Objective quality evaluation measure of symmetric and asymmetric distorted stereoscopic images

Shanbo Gu (顾珊波), Feng Shao (邵 枫)*, Gangyi Jiang (蒋刚毅), and Mei Yu (郁 梅)

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

**Corresponding author: shaofeng@nbu.edu.cn*

Received September 14, 2011; accepted October 9, 2011; posted online May 9, 2012

An objective quality evaluation measure of asymmetric and symmetric distorted stereoscopic images is proposed. In the measure, stereoscopic features are first extracted from left and right images by using singular value decomposition. Then, the relationship between the stereoscopic features and subjective scores is established by using support vector regression. Finally, the objective evaluation scores are tested on symmetric and asymmetric databases. Experimental results show that the proposed measure is more effective in quantifying image quality, compared with other two relevant quality evaluation measures.

OCIS codes: 110.0110, 100.2000, 100.6890.

doi: 10.3788/COL201210.S11101.

With the development of stereoscopic display and network technologies, stereoscopic imaging technology has attracted increasing attentions, and has a widespread prospect of applications[1]. Since stereoscopic content is acquired simultaneously with two cameras, the amount of data may be doubled compared to the corresponding two-dimensional (2D) image However, the occurrence of the compression induced artifacts, such as blurring, blocking, ringing and other distortions, will be inevitable. Therefore, how to quantify the quality of stereoscopic content is a key problem in stereoscopic video systems.

Extensive works had already been done to develop 2D image quality evaluation measures, such as Structural Similarity Index (SSIM)[2] and Singular Value Decomposition (SVD)[3]. However, still now very few efforts have been concentrated for stereoscopic images. Benoit *et al.* proposed a quality assessment measure for stereoscopic images by fusing 2D quality measures and depth information[4]. Akhter *et al.* proposed an no-reference perceptual quality assessment measure for JPEG coded stereoscopic images based on segmented local features of artifacts and disparity[5]. However, these measures imposed certain assumptions on the relationship between the objective evaluation scores and the subjective scores, and human visual system (HVS) probably does not meet these assumptions.

In addition, the performance of quality evaluation measure for stereoscopic images may be highly dependent on specific stereoscopic database. In stereoscopic image/video coding, it is possible to allocate the bitrates on the two views symmetrically or asymmetrically. Gorley *et al.* suggested that symmetric stereoscopic image compression can produce better results than asymmetric compression[6]. Saygili *et al.* suggested that symmetric coding achieved better perceived quality than asymmetric coding below the asymmetry threshold[7]. Therefore, it is necessary to objectively evaluate the quality of asymmetric and symmetric coding.

Considering the above two issues, an objective quality evaluation measure of asymmetric and symmetric distorted stereoscopic image is proposed in this letter. The proposed measure extracts stereoscopic features from left and right images by using SVD, and the relationship between stereoscopic features and subjective score is established by using support vector regression (SVR). Finally, the objective evaluation scores are tested on the symmetric and asymmetric database. The outstanding advantage of the proposed measure is that complex simulation of perceptual characteristics and mechanisms of HVS can be avoided and the objective evaluation scores have a good correspondence with the subjective scores.

SVD of an image $\mathbf{I} \in R^{M\times N}$ can be written as

$$\mathbf{I} = \mathbf{USV}^{\mathrm{T}}, \tag{1}$$

where $\mathbf{U}$ (with size of $M\times M$) and $\mathbf{V}$ (with size of $N\times N$) are orthogonal matrix, $\mathbf{S}$ is $M\times N$ diagonal matrix. The diagonal elements in $\mathbf{S}$ are called the singular values.

The singular values can well characterize the structural information in an image, and the singular values have strong stability[3]. In this work, we use the singular values as the feature basis for the task. The singular value vector is defined as $\mathbf{T} = (\sigma_1,\ \sigma_2,\ \cdots\ \sigma_i,\ \cdots\ \sigma_n)$, where n=min(M, N), and σ_i is the i-th singular values in $\mathbf{S}$. Thus, feature vector $\mathbf{X}_{\mathrm{l}}$ of left image and feature vector $\mathbf{X}_{\mathrm{r}}$ of right image are calculated as

$$\mathbf{X}_{\mathrm{l}} = \left|\mathbf{T}^{\mathrm{l}}_{\mathrm{org}} - \mathbf{T}^{\mathrm{l}}_{\mathrm{dis}}\right|, \tag{2}$$

$$\mathbf{X}_{\mathrm{r}} = \left|\mathbf{T}^{\mathrm{r}}_{\mathrm{org}} - \mathbf{T}^{\mathrm{r}}_{\mathrm{dis}}\right|, \tag{3}$$

where $\mathbf{T}^{\mathrm{l}}_{\mathrm{org}}$ and $\mathbf{T}^{\mathrm{l}}_{\mathrm{dis}}$ are the singular value vectors of the original and the distorted left images, respectively, and $\mathbf{T}^{\mathrm{r}}_{\mathrm{org}}$ and $\mathbf{T}^{\mathrm{r}}_{\mathrm{dis}}$ are the singular value vectors of the original and the distorted right images, respectively. Since the quality differences between the left and right images can ultimately affect the stereoscopic perception, the feature vector $\mathbf{X}$ of a stereoscopic image can be expressed as a linear combination of the feature vectors $\mathbf{X}_{\mathrm{l}}$ and $\mathbf{X}_{\mathrm{r}}$

$$\mathbf{X} = w_{\mathrm{l}}\mathbf{X}_{\mathrm{l}} + w_{\mathrm{r}}\mathbf{X}_{\mathrm{r}}, \tag{4}$$

where w_{l} and w_{r} are the weights for feature vectors of left and right images, respectively.

In this letter, we formulate image quality prediction as a regression problem and use SVR to find a mapping function between the features and subjective scores[8]. The feature fusion procedure is conducted in two phases:

1) Training phase: Suppose that $\mathbf{X}_p$ is the feature value vector of the p-th stereoscopic image pair in the training set, and y_p is the corresponding difference mean opinion score (DMOS) (p=1, 2, $\cdots$, p_m; p_m is the number of training image pairs). The purpose of ε-SVR is to map the original data space into a high-dimensional feature space, and form the best linear function in the high-dimensional feature space. The linear function is described as

$$f(\mathbf{X}) = \sum_{i=1}^{p_m} w_i \cdot k(\mathbf{X},\ \mathbf{X}_i) + b, \tag{5}$$

where $k(\mathbf{X},\ \mathbf{X}_i)$ is a kernel function, $\mathbf{w}=\{w_i,\ i=1,\ 2,\ \cdots, p_m\}$ is a weight vector, and b is a bias term. The aim of SVR is to find the unknown parameters $\mathbf{w}$ and b from the training data such that the error is less than a predefined value denoted by insensitivity parameter ε. In the experiment, we have used the exponential radial basis function (ERBF) as the kernel function with the form of $k(\mathbf{X},\ \mathbf{X}_i) = \exp(-\sqrt{||\mathbf{X}-\mathbf{X}_i||^2}/\gamma^2)$, where γ is a parameter controlling the radius.

2) Test phase: With the estimated parameters $\mathbf{w}$ and b from the training phase, the objective score y_q for the q-th stereoscopic image pair can be predicted. However, the current training and test data should be different. We employ the fivefold cross-validation strategy to test the performance of the proposed measure. The data of stereoscopic images in different types of distortion is split into five subsets, and one used for testing and the remaining four subsets are used for training. The experiment is repeated with each of the five subsets used for testing.

For the experiments, we have used two databases created by ourselves. The asymmetric database includes 10 original stereoscopic image pairs from which 370 distorted stereoscopic images were obtained with four types of distortion: JPEG, JPEG2000, Gaussian blur, and white noise[9]. Specifically, 70, 100, 100, and 100 distorted stereoscopic images with different levels of JPEG, JPEG2000, Gaussian blur and white noise distortions are included in the database, respectively. For asymmetric database, even though the binocular perception is dominated by the high quality view with binocular suppression, the weight value w_{r} is set to 0.5 because the left image is undistorted in the database.

The symmetric database includes 12 original stereoscopic image pairs from which 312 distorted stereoscopic images have been generated with five types of distortion. Besides the same four types of distortion with the asymmetric database, H.264 distortion is added in the database which is obtained with H.264/Advanced Video Coding (AVC) coding. Specifically, 60, 60, 60, 60, and 72 distorted stereoscopic images with different levels of JPEG, JPEG2000, Gaussian blur, white noise and H.264 distortions are included in the database, respectively. For symmetric database, according to the properties of binocular fusion and binocular summation, the weight values w_l and w_r are set to 0.50 for all types of distortion.

The DMOS values are provided for each database by subjective experiments. In the experiments, twenty non-expert adult viewers participated in the quality evaluation process, whose ages vary from 20 to 25. All the participants in this experiment met the minimum visual acuity of 20/30, stereo acuity up to 40 seconds of arc (sec-arc), and passed a color vision test. The participants were asked to rank the stereoscopic images based on their own judgment. The corresponding test methods can be found in Ref. [9].

We have compared the performance of the proposed measure with the most relevant SSIM[2] and SVD[3] measures. Since SSIM and SVD cannot directly apply to the evaluation of stereoscopic images, the left and right images are evaluated separately, and weighted with the same parameters of the proposed measure. The experimental results are reported in terms of the four criteria used for performance comparison, namely: Pearson linear correlation coefficient (CC), Spearman rank correlation coefficient (SROCC), root mean squared error (RMSE), and Outlier ratio (OR), between the subjective and the objective scores. For a perfect match between the objective and subjective scores, CC = ROCC =1 and RMSE = OR = 0. We can see from the experimental results in Table 1 that the proposed measure performs better than the existing measures in terms of the above four indicators, because the predicted scores of the proposed measure have a good correspondence with the subjective scores.

The results of CC and SROCC for different types of distortion are presented in Table. 2. Scatter plots of DMOS (y-axis) versus objective scores (x-axis) is shown in Fig. 1. The scatter plots for asymmetric database is shown in Fig. (a), and for symmetric is shown in Fig. (b). The high accuracy fitting results show the effectiveness of the proposed measure. It is also shown the extracted features of the two databases are widely different, especially for Gaussian blur and JPEG2000. The reason is that the dependencies of subjective quality on different distortion parameters for the two databases are widely different, as can be seen from the accuracy (CC) and the monotonicity (SROCC) indexes in Table 2.

In order to determine the generality of a machine-learning based image quality predictor, the cross-database validation strategy is used in this letter. We use all the images with a particular type of distortion from one database for training and use resultant measure to test all the images with the same type of distortion in another database. Suppose that $\mathrm{Model}_{\mathrm{sym}\backslash\mathrm{asy}}$ denotes symmetric database for training and asymmetric

Table 1. Performance Comparison for Asymmetric and Symmetric Databases

Database	Measure	CC	SROCC	OR	RMSE
Asymmetric	SVD	0.7359	0.7976	0.0513	7.8755
	SSIM	0.7157	0.8064	0.3297	12.6010
	Proposed	0.9447	0.9230	0.0027	4.0971
Symmetric	SVD	0.9069	0.9148	0.0192	7.2375
	SSIM	0.8287	0.8543	0.0353	9.6174
	Proposed	0.92499	0.93169	0.00641	6.5281

Table 2. CC and SROCC Results for Different Types of Distortion

Criteria Distortion	Asymmetric		Symmetric	
	CC	SROCC	CC	SROCC
Gaussian Blur	0.9247	0.8532	0.9302	0.9045
JPEG2000	0.9365	0.7737	0.9355	0.9388
JPEG	0.9066	0.8735	0.8713	0.8982
White Noise	0.9758	0.9496	0.9347	0.8852
H.264	–	–	0.8886	0.9038

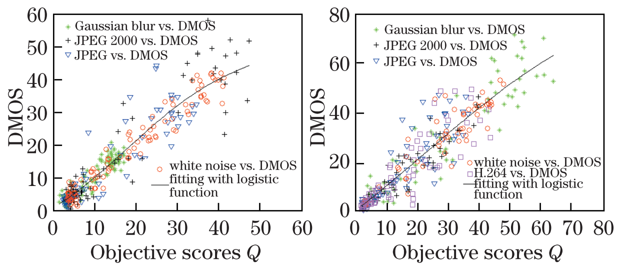

Fig. 1. Scatter plots of DMOS versus objective scores. (a) asymmetric database; (b) symmetric database.

Table 3. Cross-Database Validation Performance Comparison

Distortion\Model	$\text{Model}_{\text{sym}\backslash\text{asy}}$		$\text{Model}_{\text{asy}\backslash\text{sym}}$	
	CC	SROCC	CC	SROCC
Gaussian Blur	0.9531	0.9046	0.9513	0.9379
JPEG2000	0.7679	0.7287	0.9506	0.9282
JPEG	0.9430	0.8871	0.8272	0.8688
White Noise	0.9537	0.9552	0.8811	0.7868

database for test, and $\text{Model}_{\text{asy}\backslash\text{sym}}$ denotes asymmetric database for training and symmetric database for test. The results for the cross-database validation are presented in Table 3 (CC and SROCC values are shown). For Gaussian blur, $\text{Model}_{\text{sym}\backslash\text{asy}}$ and $\text{Model}_{\text{asy}\backslash\text{sym}}$ perform almost equivalently. For JPEG2000, $\text{Model}_{\text{asy}\backslash\text{sym}}$ outperforms $\text{Model}_{\text{sym}\backslash\text{asy}}$, and for other two distortions, $\text{Model}_{\text{sym}\backslash\text{asy}}$ outperforms $\text{Model}_{\text{asy}\backslash\text{sym}}$. The phenomenon can be explained by Table 2 that the accuracy and the monotonicity of the two databases are significantly different for some types of distortion. These results confirm that the proposed cross-database evaluation is highly dependent on the loss of perceptual quality of test database and not on the distortion contents of training database.

In conclusion, an objective quality evaluation measure of symmetric and asymmetric distorted stereoscopic images is proposed. The main contribution of the proposed measure is that the objective evaluation scores have a good correspondence with the subjective scores by establishing the relationship between stereoscopic features and subjective scores. Besides, since the stereoscopic images are well trained, the objective evaluation scores for arbitrary stereoscopic images can be automatically predicted. The experimental results show the effectiveness of the proposed measure. In order to describe stereoscopic visual perception combined with image quality and depth perception, future research will focus on no-reference stereoscopic image/video quality evaluation by fusing more perceptual features.

This work was supported by the National Natural Science Foundation of China (Nos. 60902096 and 61071120) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20093305120002).

References

1. J. Y. Son, B. Javidi, S. Yano, and K. H. Choi, J. Display Technol. **6,** 394 (2010).
2. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, IEEE Trans. on Image Processing **13,** 600 (2004).
3. A. Shnayderman, A. Gusev, and A. M. Eskicioglu, IEEE Trans. Image Processing **15,** 422 (2006).
4. A. Benoit, P. L. Callet, P. Campisi, and R. Cousseau, EURASIP J. Image and Video Processing **2008,** 659024 (2008).
5. R. Akhter, Z. M. P. Sazzad, Y. Horita, and J. Baltes, Proc. SPIE **7524,** 75240T (2010).
6. P. Gorley and N. Holliman, in *Proceedings of IS&T/SPIE Electronic Imaging* **6803,** 680305 (2008).
7. G. Saygili, C. G. Gürler, and A. M. Tekalp, IEEE Trans. on Broadcasting **57,** 593 (2011).
8. A. K. Moorthy and A. C. Bovik, IEEE Signal Processing Lett. **17,** 513 (2010).
9. X. Wang, M. Yu, Y. Yang, and G. Jiang, Proc. SPIE **7255,** 725509 (2009).