end{cases} \tag{18}$$

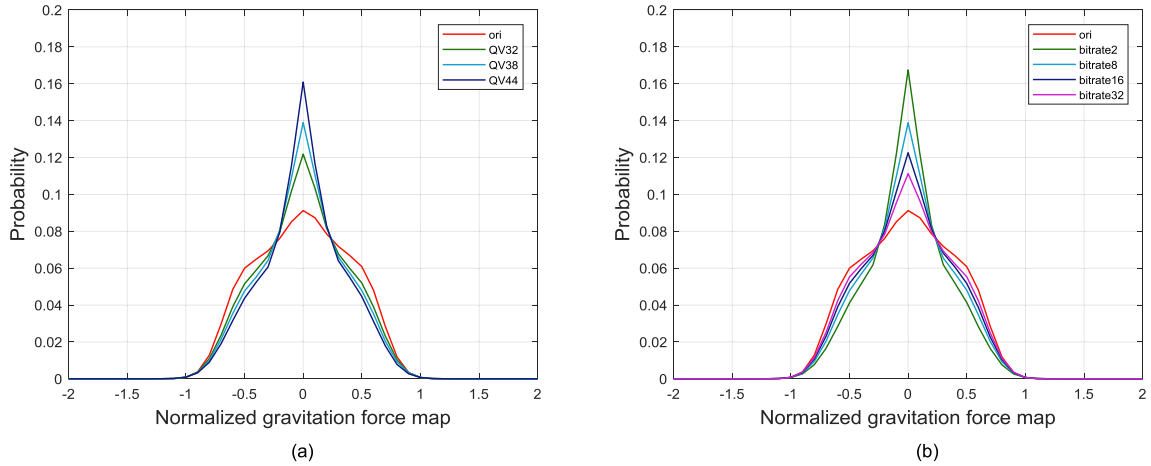


FIGURE 6. Histograms of normalized local gravitational force orientation map of the cyclopean image computed on the (a) H.264 and (b) JPEG2000 distorted versions of an original stereoscopic video at several distortion levels. ori: original stereoscopic video. QV32: quantization value 32. QV38: quantization value 38. QV44: quantization value 44. bitrate2: 2 megabits per second (Mbps). bitrate8: 8 Mbps. bitrate16: 16 Mbps. bitrate32: 32 Mbps.

where

$$\beta_l = \sigma_l \sqrt{\frac{\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{3}{\gamma}\right)}} \tag{19}$$

$$\beta_r = \sigma_r \sqrt{\frac{\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{3}{\gamma}\right)}} \tag{20}$$

and γ is the shape parameter, σ_l and σ_r are the scale parameters of left and right sides respectively. We use the best fit parameters $(\eta, \gamma, \sigma_l, \sigma_r)$ as statistical features, where

$$\eta = (\beta_l - \beta_r)\frac{\Gamma\left(\frac{2}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma}\right)} \tag{21}$$

We also consider the multi-channel characteristics of the human vision [45], and extract these statistical features from all gravitational force based maps at two scales, namely original and coarse scales.

D. QUALITY EVALUATION

In order to predict stereoscopic video quality, it is necessary to select a regression model to generate a quality score. Here we choose a support vector regression (SVR) with the radial basis function kernel to train regression models, since it achieved good performance on dealing with high dimensional regression problems [46].

The LIBSVM package [47] is utilized to build a mapping from extracted statistical features to the quality score. First, we formed a regression training set by selecting a proportion of all the distorted stereoscopic videos. The regression training set consists of extracted features and their quality scores. Then, we utilized the SVR on it to train a regression model, yielding a finite-dimensional quality evaluation vector $\vec{h}$ and a bias b. The final quality score Q is given by:

$$Q = \vec{h}\mathrm{V}_f + b \tag{22}$$

where the V_f is the stereoscopic video feature vector obtained by average pooling. We summarize the commonly used symbols in Table 1.

TABLE 1. Summary of commonly used symbols.

Symbol	Symbol description
F	Gravitational force
F_x	Horizontal direction component of F
F_y	Vertical direction component of F
$F_{\hat{x}}$	Local average of F_x
$F_{\hat{y}}$	Local average of F_y
F_M	Gravitational force magnitude
F_O	Gravitational force orientation
F_{RO}	Relative gravitational force orientation
F_{RM}	Relative gravitational force magnitude
θ	Angle between two pixels with respect to horizontal direction
p	Number of adjacent pixels exerted on a center pixel
r	Euclidean distance between two pixels.
CI	Cyclopean image
PI	Product image
DF_t	Frame difference

IV. EXPERIMENTAL RESULTS

Our proposed model was validated on three publicly available stereoscopic video quality databases: the NAMA3DS1-COSPAD1 database [38], the SVQA database [8] and the Waterloo IVC Phase I database [48]. The NAMA3DS1-COSPAD1 database consists of 10 reference stereoscopic video sequences and 100 symmetrically distorted videos sequences of frame rates 25 frames per second (fps), and

TABLE 2. Performance of compared methods on three stereoscopic video databases.

	NAMA3DS1				SVQA				Waterloo-IVC Phase I			
Metrics	SRCC	PLCC	KRCC	RMSE	SRCC	PLCC	KRCC	RMSE	SRCC	PLCC	KRCC	RMSE
PSNR	0.6118	0.4894	0.3153	1.1318	0.8274	0.7968	0.6536	0.3205	0.5333	0.7084	0.3850	15.4544
SSIM	0.7408	0.7347	0.5706	0.8090	0.7405	0.7325	0.5624	0.6770	0.4794	0.5044	0.3440	18.9060
ViS3	0.8988	0.9098	0.7157	0.4834	0.8602	0.8485	0.6828	0.5263	0.3250	0.5016	0.2300	18.9414
SpEED	0.6828	0.7246	0.5500	0.8220	0.7642	0.7430	0.5814	0.6657	0.2941	0.2692	0.2052	21.0869
SINQ	0.871	0.9253	0.7146	0.4368	0.8565	0.8179	0.6674	0.5828	0.887	0.9348	0.7400	7.6829
SJND	0.6229	0.6503	0.4575	0.8629	0.8379	0.8415	0.6650	0.5372	-	-	-	-
MNSVQM	0.8394	0.8611	-	0.5634	-	-	-	-	-	-	-	-
Modi3d	0.6233	0.606	-	0.9853	-	-	-	-	0.4265	-	-	-
Wang	0.9091	0.9147	-	-	-	-	-	-	-	-	-	-
BSVQE	0.9117	0.9411	0.7607	0.4018	0.9354	0.9427	0.7895	0.3288	0.9111	**0.9574**	0.7657	**6.2116**
SVIPOS	**0.9425**	**0.9765**	**0.819**	**0.2522**	**0.9391**	**0.9462**	**0.7995**	**0.3188**	0.9171	0.9546	**0.7707**	6.4624

resolution 1920 x 1080 pixels. The distortion types include JPEG2000, image sharpening, H.264, reduction of resolution, and downsampling with sharpening. The SVQA database contains 9 reference video sequences and 450 symmetrically and asymmetrically distorted video sequences of frame rates 25 fps, and multi-resolution (1920 $\times$ 1088, 1280 $\times$ 720, 1024 $\times$ 768). The distortion types include H.264 and blur. The Waterloo IVC Phase I contains 4 reference stereoscopic video sequences and 176 symmetrically and asymmetrically distorted video sequences of resolution 1024 $\times$ 768 pixels, and frame rates 16.67/30 fps.

Each database was randomly divided into 80% training and 20% testing subsets in our experiments. We repeated the training-testing procedure over 1000 times and utilized four standard criteria to evaluate model performance: the Pearson's Correlation Coefficient (PLCC), Spearman rank-order correlation coefficient (SRCC), Kendall Rank Order Correlation Coefficient (KRCC) and the Root Mean Squared Error (RMSE). Higher SRCC, KRCC, PLCC and lower RMSE values indicate better consistency between human judgements and prediction scores. Several widely used 2D and 3D IQA/VQA algorithms were selected in comparison experiments, including the 2D IQA algorithms PSNR and SSIM [49], the 3D IQA algorithm SINQ [28], the 2D VQA algorithms ViS3 [51] and SpEED [50], and the 3D VQA algorithms SJND [8], MNSVQM [9], Modi3d [52], Wang's model [14] and BSVQE [7]. The IQA methods were applied on each video frames and used mean pooling to generate a final score, while the VQA methods were run using the settings reported in the original paper. The software release of each of the selected algorithms except SJND [8], MNSVQM [9], Modi3d [52] and Wang's model [14] is publicly available.

A. PERFORMANCE ON THREE DATABASES

We tested the performance of SVIPOS against several widely used IQA/VQA algorithms except SJND [8], MNSVQM [9], Modi3d [51] and Wang's model [14] on the NAMA3DS1-COSPAD1, SVQA, and Waterloo IVC Phase I databases, for the codes for these six methods are not publicly available. Accordingly we listed their results reported in the original papers in the tables below. The median PLCC, SRCC, KRCC and RMSE values across 1000 random runs are tabulated in Table 2. The top performing algorithm is highlighted in bold type. It can be seen that our proposed SVIPOS performed better than all the other compared method on the NAMA3DS1 and SVQA databases, and that it delivered better performance than all the other compared algorithm except BSVQE model on the Waterloo IVC Phase I database. Compared with BSVQE algorithm, SVIPOS still achieved competitive performance on the Waterloo IVC Phase I database. This may be because the cyclopean and product images computed on the left and right videos we use are more effective for binocular perception than simple summation or difference of left and right videos, and our method used the AGGD model when computing NSS features, which may be partially suitable for representing information in synthetic stereoscopic videos of unnatural scenes [52] from the Waterloo IVC Phase I database. These results show that SVIPOS performed best in 10 of 12 performance comparison metrics. Also, we show the scatter plots between the MOS (Mean Opinion Score) values and predicted scores of SVIPOS on three databases in Fig. 7. Clearly, the prediction scores of our proposed method match well with human subjective judgements. The results shown in Table 2 and Fig. 7 may indicate that SVIPOS is able to more effectively measure the perceptual quality of stereoscopic video sequences.

B. PERFORMANCE ON EACH DISTORTION TYPE

Since the NAMA3DS1 and SVQA databases contain different distortion types, we only conducted an experiment on each distortion type from these two databases. The SRCC values are listed in Table 3. The two top-performing metrics are highlighted in bold type. The "Downsampling and Sharpening" is a combination of three distortions: reduction of resolution, image sharpening, and downsampling with sharpening. This is because each distortion contains only 10

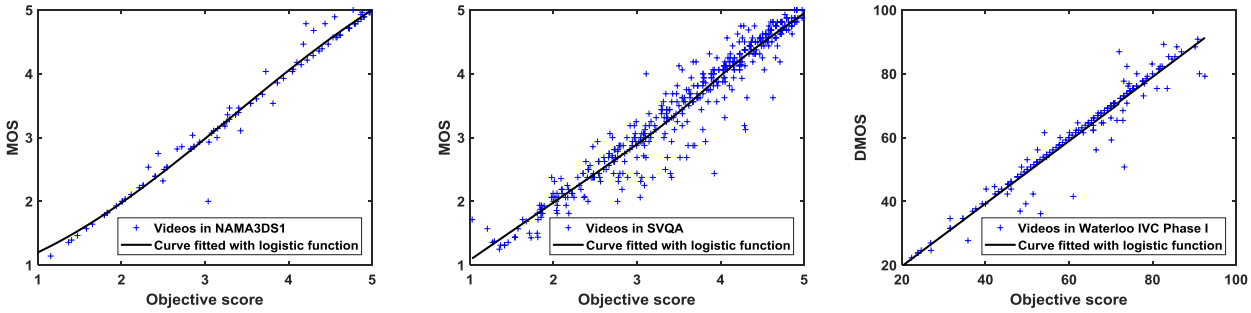


FIGURE 7. Scatter plots of MOS against predicted quality scores of SVIPOS on the (a) NAMA3DS1, (b) SVQA, and (c) Waterloo IVC Phase I databases.

TABLE 3. Median SRCC of compared methods on each distortion type.

	NAMA3DS1			SVQA	
Type	H.264	JPEG 2000	Downsampling and Sharpening	H.264	Blur
PSNR	0.5227	0.6314	0.5181	0.8450	0.8420
SSIM	0.6695	0.5915	0.6597	0.7670	0.8550
ViS3	0.8657	**0.9662**	**0.9453**	0.8826	0.8466
SpEED	**0.9529**	**0.9712**	0.1231	0.8449	0.6962
SINQ	0.7714	0.8571	0.8857	0.8147	0.9197
SJND	0.6810	0.6901	0.5071	-	-
MNSVQM	0.7715	0.8982	0.7775	-	-
Modi3D	0.6935	0.8540	-	-	-
Wang	0.8894	0.9379	-	-	-
BSVQE	0.8452	0.8383	0.8286	**0.9286**	**0.9513**
SVIPOS	**0.9429**	0.9048	**0.8857**	**0.9039**	**0.9663**

samples, and such a small amount of data will lead to unstable results.

From Table 3, we can see that SVIPOS delivered competitive performance on each distortion type from the two databases in most cases, and that other 3D VQA algorithms also achieved good results on each distortion. In particular, the 2D VQA methods ViS3 and SpEED also achieved good performance on several distortion types from the NAMA3DS1 database, while they performed significantly worse on each distortion type from the SVQA database. Since the NAMA3DS1 database only contains symmetric distortions, and the SVQA database contains symmetric and asymmetric distortions, these results may indicate that our proposed is also able to effectively measure the perceptual severity of asymmetric and symmetric distortions in stereoscopic video sequences.

C. STATISTICAL SIGNIFICANCE

To verify the statistical significance of the performance of all the compared models except SJND [8], MNSVQM [9], Modi3d [52] and Wang's model [14], we implemented the statistical tests [53] on the NAMA3DS1-COSPAD1, SVQA and Waterloo IVC Phase I databases. Fig. 8 shows the box plots of SRCC values of models across 1000 runs on three databases. It can be seen that the accuracy and stability of SVIPOS boosts best among that of all methods. We also utilized the T-test [54] to measure significance on the SRCC of all methods. The null hypothesis we adopted is that the mean correlation value in the row is equal to that of the column at 95% confidence level. Weshow the results of the t-test on three databases in Fig. 9, where the indicator "1" or "−1" means that the method in a row is statistically better, or statistically worse than the method in a column, while the indicator "0" denotes there are no statistical differences between them. The results show that SVIPOS is stable, and statistically superior to all other methods.

D. CONTRIBUTIONS OF DIFFERENT FEATURE IMAGES

To verify the contributions of each feature images, we also tested the performance of different feature images on the NAMA3DS1-COSPAD1 database. The median SRCC, PLCC, KRCC and RMSE values are listed in Table 4. We found that although features extracted from the frame difference performed slightly worse than those from the cyclopean image or the product image, it still achieved a modest improvement in performance. These results indicate that the cyclopean image and product image may be complementary to each other, and both of them are able to more effectively represent spatial information in stereoscopic video content. The better performance was achieved using features from all three parts. Clearly, all three sources are necessary and contribute to the performance of our method.

E. VARIATION WITH WINDOW SIZE

In order to study the impact of the window size of OLGF descriptor on the algorithm performance, we tested the performance of SVIPOS with different sizes on the NAMA3DS1 database. The results are listed in Table 5.

The results show that the performance of SVIPOS has a decreasing trend when the window size is increasing. This may because more pixels are farther away from the center pixel and occupy a smaller weight, the calculations of larger windows tend to be non-local.

F. ADVANTAGES OF ORIENTED LOCAL GRAVITATIONAL FORCE DESCRIPTOR

In order to further analyze the advantages of the proposed OLGF descriptor, we tested the performance of

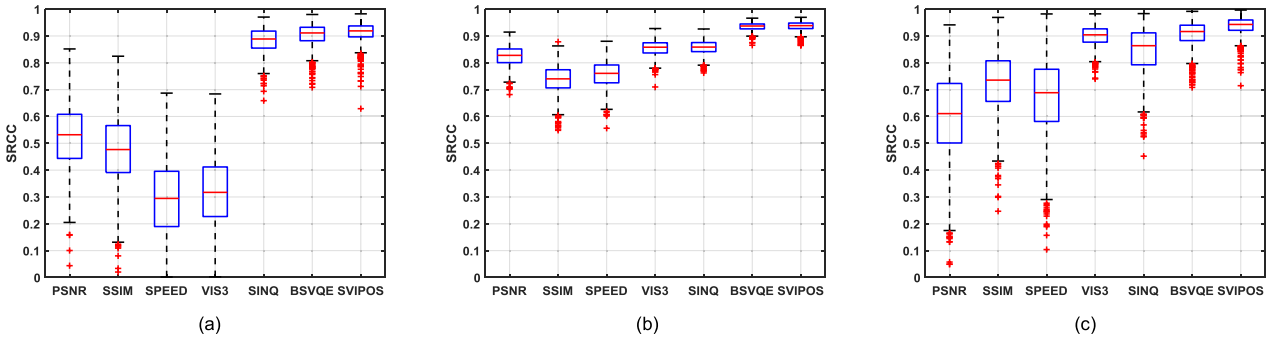
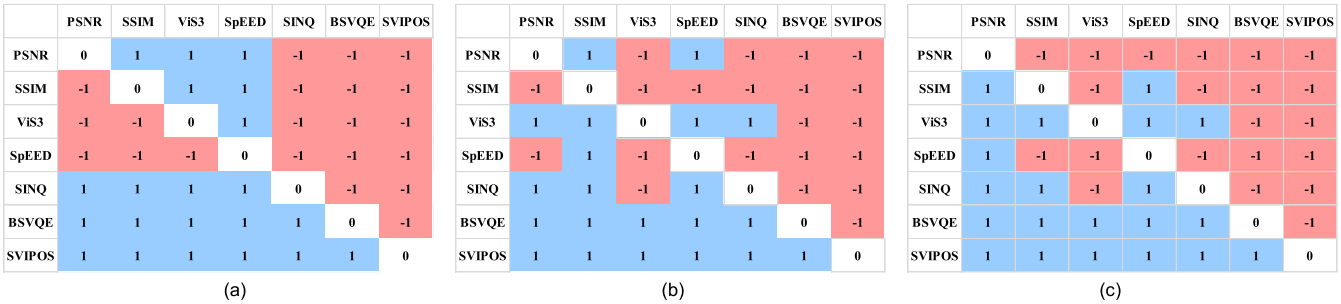


FIGURE 8. Box plots of SRCC values of models across 1000 runs on the (a) NAMA3DS1, (b) SVQA, and (c) Waterloo IVC Phase I databases.



	PSNR	SSIM	ViS3	SpEED	SINQ	BSVQE	SVIPOS
PSNR	0	1	1	1	-1	-1	-1
SSIM	-1	0	1	1	-1	-1	-1
ViS3	-1	-1	0	1	-1	-1	-1
SpEED	-1	-1	-1	0	-1	-1	-1
SINQ	1	1	1	1	0	-1	-1
BSVQE	1	1	1	1	1	0	-1
SVIPOS	1	1	1	1	1	1	0

(a)

	PSNR	SSIM	ViS3	SpEED	SINQ	BSVQE	SVIPOS
PSNR	0	1	-1	1	-1	-1	-1
SSIM	-1	0	-1	-1	-1	-1	-1
ViS3	1	1	0	1	1	-1	-1
SpEED	-1	1	-1	0	-1	-1	-1
SINQ	1	1	-1	1	0	-1	-1
BSVQE	1	1	1	1	1	0	-1
SVIPOS	1	1	1	1	1	1	0

(b)

	PSNR	SSIM	ViS3	SpEED	SINQ	BSVQE	SVIPOS
PSNR	0	-1	-1	-1	-1	-1	-1
SSIM	1	0	-1	1	-1	-1	-1
ViS3	1	1	0	1	1	-1	-1
SpEED	1	-1	-1	0	-1	-1	-1
SINQ	1	1	-1	1	0	-1	-1
BSVQE	1	1	1	1	1	0	-1
SVIPOS	1	1	1	1	1	1	0

(c)

FIGURE 9. Results of T-test between median SRCC of several models on the (a) NAMA3DS1, (b) SVQA, and (c) Waterloo IVC Phase I databases.

TABLE 4. Performance comparison of three feature parts on the NAMA3DS1 database.

	Left view	Right view	Cyclopean	Product	Cyclopean and Product	Frame difference	Frame difference and Cyclopean	Frame difference and Product	SVIPOS
SRCC	0.8952	0.8757	0.9284	0.9203	0.9331	0.8386	0.9364	0.9389	**0.9425**
PLCC	0.9369	0.9182	0.9645	0.9605	0.9695	0.8904	0.9728	0.9730	**0.9765**
KRCC	0.7468	0.7261	0.7983	0.7896	0.8079	0.6726	0.8095	0.8141	**0.8190**
RMSE	0.4165	0.4594	0.3060	0.3262	0.2913	0.5291	0.2693	0.2666	**0.2522**

TABLE 5. Performance of SVIPOS with different window sizes on the NAMA3DS1 database.

Window size	SRCC	PLCC	KRCC	RMSE
3 x 3	0.9425	0.9765	0.8190	0.2522
5 x 5	0.9381	0.9733	0.8131	0.2681
7 x 7	0.9376	0.9710	0.8156	0.2779
9 x 9	0.9373	0.9698	0.8162	0.2842

different gravitational force feature components on the NAMA3DS1 database. For simplicity we divided the feature components into three groups and only tested the performance of each group. The results are tabulated in Table 6. We can see that the relative local gravitational force magnitude and orientation responses achieved better performance than the local gravitational force magnitude and orientation (i.e., the LGF descriptor). The better performance was delivered by combining all four components. These results confirm the efficacy of our proposed OLGF descriptor.

G. DATABASE INDEPENDENCE

We tested the database independence of SVIPOS on the NAMA3DS1-COSPAD1, SVQA, and Waterloo IVC Phase I databases.Two available models SINQ and BSVQE were also tested. Table 7 tabulates the median SRCC values of three algorithms. The results of SVIPOS were highly competitive on three databases.

H. COMPUTATIONAL COMPLEXITY

We also compared the computational complexity of several algorithms on a stereoscopic video of resolution 1920 x 1080 pixels, frame rates 25 fps, and sixteen second durations (i.e., the input stereoscopic video is V = 1920 × 1080 × 400 pixels) from the NAMA3DS1 database. Let g denote the window size in the process of feature computation. For example, g in ViS3 is the size of the convolution matrix, while in BSVQE, g is the spans of the Gaussian operator, and g in SVIPOS is the neighborhood size. The results are shown in Table 8. We can see that 2D methods PSNR, SSIM, SpEED, and ViS3 are faster than 3D methods SINQ, BSVQE, and SVIPOS, because these 2D algorithms only consider spatial factors and do not take into account binocular perception mechanisms such as binocular disparity consuming a lot of time in 3D IQA/VQA methods. SINQ is a 3D IQA method which does not process the temporal information so it is

TABLE 6. Performance of different gravitational force components on the NAMA3DS1 database.

	Local gravitational force magnitude and orientation	Relative local gravitational force magnitude and orientation	OLGF
SRCC	0.9057	0.9217	0.9331
PLCC	0.9381	0.9587	0.9695
KRCC	0.7669	0.7817	0.8079
RMSE	0.3992	0.3279	0.2913

TABLE 7. Median SRCC of different models across three databases.

Train Database	NAMA3DS1		SVQA		Waterloo-IVC Phase I	
Test Database	SVQA	Waterloo-IVC Phase I	NAMA3DS1	Waterloo-IVC Phase I	NAMA3DS1	SVQA
SINQ	0.3846	0.1783	0.2109	0.1587	0.2535	0.2525
BSVQE	0.4811	**0.4480**	**0.5511**	0.3908	0.3836	0.4877
SVIPOS	**0.5130**	0.3472	0.4769	**0.5341**	**0.4787**	**0.6296**

TABLE 8. Comparison of time complexity of seven algorithms.

	Algorithm complexity	Time (min)
PSNR	$O(V)$	1.57
SSIM	$O(Vg)$	2.03
SpEED	$O(Vg)$	2.31
ViS3	$O(Vg^2 + V^2log(V))$	42.62
SINQ	$O(V^2 + Vg^2)$	91.40
BSVQE	$O(V^2g + Vg)$	156.10
SVIPOS	$O(V^2 + Vg^2)$	113.46

faster than SVQA methods SVIPOS and BSVQE. Although our proposed SVIPOS method is slower than compared 2D methods and 3D IQA method SINQ, it is faster than the SVQA method BSVQE by a large margin.

V. CONCLUSION

We have developed a novel NR SVQA model using a novel oriented local gravitational force descriptor. The proposed OLGF is utilized to effectively represent discriminative information in stereoscopic video content, and accounts for local gravitation force magnitude and orientation and their relevant potential. Deploying its two novel feature components may further boost performance. Compared with other methods that only use the summation or difference of left and right views, we use the cyclopean and product images to measure correlations between the left and right videos more effectively. The experimental results tested on three stereoscopic video databases show that SVIPOS has outstanding performance. We believe that contributions of relative local gravitational force magnitude and orientation features and a more effective motion model may merit further research. In the future we will incorporate local entropy to build a general motion model for performance improvement, and apply our proposed OLGF to other image and video processing applications, especially in blind image quality predictions.

REFERENCES

[1] S. Winkler and D. Min, "Stereo/multiview picture quality: Overview and recent advances," *Signal Process., Image Commun.*, vol. 28, no. 10, pp. 1358–1373, Nov. 2013.

[2] C.-C. Su, A. K. Moorthy, and A. C. Bovik, "Visual quality assessment of stereoscopic image and video: Challenges, advances, and future trends," in *Visual Signal Quality Assessment*. Cham, Switzerland, Springer, 2015, pp. 185–212.

[3] D. K. Broberg, "Infrastructures for home delivery, interfacing, captioning, and viewing of 3-D content," *Proc. IEEE*, vol. 99, no. 4, pp. 684–693, Apr. 2011.

[4] *Subjective Methods for the Assessment of Stereoscopic 3DTV Systems*, document ITU-R BT.2021, Int. Telecommun. Union, Geneva, Switzerland, Feb. 2015.

[5] F. Zhang and D. R. Bull, "A perception-based hybrid model for video quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 6, pp. 1017–1028, Jun. 2016.

[6] J. Yang, C. Ji, B. Jiang, W. Lu, and Q. Meng, "No reference quality assessment of stereo video based on saliency and sparsity," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 341–353, Jun. 2018.

[7] Z. Chen, W. Zhou, and W. Li, "Blind stereoscopic video quality assessment: From depth perception to overall experience," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 721–734, Feb. 2018.

[8] F. Qi, D. Zhao, X. Fan, and T. Jiang, "Stereoscopic video quality assessment based on visual attention and just-noticeable difference models," *Signal, Image Video Process.*, vol. 10, no. 4, pp. 737–744, Apr. 2016.

[9] G. Jiang, S. Liu, M. Yu, F. Shao, Z. Peng, and F. Chen, "No reference stereo video quality assessment based on motion feature in tensor decomposition domain," *J. Vis. Commun. Image Represent.*, vol. 50, pp. 247–262, Jan. 2018.

[10] Y. Fang, X. Sui, and J. Wang, "A spatial-temporal weighted method for asymmetrically distorted stereo video quality assessment," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2019, pp. 1–5.

[11] J. Yang, H. Wang, W. Lu, B. Li, A. Badii, and Q. Meng, "A no-reference optical flow-based quality evaluator for stereoscopic videos in curvelet domain," *Inf. Sci.*, vol. 414, pp. 133–146, Nov. 2017.

[12] C. Galkandage, J. Calic, S. Dogan, and J.-Y. Guillemaut, "Stereoscopic video quality assessment using binocular energy," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 102–112, Feb. 2017.

[13] B. Appina, A. Jalli, S. S. Battula, and S. S. Channappayya, "No-reference stereoscopic video quality assessment algorithm using joint motion and depth statistics," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2800–2804.

[14] Y. Wang, Y. Shuai, Y. Zhu, J. Zhang, and P. An, "Jointly learning perceptually heterogeneous features for blind 3D video quality assessment," *Neurocomputing*, vol. 332, pp. 298–304, Mar. 2019.

[15] M.-J. Chen, D.-K. Kwon, and A. C. Bovik, "Study of subject agreement on stereoscopic video quality," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, Apr. 2012, pp. 173–176.

[16] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Signal Process., Image Commun.*, vol. 40, no. 1, pp. 1–15, Jan. 2016.
[17] *Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference*, document Rec. ITU-T J.144, International Telecommunications Union, 2004.
[18] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.
[19] W. Lu, R. He, J. Yang, C. Jia, and X. Gao, "A spatiotemporal model of video quality assessment via 3D gradient differencing," *Inf. Sci.*, vol. 478, pp. 141–151, Apr. 2019.
[20] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 541–545, Apr. 2016.
[21] J. Yang, P. An, L. Shen, and Y. Wang, "No-reference stereo image quality assessment by learning dictionaries and color visual characteristics," *IEEE Access*, vol. 7, pp. 173657–173669, 2019.
[22] W. Ji, J. Wu, M. Zhang, Z. Liu, G. Shi, and X. Xie, "Blind image quality assessment with joint entropy degradation," *IEEE Access*, vol. 7, pp. 30925–30936, 2019.
[23] D. Bhattacharjee and H. Roy, "Pattern of local gravitational force (PLGF): A novel local image descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Aug. 1, 2019, doi: 10.1109/TPAMI.2019.2930192.
[24] F. Shao, W. Lin, G. Jiang, and Q. Dai, "Models of monocular and binocular visual perception in quality assessment of stereoscopic images," *IEEE Trans. Comput. Imag.*, vol. 2, no. 2, pp. 123–135, Jun. 2016.
[25] L. Shen, R. Fang, Y. Yao, X. Geng, and D. Wu, "No-reference stereoscopic image quality assessment based on image distortion and stereo perceptual information," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 1, pp. 59–72, Feb. 2019.
[26] Y. Liu, J. Yang, Q. Meng, Z. Lv, Z. Song, and Z. Gao, "Stereoscopic image quality assessment method based on binocular combination saliency model," *Signal Process.*, vol. 125, pp. 237–248, Aug. 2016.
[27] G. Yue, C. Hou, Q. Jiang, and Y. Yang, "Blind stereoscopic 3D image quality assessment via analysis of naturalness, structure, and binocular asymmetry," *Signal Process.*, vol. 150, pp. 204–214, Sep. 2018.
[28] L. Liu, B. Liu, C.-C. Su, H. Huang, and A. C. Bovik, "Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment," *Signal Process., Image Commun.*, vol. 58, pp. 287–299, Oct. 2017.
[29] P. Chen, L. Li, L. Ma, J. Wu, and G. Shi, "RIRNet: Recurrent-in-recurrent network for video quality assessment," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 834–842.
[30] J. Wu, Y. Liu, W. Dong, G. Shi, and W. Lin, "Quality assessment for video with degradation along salient trajectories," *IEEE Trans. Multimedia*, vol. 21, no. 11, pp. 2738–2749, Nov. 2019.
[31] Y. Huang, L. Li, Y. Zhou, and B. Hu, "No-reference quality assessment for live broadcasting videos in temporal and spatial domains," *IET Image Process.*, vol. 14, no. 4, pp. 774–781, Mar. 2020.
[32] Y. Zhang, X. Gao, L. He, W. Lu, and R. He, "Blind video quality assessment with weakly supervised learning and resampling strategy," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2244–2255, Aug. 2019.
[33] M. A. Saad and A. C. Bovik, "Blind quality assessment of videos using a model of natural scene statistics and motion coherency," in *Proc. 46th Conf. Rec. Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2012, pp. 332–336.
[34] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, Apr. 2013.
[35] L. Liu, T. Wang, H. Huang, and A. C. Bovik, "Video quality assessment using space–time slice mappings," *Signal Process., Image Commun.*, vol. 82, Mar. 2020, Art. no. 115749.
[36] H. Roy and D. Bhattacharjee, "Local-gravity-face (LG-face) for illumination-invariant and heterogeneous face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 7, pp. 1412–1424, Jul. 2016.
[37] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, Jan. 1962.
[38] M. Urvoy, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricorde, P. Le Callet, J. Gutierrez, and N. Garcia, "NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences," in *Proc. 4th Int. Workshop Qual. Multimedia Exper.*, Jul. 2012, pp. 109–114.
[39] S. Yasakethu, C. Hewage, W. Fernando, and A. Kondoz, "Quality analysis for 3D video using 2D video quality models," *IEEE Trans. Consum. Electron.*, vol. 54, no. 4, pp. 1969–1976, Nov. 2008.
[40] 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, Oct. 2013.
[41] L. Wang, Y. Kaneoke, and R. Kakigi, "Spatiotemporal separability in the human cortical response to visual motion speed: A magnetoencephalography study," *Neurosci. Res.*, vol. 47, no. 1, pp. 109–116, Sep. 2003.
[42] W. Zhou, L. Yu, W. Qiu, T. Luo, Z. Wang, and M.-W. Wu, "Utilizing binocular vision to facilitate completely blind 3D image quality measurement," *Signal Process.*, vol. 129, pp. 130–136, Dec. 2016.
[43] K. Gu, J. Qiao, S. Lee, H. Liu, W. Lin, and P. Le Callet, "Multiscale natural scene statistical analysis for no-reference quality evaluation of DIBR-synthesized views," *IEEE Trans. Broadcast.*, vol. 66, no. 1, pp. 127–139, Mar. 2020.
[44] 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.
[45] N.-E. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 2281–2284.
[46] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
[47] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, Apr. 2011, Art. no. 27. [Online]. Available: http://www.csie.ntu.edu.tw/?cjlin/libsvm
[48] J. Wang, S. Wang, and Z. Wang, "Asymmetrically compressed stereoscopic 3D videos: Quality assessment and rate-distortion performance evaluation," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1330–1343, Mar. 2017.
[49] 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.
[50] C. G. Bampis, P. Gupta, R. Soundararajan, and A. C. Bovik, "SpEED-QA: Spatial efficient entropic differencing for image and video quality," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1333–1337, Sep. 2017.
[51] P. V. Vu and D. M. Chandler, "ViS3: An algorithm for video quality assessment via analysis of spatial and spatiotemporal slices," *J. Electron. Imag.*, vol. 23, no. 1, Jan. 2014, Art. no. 013016.
[52] B. Appina, S. V. R. Dendi, K. Manasa, S. S. Channappayya, and A. C. Bovik, "Study of subjective quality and objective blind quality prediction of stereoscopic videos," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5027–5040, Oct. 2019.
[53] *Statistical Analysis, Evaluation and Reporting Guidelines of Quality Measurements*, document ITU-T P.1401, 2012.
[54] *Hybrid Perceptual Bitstream Models for Objective Video Quality Measurements*, document ITU-T J.343, 2014.

YUJIAN HOU received the B.S. degree from the School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China, in 2018, where he is currently pursuing the M.S. degree with the School of Computer Science and Technology. His research interests include 2D/3D video quality assessment and machine learning.

LIXIONG LIU (Member, IEEE) was a Visiting Researcher with the Laboratory for Image and Video Engineering, The University of Texas at Austin, USA, from December 2009 to January 2011. He is currently an Associate Professor with the School of Computer Science, Beijing Institute of Technology, Beijing, China. His research interests include image analysis and image quality assessment.

YONGMEI ZHANG received the B.S. degree from Peking University in 1990 and the M.Sc. and Ph.D. degrees from the North University of China. She was a Postdoctoral Researcher at Beihang University in 2008. In 2012, she was a Visiting Scholar at Peking University. She is currently a Professor with the North China University of Technology. Her research interests include artificial intelligence, image processing, and so on. She is a Senior Member of CCF.

QINGBING SANG received the B.S. degree in computer science from the China University of Geosciences, Wuhan, China, in 1996, and the M.S. and Ph.D. degrees in pattern recognition from the Jiangnan University, Wuxi, China, in 2005 and 2013, respectively. He was a Visiting Scholar with the LIVE Laboratory, The University of Texas at Austin, from August 2010 to August 2011. He is currently an Associate Professor with the School of Artificial Intelligence and Computer, Jiangnan University. His research interests include image processing, computer vision, and machine learning. He has published more than 30 technical articles in these areas.