Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (15): 3061-3066, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.3061.3066

New Stereoscopic Image Quality Assessment Metric Based on Three Dimensional-discrete Cosine Transform for 3d Media

[1]Xianwei Tang, [1,2]Mei Yu, [1,2]Gangyi Jiang, [1]Zongju Peng, [1]Feng Shao and [1]Xiaodong Wang
[1]Faculty of Information Science and Engineering, Ningbo University, Ningbo, China
[2]National Key Lab of Software New Technology, Nanjing University, Nanjing, China

Abstract: This study presents a new stereoscopic image quality assessment metric based on three dimensional-discrete cosine transform (3D-DCT) for 3D media. In the proposed metric, the factors affecting stereoscopic perception are first combined to construct a three dimensional signal (called as 3D stack). 3D stack's coefficients are calculated in the 3D-DCT domain. Then, Comparisons between reference and distorted stereoscopic images are made in terms of three important coefficients after perceptual weighting. Finally, we get the objective assessment scores of the distorted stereoscopic images. Experimental results show that the proposed metric outperforms relevant existing metrics in respect to correlation with the subjective evaluation and can predict the stereoscopic image quality well.

Key words: Stereoscopic image, quality assessment, three dimensional-discrete cosine transform (3D-DCT)

INTRODUCTION

With the rapid development of stereoscopic display technologies, three-dimensional (3D) image processing technologies have attracted public attentions and have widespread prospect of applications (Shao *et al*., 2013). It is inevitable to introduce distortions for images after a series of links such as acquisition and processing. Stereoscopic Images Quality Assessment (SIQA) is an evaluation technique for stereoscopic images which passed through a processing system (Seo *et al*., 2012). It has a great significance in measuring performance of stereoscopic image processing techniques.

SIQA metrics can be divided into two categories, namely subjective assessment and objective assessment. Subjective assessment is known to be the most reliable method because it concerns how a pair of stereoscopic image is perceived by a viewer. However, it is time consuming and expensive and it cannot be embedded into the system for real-time evaluation. So, objective assessment metrics are needed. Objective evaluations of the existing SIQA methods are mainly the following two types: (1) "directly 2D metrics", in which the 2D image quality assessment algorithm (2D-IQA) was applied on the left and right images separately and the estimated quality scores averaged to produce single measure of 3D quality (Lambooij *et al*., 2011). These metrics consider neither stereoscopic perception nor interaction between the two viewpoints, so evaluation results of these metrics are not satisfactory. (2) "2D based metrics", these metrics are simple extensions of 2D-IQA algorithms with some additional "features" extracted from depth (Ma *et al*., 2013) or taking the properties of the HVS into consideration. Yang (2009) assessed stereoscopic images from the perspective of image quality and stereoscopic sense with the PSNR to measure the image quality and considering disparity distribution to measure the stereoscopic sense. Some researchers proposed 3D-QA metrics based on the traditional 2D-IQA with several main HVS properties been considered, such as the contrast sensitivity function, masking effect and multi-channel mechanism (Gorley and Holliman, 2008; Shen *et al*., 2009; Zhu and Wang, 2009). The proposed metrics are more perceptually consistent compared with the traditional 2D-IQA. However, these metrics still do not break through the "directly 2D metrics". Also, they involve the combination between image quality and depth perception and the combination between the two viewpoints for image quality which may lead to another problem of parameter optimization. There are also some non full-reference 3D QA metrics for stereoscopic images, for example, Hewage and Martini (2010) proposed a reduced-reference quality metric for 3D depth map transmission based on edge detection. Akhter *et al*. (2010) developed a no-reference perceptual quality assessment for JPEG coded stereoscopic image based on segmented local features of artifacts and disparity.

The systems used to display stereoscopic images present alternatively to the left and the right eyes two slightly different images in such a way that the Human

Corresponding Author: Xianwei Tang, Faculty of Information Science and Engineering, Ningbo University, Ningbo, China

Visual System (HVS) gets a perception of depth. Therefore, we can infer that left view, right view and the difference between the left and right views are the three main factors affecting the quality of stereo images. In this paper, a new objective evaluation metric for stereo images is proposed. The proposed method evaluates stereoscopic images quality in the form of 3D stack constructed by the three main factors.

PROPOSED METRIC

The flowchart of the proposed metric is given in Fig. 1. For each reference block L_i in the left reference view, the corresponding block L'_i in the left distorted view is selected. In the right reference view, the best matched block of L_i is found, namely R_i and the corresponding block R'_i in the right distorted view is selected. In the reference views, L_i, R_i and D_i which reflects the differential information between L_i and R_i are stacked into a 3D structure B_i which then undergoes 3D-DCT. The same procedure is done for the structure B'_i associated with the structure B_i in the distorted views. Then, perceptual weighting is made to three coefficients in DCT domain according to HVS properties. Next, differences between the two sets of coefficients are calculated to measure the quality of 3D stack B'_i. Finally, we get the quality index of stereo image by combining all 3D structures together.

Constructing 3D structure: For a reference stereo image pair, the left reference view I_L is first divided into non-overlapping blocks with size 8×8, as shown in Fig. 2a, L_i (the black area in Fig. 2a with vertex coordinate of the upper left (r, c)) is a block of I_L. A matching region (MR) for L_i is specified in the right reference view I_R with horizontal range from c-16 to c+15, as shown in Fig. 2b, Block-Matching (BM) is applied to find the best

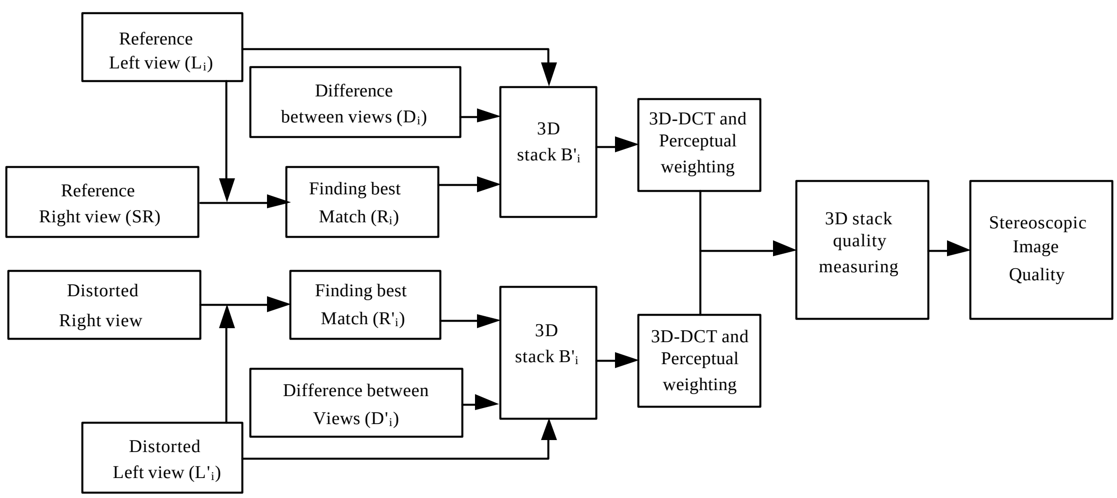

Fig. 1: Flowchart of proposed metric

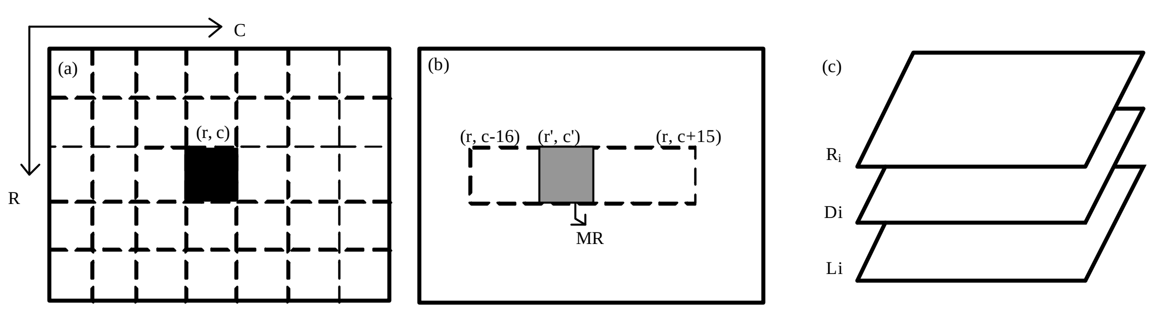

Fig. 2(a-c): Constructing the 3D stack, (a) Block segmentation, (b) Searching for the best match and (c) Grouping horizontal block match of L_i in the MR of I_R. Sum of Absolute Differences (SAD) is used as matching metric

$$SAD = \sum_{h=1}^{8}\sum_{l=1}^{8}|L_i(h,l) - S_i(h,l)| \tag{1}$$

where $L_i(h,l)$ and $S_i(h,l)$ denote the pixel intensity of the reference block and searched block respectively. After using BM, the best horizontal block match R_i is found with vertex coordinate of the upper left (r', c') and marked as gray area in Fig. 2b. The difference between the left and right views associated with L_i can represented as disparity of reference stereo image pair or absolute difference block D_i

$$D_i = |L_i - R_i| \tag{2}$$

where L_i, R_i and D_i are in size of 8×8. Finally, L_i, R_i and D_i are grouped into a 3D stack B_i with size 8×8×3, as shown in Fig. 2c.

3D structure B'_i associated with B_i in the distorted stereoscopic image is constructed similarly except that it doesn't need the process of BM. The best matched block R'_i of L'_i in the distorted stereoscopic image pair can be determined through the coordinate (r', c') directly.

3D stack quality prediction: Two dimensional DCT is widely used in the field of signal processing due to its capabilities to decorrelate data and achieve highly sparse representation. Li *et al*., (2013) pointed out that 3D-DCT has a better performance in decorrelation and can generate a compact energy spectrum. Therefore, we first calculate the coefficients of B_i and B'_i in 3D-DCT space by 3D-DCT, respectively. Then, we choose the three most important coefficients to represent the information about B_i and B'_i, respectively. Finally, we get the quality of the distorted 3D stack B'_i by measuring the difference between the two sets of coefficients. For a 3D signal (f (x, y, z))$N_1 \times N_2 \times N_3$), the 3D-DCT is defined by:

$$F(\mu,\nu,\omega) = \alpha_1(\mu)\alpha_2(\nu)\alpha_3(\omega)\sum_{x=0}^{N_1-1}\sum_{y=0}^{N_2-1}\sum_{z=0}^{N_3-1} f(x,y,z) \cdot \left\{\cos\left[\frac{(2x+1)\mu\pi}{2N_1}\right]\cos\left[\frac{(2y+1)\nu\pi}{2N_2}\right]\cos\left[\frac{(2z+1)\omega\pi}{2N_3}\right]\right\} \tag{3}$$

where, $\mu \in \{0, 1, \ldots, N_1-1\}$, $\nu \in \{0, 1, \ldots, N_2-1\}$, $\omega \in \{0, 1, \ldots, N_3-1\}$ and $\alpha_k(\mu)$ is defined as:

$$\alpha_k(\mu) = \begin{cases} \sqrt{1/N_k}, & \text{if } \mu = 0; \\ \sqrt{2/N_k}, & \text{otherwise;} \end{cases} \tag{4}$$

where, k is a positive integer. Therefore, the coefficients in 3D-DCT space of the 3D stacks B_i and B'_i with the size 8×8×3 can be expressed as $F_i(\mu, \nu, \omega)$ and $F'_i(\mu, \nu, \omega)$ respectively, here, $0 = \mu$, $\nu<8$, $0 = \omega<3$. The 3D-DCT coefficients, $F_i(\mu, \nu, \omega)$ and $F'_i(\mu, \nu, \omega)$ can be calculated first in the (x, y) directions by 2D-DCT then in the z direction by 1D-DCT, also, we can use a fast algorithm to get the 3D-DCT coefficients directly.

Here we use only three coefficients to represent the characteristics of the 3D stack, namely, the DC coefficient F(0, 0, 0), low-frequency coefficients F(0, 1, 0) and F(1, 0, 0). In order to better reflect the characteristics of HVS, we have weighted these coefficients with perceptual weights. Since the JPEG quantization table (Wallace, 1991) is made according to HVS's sensitivity to different locations in the DCT space, we can define the perceptual weights by inversing the JPEG quantization table. The perceptual weights of the three coefficients (F(0, 0, 0), F(0, 1, 0), F(1, 0, 0)) are 0.0625, 0.0909 and 0.0833, respectively. Then, we define a feature vector $\Gamma_i = (\tau_1, \tau_2, \tau_3)$ based on the three coefficients to reflect the feature information about the 3D stack B_i, where $\tau_1 = 0.0625 \times F_i(0, 0, 0)$, $\tau_2 = 0.0909 \times F_i(0, 1, 0)$, $\tau_3 = 0.0833 \times F_i(1,0,0)$. Similarly, Ã'$_i$ = (τ'_1, τ'_2, τ'_3) is the feature vector of B'_i, where $\tau'_1 = 0.0625 \times F'_i(0, 0, 0)$, $\tau'_2 = 0.0909 \times F'_i(0, 1, 0)$, $\tau'_3 = 0.0833 \times F'_i(1,0,0)$. Finally, the local quality of B'_i in the distorted stereoscopic image pair is computed by:

$$Q_i = \sqrt{\frac{\sum_{k=1}^{3}(\tau_k - \tau'_k)^2}{3}} \tag{5}$$

Overall stereoscopic image quality prediction: Given a distorted stereo image pair, we can construct N_{blk} 3D stacks. For each 3D stack B'_i, L'_i and R'_i are the first and the third component of B'_i respectively, U'_L, U'_R are means of the luminance of block L'_i and R'_i, respectively. Therefore, we define $U' = (U'_L + U'_R)/2$ as the mean of the luminance of B'_i. Wang *et al*. (2004) pointed that dark regions usually do not attract fixations and should be assigned smaller weightings. According to this conclusion, the local weighting is adjusted as:

$$w_i = \begin{cases} 0, & U' \le 40; \\ (U' - 40)/10, \quad 40 < U' \le 50; & \\ 1, & U' > 50; \end{cases} \tag{6}$$

Finally, the overall of the entire distorted stereoscopic image pair is given by:

$$Q = \sum_{i=1}^{N_{blk}} Q_i \cdot w_i \Big/ \sum_{i=1}^{N_{blk}} w_i \tag{7}$$

where, Q_i is the quality prediction of the *i*-th 3D stack in the distorted stereoscopic image pair and N_{blk} is the numbers of the 3D stacks.

EXPERIMENTAL RESULTS AND ANALYSIS

Database description: We used stereoscopic images from two databases in this study. The first database (Wang *et al*., 2009) (we refer to it as the NBU's database) consists of 12 reference stereoscopic images at different spatial resolution. The distorted stereoscopic images are generated with five distortions: Gaussian blurring (GBlur) with five quality levels, H.264 compression with six quality levels, JPEG2000 compression (JP2K) with five quality levels, JPEG compression with five quality levels and white Gaussian noise (WN) with five quality levels. Thus, finally 312 stereoscopic images were rated by 26 subjects. Subjective scores are provided in the form of Difference Mean Opinion Scores (DMOS). The second database we used is the LIVE 3D-IQA database (Moorthy *et al*., 2013) which is a publicly available database. It contains 20 reference images, 5 distortion categories and a total of 365 distorted images along with the associated DMOS. Meanwhile, LIVE 3D-IQA database incorporates "true" depth information and provides disparity map. Furthermore, the authors employed several 3D image quality metrics to provide the benchmark performances in this database, such as: Yang's, Gorley's, Shen's, Zhu's, Hewage's and Akhter's which enables performance comparisons of different algorithms.

Performance comparison and discussion: The experimental results are reported in terms of four criteria, namely, Pearson linear correlation coefficient CC (for prediction accuracy), Spearman rank order correlation coefficient SROCC (for monotonicity), root mean squared error RMSE (for prediction accuracy) and outlier ratio OR (for prediction consistency) which are commonly used for performance comparison between the subjective score and the objective prediction. A better quality metric will have higher CC, SROCC and lower RMSE and OR. A monotonic function should be employed to provide a nonlinear mapping between the objective/subjective scores before calculating the four criteria (Brunnstrom *et al*., 2009).

For NBU's database, the logistic function used for nonlinear mapping is a four-parameter logistic function:

$$DMOS_p = \frac{a-b}{1+\exp\left[-(Q-c)/\text{abs}(d)\right]}+b \tag{8}$$

The performance of the proposed metric is shown in Table 1 and Fig. 3a shows the scatter plot of subjective ratings (MOS) versus the proposed metric(Q). For the scatter plot, the density of data points closely to the fitting curve represents the consistency of the predictive method and the subjective evaluation.

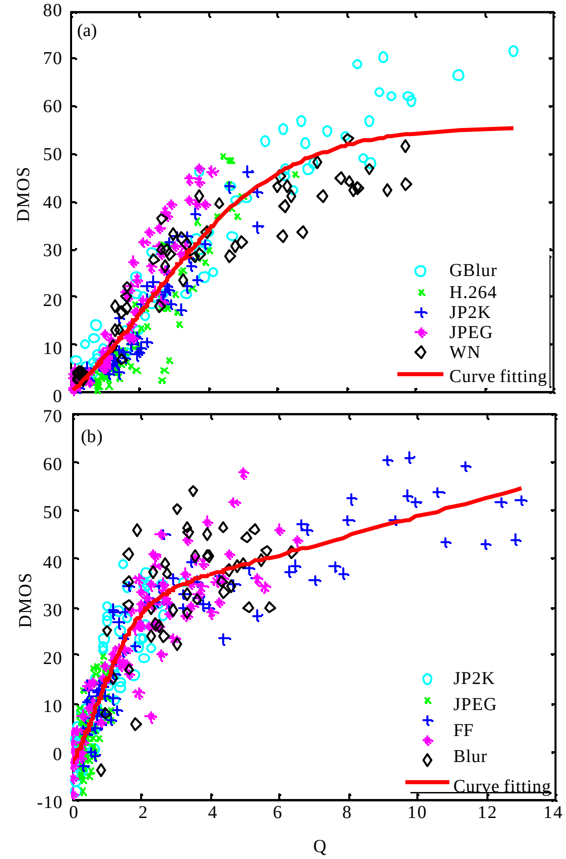

Fig. 3(a-c): Scatter plots for the quality predictions in two databases: (a) about NBU's database, (b) about LIVE's database, (a) About NBU's database, (b) About LIVE's database

Table 1: Proposed metric's performance on NBU's database

	CC	SROCC	RMSE	OR
GBlur	0.9696	0.9632	5.1634	0
H.264	0.9401	0.9108	4.7829	0.0139
JP2K	0.9486	0.9508	3.7937	0
JPEG	0.9551	0.9575	4.2187	0
WN	0.9627	0.9381	4.2282	0
ALL	0.9314	0.9298	6.2531	0.0096

From Table 1, we can see that the CC is higher than 0.93 for all the distortion types (ALL) and 0.94 for each individual distortion type and the data points in the scatter plot are very closely to the fitting curve which demonstrate that the proposed metric can predict the quality of stereoscopic images well.

For LIVE's database, the function used there for nonlinear mapping is a five-parameter logistic function which has also been used in the comparison metrics. The five-parameter logistic function is:

$$DMOS_p = \beta_1 \bullet \text{logistic}(\beta_2, (Q-\beta_3))+\beta_4 \bullet Q+\beta_5 \tag{9}$$

Table 2: Performance comparison for considered metrics (the cases in bold denote the best performance)

	Criteria	JP2K	JPEG	WN	FF	Blur	ALL
Yang's	CC	0.2012	0.2738	0.8701	0.2824	0.6261	0.3909
	SROCC	0.1501	0.1328	0.8471	0.1426	0.3266	0.0785
	RMSE	12.6979	6.2894	8.2002	11.9462	12.1291	15.2481
Gorley's	CC	0.4853	0.3124	0.7961	0.3648	**0.8527**	0.4511
	SROCC	0.4203	0.0152	0.7408	0.3663	**0.7498**	0.1419
	RMSE	11.3237	6.2119	10.1979	11.5691	**7.5622**	14.6350
Shen's	CC	0.5039	0.3899	0.8988	0.4830	0.6846	0.5743
	SROCC	0.2133	0.2440	0.8917	0.2665	0.6586	0.0679
	RMSE	12.2754	6.0216	7.2939	10.8820	10.5547	13.5473
Zhu's	CC	0.8073	0.3790	0.5178	0.5038	0.7770	0.6263
	SROCC	0.7708	0.2929	0.4651	0.4752	0.7935	0.6388
	RMSE	7.6813	6.0684	14.7201	10.7362	9.1270	12.7828
Hewage's	CC	0.9043	0.5305	0.8955	0.6698	0.7984	0.8303
	SROCC	0.8558	0.5001	0.8963	0.5447	0.6900	0.8140
	RMSE	5.5300	5.5431	7.4056	9.2263	8.7480	9.1393
Akhter's	CC	0.9059	**0.7294**	**0.9047**	0.6603	0.6177	0.4270
	SROCC	0.8657	**0.6754**	0.9137	0.6393	0.5549	0.3827
	RMSE	5.4836	**4.4736**	**7.0929**	9.3321	11.3872	14.8274
Proposed	CC	**0.9151**	0.6396	0.8798	**0.9091**	0.6785	**0.9232**
	SROCC	**0.8885**	0.6272	**0.9273**	**0.8834**	0.5254	**0.9162**
	RMSE	**5.2229**	5.0724	7.9181	**6.2458**	8.7057	**6.3032**

$$\text{logistic}(\tau, Q) = \frac{1}{2} - \frac{1}{1 + \exp(\tau \cdot Q)} \quad (10)$$

Figure 3b shows the scatter plot of the proposed metric. We compared the results of the proposed evaluation method with six algorithms. The six algorithms are implemented and employed by the authors of the database for the purpose of providing a benchmark performance in the database. The performance of each algorithm under each distortion type with the database is listed in Table 2. From the table, we can see that none of the algorithms achieves the best performance under each individual type of distortions. The proposed algorithm outperforms the others' under the distortions of JP2K and FF, the Akhter's algorithm achieves the best performance under the distortion of WN and JPEG and the Gorley's algorithm is good at dealing with the distortion of Blur. However, the proposed metric is the best algorithm in terms of the overall performance. The values of the CC, SROCC and RMSE of the proposed algorithm are 0.9233, 0.9162 and 6.3032, respectively which along with the scatter plot show that the proposed metric can predict the quality of stereoscopic images well.

CONCLUSION

In this study, a new objective image quality metric for stereoscopic images is proposed. The main contribution of the proposed method is that it presents a solution to the problems existing in the field of stereoscopic image quality assessment. It not only takes stereoscopic factors into consideration, but also avoids weighting the image quality and the depth perception which are considered the two indispensable parts of the currently typical methods. In the metric, factors affecting stereoscopic perception are first combined to construct a three dimensional signal. Then, the coefficients of this signal are calculated in the 3D-DCT space. Next, Comparisons between reference and distorted stereoscopic images are made in terms of three important coefficients after perceptual weighting. Finally, we get the objective assessment values of distorted stereoscopic images.

The proposed metric has been validated using two stereoscopic image databases, including a publicly database. It is found to perform better than the relevant metrics used as benchmark on the public database. The results of the validation on the two databases show that the proposed metric can predict the quality of the stereoscopic image accurately. In the future work, we will consider extending this metric to stereoscopic videos quality assessment to further explore 3D quality assessment.

ACKNOWLEDEGMENTS

The research work was supported by National Natural Science Foundation of China under Grant Nos. 61271270, 61171163, 61271021, 61111140392.

REFERENCES

Akhter, R., Z.M. Parvez Sazzad, Y. Horita and J. Baltes, 2010. No-reference stereoscopic image quality assessment. Proceedings of the SPIE 21st Conference on Stereoscopic Displays and Applications, January 17, 2010, San Jose, CA., USA.

Brunnstrom, K., D. Hands, F. Speranza and A. Webster, 2009. VQEG validation and ITU standardization of objective perceptual video quality metrics. IEEE Signal Process. Mag., 26: 96-101.

Gorley, P. and N. Holliman, 2008. Stereoscopic image quality metrics and compression. Proceedings of the SPIE 19th Conference on Stereoscopic Displays and Applications, January 27, 2008, San Jose, CA., USA.

Hewage, C.T.E.R. and M.G. Martini, 2010. Reduced-reference quality metric for 3D depth map transmission. Proceedings of Conference on the True Vision-Capture, Transmission and Display of 3D Video, June 7-9, 2010, Tampere, Finland, pp: 1-4.

Lambooij, M., W. IJsselsteijn, D.G. Bouwhuis and I. Heynderickx, 2011. Evaluation of stereoscopic images: Beyond 2D quality. IEEE Trans. Broadcast, 57: 432-444.

Li, X., A. Dick, C. Shen, A.V.D. Hengel and H. Wang, 2013. Incremental learning of 3d-dct compact representations for robust visual tracking. IEEE Trans. Pattern Anal., 35: 863-881.

Ma, L., C. Deng, N.K. Ngi and W. Lin, 2013. Recent advances and challenges of visual signal quality assessment. China Commun., 10: 62-78.

Moorthy, A.K., C.C. Su, A. Mittal and A.C. Bovik, 2013. Subjective evaluation of stereoscopic image quality. Signal Process. Image Commun., 28: 870-883.

Seo, J., X. Liu, D. Kim and K. Sohn, 2012. An objective video quality metric for compressed stereoscopic video. Circ. Syst. Signal Process., 31: 1089-1107.

Shao, F., W. Lin, S. Gu, G. Jiang and T. Srikanthan, 2013. Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics. IEEE Trans. Image Process., 22: 1940-1953.

Shen, L., J. Yang and Z. Zhang, 2009. Stereo picture quality estimation based on a multiple channel HVS model. Proceedings of the 2nd International Congress on Image and Signal Processing, October 17-19, 2009, Tianjin, China.

Wallace, G.K., 1991. The JPEG still picture compression standard. Commun. ACM, 38: 18-34.

Wang, X., M. Yu, Y. Yang and G. Jiang, 2009. Research on subjective stereoscopic image quality assessment. Proceedings of the SPIE Conference on Multimedia Content Access: Algorithms and Systems III, January 18, 2009, San Jose, CA., USA.

Wang, Z., L. Lu and A.C. Bovik, 2004. Video quality assessment based on structural distortion measurement. Signal Process Image, 19: 121-132.

Yang, J., C. Hou, Y. Zhou, Z. Zhang and J. Guo, 2009. Objective quality assessment method of stereo images. Proceedings of the 3DTV Conference on The True Vision-Capture, Transmission and Display of 3D Video, May 4-6, 2009, Potsdam, Germany, pp: 1-4.

Zhu, Z. and Y. Wang, 2009. Perceptual distortion metric for stereo video quality evaluation. WSEAS Trans. Signal Process., 5: 241-250.