Quality Assessment for Stereoscopic Images by Distortion Separation

Chaozheng Hu, Feng Shao*, Gangyi Jiang, Mei Yu, Fucui Li, Zongju Peng
Faculty of Information Science and Engineering, Ningbo University, 315211, Ningbo, China
Email: shaofeng@nbu.edu.cn

***Abstract*—Quality assessment for stereoscopic images is a challenging issue in three-dimensional research. In this paper, we present an objective quality assessment method for stereoscopic images by distortion separation. In the method, we separate the distortion information for the distorted stereoscopic image (i.e., decompose the distorted stereoscopic image into restored and disturbed stereoscopic images), and use phase-amplitude description model and singular value decomposition model to evaluate them respectively. The experimental results show that compared with other schemes, the proposed method can achieve much higher consistency with the subjective assessments.**

***Index Terms*—stereoscopic image quality assessment, distortion separation, phase-amplitude description, singular value decomposition**

I. Introduction

With the great advancement of three-dimensional (3D) related technologies [1], 3D video [2] applications have drawn increasing attention in recent years. Since perceptual issue in 3D are completely different with that in 2D case, the necessity for designing 3D image quality assessment (3D-IQA) or stereoscopic image quality assessment (SIQA) approach is increasingly important [3].

In contrast to the 2D case, 3D quality of experience (QoE) needs to consider the various factors that contribute to the overall visual experience in 3D vision [4], e.g., depth perception, visual comfort, etc. Therefore, the direct use of 2D image quality assessment (2D-IQA) in measuring 3D image quality may not be straightforward, because the above perceptual attributes are not considered. A straightforward way of applying the state-of-the-art 2D-IQA methods to 3D-IQA is to evaluate the two views of the stereoscopic image and the estimated disparity map separately, and then combine them into an overall score. Boev *et al*. combined the monoscopic quality component and the stereoscopic quality component for developing a stereo-video quality metric [5]. Gorley *et al*. proposed a Stereo Band Limited Contrast (SBLC) algorithm to evaluate the stereoscopic image quality [6]. You *et al*. investigated the capabilities of some common 2D quality metrics, and integrated the disparity information into quality assessment [7]. Benoit *et al*. presented a linear combination for disparity distortion and the measurement of 2D image quality on both views [8]. However, the quality of stereoscopic image is not a simple combination of the qualities of left and right images, and it is not effective to assess the quality of disparity maps using 2D-IQA methods [9].

Many 3D-IQA methods were proposed by taking binocular properties into account. Maalouf *et al.* computed the cyclopean image from left and right images to simulate the brain perception, and used contrast sensitivity coefficients of cyclopean image as the basis of evaluation [10]. Jin *et al*. grouped the similar blocks from left and right views of stereoscopic image into a 3D stack, and evaluated the quality by 3D-DCT and considering contrast sensitive function and luminance masking [11]. Wang *et al*. proposed a metric by considering the binocular spatial sensitivity to reflect the binocular fusion and suppression properties [12], but the process of the binocular perception were not considered since only a weighted average of left and right views was used. Bensalma *et al*. proposed a Binocular Energy Quality Metric (BEQM) by modeling the simple cells responsible for the local spatial frequency analysis and the complex cells responsible for the generation of the binocular energy [13].

In general, distortion in an image will cause the following two cases: 1) losing some visual information; 2) adding some noticeable artifacts. Different types of quality degradation will have different influence on the perceptual quality. In this paper, we try to separate the distortion from the distorted stereoscopic image, and use different models to evaluate the restored and disturbed stereoscopic images respectively. The rest of the paper is organized as follows. Firstly, the proposed objective quality assessment metric is described in Section II. Then, experimental results are shown in Section III. Finally, conclusions and future work are given.

II. Proposed Stereoscopic Image Quality Assessment Metric

The framework of the proposed quality assessment metric is illustrated in Fig.1. Given the original and distorted stereoscopic images (case of left and right images), the distorted image is first decomposed into a restored and a disturbed images by distortion separation strategy, and phase-amplitude description (PAD) model and singular value decomposition (SVD) model are used to measure their similarities with the original image respectively. Finally, binocular combination is made to get a total quality score.

doi:10.4304/jsw.9.1.37-43

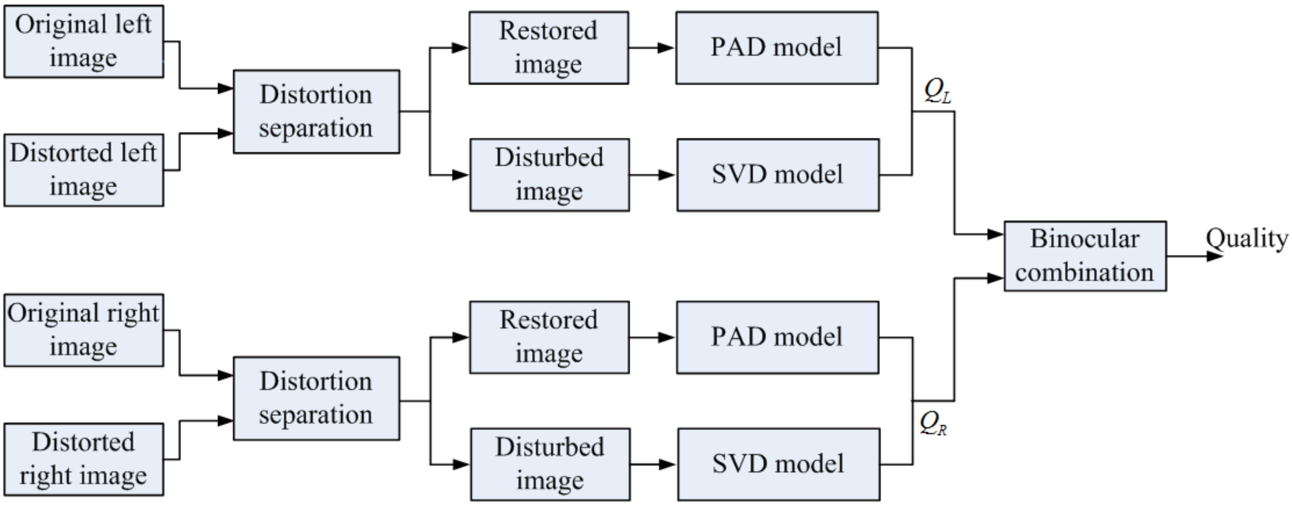


Figure 1. The framework of the proposed quality assessment metric.

A. Distortion Separation From Stereoscopic Image

It is known that distortion in image will cause two cases: 1) losing some visual information; 2) adding some noticeable artifacts. In this work, we classify the types of distortions into two group, information-loss distortion and artifact-additive distortion. Specifically, we separate the distorted image into a restored image and a disturbed image, and measure the detail and redundancy degradation. Firstly, considering that wavelet transform decomposes image into different frequency, we use discrete wavelet transform (DWT) to decompose the original and distorted images into a set of subbands. In this work, we adopt block-based image restoration in wavelet domain. Supposed that $D(\lambda,\theta,i,j)$ (8×8 block in the experiment) denotes the DWT coefficients on different scales and along different orientations of the (i,j)-th block of the distorted image (denoted by spatial scale index λ and orientation index θ), and $O(\lambda,\theta,i,j)$ denotes the corresponding DWT coefficients of the original image. The scale factors are given by [14]

$$k(\lambda,\theta,i,j) = clip\left(\frac{D(\lambda,\theta,i,j)}{O(\lambda,\theta,i,j)+10^{-30}},0,1\right) \quad (1)$$

The DWT coefficients of the restored image can be obtained by

$$R(\lambda,\theta,i,j) = \begin{cases} D(\lambda,\theta,i,j), & \theta=1 \\ k(\lambda,\theta,i,j)\times O(\lambda,\theta,i,j), & otherwise \end{cases} \quad (2)$$

Since only DWT coefficients in the orientations ($\theta\neq1$) are restored, the detail information of the original image is preserved in the restored image while the added redundancy information is discarded.

The DWT coefficients of the redundancy image is obtained by

$$A(\lambda,\theta,i,j) = D(\lambda,\theta,i,j) - R(\lambda,\theta,i,j) \quad (3)$$

The disturbed image is described by

$$D'(\lambda,\theta,i,j) = O(\lambda,\theta,i,j) + A(\lambda,\theta,i,j) \quad (4)$$

Finally, the restored images and the disturbed images are generated by inverse-transforming their respective DWT coefficients. Fig.2 illustrates the results of the proposed separation method for different distortion types.

B. Quality Assessment Metric

For the restored image, the detail is preserved while the redundancy is discarded. Therefore, structural similarity between the original and restored images is expected to give a reasonable estimation of quality degradation. We get the local phase (LP) and local amplitude (LA) referring to the method in [15]. Then, the phase and magnitude similarities for each pixel in the left image are defined as

$$S^l_{LP}(x,y) = \frac{2\times LP^{(org)}(x,y)\times LP^{(res)}(x,y)+C_1}{LP^{(org)}(x,y)^2+LP^{(res)}(x,y)^2+C_1} \quad (5)$$

$$S^l_{LA}(x,y) = \frac{2\times LA^{(org)}(x,y)\times LA^{(res)}(x,y)+C_2}{LA^{(org)}(x,y)^2+LA^{(res)}(x,y)^2+C_2} \quad (6)$$

where C_1 and C_2 are constants to avoid the denominator being zero. Finally, the final quality score for the left image is obtained by summing the scores of all pixels

$$Q^l_{PAD} = \left(\frac{1}{H\times W}\sum_{y=1}^{H}\sum_{x=1}^{W}S^l_{LP}(x,y)\right)\cdot\left(\frac{1}{H\times W}\sum_{y=1}^{H}\sum_{x=1}^{W}S^l_{LA}(x,y)\right) \quad (7)$$

For the disturbed image, the redundancy information is added on the original image. Therefore, energy similarity between the original and restored images is used to measure the quality degradation. In this work, we use the singular values as feature basis for the task [16]. The energy change between the original and disturbed images in the singular values is calculated

$$\tau_k = \frac{\left\langle \left|\mathbf{S}_k^{(org)}-\mathbf{S}_k^{(res)}\right|\cdot\mathbf{S}_k^{(org)}\right\rangle}{\left\langle \mathbf{S}_k^{(org)}\cdot\mathbf{S}_k^{(res)}\right\rangle} \quad (8)$$

where $\mathbf{S}_k^{(org)}$ and $\mathbf{S}_k^{(res)}$ denote the singular value vectors of the original and restored images, respectively, and $\langle\cdot\rangle$ denotes the inner product. Finally, the final quality score for the left image is obtained by averaging the changes over all the blocks

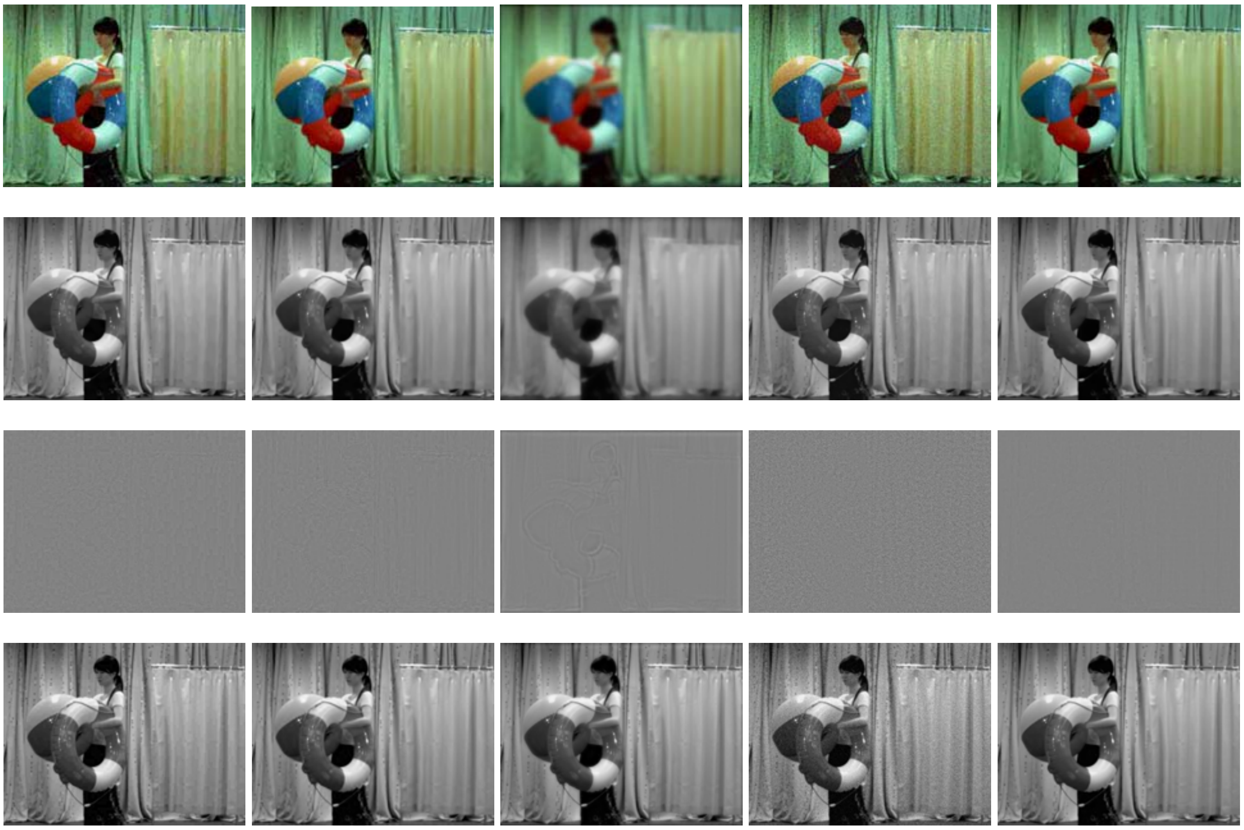


Figure 2. Results of the proposed separation method for different types of distortion, from top to bottom row: the distorted images, the restored images, the redundancy images, and the disturbed images. For left to right: JPEG, JPEG2000, Gaussian Blur, White Noise and H.264.

$$Q_{SVD}^{l}=\frac{1}{N_{\text{block}}}\sum_{k=1}^{N_{\text{block}}}\tau_k \tag{9}$$

Considering that the detail and redundancy losses in the distorted image are superimposed, the above two quality scores Q_{PA}^{l} and Q_{SVD}^{l} are combined into an overall score by a linear weighted sum method, i.e.,

$$\begin{aligned} &Q_L = w_1\cdot Q_{PAD}^{l} + w_2\cdot Q_{SVD}^{l} \\ &s.t.\ w_1 + w_2 = 1 \end{aligned} \tag{10}$$

where w_1 and w_2 are parameters used to adjust the relatively importance of redundancy adjunction and detail loss in the quality degradation. In this paper, the parameters can be determined by training. Similarly, the quality score Q_R for the right image can be measured by the same manner.

C. Binocular Combination

After having obtained the quality scores Q_L and Q_R, the next step is to combine the two quality scores into a final score. The direct way is to combine the quality scores Q_L and Q_R by average weighting. However, the weight-averaged method is not effective because binocular combination property is not well considered. In this work, we use two-stage gain control model to combine the two quality scores [17]

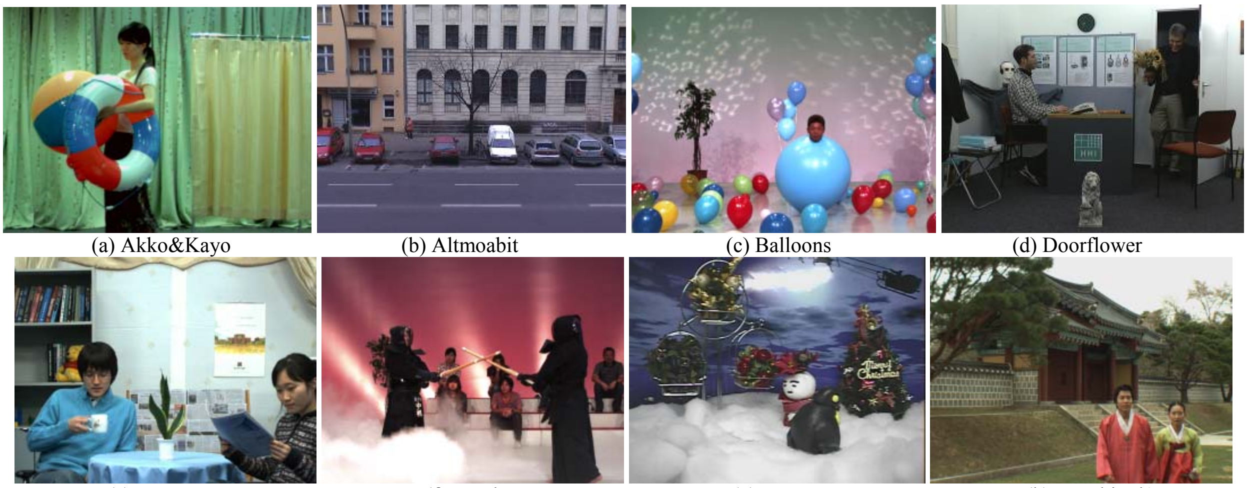
$$Q(L,R)=\frac{\left(Stage1(Q_L)+Stage1(Q_R)\right)^{p}}{z+\left(Stage1(Q_L)+Stage1(Q_R)\right)^{q}} \tag{11}$$

where $Stage1(Q_L)=\frac{(Q_L)^m}{s+Q_L+Q_R}$, $Stage1(Q_R)=\frac{(Q_R)^m}{s+Q_L+Q_R}$, and p, q, m, z are model parameters. In the experiment, the same parameter setting with [17] is used.

III. Experimental Results

A. Stereoscopic Image Quality Database

In the experiment, we have used the database presented in [18]. Twenty-six non-expert adult viewers were participated in the subjective evaluation of the database. According to Double Stimulus Continuous Quality Scale (DSCQS) testing method described in ITU-R recommendation BT.500-11, the subjective ratings for the distorted stereoscopic images were obtained on a scale of 0-10. The database includes 12 original stereoscopic image pairs, from which 312 distorted stereoscopic images are generated with five types of distortion: JPEG, JPEG2000, Gaussian Blur, White Noise and H.264. The symmetric distortions are added on left and right images. More specifically, there are 60, 60, 60, 60 and 72 distorted stereoscopic images in the database with JPEG, JPEG2000, Gaussian Blur, White Noise and H.264 distortions, respectively; there are different distortion levels for each distortion type. The corresponding differential mean opinion score (DMOS) values are



(a) Akko&Kayo (b) Altmoabit (c) Balloons (d) Doorflower

(e) Newspaper (f) Kendo (g) Xmas (h) Lovebierd1

Figure 3. Eight selected reference left images in the 3D database [18].

provided. Eight selected reference left images used in the database are shown in Fig.3.

B. Performance Determination

In the experiment, four commonly used performance indicators are employed to further evaluate the metric: Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC), Kendall rank-order correlation coefficient (KROCC), and root mean squared error (RMSE), between the objective scores after nonlinear regression and the subject scores. Among these four criteria, SROCC and KROCC are employed to assess prediction monotonicity, and PLCC and RMSE are used to evaluate prediction accuracy. For a perfect match between the objective and subjective scores, PLCC=SROCC=KROCC=1 and RMSE=0. To obtain the relationship between the objective scores and the subjective scores, we use the nonlinear regression with four-parameter logistic function by

$$DMOS_p = \frac{\beta_1 - \beta_2}{1 + \exp(-(x - \beta_3)/\beta_4)} + \beta_2 \qquad (12)$$

where β_1, β_2, β_3 and β_4 are determined by using the subjective scores and the objective scores.

In the experiment, in order to determine the parameters w_1 and w_2, we select a subset of the database to train the parameters by optimizing the PLCC values between the objective and subjective scores. The final parameter determination results is w_1=0.9208, w_2=0.0792. It is obvious that the restored image is more important than the disturbed image in measuring the quality degradation.

C. Overall Assessment Performance

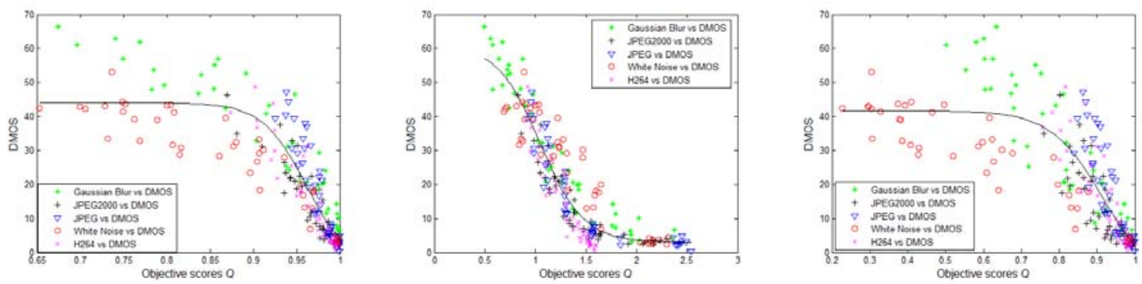
In order to evaluate the performance of the proposed scheme, we compare the evaluation results with two 2D-IQA metrics MSSIM [19], SVD [20], and one SIQA metric (named as Wang-SIQA) [12]. The former two schemes directly estimate the quality of each view separately and generate a weighted average score. The results of PLCC, SROCC, KROCC and RMSE are presented in Table I. From the table we can see that the proposed scheme outperforms the other schemes. For MSSIM and SVD metrics, since they are directly extended from the 2D case and do not take the binocular properties into account, the overall performance is far worse than the proposed scheme. For Wang-SIQA metric, even though it may be effective for some individual distortion types, the overall assessment performance is not very high; the reason is that uniform assessment is adopted for the left and right images, while in the proposed scheme, the similarity measured from the restored image and disturbed image respectively will have a good correspondence with the subjective scores. Fig.4 gives the scatter plots for the MSSIM, SVD and Wang-SIQA metrics. Fig.5 gives the scatter plots for the independent and overall distortion types for the proposed scheme, respectively. The vertical axis denotes the subjective ratings of the perceived distortions and the horizontal axis denotes the predicted objective scores. From the figures, the high accuracy fitting results show the effectiveness of the proposed scheme.

IV. Conclusions

This paper presents a quality assessment method for stereoscopic images by distortion separation. The prominent advantage of the proposed method is that we separate the distortion from the distorted stereoscopic image, and use different singular value decomposition (SVD) model and phase-amplitude description (PAD) model to evaluate the restored and disturbed stereoscopic images respectively. The experimental results show that the proposed method can achieve much higher consistency with the subjective assessments. In this research, only simple image separation model is used for stereoscopic image without considering the binocular characteristics. In the future work, more comprehensive study of various distortions affecting depth perception, visual comfort is needed, and these cues should be fully considered in the separation model.

TABLE 1.
PERFORMANCE COMPARISONS OF DIFFERENT SCHEMES .

Algorithms		Distortion types					
		JPEG	JP2K	GB	WN	H.264	All
PLCC	MSSIM[19]	0.9399	0.9256	0.9525	0.9486	0.9405	0.9028
	SVD[20]	0.9516	0.9523	0.9715	0.9417	0.9639	0.9295
	Wang-SIQA[12]	0.8837	0.8261	0.9221	0.9385	0.9097	0.8564
	Proposed	**0.9840**	**0.9599**	**0.9793**	**0.9564**	**0.9743**	**0.9348**
SROCC	MSSIM[19]	0.9340	0.9248	0.9607	0.9261	0.9314	0.9138
	SVD[20]	0.9443	0.9461	0.9645	0.8993	0.9568	0.9042
	Wang-SIQA[12]	0.8998	0.8737	0.9355	0.9088	0.9002	0.8887
	Proposed	**0.9849**	**0.9692**	**0.9741**	**0.9447**	**0.9637**	**0.9313**
KROCC	MSSIM[19]	0.7509	0.7616	0.8277	0.7535	0.7681	0.7307
	SVD[20]	0.7852	0.8439	0.7899	0.7091	0.8339	0.7271
	Wang-SIQA[12]	0.7044	0.6788	0.7731	0.7394	0.7205	0.6917
	Proposed	**0.9045**	**0.8566**	**0.8742**	**0.8020**	**0.8451**	**0.7709**
RMSE	MSSIM[19]	4.7399	4.2306	6.0480	4.7454	4.3388	7.0088
	SVD[20]	4.2663	3.4077	4.6954	5.0455	3.4038	6.0069
	Wang-SIQA[12]	6.4974	6.2955	7.6666	5.1829	5.3026	8.4101
	Proposed	**2.4706**	**3.1321**	**4.0169**	**4.4295**	**2.8802**	**5.7838**



(a) MSSIM[19] (b) SVD[20] (c) Wang-SIQA[12]
Figure 4. Scatter plots of objective scores vs. subjective scores for MSSIM, SVD and Wang-SIQA schemes.

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of China (grant 61071120, 61271021, 612712700), and the Natural Science Foundation of Ningbo (grant 2012A610039).

REFERENCES

[1] A. Smolic, P. Kauff, S. Knorr, et al, "Three-dimensional video postproduction and processing," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 607-625, April 2011.

[2] Peng Geng, He Jiang, Zhigang Zhang, et al, "A Video Denoising Method with 3D Surfacelet Transform Based on Block matching and Grouping," *Journal of Computers*, vol. 7, no. 5, pp. 1065-1066, 2012.

[3] H. T. Quan, P. Le Callet, M. Barkowsky, "Video quality assessment: from 2D to 3D – challenges and future trends," in *Proc. of IEEE International Conference on Image Processing*, pp. 4025-4028, Hong Kong, Sep. 2010.

[4] L. Goldmann, T. Ebrahimi, "Towards reliable and reproducible 3D video quality assessment," in *Proc. of SPIE: Three-Dimensional Imaging, Visualization, and Display*, vol. 8043, no. 804302, San Jose, California, USA, April 2011.

[5] A. Boev, A. Gotchev, K. Egiazarian, et al, "Towards compound stereo-video quality metric: a specific encoder-based framework," in *Proc. of IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 218-222, Denver, Colorado, 2006.

[6] P. Gorley, and N. Holliman, "Stereoscopic image quality metrics and compression," in *Proc. of SPIE: Stereoscopic Displays and Applications XIX*, San Jose, CA, USA, vol. 6803, no. 680305, 2008.

[7] J. You, L. Xing, A. Perkis, at al, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in *Proc. of*

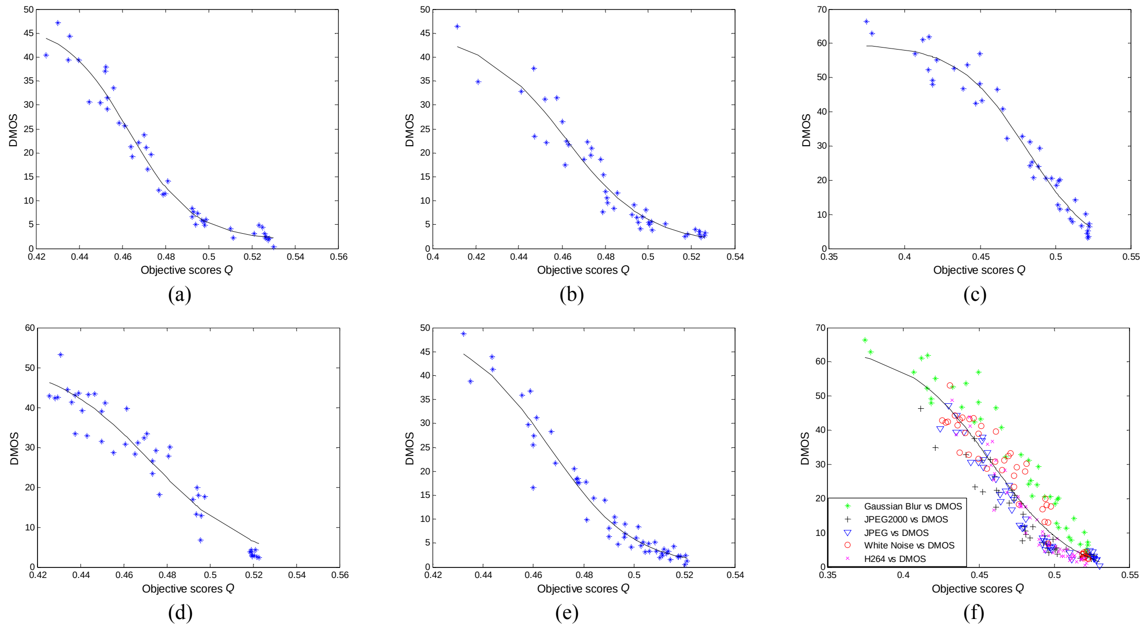


Figure 5. Scatter plots of objective scores vs. subjective scores for the proposed scheme: (a) JPEG; (b) JPEG2000; (c) Gaussian Blur; (d) White Noise; (e) H.264; (f) All distortions.

International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ, USA, 2010

[8] A. Benoit, P. Le Callet, P. Campisi, et al, "Using disparity for quality assessment of stereoscopic images," in *Proc. of IEEE Intentional conference on Image Processing*, pp. 389-392, San Diego, CA, USA, Oct. 2008.

[9] Yiyong Han, Jujun Zhang, Benkang Chang, et al, "Novel Fused Image Quality Measures Based on Structural Similarity," *Journal of Computers*, vol. 7, no. 3, pp. 563-566, 2012.

[10] A. Maalouf, M. C. Larabi, "CYCLOP: A stereo color image quality assessment metric," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1161-1164, Prague, Czech Republic, May 2011.

[11] L. Jin, A. Boev, A. Gotchev, et al, "3D-DCT based perceptual quality assessment of stereo video," in *Proc. of IEEE Intentional conference on Image Processing*, pp. 2521-2524, Brussels, Belgium, Sep. 2011.

[12] X. Wang, S. Kwong, Y. Zhang, "Considering binocular spatial sensitivity in stereoscopic image quality assessment," in *Proc. of IEEE Visual Communications and Image Processing*, Tainan City, Taiwan, Nov. 2011.

[13] R. Bensalma, M. C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Systems and Signal Processing*, to appear, 2012.

[14] S. Li, F. Zhang, L. Ma, et al, "Image quality assessment by separately evaluating detail losses and additive impairments," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 935-949, Oct. 2011.

[15] L. Zhang, L. Zhang, X. Mou, et al, "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. on Image Processing*, vol. 20, no. 8, pp. 2378-2386, Aug. 2011.

[16] Yanfang Hou, Hongmei Feng. "Study of Modulation Recognition Algorithm Based on Wavelet Transform and Neural Network," *Journal of Computers*, vol. 6, no. 7, pp. 1511-1518, 2011.

[17] T. S. Meese, M. A. Georgeson, D. H. Baker, "Binocular contrast vision at and above threshold," *Journal of Vision*, vol. 6, no. 11, pp. 1224-1243, Oct. 2006.

[18] J. Zhou, G. Jiang, X. Mao, et al, "Subjective quality analyses of stereoscopic images in 3DTV system," in *Proc. of IEEE Visual Communications and Image Processing*, Tainan City, Taiwan, Nov. 2011.

[19] 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, April 2004.

[20] A. Shnayderman, A. Gusev, A. M. Eskicioglu, "An SVD-based grayscale image quality measure for local and global assessment," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 422-429, 2006.

Chaozheng Hu received his B.S.degree from Department of Computer and Information, Anqing Normal College, Anqing, China, in 2011. He is now working towards his M.S. degree in Ningbo University, Ningbo, China. His current research interests include image and video processing and stereoscopic image quality assessment.

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 in 1992, and received his Ph.D. degree from Ajou University, Korea, in 2000. He is now a professor in Faculty of Information Science and Engineering, Ningbo University, China. His research interests mainly include video compression, multi-view video coding, etc.

Mei Yu received her M.S. degree from Hangzhou Institute of Electronics Engineering, China, in 1993, and Ph.D. degree form Ajou University, Korea, in 2000. She is now a professor in

Faculty of Information Science and Engineering, Ningbo University, China. Her research interests include video coding and video perception.

Fucui Li received her M.S. degree from Hefei University of Technology, China, in 2004. She is now pursuing the Ph.D degree at Ningbo University, China. Her research interests include video compression and communications, multi-view video coding and error concealing.

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.