IMAGE QUALITY ASSESSMENT BASED ON STRUCTURE VARIANCE CLASSIFICATION

Yibing Zhan[1,2] and Rong Zhang[1,2]

[1]Dept. of Electronic Engineering and Information Science, Univ. of Sci. and Tech. of China, Hefei, China
[2]Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Hefei, China

ABSTRACT

In this paper, we find that the structure variance of images could be divided into four classifications, slight deformations, additive impairments, detail losses, and confusing contents, and what's more, for each classification, subjective evaluation is different. According this, we propose a novel image quality assessment (IQA) method based on structure variance classification. The proposed method classifies the structure variance of each patch into one of the four classifications using binary logic and then summarizes the areas of different classifications. To get more comprehensive evaluation, the proposed method also incorporates the measurements of differences between extracted features. Our method is tested on five public databases and compared with seven state-of-art methods. The experimental results demonstrate that our method can achieve higher consistency in relation to the subjective evaluation compared to the state-of-art IQA methods.

***Index Terms*—**Image Quality Assessment (IQA), Structure Variance Classification, Binary Logic

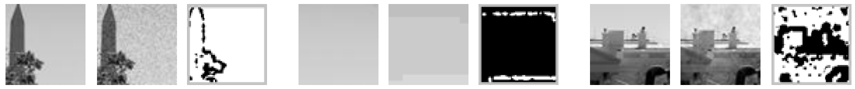
1. INTRODUCTION

In the past few decades, numerous objective IQA methods, especially the full reference IQA (FR-IQA) methods [1], have been put forward for efficient evaluation of image quality. Most FR-IQA methods own their unique models to handle image distortion and to achieve high consistency with the subjective scores. The simplest FR-IQA methods, mean-squared error (MSE) and peak signal-to-noise ratio (PSNR) compare the intensity of images. Structure similarity index (SSIM) [2] predicts images' fidelity by calculating the loss of image structure, luminance and contrast. The improvement methods of SSIM, such as Multi-scale SSIM (MS-SSIM) [3] and Information content weighting SSIM (IW-SSIM) [4], are proposed according to the characteristics of human visual system (HVS) [1]. Visual information fidelity (VIF) [5] regards IQA as an information transfer process. Most apparent distortion (MAD) [6] divides image into high-quality and low-quality situations with different treatments to obtain the distortion. Detail Losses and Additive Impairments (DLAI) [7] separates the structure variance of images into detail losses and additive impairments to achieve the final evaluation.

In terms of image quality, structure information is far more important than gray/color-intensity information and is the major consideration of most advanced IQA methods. However, most of the methods get the structure distortion by directly subtracting structure feature values or calculating the correlation coefficient of the values without analyzing the details of the structure variances. But natural images, which contain rich structure information, are not precise numerical storage in computers. People convert images to various structural variances and analyze these variances to evaluate the image quality. Thereby, the model used in DLAI attracts our attention, for it discusses the structure variance from the point of human understanding. However, when collecting the structure variances of images and using the model to classify them, we find that parts of the structure variances would not be proper to be labeled as detail losses or additive impairments visually. Therefore, in this paper, we propose a more complete model, which divides the structure variances into four classifications, slight deformations, additive impairments, detail losses, and confusing contents. And to realize the model, we introduce the strategy of binary logic to analyze the structural features obtained from images to count the degree of the visual information changes, and all the pixels in images can be classified into one of the four classifications. Then, by calculating the areas of different classifications and the differences of the extracted features, a novel IQA method has been proposed. The results on five public databases show that the method leads to a promising assessment performance.

2. THE BASIC MODEL

The model in DLAI that the structure variance could be classified into detail losses and additive impairments really attracts our attention. To prove the reliability, we conducted an experiment requiring ten people to classify the patches owning structure variances into the two classifications. The patches were extracted from the image "1600" of CSIQ[6].



(a)Examples of additive impairments

(b)Examples of detail losses

(c)Examples of slight deformations

Fig. 1. Examples of the patches used in the experiments and their classifications. Each groups of (a), (b) and (c) has three kind images. The left and middle images are original and distorted images for each. The right images are the map using structure component of SSIM, in which the white parts means the value of the pixel is lower than 0.9.

Some patches could be described by the model perfectly (Fig.1(a) and (b)), whereas, some patches (Fig.1(c)) could not be labeled well.

According to the appearance of structure variances, we propose a more complete model which divides the structure variances into four classifications: slight deformations, additive impairments, detail losses, and confusing contents. Slight deformations (Fig.1(c)) refer to the similar visual changes between original and distorted images. Detail losses (Fig.1(b)) and additive impairments (Fig.1(a)) refer to the loss of visual information and redundant visual information respectively. Confusing contents (Fig.2) refer to the structure information of the original and distorted images is unrelated, in case that there exist situations that could not be applied into the three classifications mentioned above.

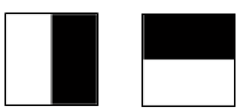


Fig. 2. Examples of confusing contents. The structure variance of the two pictures is improper to be described by the other three classifications.

To prove the validity, another experiment was conducted. People were asked to classify the same patches of the first experiment (Fig.1) using the new model. This time we gladly found that all the patches could be classified well. What's more, the results show that there exist some relationships between the classifications and the subjective evaluation (Fig.3). The patches from the images with high quality were intended to be classified into slight deformations. Most of the detail losses patches were extracted from the images with bad quality and most of the patches labeled as additive impairments were from the images of middle quality. Few patches were labeled as confusing contents. It could be concluded that the quality of detail losses is the lowest, for the original information is missing; the quality of the additive impairments is the middle, for the noise do not hurt the original contents; the quality of the slight deformations is the highest, for the original information is nearly not affected.

Fig. 3. Structure variances with different DMOS [1]. From left to right, the images are original, slight deformation, additive impairments, and detail losses. The DMOS of the distorted images are 0.206, 0.467, and 0.750.

3. THE BINARY LOGIC

To realize the classification process, we utilize a binary logic strategy, which mimics the threshold behavior of neuronal characteristic [8]-[10]. The mathematical expression of BL is defined as

$$lf = \begin{cases} 1 & f > \mathrm{t} \\ F(f) & f \leq \mathrm{t} \end{cases} \tag{1}$$

where f represents the feature. lf represents the logic feature, which depicts the possibility whether the related feature exist. t represents the threshold. F represents a membership function, which establishes a transition zone between 0 and t. If f is greater than t, lf is set as 1, which means the extracted feature exists. If f is lower than or equal to t, lf is set as a value between 0 and 1 to reflect the existence probability of the extracted feature.

By comparing the logic features from (1), we can obtain the judgment whether the shared feature exists in both of the original and distorted image. The process is given as:

$$j = \begin{cases} 1 & lf^o - lf^d > 0.5 \\ -1 & lf^o - lf^d < -0.5 \\ 0 & \text{else} \end{cases} \tag{2}$$

where j represents the judgment. o and d represent the original and distorted image. "1" depicts the distorted image loses the structure information that belong to the original image. "-1" means the distorted image adds some structure information that doesn't exist in the original image and do not hurt the original structure information. "0" reflects the extracted feature exists in both images.

The process which divides the structure variance into the four classifications could be completed by analyzing numerous judgments made by different structural features from the binary logic. Firstly, these judgments are counted separately according to their values, which are defined as:

$$v_a = \sum_{j_k=-1}^{k} 1, \quad v_l = \sum_{j_k=1}^{k} 1, \quad v_u = \sum_{j_k=0}^{k} 1 \tag{3}$$

where k represents the k-th feature. v_a, v_l , and v_u represent the votes for distorted images with additive, losing, and unchanged structure information for each.

Next, by analyzing v_a, v_l, and v_u, we can make the conclusion which classification the structure variance belongs to. The process is given as follows:

$$r = \begin{cases} \text{no distortion} & v_u = \mathrm{n} \\ \text{slight} & \mathrm{n} > v_u > \mathrm{t_s} \\ \text{additive} & v_a > \mathrm{t_d},\ v_l < \mathrm{t_u} \\ \text{losses} & v_l > \mathrm{t_d},\ v_a < \mathrm{t_u} \\ \text{confusing} & \text{else} \end{cases} \tag{4}$$

where r represents the conclusion. n is the total number of the utilized features. t_s, t_d, and t_u are the thresholds and are adjusted according to n. If v_u equals to n, it means the place has no distortion. If v_u is higher than t_s and lower than n, it means the majority votes for unchanged structure information and the place is slight deformations. If v_a is higher than t_d and v_l is lower than t_u, it means the majority votes for distorted images with additional structure information and the place is labeled as additive impairments. If v_l is higher than t_d and v_a is lower than t_u, it means the majority votes for distorted images losing some structure information and detail losses are the majority choices. The rest situations are labeled as confusing contents.

4. THE PROPOSED METHOD

Our proposed method evaluates the image distortion from two aspects: areas of structure variance classification and differences between extracted features. Remainder of this part gives the details of the proposed method.

4.1. Areas of structure variance classifications

Two kinds of features (Fig.4), Laws' mask and texture gradient features [11]-[13] are used in our method. We convolve the original and the distorted images with the filter banks as:

$$f_n = | I * Flter_n |, \ n = 1, 2, \cdots, 14 \tag{5}$$

By using (1), the extracted features are transformed into logic features. The membership function used is as follows.

$$\begin{bmatrix} -0.0156 & 0 & 0.0156 \\ -0.0313 & 0 & 0.0313 \\ -0.0156 & 0 & 0.0156 \end{bmatrix} \quad \begin{bmatrix} -0.0156 & -0.0313 & -0.0156 \\ 0 & 0 & 0 \\ 0.0156 & 0.0313 & 0.0156 \end{bmatrix} \quad \begin{bmatrix} 0.0156 & 0 & -0.0156 \\ 0 & 0 & 0 \\ -0.0156 & 0 & 0.0156 \end{bmatrix}$$

$$\begin{bmatrix} 0.0156 & -0.0313 & 0.0156 \\ 0 & 0 & 0 \\ -0.0156 & 0.0313 & -0.0156 \end{bmatrix} \quad \begin{bmatrix} 0.0156 & 0 & -0.0156 \\ -0.0313 & 0 & 0.0313 \\ 0.0156 & 0 & -0.0156 \end{bmatrix} \quad \begin{bmatrix} 0.0156 & -0.0313 & 0.0156 \\ -0.0313 & 0.0625 & -0.0313 \\ 0.0156 & -0.0313 & 0.0156 \end{bmatrix}$$

$$\begin{bmatrix} -0.0156 & 0.0313 & -0.0156 \\ -0.0313 & 0.0625 & -0.0313 \\ -0.0156 & 0.0313 & -0.0156 \end{bmatrix} \quad \begin{bmatrix} -0.0156 & -0.0313 & -0.0156 \\ 0.0313 & 0.0625 & 0.0313 \\ -0.0156 & -0.0313 & -0.0156 \end{bmatrix}$$

(a) Law's mask filters

$$\begin{bmatrix} -0.05 & -0.05 & 0 & 0.05 & 0.05 \\ -0.05 & -0.05 & 0 & 0.05 & 0.05 \\ -0.05 & -0.05 & 0 & 0.05 & 0.05 \\ -0.05 & -0.05 & 0 & 0.05 & 0.05 \\ -0.05 & -0.05 & 0 & 0.05 & 0.05 \end{bmatrix} \quad \begin{bmatrix} 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\ 0 & 0 & 0 & 0 & 0 \\ -0.05 & -0.05 & -0.05 & -0.05 & -0.05 \\ -0.05 & -0.05 & -0.05 & -0.05 & -0.05 \end{bmatrix}$$

$$\begin{bmatrix} -0.0454 & 0.145 & 0.0454 & 0.0454 & 0.0454 \\ -0.0454 & -0.0354 & 0.0417 & 0.0454 & 0.0454 \\ -0.0454 & -0.0454 & 0 & 0.0454 & 0.0454 \\ -0.0454 & -0.0454 & -0.0417 & 0.0354 & 0.0454 \\ -0.0454 & -0.0454 & -0.0454 & -0.0145 & 0.0454 \end{bmatrix} \quad \begin{bmatrix} 0.0454 & 0.0454 & 0.0454 & 0.0145 & -0.0454 \\ 0.0454 & 0.0454 & 0.0417 & -0.0354 & -0.0454 \\ 0.0454 & 0.0454 & 0 & -0.0454 & -0.0454 \\ 0.0454 & 0.0354 & -0.0417 & -0.0454 & -0.0454 \\ 0.0454 & -0.0145 & -0.0454 & -0.0454 & -0.0454 \end{bmatrix}$$

$$\begin{bmatrix} 0.0454 & 0.0454 & 0.0454 & 0.0454 & 0.0454 \\ 0.0454 & 0.0454 & 0.0454 & 0.0354 & -0.0145 \\ 0.0454 & 0.0417 & 0 & -0.0417 & -0.0454 \\ 0.0145 & -0.0354 & -0.0454 & -0.0454 & -0.0454 \\ -0.0454 & -0.0454 & -0.0454 & -0.0454 & -0.0454 \end{bmatrix} \quad \begin{bmatrix} 0.0454 & 0.0454 & 0.0454 & 0.0454 & 0.0454 \\ -0.0145 & 0.0354 & 0.0454 & 0.0454 & 0.0454 \\ -0.0454 & -0.0417 & 0 & 0.0417 & 0.0454 \\ -0.0454 & -0.0454 & -0.0454 & -0.0354 & 0.0145 \\ -0.0454 & -0.0454 & -0.0454 & -0.0454 & -0.0454 \end{bmatrix}$$

(b) Gradient filters

Fig. 4. Feature filters.

$$F(f) = 2 / (1 + e^{(-f*6)}) - 1 \tag{6}$$

For (6) has an upper bound, there is no need for us to set the threshold in (1), which changes as:

$$lf_n = F(f_n) \tag{7}$$

Then, by using (2) to compare the logic features from (7), the judgments are obtained. These judgments are analyzed by (3)-(4) to divide the structure distortion into different structure variance classifications.

After getting the structure variance classifications, the areas of these classifications are calculated individually as:

$$A_r = \sum_r 1, r = \text{slight}, \text{additive}, \text{losses}, \text{confusing} \tag{8}$$

where A_r represents the areas. Then, by combining these areas with different weights, the area result is obtained.

$$S = \sum \frac{A_r * \alpha_r}{m * n} \tag{9}$$

Here, S is the result and α_r represents the weights for different classifications. $m*n$ is the whole image area.

Additionally, a multi-scale strategy, which follows MS-SSIM, is introduced. The proposed method resizes an image 4 times, getting same content images with 5 different sizes. To be more specific, each time the resized image halves the original one. The 'imresize' function and 'bicubic' method in MATLAB are used to do the resizing process. The final result is improved as：

$$S_{mlt} = \sum (S^i * \beta^i) \tag{10}$$

where S_{mlt} is the final results. S^i means the area result in i-th scale. β^i is the i-th scale weight. The values are $\beta^i = [0.0448, 0.2856, 0.3001, 0.2363, 0.1333]$,which have been used in MS-SSIM.

4.2. Differences between extracted features

We use the texture gradient features (Fig.4(b)) to evaluate the differences, which is defined as

$$d(i,j) = \sum_n (f_n^o(i,j) - f_n^d(i,j))^2, n = 1, 2, \cdots, 6 \tag{11}$$

TABLE I Performance comparison of IQA methods on databases

Database	criteria	PSNR	SSIM	MS-SSIM	IW-SSIM	VIF	MAD	DLAI	Our
CSIQ	PLCC	0.800	0.804	0.867	0.914	0.928	**0.950**	**0.928**	**0.938**
	RMSE	0.158	0.157	0.131	0.107	0.098	**0.082**	**0.098**	**0.091**
	SROCC	0.806	0.820	0.877	0.921	0.920	**0.946**	**0.933**	**0.934**
IVC	PLCC	0.703	0.912	0.911	**0.923**	0.903	**0.921**	0.913	**0.919**
	RMSE	0.878	0.500	0.503	**0.469**	0.524	**0.481**	0.496	**0.485**
	SROCC	0.691	0.902	0.898	**0.913**	0.896	**0.915**	0.903	**0.908**
LIVE	PLCC	0.872	0.945	0.949	0.952	**0.960**	**0.968**	0.936	**0.960**
	RMSE	13.402	8.95	8.619	8.375	**7.667**	**6.929**	9.627	**7.673**
	SROCC	0.876	0.948	0.951	0.957	**0.963**	**0.967**	0.946	**0.962**
MICT	PLCC	0.635	0.889	0.893	**0.925**	0.914	**0.941**	**0.942**	0.917
	RMSE	0.979	0.574	0.564	**0.476**	0.508	**0.431**	**0.421**	0.503
	SROCC	0.613	0.880	0.887	**0.920**	0.908	**0.936**	**0.937**	0.917
TID 2008	PLCC	0.573	0.773	0.845	**0.858**	0.809	0.831	**0.869**	**0.869**
	RMSE	1.101	0.851	0.717	**0.690**	0.789	0.747	**0.662**	**0.665**
	SROCC	0.553	0.775	0.854	**0.856**	0.750	0.834	**0.861**	**0.851**
Average	PLCC	0.717	0.865	0.89	0.91	0.903	**0.922**	**0.918**	**0.921**
	SROCC	0.708	0.865	0.893	0.913	0.887	**0.920**	**0.916**	**0.914**

where $d(i,j)$ represents the difference at pixel (i,j).

The differences of the gradient features of images are measured as

$$D = \sqrt{\sum_{i,j} d(i,j)} \tag{12}$$

where D is the difference result.

4.3. Final results

The final evaluation is calculated by the combination of the areas (S_{mlt}) and the differences assessment (D) as:

$$Result = S_{mlt} * D \tag{13}$$

5. THE RESULTS

5.1. Determination of parameters

The parameter, α_r is needed to be determined. To this end, we tuned the parameters based on A57 [14] database. The parameters value leading to a higher SROCC would be chosen. As a result, the parameters were set as: $\alpha_r = [0.5, 3.5, 9.0, 3.0]$. The value of α_r also reflects that the conclusion made in section 2 is correct. To achieve proper structure variance classifications, the parameters in (4) are: $\mathrm{n} = 14$, $\mathrm{t_s} = 10$, $\mathrm{t_d} = 2$, and $\mathrm{t_u} = 2$.

5.2. Databases and performance metrics

Five publicly IQA databases, MICT [15], IVC [16], LIVE [17], CSIQ [6], and TID2008 [18] are used for algorithm validation and comparison. Three performance metrics, Pearson linear coefficient (PLCC), root mean square error (RMSE), and Spearman rank order correlation (SROCC), are used to evaluate the methods [17]. To compute the PLCC and RMSE, we need to apply a regression analysis, whose mapping function is as follows:

$$f(x) = a_1 * (0.5 - 1/(1 + e^{a2*(x-a3)})) + a_4 * x + a_5 \tag{14}$$

where a_i are the parameters to be fitted. A higher value of SROCC, PLCC and a lower value of RMSE means a better objective method.

5.3. The performance of our method and comparison with other algorithms

In this section, the performances of our method are compared with seven state-of-art IQA methods. The IQA methods are PSNR, SSIM [2], MS-SSIM [3], IW-SSIM [4], VIF [5], MAD [6], and DLAI [7]. The results of SROCC, PLCC, and RMSE of the methods are listed in Table I. For each performance measure, the three IQA methods producing the best results are highlighted in boldface.

In Table I, it can be seen that our method performs consistently well across all the databases. The proposed method ranks the top three in almost all five databases. Moreover, for the average scores, our method gets great performance. According to the results, our method has good robustness and universality, showing more compliance with human perception.

6. CONCLUSION

This paper proposes a model that the structure variance could be divided into four classifications and a novel IQA method based on the structure variance classification. The results on the public IQA databases show our method leads to a promising assessment performance.

7. ACKNOWLEDGENENT

This work is supported by National Nature Science Foundation of China under Grant No.61331020.

8. REFERENCES

[1]Wang, Zhou, and Alan C. Bovik. "Modern image quality assessment." *Synthesis Lectures on Image, Video, and Multimedia Processing* 2.1 (2006): 1-156.

[2]Wang, Zhou, et al. "Image quality assessment: From error visibility to structural similarity." *Image Processing, IEEE Transactions on* 13.4 (2004): 600-612.

[3]Wang, Zhou, Eero P. Simoncelli, and Alan C. Bovik. "Multiscale structural similarity for image quality assessment." *Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on*. Vol. 2. IEEE, 2003.

[4]Wang, Zhou, and Qiang Li. "Information content weighting for perceptual image quality assessment." *Image Processing, IEEE Transactions on* 20.5 (2011): 1185-1198.

[5]Sheikh, Hamid R., and Alan C. Bovik. "Image information and visual quality." *Image Processing, IEEE Transactions on* 15.2 (2006): 430-444.

[6]Larson, Eric C., and Damon M. Chandler. "Most apparent distortion: full-reference image quality assessment and the role of strategy." *Journal of Electronic Imaging* 19.1 (2010): 011006-011006.

[7]Li S, Zhang F, Ma L, et al. Image quality assessment by separately evaluating detail losses and additive impairments[J]. Multimedia, IEEE Transactions on, 2011, 13(5): 935-949.

[8]Mathis, J., D. De Quervain, and C. W. Hess. "Dependence of the transcranially induced silent period on theinstruction set'and the individual reaction time." *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control* 109.5 (1998): 426-435.

[9]Izhikevich, Eugene M. "Simple model of spiking neurons." *Neural Networks, IEEE Transactions on* 14.6 (2003): 1569-1572.

[10]Rinzel, John, and G. Bard Ermentrout. "Analysis of neural excitability and oscillations." *Methods in neuronal modeling* 2 (1998): 251-292.

[11]Davies, E. R. "Laws' texture energy in TEXTURE." *Machine vision: theory, algorithms, practicalities* (1997).

[12]Saxena, Ashutosh, Sung H. Chung, and Andrew Y. Ng. "3-d depth reconstruction from a single still image." *International Journal of Computer Vision* 76.1 (2008): 53-69.

[13]Saxena, Ashutosh, Sung H. Chung, and Andrew Ng. "Learning depth from single monocular images." *Advances in Neural Information Processing Systems*. 2005.

[14]Chandler, Damon M., and Sheila S. Hemami. "VSNR: A Visual Signal-to-Noise Ratio for Natural Images Based on Near-Threshold and Suprathreshold Vision."

[15]Horita, Yuukou, et al. "Mict image quality evaluation database." *2012-06-03]. http://mict, eng. u-toyama. ac. jp/mictdb, html* (2011).

[16]Ninassi, A., P. Le Callet, and F. Autrusseau. "Subjective quality assessment-IVC database." *2011-10-2]. http://www2. irccyn, ec-nanates, fr/ivcdb*.

[17]Sheikh, Hamid R., Muhammad F. Sabir, and Alan C. Bovik. "A statistical evaluation of recent full reference image quality assessment algorithms." *Image Processing, IEEE Transactions on* 15.11 (2006): 3440-3451.

[18]Ponomarenko, Nikolay, et al. "Tid2008-a database for evaluation of full-reference visual quality assessment metrics." *Advances of Modern Radioelectronics* 10.4 (2009): 30-45.

[19]Ponomarenko, Nikolay, et al. "Color image database TID2013: Peculiarities and preliminary results." *Visual Information Processing (EUVIP), 2013 4th European Workshop on*. IEEE, 2013.

[20]Video Quality Experts Group. Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, phase II[J]. VQEG, Aug, 2003. Available: http://www.vqeg.org/

[21]Weisi Lin, C.-C. Jay Kuo. "Perceptual visual quality metrics: A survey." *J. Visual Communication and Image Representation* 22(4): 297-312 (2011)

[22]Anzhou Hu, Rong Zhang, Dong Yin, and Yibing Zhan. "Image quality assessment using a SVD-based structural projection." *Signal Processing: Image Communication*. 29(2014)293–302.