VISUAL QUALITY PREDICTION ON DISTORTED STEREOSCOPIC IMAGES

Ching-Ti Lin[1], *Tsung-Jung Liu*[1], *Kuan-Hsien Liu*[2]

[1]Department of Electrical Engineering, National Chung Hsing University
[2]Department of Computer Science and Information Engineering, National Taichung University of Science and Technology

ABSTRACT

We propose a no-reference 3D image quality assessment (IQA) model that can automatically evaluate stereoscopic images. First, the model extracts statistical features from 2D single-view images (i.e., the stereopair) and their pseudo-disparity (i.e., absolute difference) map. Then the features are used to train an IQA model to predict the image quality score by the regression module of support vector machine (SVM). The model we proposed is tested on LIVE 3D Image Quality Database Phase I, which contains only symmetric-distorted stereoscopic images, and LIVE 3D Image Quality Database Phase II, which contains both symmetric-distorted and asymmetric-distorted stereoscopic images. The experimental results on both LIVE 3D Database Phase I and II show that our proposed model leads to significantly improved performance on quality prediction of stereoscopic images.

Index Terms— Disparity, image quality assessment, no reference, stereoscopic images, 3D images

1. INTRODUCTION

Image quality assessment (IQA) is important in image processing applications. In general, IQA can be divided into two classes [1, 2]: subjective IQA and objective IQA. The subjective IQA method is evaluated by the human subjective sensation and self-experience, and it requires a pre-designed laboratory environment following the viewing test. It can be seen that the evaluation of subjective IQA is very time-consuming and also requires a sufficient number of people to achieve the reasonable results. Rather than subjective IQA, objective IQA method refers to automatic quality assessment using an algorithm to predict the image quality. It is usually divided into three types: full-reference (FR), reduced-reference (RR) and no-reference (NR). The NR or blind IQA method does not perform well because the input only considers the distorted image and without reference available. The FR IQA method requires the distorted image and its associated reference image. RR IQA method is similar to FR IQA method with only one difference, meaning RR method only needs partial information of the reference image.

In this paper, we propose a NR IQA model which calculates the features from 2D single-view images (left-eye image and right-eye image) and the pseudo-disparity image (i.e., the absolute difference between two single-view images). Then we use a support vector regression (SVR) model combined with the above features and corresponding ground truths (i.e., the difference mean opinion score (DMOS)) to predict the quality scores of stereoscopic images. The LIVE 3D Image Quality Database Phase I [3, 4] and Phase II [5] are used to verify the performance of our proposed 3D NR IQA model.

2. PREVIOUS WORK

In these years, many researchers have proposed a number of objective stereoscopic IQA models. These models can also be divided into three groups (FR, RR, and NR) based on the availability of references. Chen *et al.* [6] developed a 3D FR quality assessment (QA) model and stated that modelling binocular effects can yield better performance when predicting the 3D quality scores. Shao *et al.* [7] proposed a 3D FR IQA model using binocular receptive field properties. First, they learn a multi-scale dictionary to compute the sparse feature similarity and global luminance similarity. Then they calculate the sparse energy and sparse complexity and use the binocular combination to form the quality score. You *et al.* [8] proposed a 3D FR IQA model by applying 2D QA algorithms on the stereopair and disparity image. And they showed that applying SSIM on stereopair and MAD on the disparity image have better performance in predicting 3D quality scores.

Chen *et al.* [9] developed a NR binocular IQA model that operates on static stereoscopic images. Their model deploys 2D and 3D features extracted from stereopairs to assess the perceptual quality. The experimental results show their method outperforms the conventional 2D FR IQA algorithm and the 3D FR IQA algorithm on asymmetrically distorted stereopairs. Wang *et al.* [10] proposed a binocular rivalry-inspired multi-scale model to predict the quality score of asymmetrical stereoscopic images from the single-view images. Their method successfully eliminates the prediction bias, leading to significant improvement of the quality prediction for stereoscopic image, without identifying image distortion types.

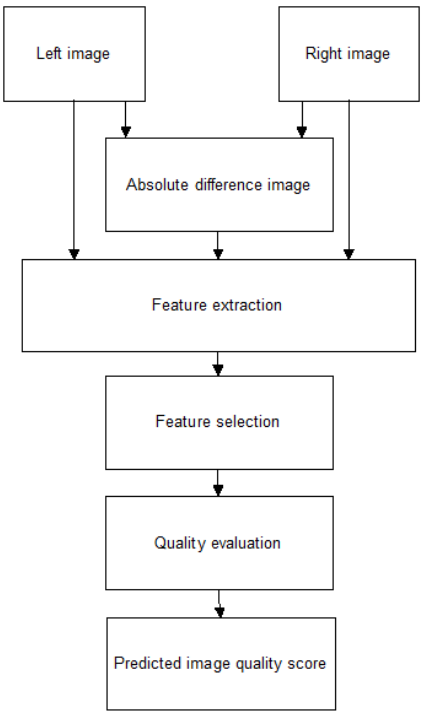

Fig. 1. Block diagram of the proposed 3D NR IQA model

3. PROPOSED BLIND QUALITY ASSESSMENT METHOD FOR STEREOSCOPIC IMAGES

The proposed blind IQA method in Fig. 1 consists of the following steps: 1) feature extraction; 2) feature selection; 3) quality evaluation. The details of these steps are described below:

3.1. Feature extraction

First, the pseudo-disparity (i.e., absolute difference) map between 2D single-view images is computed by:

$$I_D = |I_{\mathrm{L}} - I_{\mathrm{R}}| \tag{1}$$

where I_{L} is the left-view image, and I_{R} is the right-view image. Then, for each stereoscopic image pair (left-view and right-view images) and pseudo-disparity map, we compute the following 15 types of features from them. To simplify the symbol expression, we use D to denote the image set I_L, I_R, I_D.

The 1st feature (f1) is the mean of D, given by:

$$\mu_D = \mathrm{E}\left[\mathrm{D}\right] \tag{2}$$

The 2nd feature (f2) is the standard deviation of D, as given by:

$$\sigma_D = \sqrt{E\left[(D-\mu_D)^2\right]} \tag{3}$$

The 3rd feature (f3) is the kurtosis of D, which is computed by:

$$K_D = \frac{E\left[(D-\mu_D)^4\right]}{\left(E\left[(D-\mu_D)^2\right]\right)^2} \tag{4}$$

The 4th feature (f4) is the skewness of D, which can be computed by:

$$S_D = \frac{E\left[(D-\mu_D)^3\right]}{\left(E\left[(D-\mu_D)^2\right]\right)^{\frac{3}{2}}} \tag{5}$$

For the 5th to 8th features, they are extracted from the differential map of D. The 5th feature (f5) is the differential mean of D [11], which is given by:

$$\mu_{\delta D} = \mathrm{E}\left[\delta \mathrm{D}\right] \tag{6}$$

where δD is computed using a Laplacian operator.

The 6th feature (f6) is the differential standard deviation of D, given by:

$$\sigma_{\delta D} = \sqrt{E\left[(\delta D-\mu_{\delta D})^2\right]} \tag{7}$$

The 7th feature (f7) is the differential kurtosis of D, which is determined by:

$$K_{\delta D} = \frac{E\left[(\delta D-\mu_{\delta D})^4\right]}{\left(E\left[(\delta D-\mu_{\delta D})^2\right]\right)^2} \tag{8}$$

The 8th feature (f8) is the differential skewness of D, given by:

$$S_{\delta D} = \frac{E\left[(\delta D-\mu_{\delta D})^3\right]}{\left(E\left[(\delta D-\mu_{\delta D})^2\right]\right)^{\frac{3}{2}}} \tag{9}$$

The 9th feature (f9) is the mean of gradient magnitude of D, which can be calculated by:

$$\mathrm{E}_{\nabla D} = \mathrm{E}\left[\left|\left[\begin{array}{c}\frac{\partial D}{\partial i}\\ \frac{\partial D}{\partial j}\end{array}\right]\right|\right] \tag{10}$$

where i and j are spatial indices.

The 10th feature (f10) is the local kurtosis of D [12]. We can compute its value by:

$$LK_D(i,j) = \sum_{p=-P}^{p=P}\sum_{q=-Q}^{q=Q} \frac{E\left[(D(i+p,j+q)-\mu_D(i+p,j+q))^4\right]}{\left(E\left[(D(i+p,j+q)-\mu_D(i+p,j+q))^2\right]\right)^2} \tag{11}$$

where i and $j$are spatial indices, P and Q are the window size of filters. Here, we choose $P = Q = 5$.

The 11th feature (f11) is the local power spectrum slope [12] of D. Averaged power spectrum, intuitively, represents the strength of change. Blur attenuates high frequency components and therefore makes the power spectrum fall off much faster than its sharp counterpart.

The 12th feature (f12) is the threshold value of edge detection [13] on D. Edge detection can greatly reduce the amount of image data, and remove irrelevant information, leaving only the image of important architectural properties, such as discontinuity on the depth of field, discontinuity in the direction of the image surface, changes in image attributes, and changes in scene illumination.

The 13th feature (f13) is generalized Gaussian distribution (GGD) parameters, including γ, a, b, which can be described by:

$$f_x\left(x;\mu,\sigma^2,\gamma\right) = \mathrm{a}e^{-[b|x-\mu|]^\gamma}, \tag{12}$$

where μ, σ^2 and γ are mean, variance, and shape parameter of the distribution, respectively. The positive constants a and b are given by:

$$a = \frac{b\gamma}{2\Gamma\left(1/\gamma\right)} \tag{13}$$

and

$$b = \frac{1}{\sigma}\sqrt{\frac{\Gamma\left(3/\gamma\right)}{\Gamma\left(1/\gamma\right)}}, \tag{14}$$

where$\Gamma\left(.\right)$ is a gamma function:

$$\Gamma\left(\mathrm{x}\right) = \int_0^\infty t^{x-1}e^{-t}dt, x > 0. \tag{15}$$

The parameter γ is estimated using the method in [14].

The 14th feature (f14) is the blockiness value [13]. The blocking effect is generated by the excessive block coding principle. The coding is applied to the pixels of a block, and the transform coefficients of each block are quantized in order to achieve destructive data compression. When the lower bit rate is used, the more the coefficients are coarsely stored and are quantized to zeroes. Because the process of quantization is applied separately to each block, the quantization of adjacent block coefficients is different. This phenomenon causes discontinuities in the boundaries of the block and becomes most pronounced in areas where the color change is gentle, since there are fewer image contents that can cover this phenomenon.

The 15th feature (f15) is the eigenvalues of D. We use singular value decomposition (SVD) [15, 16] to calculate eigenvalues from the image set D.

To summarize, the features used in this work are derived from commonly-used statistical properties, which are simple and can reduce the computation cost of our proposed system.

3.2. Feature selection

Since the size of these computed features are large compared to the size of the dataset, over-fitting is a possibility. Hence, we use forward feature selection (FFS) [17] to reduce the feature dimension. First we choose the feature that correlates the best with perceptual scores on the training set; then, we find the next feature which combines with the chosen feature in previous step and the combined one has the best correlation with perceptual scores. This process continues until the increment of correlation coefficients is less than 0.001. The step-by-step selection result is shown in Fig. 2. Thus, the final selected features are listed by the order as follows: {f15, f14, f9, f4, f5, f12, f6, f7, f8}.

3.3. Quality evaluation

In order to map the feature onto the perceptual quality scores, we use a support vector regression (SVR) model. SVR has previously been applied to solve many problems, such as image quality assessment [18, 19, 20], and facial image processing [21, 22, 23]. In the training stage, the selected features are combined into the feature vector, and then the feature vector of each image is mapped to the corresponding subjective score by the regression module of LIBSVM [24]. Then, in the testing stage, this trained mapping model is used to generate the predicted quality score of the test image. Here, the linear kernel is used in SVR for all conducted experiments because of its low computations.

4. EXPERIMENTAL RESULTS

Databases with subjective data (i.e., the ground truth of image quality) facilitate the development of objective metrics. There are 2 stereoscopic image databases that are publicly available and commonly recognized in the 3D IQA research, including LIVE 3D Image Quality Database Phase I [3, 4]and LIVE 3D Image Quality Database Phase II [5]. Phase I contains 365 images which have symmetrical distortions while Phase II has 360 images which have symmetrical and asymmetrical distortions. Here, if left and right view images both have the same distortion levels, then it will be considered as symmetrical distorted stereopair, while it will be considered as asymmetrical distorted stereopair if left and right view images both have different distortion levels. In addition, Phase I and Phase II both have the same distortion types: JPEG, JPEG2000 (JP2K) compressions, additive white Gaussian noise (WN), Gaussian blur (Blur) and fast-fading (FF) model based on the Rayleigh fading channel.

We use the LIVE 3D Image Quality Database Phase I and Phase II to validate the performance of our proposed model. Because this method requires a training procedure to calibrate the regression module, the LIVE 3D Image Quality Database Phase I and Phase II are divided into two randomly selected subsets (i.e., 80% for the training set and 20% for the testing set). The above step along with the cross-validation can avoid both the overlap between the training set and the testing set, and over-fitting phenomenon. In addition, we execute this random split of data for 1000 times to ensure that the experimental results do not rely on the characteristics extracted from the known image contents (i.e., scenes) in order to avoid artificially improved performance, and the median results are reported as the final performance.

Two performance indicators (Pearson linear correlation coefficient (PLCC) [25] and Spearman rank ordered correlation coefficient (SROCC)) are used to evaluate the performance of proposed model with respect to the similarity between predicted scores and subjective scores. The PLCC and SROCC measure the accuracy, monotonicity between the predictive scores and the subjective scores. When PLCC and SROCC are closer to 1, the correlation between the assessment scores and the true scores is better. The performance results of our proposed model are listed in Tables 1-4, and the best results are shown in bold. In the proposed model, we

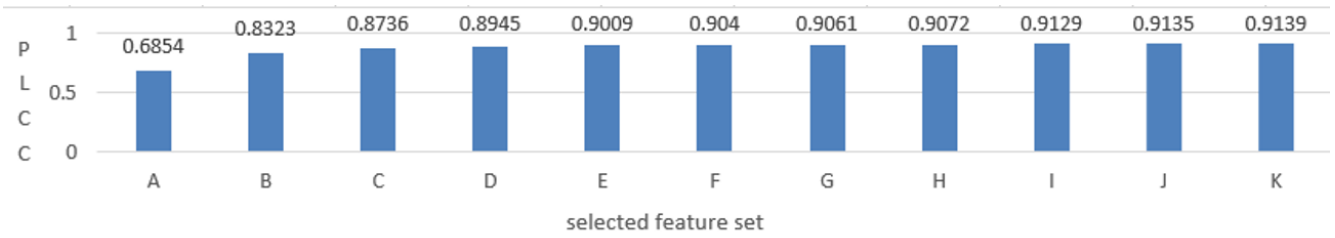

Fig. 2. The FFS result. Set A includes f15. Set B includes Set A and f14. Set C includes Set B and f9. Set D includes Set C and f4. Set E includes Set D and f5. Set F includes Set E and f12. Set G includes Set F and f6. Set H includes Set G and f7. Set I includes Set H and f8. Set J includes Set I and f3. Set K includes Set J and f13.

Table 1. Median SROCC across 1000 Train-Test Combinations on the LIVE 3D IQA Database Phase I

Model	JPEG	JP2K	WN	FF	Blur	ALL
Chen [6]	0.582	0.896	**0.948**	0.688	**0.926**	0.916
Shao [7]	0.495	0.895	0.941	**0.940**	0.796	0.925
You [8]	0.439	0.860	0.940	0.588	0.882	0.878
Chen [9]	0.617	0.863	0.919	0.652	0.878	0.891
Proposed model	**0.772**	**0.902**	0.929	0.820	0.903	**0.937**

Table 2. Median PLCC across 1000 Train-Test Combinations on the LIVE 3D IQA Database Phase I

Model	JPEG	JP2K	WN	FF	Blur	ALL
Chen[6]	0.634	0.916	0.944	0.758	0.942	0.917
Shao[7]	0.520	0.921	**0.945**	0.859	**0.959**	0.935
You[8]	0.487	0.877	0.941	0.730	0.919	0.881
Chen[9]	0.695	0.907	0.917	0.735	0.917	0.895
Proposed model	**0.761**	**0.934**	0.933	**0.861**	0.932	**0.942**

do not need any reference images. We only have to consider the distorted images by taking the information of the pseudo-disparity map and 2D single-view images as inputs, and then compute the feature vectors. Finally, LIBSVM [24] is used to train these features to predict the image quality scores.

In Tables 1-2, SROCC and PLCC are calculated from the predicted and perceptual quality scores from the symmetrical distortion images, respectively. As shown in Tables 1-2, our method has the best PLCC and SROCC values compared to the other four IQA methods in JPEG and JP2K distortion types, and the overall performance results are also ranked the first comparing to the other four methods. In Tables 3-4, we calculate the SROCC and PLCC between the predicted scores and subjective scores for our proposed model and other compared IQA methods. Our method has the 2nd best performance for symmetrically and asymmetrically distorted images among all IQA models. Therefore, it can be concluded that the proposed method works well for both symmetrical and asymmetrical distortion types in stereoscopic images. Furthermore, in Table 5, we compare the performance results under four different train-test proportions. 50%-50%, 60%-40%, 70%-30% and 80%-20% of samples are used for training and testing, respectively. This process is also repeated 1000 times for random train-test split

Table 3. Median SROCC across 1000 Train-Test Combinations on the LIVE 3D IQA Database Phase II

Model	JPEG	JP2K	WN	FF	Blur	ALL
Chen[6]	0.834	0.814	0.940	0.884	0.908	0.889
Shao[7]	0.733	0.785	**0.965**	0.891	0.920	0.849
You[8]	0.795	0.894	0.909	0.891	0.813	0.786
Chen[9]	**0.867**	0.867	0.950	**0.933**	0.900	0.880
Wang[10]	N/A	N/A	N/A	N/A	N/A	**0.919**
Proposed model	0.774	**0.897**	0.903	0.881	**0.922**	0.906

Table 4. Median PLCC across 1000 Train-Test Combinations on the LIVE 3D IQA Database Phase II

Model	JPEG	JP2K	WN	FF	Blur	ALL
Chen[6]	0.862	0.834	**0.957**	0.901	0.963	0.900
Shao[7]	0.747	0.782	0.946	0.905	0.958	0.863
You[8]	0.830	**0.905**	0.912	0.915	0.784	0.800
Chen[9]	**0.901**	0.899	0.947	**0.932**	0.941	0.895
Wang[10]	N/A	N/A	N/A	N/A	N/A	**0.916**
Proposed model	0.812	0.901	0.895	0.896	**0.979**	0.913

combinations, and the median results are reported as the final performance. The results show the performance of the proposed system still stays the same even though the training ratio decreases to 50%.

5. CONCLUSION

In this work, we propose a no-reference (i.e., blind) stereoscopic image quality assessment model based on the statistical features extracted from 2D single-view images and the pseudo-disparity (absolute difference) image. After generating the features, we combine the features with SVM regression model to evaluate the quality of the test images to obtain predicted quality scores. Experimental results in two authoritative databases show that our proposed model achieves state-of-the-art performance on prediction accuracy.

Table 5. Performance Results for 4 Different Train-Test Proportions of Images in 2 Databases

Databases	Train-Test (%)	50-50	60-40	70-30	80-20
LIVE-I	PLCC	0.937	0.940	0.942	0.942
	SROCC	0.933	0.935	0.938	0.937
LIVE-II	PLCC	0.904	0.909	0.913	0.913
	SROCC	0.899	0.904	0.907	0.906

6. REFERENCES

[1] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "Recent developments and future trends in visual quality assessment," in *Proceedings of Asia-Pacific Signal and Information Processing Association Annual Submit and Conference*, pp. 18–21, 2011.

[2] T.-J. Liu, Y.-C. Lin, W. Lin, and C.-C. J. Kuo, "Visual quality assessment: recent developments, coding applications and future trends," *APSIPA Transactions on Signal and Information Processing*, vol. 2, e4 2013.

[3] "LIVE 3D Image Quality database phase I. [online]. Available:." `http://live.ece.utexas.edu/research/quality/live_3dimage_phase1.html`.

[4] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 870–883, 2013.

[5] "LIVE 3D Image Quality database phase II. [online]. Available:." `http://live.ece.utexas.edu/research/quality/live_3dimage_phase2.html1`.

[6] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, 2013.

[7] F. Shao, K. Li, W. Lin, G. Jiang, M. Yu, and Q. Dai, "Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 2971–2983, 2015.

[8] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2d image quality metrics and disparity analysis," *Proc. Int. Workshop Video Process. Qaulity Metrics Consum. Elect*, pp. 1–6, 2010.

[9] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3379–3391, 2013.

[10] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3d images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3400–3414, 2015.

[11] A. Mittal, A. K. Moorthy, J. Ghosh, and A. C. Bovik, "Algorithmic assessment of 3d quality of experience for images and videos," in *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE*, pp. 338–343, IEEE, 2011.

[12] J. Shi, L. Xu, and J. Jia, "Discriminative blur detection features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2965–2972, 2014.

[13] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, pp. 591–602, April 2014.

[14] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, pp. 52–56, Feb 1995.

[15] X. Liu, L. Zhang, and K. Lu, "A 3d image quality assessment method based on vector information and svd of quaternion matrix under cloud computing environment," *IEEE Transactions on Cloud Computing*, 2015.

[16] H. Andrews and C. Patterson, "Singular value decompositions and digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, pp. 26–53, Feb 1976.

[17] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *j. mach. learn. res*, vol. 3, pp. 1157–1182, March 2003.

[18] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "A multi-metric fusion approach to visual quality assessment," in *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, pp. 72–77, IEEE, 2011.

[19] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "Image quality assessment using multi-method fusion," *Image Processing, IEEE Transactions on*, vol. 22, no. 5, pp. 1793–1807, 2013.

[20] T.-J. Liu, K.-H. Liu, J. Y. Lin, W. Lin, and C.-C. J. Kuo, "A paraboost method to image quality assessment," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 1, pp. 107–121, 2017.

[21] K.-H. Liu, S. Yan, and C.-C. J. Kuo, "Age group classification via structured fusion of uncertainty-driven shape features and selected surface features," in *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*, pp. 445–452, IEEE, 2014.

[22] K.-H. Liu, T.-J. Liu, H.-H. Liu, and S.-C. Pei, "Facial makeup detection via selected gradient orientation of entropy information," in *Image Processing (ICIP), 2015 IEEE International Conference on*, pp. 4067–4071, IEEE, 2015.

[23] T.-J. Liu, K.-H. Liu, H.-H. Liu, and S.-C. Pei, "Age estimation via fusion of multiple binary age grouping systems," in *Image Processing (ICIP), 2016 IEEE International Conference on*, pp. 609–613, IEEE, 2016.

[24] C. C. Chang and C. J. Lin, "Libsvm: A Library for Support Vector Machines [online]. Available:," 2001. `http://www.csie.ntu.edu.tw/cjlin/libsvm/`.

[25] T.-J. Liu, K.-H. Liu, H.-H. Liu, and S.-C. Pei, "Comparison of subjective viewing test methods for image quality assessment," in *Image Processing (ICIP), 2015 IEEE International Conference on*, pp. 3155–3159, IEEE, 2015.