

# New Stereo Video Quality Assessment Metric for Three Dimensional Video Systems

Xianwei Tang

Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China

Email: tangxw\_love@126.com

Mei Yu \*

Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China

National Key Lab of Software New Technology, Nanjing University, Nanjing 210023, China

Email: yumei2@126.com

Xin Jin, Zongju Peng, Feng Shao, Gangyi Jiang

Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China

Email: Jin\_Xin1006@163.com, {pengzongju, shaofeng, jianggangyi}@nbu.edu.com

**Abstract**—Since traditional two dimensional video quality metrics are not suitable for evaluating stereo video, a new stereo video quality metric based on support vector regression (SVR) for three dimensional video systems is proposed in this paper. Firstly, three factors affecting the quality of stereo video are analyzed, namely left-view video quality, right-view video quality and difference between two views. Then, each of these factors is calculated respectively with measuring the singular vectors' similarity between original image blocks and distorted image blocks and with temporal weighting using fluctuations of video frames' spatial quality. Then, SVR is used to learn the relationships between the factors and the overall perceived quality of stereo video. Finally, the relationship is applied to predict the objective quality of stereo video. By applying the proposed metric to stereo video database, Person's linear correlation coefficient is higher than 0.82 for all types of distortions and 0.81 for individual distortion, which outperforms the other existing metrics. Experimental results also show that, the proposed metric can achieve stable performance for all kinds of distortions and predict human perception well.

**Index Terms**—three dimensional video system, stereo video quality assessment, difference between left and right views, singular value decomposition

## I. INTRODUCTION

Three dimensional video (3DV) systems can create the perception of depth by binocular disparity, providing viewers with a more realistic feeling of the scene, and it is expected as development direction of the next generation of digital television [1]-[2]. Since video processing such as acquisition, coding and transmission will undoubtedly introduce distortions to videos, video quality assessment (VQA) is necessary for video processing and transmission systems [3].

VQA metrics can be divided into two categories, namely subjective assessment and objective assessment.

Subjective assessment is known to be the most reliable metric because it concerns how video is perceived by a viewer. However, it is time consuming and expensive, and it cannot be embedded into video system for real-time evaluation. Thus, objective assessment is required. Recently, development in two dimensional (2D) video quality evaluation were achieved [4]-[6], and some image quality assessment (IQA) metrics, based on structural similarity such as Structural SIMilarity (SSIM) [7] and Multi-scale Structural SIMilarity (MSSIM) [8], were proposed and became representative 2D IQA evaluation methods. However, evaluation of three dimensional video quality is a new research field. Although there are already some works on stereo video quality assessment (SVQA), most of them are subjective assessments [9]-[10], moreover, the existing objective SVQA metrics are mainly simply extended from 2D VQA metrics. Yasakethu et al. pointed out that simply using 2D metrics to evaluate the quality of each frame of stereo video and with the average value of all frames as quality of stereo video cannot predict the stereo video quality well [11], furthermore, they concluded that the traditional VQA metrics are not suitable for stereo video through experiments.

3DV display systems present human's left and right eyes with two slightly different videos in such a way that human visual system (HVS) gets perception of depth [12]. It means that left view, right view and the difference between the two views are three main factors affecting perceptual quality of stereo videos. In this paper, support vector regression (SVR) is used to combine the three factors so as to assess stereo video quality, because of its advantage of tackling nonlinear problems via the use of kernels. A new SVQA metric is proposed for 3DV systems. In the proposed metric, the factors affecting the quality of stereo video are first calculated; then, SVR is used to train the relationships between the factors and the subjective quality assessment scores of stereo video.

Finally, the SVQA scores are predicted by the relationships and factors.

The rest of the paper is organized as follows. Section 2 describes the details of the proposed metric. Experimental results with stereo videos from a public database are given in Section 3, and finally Section 4 concludes this paper.

## II. PROPOSED STEREOS VIDEO QUALITY ASSESSMENT METRIC

Fig. 1 shows the block diagram of the proposed SVQA metric. Three main factors affecting the quality of stereo videos are first analyzed quantitatively, including, left-view quality score, right-view quality score and similarity of the difference between the two views. Here, SVR is used to obtain the relationship between the three factors and the subjective evaluation scores of stereo video, so that objective quality score of stereo video is finally predicted with the three factors and the relationship.

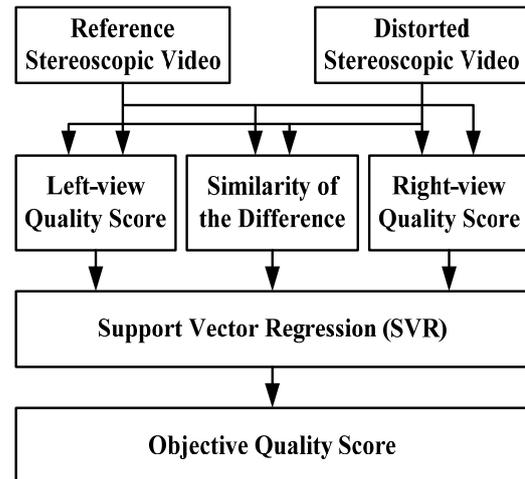


Figure 1. Block diagram of the proposed SQVA metric

### A. Single View Quality and Similarity of Difference between Two Views

The process of calculating quality scores of left view, right view and similarity of the difference between the two views are treated as a two-stage process: 1) measuring the spatial quality of the single view (left-view and right-view) and the similarity of the difference between the two views at each moment, and 2) weighting all the spatial quality scores of single view into an overall single view quality score and similarity of the difference between two views at each moment into an overall similarity.

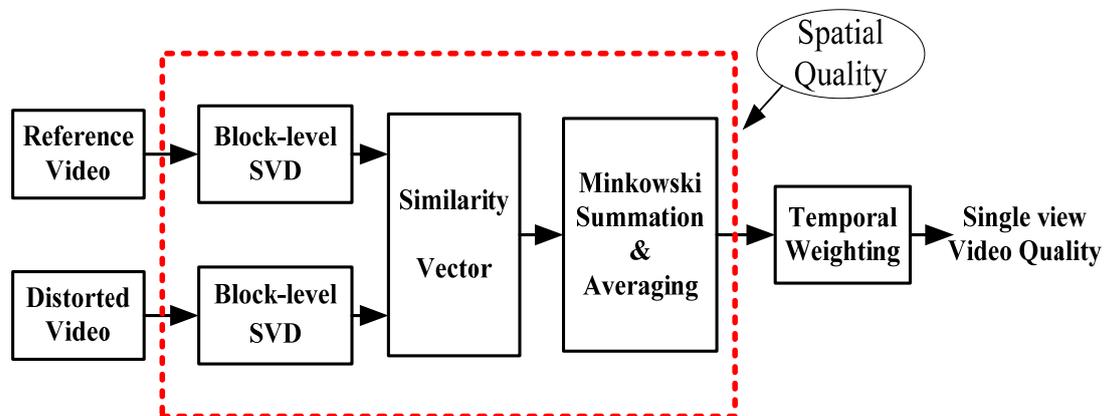


Figure 2. Framework for assessing single view video quality

The framework for assessing single view video quality is shown in Fig. 2. The first stage, namely spatial quality evaluation, uses a modified singular value decomposition (SVD) based image quality metric for calculating the spatial quality score of each frame. Then, the spatial quality score of each frame is weighted into an overall single video view quality score according to the fluctuations of each frames' spatial quality.

#### 1) Spatial Quality Evaluation

SVD has been used for image quality assessment for a long time, and the difference of singular values between the original and distorted images are typically used to measure the quality of the distorted image. It is obvious that using singular values for quality evaluation

considered only luminance, and the structural information, vital for HVS, has not been considered [14]. By contrast, singular vectors denote clear physical meaning for representing structural information. Therefore, some researchers developed image quality metrics based on the distortion of singular vectors of the whole image. However, the computational complexity of calculating singular values will present cubic growth as the dimension of matrix increasing. Thus, the computational complexity of SVD for images and videos with high-resolution cannot be ignored. To reduce the computational complexity, we calculate the spatial quality of each frame with performing SVD on image blocks instead of the whole images. The detailed process of

spatial quality evaluation is described as follows.

Let  $k$  denote the frame index such that  $k = 1$  to  $N_f$ , and assume that there are  $N_f$  frames in a single view video sequence of stereo video. For an original video frame  $\mathbf{A}$  and the corresponding distorted frame  $\mathbf{A}^{(d)}$  with the size of  $M \times N$ ,  $\mathbf{A}$  and  $\mathbf{A}^{(d)}$  are firstly divided into non-overlapping blocks with the size of  $r \times c$  in the same way, thus, there are a total of  $N_{block}$  blocks, and  $N_{block} = (M / r) \times (N / c)$ . Here, both of  $r$  and  $c$  are set to 120. Then, similarity of singular vectors of each block is measured. Finally, the results of all blocks are averaged to get a spatial quality evaluation of each frame.

Let  $\mathbf{I}_i$  and  $\mathbf{I}_i^{(d)}$  be the  $i^{th}$  ( $i = 1, 2, \dots, N_{block}$ ) block of  $\mathbf{A}$  and  $\mathbf{A}^{(d)}$ , respectively. Then, the procedure of calculating the similarity of singular vectors of the  $i^{th}$  block is described as follows:

Firstly, singular value decomposition of luminance components of  $\mathbf{I}_i$  and  $\mathbf{I}_i^{(d)}$  with the size of  $r \times c$  can be expressed as

$$\mathbf{I}_i = \mathbf{U}\boldsymbol{\sigma}\mathbf{V}^T = [\mathbf{u}_1\mathbf{u}_2\dots\mathbf{u}_r] \times \text{diag}(\sigma_1\sigma_2\dots\sigma_t) \times [\mathbf{v}_1\mathbf{v}_2\dots\mathbf{v}_c] \quad (1)$$

$$\begin{aligned} \mathbf{I}_i^{(d)} &= \mathbf{U}^{(d)}\boldsymbol{\sigma}^{(d)}\mathbf{V}^{(d)T} \\ &= [\mathbf{u}_1^{(d)}\mathbf{u}_2^{(d)}\dots\mathbf{u}_r^{(d)}] \times \text{diag}(\sigma_1^{(d)}\sigma_2^{(d)}\dots\sigma_t^{(d)}) \times [\mathbf{v}_1^{(d)}\mathbf{v}_2^{(d)}\dots\mathbf{v}_c^{(d)}] \end{aligned} \quad (2)$$

where  $\mathbf{U}$ ,  $\mathbf{V}$  and  $\boldsymbol{\sigma}$  represent the left singular vector matrix, the right singular vector matrix, and the diagonal matrix of singular values of the original block  $\mathbf{I}_i$ ;  $\mathbf{u}_m$  and  $\mathbf{v}_n$  are column vectors while  $\sigma_q$  is a singular value ( $m = 1, 2, \dots, r$ ;  $n = 1, 2, \dots, c$ ;  $q = 1, 2, \dots, t$  and  $t = \min(r, c)$ ), the singular values appear in descending order, i.e.,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_t$ . In the case of singular vectors the magnitudes of each vector is unity. Similarly, the variables in Eq. (2) denote the components of  $\mathbf{I}_i^{(d)}$  in the same way as those of  $\mathbf{I}_i$  in Eq. (1).

Then, let  $\alpha_j$  be the dot product between the original and the distorted  $j^{th}$  left singular vectors and  $\beta_j$  be the dot product for the corresponding right singular vectors of the  $i^{th}$  block. Thus,

$$\alpha_j = \mathbf{u}_j \cdot \mathbf{u}_j^{(d)} \quad (3)$$

$$\beta_j = \mathbf{v}_j \cdot \mathbf{v}_j^{(d)} \quad (4)$$

As mentioned above, the magnitude of each singular vector is unity, i.e.,  $|\mathbf{u}_j| = |\mathbf{u}_j^{(d)}| = |\mathbf{v}_j| = |\mathbf{v}_j^{(d)}| = 1$ . Let  $\theta_u$  be the angle between  $\mathbf{u}_j$  and  $\mathbf{u}_j^{(d)}$ , and  $\theta_v$  be the angle between  $\mathbf{v}_j$  and  $\mathbf{v}_j^{(d)}$ . Eqs. (3) and (4) can be further expressed as

$$\alpha_j = \mathbf{u}_j \cdot \mathbf{u}_j^{(d)} = |\mathbf{u}_j| |\mathbf{u}_j^{(d)}| \cos(\theta_u) = \cos(\theta_u) \quad (5)$$

$$\beta_j = \mathbf{v}_j \cdot \mathbf{v}_j^{(d)} = |\mathbf{v}_j| |\mathbf{v}_j^{(d)}| \cos(\theta_v) = \cos(\theta_v) \quad (6)$$

From Eqs. (5) and (6), it can be seen that the dot product of singular vectors between the original block

and the corresponding distorted block equals to the cosine of the angle between singular vectors. Therefore, spatial quality of the distorted image block can be measured as dot products of singular vectors between the distorted block and the corresponding original block. Next, we define a similarity vector  $\boldsymbol{\Gamma}_i = \{\gamma_j; j = 1, 2, \dots, t\}$  for the distorted block to measure the distortion of the distorted block compared to the original block, where  $\gamma_j = |\alpha_j + \beta_j|$ . The Minkowski metric and a logarithmic scale are used to calculate the similarity  $S_i$  of the all singular vectors for the  $i^{th}$  block.

$$S_i = \ln(1 + (\sum_{j=1}^t \gamma_j^2)^{\frac{1}{2}}) \quad (7)$$

Finally, similarities of all the blocks are averaged into spatial quality score for the  $k^{th}$  video frame

$$Q_k = \frac{1}{N_{block}} \sum_{i=1}^{N_{block}} S_i \quad (8)$$

## 2) Similarity of Differences between Left and Right Views

Human brain can generate binocular perception from the difference between views, and the difference between views is also an important factor influencing the quality of stereo video. Absolute disparity image can be used to reflect the difference between views [15], where the absolute disparity image can be expressed as  $\mathbf{D} = |\mathbf{R} - \mathbf{L}|$ , where  $\mathbf{L}$  and  $\mathbf{R}$  are the left and right views of stereo image respectively. The more similar the absolute disparity images of the reference and the distorted stereo images are, the better the stereo scene presents. Therefore, the absolute disparity image  $\mathbf{D}$  at the  $k^{th}$  moment are considered as the difference between two views, and the 2D image quality assessment metric SSIM is directly used to measure the similarity between the absolute disparity image  $\mathbf{D}$  of the reference and distorted stereo video frames. Then, the similarity  $Q_{Dk}$  of the difference between two views at the  $k^{th}$  moment is obtained.

## 3) Temporal Weighting

To demonstrate that fluctuations of spatial quality of video frames over time can influence the overall quality of stereo video, as an example, we show the spatial quality of each frame of single view video of stereo videos with different mean opinion scores (MOS) in Fig. 3. These two videos with MOS 4.3571 and 3.75 respectively have been taken from the public stereo video database of IRCCyN [16], which has also been used as the test database for evaluating the performance of the proposed metric. Fig. 3 (a) and Fig. 3 (b) are the cases of the left view and right view of the two stereo videos, respectively. It is noted that a higher MOS

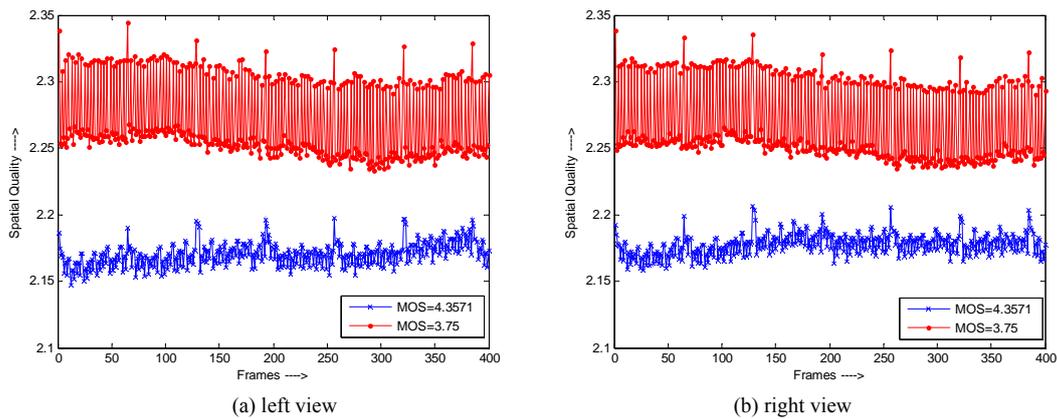


Figure 3. Example of fluctuations of video frame's quality

implies a higher video quality, thus, the stereo video with MOS = 4.3571 has a higher quality than the one with MOS = 3.75. However, it can also be seen that the lower quality video (MOS = 3.75) has all frames with higher spatial quality as compared to the video with MOS = 4.3571 in Fig. 3. That is to say, simple averaging cannot determine the single view video quality accurately. It is noted that the lower quality video (MOS = 3.75) has a bigger spatial quality variation over time compared to the higher quality video (MOS = 4.3571), which means that fluctuations of spatial quality of video frames over time should be considered in temporal weighting.

In the process of temporal weighting, a weighting factor  $w_k$  for the  $k^{\text{th}}$  single view video frame is defined as

$$w_k = \frac{1}{2F} \sum_{T=1}^F (|Q_k - Q_{k-T}| + |Q_k - Q_{k+T}|) \quad (9)$$

where  $Q_k$  denotes the spatial quality of the  $k^{\text{th}}$  frame of single view video, and  $2F$  denotes the local scope when calculating the fluctuation of the spatial quality for the  $k^{\text{th}}$  video frame respect to the nearby frames. Here,  $F$  is set as 2, i.e., two frames on either side of the current  $k^{\text{th}}$  frame to compute  $w_k$ , which denotes the variation of spatial quality of the  $k^{\text{th}}$  frame relative to the neighboring  $2F$  frames. Finally, the single view video quality score  $Q$  is calculated through temporal weighting as

$$Q = \frac{Q_1 + Q_2 + Q_{N_f-1} + Q_{N_f-2}}{4 + \sum_{k=3}^{N_f-2} w_k} + \frac{\sum_{k=3}^{N_f-2} w_k \times Q_k}{4 + \sum_{k=3}^{N_f-2} w_k} \quad (10)$$

When Eq.(10) is used to calculate the left view video quality  $Q_L$  and right view video quality  $Q_R$ , the  $Q$  is referred as  $Q_L$  and  $Q_R$  respectively. Similarly, the similarity,  $Q_D$ , of the difference between the two views can be calculated by using Eq.(10) and temporal weighting to  $Q_{Dk}$ .

Thus, for a given stereo video, we can calculate the three key factors affecting the quality of stereo video, namely, left-view quality score  $Q_L$ , right-view quality score  $Q_R$  and the similarity of the difference between two views  $Q_D$ .

### B. Overall Stereo Video Quality

We have three factors namely  $Q_L$ ,  $Q_R$  and  $Q_D$  contributing to the overall quality of stereo video. Since the HVS is extremely complicated, simply weighting of the three factors cannot reflect the relationship among them reasonably. Here, we use the SVR algorithm to combine the three factors together to get the overall stereo video quality.

Let  $\{x_1, x_2, \dots, x_l\}$  and  $\{y_1, y_2, \dots, y_l\}$  denote the training set. Here, each  $x_i = (Q_{Li}, Q_{Ri}, Q_{Di})$  represents the three dimensional vector consisting of the left-view quality score, right-view quality score and the similarity of difference between two views, and each  $y_i$  is the associated subjective score (i.e., target value) for the  $i^{\text{th}}$  stereo video. In SVR, the input sample  $x$  is first mapped into a higher dimensional space via a kernel function  $\phi(x)$ , then, a linear model is established in the high dimensional feature space to learn the regression function

$$f(x) = W^T \phi(x) + b \quad (11)$$

where  $W = (w_1, w_2, w_3)$  represent the weighting vector and  $b$  is the bias (constant). Given the training data  $(x_1, y_1), \dots, (x_l, y_l)$ , we need to find the  $W$  and  $b$  from the training data. When  $\epsilon$ -insensitive loss function is adopted as the loss function, the corresponding SVR also named as  $\epsilon$ -SVR, then,  $f(x)$  in Eq.(11) should meet the following inequality

$$|y_i - f(x_i)| \leq \epsilon \quad (12)$$

the relative optimization problem can be expressed by

$$\left. \begin{aligned} \min_{W, b, \epsilon, \epsilon^*} \quad & \frac{1}{2} \|W\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{subject to:} \quad & y_i - (W^T \phi(x_i) + b) \leq \epsilon + \xi_i, \quad i = 1, 2, \dots, l \\ & (W^T \phi(x_i) + b) - y_i \leq \epsilon + \xi_i^*, \quad i = 1, 2, \dots, l \\ & \xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \dots, l \end{aligned} \right\} \quad (13)$$

the optimization problem can be converted to its dual problem by exploiting Lagrangian function to be solved. The regression function to be learned then becomes

$$f(x) = W^T \phi(x) + b = \sum_{i=1}^{n_{sv}} (\eta_i^* - \eta_i) K(x_i, x) + b \quad (14)$$

where  $K(\mathbf{x}_i, \mathbf{x})$  is the kernel function,  $\eta_i^*$  and  $\eta_i (i = 1, 2, \dots, l, 0 \leq \eta_i^*, \eta_i \leq C)$  are the Lagrange multipliers used in the Lagrange function optimization, the samples that come with non-vanishing coefficients ( i.e., nonzero  $\eta_i^*$  and  $\eta_i$  ) are support vectors, and  $n_{sv}$  is the number of support vectors. In SVR, the weighting vector  $\mathbf{W}$  is defined only by the support vectors. Here, a well known library for support vector machines is chosen for SVR, namely LIBSVM [18], a polynomial kernel of order 2 is used for the nonlinear mapping between the three-dimensional input and the subjective quality score, and the SVR parameters  $\epsilon$  and  $C$  are set as 0.065 and 80, respectively.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Database Description

A publicly available stereo video database [16] (referred as NAMA3DS1-COSPAD1) with 10 original Full-HD (with resolution 1920×1080) stereo videos and 100 corresponding distorted stereo videos is used to evaluate the performance of the proposed SVQA metric. Left views of 10 original stereo videos sequences are shown in Fig. 4. The 100 distorted stereo videos are generated under 10 hypothetical reference conditions (HRCs), the detailed HRCs are shown in Table 1, including five distortions, video coding (H.264), still image coding (JPEG2K), downsampling (DS), edge enhancement (EE), downsampling and edge enhancement (DS+EE). Finally, 100 stereo videos were rated by 29

subjects. Subjective scores had been made available as mean opinion scores (MOSs).

TABLE I  
LIST OF HRCs OF NAMA3DS1-COSPAD DATASET

HRC	Impairments & Degradations	
	Type	Parameters
1	Video coding(H.264)	QP32
2	Video coding(H.264)	QP38
3	Video coding(H.264)	QP44
4	Still image coding(JPEG2k)	2Mb/s
5	Still image coding(JPEG2k)	8Mb/s
6	Still image coding(JPEG2k)	16Mb/s
7	Still image coding(JPEG2k)	32Mb/s
8	Reduction of resolution	↓4downsampling
9	Image sharpening	Edge enhancement
10	Downsampling & sharpening	HRC8 + HRC9

To get more objective evaluation of the proposed metric, the samples are divided into two sets with the same number of elements, one set is used for testing, and the other for training. More specifically, the data is first split into ten groups such that stereo video contents in one group do not appear in any of the remain groups, one stereo video content is defined as all the distorted versions of an original stereo video. Then, five of the ten groups are selected to constitute the training set, and the remaining five groups are used for testing, thus, there are a total of  $C_{10}^5=252$  cases for selecting. The average accuracy of the tests over the 252 cases is taken as the performance of the evaluation.



Figure 4. Left views of sequences in the stereo video database

Experimental results are reported in terms of three criteria, namely, Pearson linear correlation coefficient(CC) for prediction accuracy, Spearman rank order correlation coefficient (SROCC) for monotonicity, and root mean squared error (RMSE) for prediction accuracy. The three criteria are commonly used for performance comparison between subjective score and objective prediction in VQA and SVQA. A better SVQA metric will have higher CC and SROCC values

and lower RMSE value. For a perfect SVQA metric,  $CC = SROCC = 1$ , and  $RMSE = 0$ . In order to eliminate the nonlinear factors introduced in the process of subjective quality assessment, a nonlinear regression [19] should be performed on the objective assessment values  $Q$  of the model before the calculation of CC and RMSE, here, a four parameters polynomial function is used as follows

$$y = \alpha_1 x^3 + \alpha_2 x^2 + \alpha_3 x + \alpha_4 \tag{15}$$

TABLE II  
PERFORMANCE COMPARISON (CASES IN BOLD DENOTE THE BEST PERFORMANCE)

Measure	Metric	ALL	H.264	JPEG2K	DS	EE	DS+EE
CC	MSE	0.5547	0.6457	0.8117	0.5101	0.6609	0.6470
	SSIM	0.7634	0.7616	<b>0.9355</b>	0.4306	0.6279	0.7474
	MSSIM	0.7730	0.6817	0.9180	0.4403	0.6206	0.5953
	Proposed	<b>0.8220</b>	<b>0.8560</b>	0.9340	<b>0.8638</b>	<b>0.8172</b>	<b>0.8463</b>
SROCC	MSE	0.5679	0.6126	0.7995	0.5714	0.6930	0.5854
	SSIM	0.7639	0.7804	<b>0.9093</b>	0.4863	0.6140	0.7378
	MSSIM	0.7859	0.6469	0.9042	0.5106	0.5654	0.5854
	Proposed	<b>0.8237</b>	<b>0.8405</b>	0.8781	<b>0.8294</b>	<b>0.7851</b>	<b>0.7721</b>
RMSE	MSE	0.9451	0.8838	0.7277	0.2292	0.1424	0.2145
	SSIM	0.7337	0.7501	0.4402	0.2405	0.1477	0.1869
	MSSIM	0.7207	0.8469	0.4941	0.2393	0.1488	0.2260
	Proposed	<b>0.6335</b>	<b>0.5814</b>	<b>0.4009</b>	<b>0.0804</b>	<b>0.1090</b>	<b>0.1108</b>

### B. Performance Comparison and Discussion

There are only a few objective SVQA metrics in the literatures, and their performances are given on non-public database with almost only a single type of distortion, which makes it difficult to reproduce their results objectively. In order to verify the effectiveness of the proposed method, we compare the performance of the proposed metric with the classical 2D metrics with IRCCyN NAMA3DS1-COSPAD1 as the test database. These classical 2D metrics include the widely used MSE, SSIM, and MSSIM, and the objective evaluation scores by averaging the obtained results of each view. The performance of each metric under different distortions with the database is listed in Table 2.

From the table, it is clear that the proposed metric outperforms the others' under all the individual distortion type except JPEG2K, the performance of the proposed metric is slightly worse than the SSIM under JPEG2K. This may be explained by that JPEG2K distortion mainly leads to blurring artifact especially in edge regions thus it generates structural distortion and SSIM can deal with structural distortion very well. However, the proposed metric achieves the best performance under the ALL distortion (including all the 100 distorted stereo videos) with values of the CC, SROCC and RMSE are 0.8220, 0.8237 and 0.6335 respectively. Moreover, the proposed metric can achieve stable performance for all kinds of distortion types and can predict stereo videos quality very well.

### IV. CONCLUSIONS

In this paper, a new objective stereo video quality (SVQA) metric based on Support Vector Regression (SVR) is proposed. Three main factors affecting the quality of stereo video are first analyzed, namely, left-view video quality, right-view video quality and difference between the two views. Then, each factor is calculated with consideration of the fluctuation of spatial quality of each frame. Finally, SVR is used to learn the relationship between the factors and the subjective quality

scores of stereo videos, and to predict the objective quality of the stereo videos. Experimental results show that, the proposed metric can predict the human eye's subjective perception of stereo video well, and can achieve stable evaluation.

In the future work, we will consider the difference between views more precisely, and more complicated perceptual temporal weighting will be considered to adapt to more realistic three dimensional video systems, such as poor quality video frames appeared periodically, a small number of video frames with very poor quality exist, thereby further enhancing the proposed metric.

### ACKNOWLEDGMENT

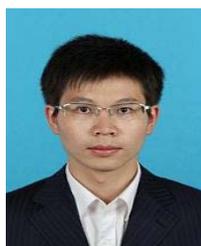
The research work was supported by National Natural Science Foundation of China (Grant Nos. 61271270, U1301257, 61311140262, 61171163, 61271021), and Zhejiang Scientific and Technical Key Innovation Team in China (Grant No. 2010R50009).

### REFERENCES

- [1] A. Gotchev, G. B. Akar, T. Capin, D. Strohmeier, and A. Boev, "Three-dimensional media for mobile devices," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 708-741, April 2011.
- [2] F. Shao, G. Jiang, W. Lin, M. Yu, Q. Dai, Joint Bit Allocation and Rate Control for Coding Multi-View Video Plus Depth Based 3D Video, *IEEE Transactions on Multimedia*, 15(8):1843-1854, 2013
- [3] S. Chikkerur, V. Sundaram, Martin Reisslein, et al., "Objective video quality assessment methods: a classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165-182, 2011.
- [4] L. Ma, C. Deng, N. King, et al., "Recent advances and challenges of visual signal quality assessment," *China Communications*, vol. 10, no. 5, pp. 62-78, 2013.
- [5] M. Vranješ, S. Rimac-Drlje, K. Grgi, "Review of objective video quality metrics and performance comparison using different databases," *Signal Processing: Image Communication*, vol. 28, no. 1, pp. 1-19, 2013.
- [6] L. Ma, C. Deng, K.N. Ngan, et al., "Overview of quality

assessment for visual signals and newly emerged trends,” *IEEE International Symposium on Circuits and Systems*, Beijing, China, 2013, pp. 1091-1094.

- [7] Z. Wang, A.C. Bovik, H. R. Sheikh, et al., “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.
- [8] Z. Wang, E.P. Simoncelli, and A.C. Bovik, “Multi-scale Structural Similarity for Image Quality Assessment,” *Thirty-seventh Asilomar Conference on Signals, Systems and Computers*, 2003, vol. 2, pp.1398-1402.
- [9] P. Aflaki, M.M. Hannuksela, M. Gabbouj, et al., “Subjective quality assessment of asymmetric stereoscopic 3D video,” *Signal, Image and Video Processing*, 2013, DOI 10.1007/s11760-013-0439-0.
- [10] H. Malekmohamadi, W.A.C. Fernando, and A.M. Kondoz, “Content-based subjective quality prediction in stereoscopic videos with machine learning,” *Electronics Letters*, vol. 48, no. 21, pp. 1344-1346, 2012.
- [11] S.L.P. Yasakethu, C.T.E.R. Hewage, W.A.C .Fernando, et al., “Quality analysis for 3d video using 2d video quality models,” *IEEE Transactions on Consumer Electronics*, vol. 54, no. 4, pp. 1969-1976, 2008.
- [12] H. Urey, K.V. Chellappan, E. Erden, et al., “State of the art in stereoscopic and autostereoscopic displays,” *Proceedings of the IEEE*, vol. 99, no. 4, pp. 540-555, 2011.
- [13] Q. Meng, X. Ma, and Y. Zhou, “Prediction of Mine Gas Emission Rate using Support Vector Regression and Chaotic Particle Swarm Optimization Algorithm”, *Journal of Computers*, vol. 8, no. 11, pp. 2908-2915, 2013.
- [14] M. Narwaria, and W. Lin, “Scalable image quality assessment based on structural vectors,” *IEEE International workshop on Multimedia Signal processing*, Rio De Janeiro, 2009, pp.1-6.
- [15] J. Yang, C. Hou, Y. Zhou, et al., “Objective quality assessment method of stereo images,” *3DTV Conference*, Potsdam, Germany, 2009.
- [16] M. Urvoy, M. Barkowsky, R. Cousseau, et al., “NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences,” *International workshop on Quality of Multimedia Experience*, Yarra Valley, VIC, 2012, pp. 109-114.
- [17] X. Liu, S. Yuan, and L. Li, “Prediction of Temperature Time Series Based on Wavelet Transform and Support Vector Machine”, *Journal of Computers*, vol. 7, no. 8, pp. 1911-1918, 2012.
- [18] C. Chang, and C. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1-27, 2011.
- [19] K. Brunnström, D. Hands, and F. Speranza, “VQEG validation and ITU standardization of objective perceptual video quality metrics,” *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 96-101, 2009.



**Xianwei Tang** received his BSc degree in Electronic Information Science and Technology from Hunan University of Arts and Science, ChangDe, China, in 2010. He is currently working towards

his MSc degree in Faculty of Information Science and Engineering Ningbo University. His research interests include image processing, video coding, stereo video quality assessment, etc.



**Mei Yu** received the M.S. degree from Hangzhou Institute of Electronics Engineering, Hangzhou, China, in1993, and the Ph.D. form Ajou University, Korea, in 2000. Then, she joined in Ningbo University, Ningbo, China. Since 2005, she works as a Professor in Faculty of Information Science and Engineering, Ningbo University, China.

Her research interests include image/video coding, stereoscopic image/video quality assessment and visual perception.



**Xin Jin** received her BSc degree in communication engineering from China Ji Liang University, Hangzhou, China, in 2012. She is currently working towards her MSc degree in Faculty of Information Science and Engineering Ningbo University. Her research interests include image processing, image quality assessment and noise classification.



**Zongju Peng** received his B.S. degree in computer science from Sichuan University, Chengdu, China, in 1998, and Ph.D. degree from Institute of Computing Technology, Chinese Academy of Science in 2010. He is currently an associate professor in Faculty of Information Science and Engineering, Ningbo University, China.

His research concentrates on image/video compression.



**Feng Shao** received his B.S. and Ph.D degrees from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively, all in Electronic Science and Technology. He is currently an associate professor in Faculty of Information Science and Engineering, Ningbo University, China. His research interests include video coding, image processing and perception, etc.



**Gangyi Jiang** received his M.S. degree from Hangzhou University, Hangzhou, China, in 1992, and received the Ph.D. degree from Ajou University, Korea, in 2000. From 2000, he joined the Faculty of Information Science and Engineering, Ningbo University, China, as a Professor. His research interests include digital video compression and communications, 3D image/video

processing and visual perception.