Reduced-reference image quality assessment method based on wavelet feature extraction and fusion

Huiqing Zhang¹,², Yueqing Li¹*, Shuo Li¹,², Yutao Liu³

¹ Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
² Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China
³ The Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China

*Corresponding author’s e-mail: liyueqing1030@163.com

Abstract. Virtual reality has many characteristics such as multi-sensation, interactivity, and presence. Virtual reality images, also known as 360-degree panoramic images, provide users with an immersive experience and have been widely used in education, medical, military and many other fields in recent years. 360-degree panoramic images have high resolution and are subject to varying degrees of distortion during transmission and storage. Therefore, the evaluation of virtual reality images has practical significance. To this end, we propose a method for evaluating the quality of images in the new CVIQD database. The proposed image quality assessment model is developed based on Haar wavelet and Db4 wavelet feature extraction and fusion. The method is compared with a single wavelet feature extraction method. Experiments show that the method has better performance.

1. Introduction

During recent years, the virtual reality (VR) technology has obtained rapid progress and been more and more broadly used in many important application scenarios such as virtual conferencing and virtual roaming. The majority of existing relevant works were mainly devoted to the quality evaluation of images created by the depth-image-based rendering (DIBR) technology. Li et al. proposed to measure the quality of DIBR-images by combining the detection of disoccluded regions and global sharpness with a weighted pooling function [1]. Gu et al. developed AR-plus thresholding (APT) metric by using the autoregression-based local image description to capture the geometry distortion to reflect the quality of DIBR-synthesized images [2]. Motivated by the above two researches, Yue et al. devised a quality assessment method by utilizing local similarity to capture the geometric distortions and scale invariance to measure the global sharpness [3]. Jakhetiya et al. attributed the geometrically distorted pixels in DIBR-synthesized images as outliers and introduced the 3 sigma rule-based robust outlyingness ratio for quality evaluation [4]. Jakhetiya et al. [5] aimed at the problems of image quality evaluation algorithm based on the principle of free energy, proposed a concise reference IQA algorithm with strong correlation, which has good SCI quality estimation ability and can process text area better. View synthesis and compression are the main sources affecting the perceived quality of free view video. Ling [6] et al. quantified the influence of the detection track on the perceived quality, and proposed a full reference video quality evaluation method. Wang et al. [7] used edge similarity to quantify the geometric distortion of DIBR composite image. Gu et al. [8] considered the self-similarity...
of geometrically distorted damage natural images introduced by DIBR, and used its features to show regular changes.

The final recipient of the image is the human visual perception system. Therefore, the proposed image quality assessment method is consistent with the characteristics of the human eye to obtain image information. Wavelet transform is closely related to visual perception of cortical properties, and its characteristics are suitable for quality assessment. This paper proposes a method based on wavelet domain feature extraction and fusion to characterize the 360-degree panoramic image spectrum, energy and information content based on wavelet coefficient size, variance and entropy [9]: 1) Extract image features based on wavelet decomposition, that is, the coefficient amplitude, variance and entropy, the image features extracted by the two wavelets are weighted and fused to obtain the fused features. 2) Calculate the difference vector between the merged feature and the original image feature. 3) Adopt the leave-one-cross-validation method, and obtain the quality prediction model after training and testing. 4) The evaluation results are judged by Pearson linear correlation coefficient (PLCC), Spearman rank correlation coefficient (SROCC) and Kendall rank correlation coefficient (KRCC). The results are compared with the results of single wavelet extraction features. The experimental results show that the method has better 360 degree panoramic images quality assessment results.

2. Wavelet domain feature extraction

2.1. Principle of extracting features based on wavelet transform

The essence of wavelet transform is the filtering process of the original signal. The decomposition result differs depending on the wavelet function selected. However, due to the "constant Q" characteristic, no matter how the wavelet function is selected, the filter center frequency and bandwidth used for each decomposition scale are fixed. The smoothed and detailed signals in each scale space can provide time-frequency local information of the original signal, especially the composition information of the signals in different frequency bands. By solving the energy of the signals on different decomposition scales, the energy values can be arranged in a scale order to form a feature vector for identification.

2.2. Wavelet base selection

Unlike the Fourier transform, the wavelet base has no uniqueness, which is irregular. Different wavelet fundamental waveforms are very different. The results obtained by using different wavelet bases for the same signal are often different. Therefore, the choice of the optimal wavelet basis function is an important and complex problem. It is subject to the comprehensive constraints of the uncertainty principle, the nature of the wavelet basis function and the characteristics of the specific application. According to the principle of wavelet transform, the choice of wavelet basis usually considers five factors: orthogonality, compact support, regularity, symmetry and vanishing moment.

If the support length is too long, a boundary problem will occur. The support length is too short and the vanishing moment is too low, which is not conducive to the concentration of signal energy. Wavelets with symmetry can effectively avoid phase distortion in image processing. A well-regulated wavelet can achieve better smoothing and reduce the visual impact of quantization or error.

The Haar wavelet function is the simplest orthogonal function. Compared with other orthogonal functions, it has the characteristics of simple structure and convenient calculation. The advantage of orthogonal wavelet is that wavelet transform can decompose the signal into non-overlapping sub-bands, which can perform efficient discrete wavelet transform. However, in all orthogonal wavelets, Haar wavelet has the shortest support, only the first-order vanishing moment, which is not conducive to energy concentration, and is not suitable for approximating smooth functions, and its application is limited.

The Daubechies (dbN) function is a wavelet function constructed by the world-famous wavelet analysis scholar Inrid Daubechies. The Db4 wavelet has approximate symmetry, orthogonality, and a suitable length of support length, which can provide more efficient analysis and synthesis than the
Haar function. However, except for the Haar wavelet, no orthogonal wavelet satisfies the symmetry condition, which will cause distortion after decomposition and reconstruction. In some occasions where symmetry is required, such as image decomposition and reconstruction, the result is unsatisfactory.

Combining the characteristics of the above two wavelets, the Haar wavelet and the Db4 wavelet are complementary: the Haar wavelet has orthogonality but no good tightness. The Db4 wavelet has no orthogonality but has the proper tightness. Therefore, the proposed algorithm selects Haar wavelet and Db4 wavelet to extract 360-degree panoramic image features and perform weighted fusion on features.

2.3. wavelet decomposition layer selection
In this algorithm, Haar wavelet and Db4 wavelet are respectively used to decompose 2, 3, and 4 layers respectively. Pearson linear correlation coefficient, Spearman rank correlation coefficient and Kendall rank correlation coefficient are selected as performance indicators. It is found that the number of decomposition layers is 2 layers when the two wavelet bases achieve the best performance. Therefore, the number of wavelet decomposition layers is selected to be two layers. Detailed performance indicator data can be found in the experimental section 3.2 below.

3. Wavelet domain feature extraction and fusion semi-reference VR image quality evaluation method

3.1. Image quality evaluation method
According to the degree of acquisition of the original image information, the image quality assessment methods are divided into three types: a full reference method, a reduced-reference method, and a no reference method.

The reduced-reference method estimates the visual perceptual quality of the missing image by extracting partial information of the original image. The advantage of the reduced-reference method is that the amount of data transmitted is reduced, and at the same time, a better evaluation effect is obtained. The disadvantage is that the method is sensitive to the extracted features, and feature extraction is the key to affecting performance indicators.

Considering the high resolution of 360-degree panoramic image, the distortion in compression and transmission, and the accessibility of the original image, the reduced-reference method is more suitable. Through the wavelet domain feature extraction, the features that conform to the characteristics of human vision can be obtained, and only some of the feature information of the original image needs to be extracted to obtain a better evaluation effect.

3.2. Image quality evaluation method based on wavelet domain feature extraction and fusion
The algorithm includes wavelet domain feature extraction, feature-weighted fusion of two wavelets to obtain new features and training and testing of new features. Wavelet decomposition has directionality, and the features extracted by the two wavelets are fused. It is possible to have different visual characteristics for high-frequency components of different directions in the human eye, and obtain objective image quality evaluation which is more in line with subjective evaluation of human eyes method.

The five groups of images contained in the database are divided into two groups: a training group containing four sets of images, and a set of images as test groups, that is, a cross-validation method. The algorithm flow chart is shown in Figure 1.

3.2.1. Wavelet domain feature extraction
Wavelet domain feature extraction is realized by scale-space orientation decomposition. Based on wavelet coefficients, the coefficients, variances and entropy of the coefficients are extracted to reflect the characteristics of image information.
First, calculate the mean of the coefficients of the logarithmic domain wavelet, assuming that the coefficient is the magnitude of the coefficient at the kth subband \((i, j)\) is:

\[
X_n(i, j) = \sum_{i} \sum_{j} W_{n}^{j} H_{i}^{N} j_{n} i_{N} x_{1} 1
\]

where \(n\) is the nth wavelet layer, \(N_H\) is the height of the nth layer wavelet, and \(N_W\) is the width of the nth layer wavelet. Then, the variance of the coefficients is calculated to reflect the fluctuations in energy:

\[
\sigma_n = \frac{1}{N_H N_W} \sum_{i} \sum_{j} \sum_{i} \sum_{j} \log |X_n(i, j)| \frac{\sum_{i} \sum_{j} X_n(i, j)^2}{N_H N_W} - \frac{\sum_{i} \sum_{j} X_n(i, j)^2}{N_H N_W}
\]

Finally, the computational entropy is used to reflect the information in the wavelet domain:

\[
S_n = \sum_{i} \sum_{j} P(X_n(i, j)) \log P(X_n(i, j))
\]

where \(P(X_n(i, j))\) represents the probability distribution of the nth layer wavelet coefficients.

The coefficient magnitude, variance, and entropy are extracted from the original image and the distorted image to reflect the image features. And calculating the difference vectors between the distorted image and the original image.

**3.2.2. Haar wavelet and Db4 wavelet feature fusion**

After Haar wavelet and Db4 wavelet are used to extract the different decomposition layer features of the original image and the distorted image in each group, the pre-feature coefficient obtained by Haar wavelet decomposition is 1 and the Db4 wavelet decomposition is obtained under the same decomposition layer number. The pre-feature coefficient changes from 0.01 to 100, and the two are added in series to obtain the fused features. The characteristics of the four groups of fused images in the test group are trained to obtain the prediction model, and then the prediction model is used to predict the fifth group. For the test, the Pearson linear correlation coefficient (PLCC), the Spearman rank correlation coefficient (SROCC) and the Kendall rank correlation coefficient (KRCC) are selected as the best method.
4. Experimental results

4.1. Database and algorithm introduction
The database used in this experiment is CVIQD (Compressed VR Image Quality Database) constructed by Sun et al., which contains 5 source VR images and 165 compressed images encoding by JPEG, H.264/AVC and H.265/HEVC.

In this paper, Pearson linear correlation coefficient (PLCC), Spearman rank correlation coefficient (SROCC) and Kendall rank correlation coefficient (KRCC) are selected as the performance indicators of the judgment algorithm. The closer the values of these three indicators are to 1, the more the prediction result of the algorithm is [28].

The experimental process is as follows: 1) First divide the five groups of images in the database into training group and test group, and use the leave-one cross-validation method. Four of them are training groups, accounting for 80% of the images in the database, and one group is the test group, accounting for 20%. 2) The training group is trained by the two wavelet extraction features and the fusion feature obtained by weighted fusion and the subjective image quality score. The coefficient of the feature extracted by the fixed Haar wavelet is 1, and the coefficient of the extracted feature of Db4 wavelet is from 0.01 to 100 changes, or the coefficient of the feature extracted by the fixed Db4 wavelet is 1, and the coefficient of the feature extracted by the Haar wavelet is changed from 0.01 to 100, and the two are added to obtain a fusion feature. A model is obtained that converts the fusion feature to the objective quality score. 3) Test the test group with the resulting conversion model. 4) Calculate the PLCC, SROCC and KRCC index values between the predicted image quality score and the subjective image quality score, and perform 1000 tests, taking the median as the final performance index value.

4.2. Experimental results and analysis and comparison
In this paper, the Haar wavelet and Db4 wavelet with two layers of decomposition are selected for feature extraction and fusion to establish a reduced-reference image quality evaluation method for predictive models. Finally, the Haar wavelet characteristic coefficient is determined to be 1, and the Db4 wavelet characteristic coefficient is 0.67 to obtain the best prediction result. First, the results of this algorithm are compared with the results of the model established using five common single wavelets, namely Haar, Jpeg9.7, Bior6.8, Db4, Sym4, and extracted features, as shown in Table 1.

| Layers | Metric | Haar | Jpeg9.7 | Bior6.8 | Db4 | Sym4 |
|--------|--------|------|---------|---------|-----|------|
| 2 layers | PLCC | 0.9218 | 0.9335 | 0.9122 | 0.9215 | 0.9394 |
| | SROCC | 0.9445 | 0.9305 | 0.9156 | 0.8987 | 0.9434 |
| | KRCC | 0.8148 | 0.7954 | 0.7700 | 0.7354 | 0.8236 |
| 3 layers | PLCC | 0.9009 | 0.9329 | 0.8950 | 0.9441 | 0.9540 |
| | SROCC | 0.9329 | 0.9255 | 0.9245 | 0.9183 | 0.9331 |
| | KRCC | 0.8040 | 0.7700 | 0.7843 | 0.8208 | 0.8144 |
| 4 layers | PLCC | 0.8985 | 0.9177 | 0.8688 | 0.9258 | 0.9292 |
| | SROCC | 0.9393 | 0.9097 | 0.8723 | 0.9193 | 0.9386 |
| | KRCC | 0.8148 | 0.7464 | 0.6948 | 0.7716 | 0.8063 |

It can be seen from the above Table 1. that the five common wavelets obtain better performance index results when the number of decomposition layers is two. As the number of decomposition layers increases, the performance index results deteriorate, so the layer of wavelet decomposition of the proposed algorithm is 2.
In the experiment, we also carried out the experiment of characteristic fusion of Haar wavelet and Db4 wavelet and the other three wavelets when the decomposition layer number is 2 layers. The optimal result is selected to determine the weighting coefficient to compare with the algorithm. The results of various fusions are shown in Table 2.

|                      | PLCC | SROCC | KRCC |
|----------------------|------|-------|------|
| Haar+Bior6.8         | 0.9228 | 0.9475 | 0.8220 |
| Haar+Sym4            | 0.9435 | 0.9536 | 0.8328 |
| **Haar+Db4**         | **0.9511** | **0.9535** | **0.8169** |
| Haar+Jpeg9.7         | 0.9369 | 0.9449 | 0.8171 |
| Db4+Bior6.8          | 0.9109 | 0.9156 | 0.7700 |
| Db4+Sym4             | 0.9386 | 0.9441 | 0.8201 |
| Db4+Jpeg9.7          | 0.9359 | 0.9302 | 0.7918 |

It can be seen from Table 2 that after the Haar wavelet and the Db4 wavelet are weighted and fused with the features of other wavelets, the optimal three performance indexes are worse than the performance indexes of Haar extracted by the Db4 wavelet.

We also compare the proposed algorithm with six quality models, including classical PSNR, SSIM, FSIM algorithm and SQMS, ADD-SSIM and PSIM algorithms proposed in recent years. The results are shown in Table 3 below.

|                      | PLCC | SROCC | KRCC |
|----------------------|------|-------|------|
| FSIM                 | 0.9356 | 0.9325 | 0.7753 |
| PSIM                 | 0.8836 | 0.8642 | 0.6813 |
| PSNR                 | 0.8126 | 0.7715 | 0.5865 |
| SQMS                 | 0.7322 | 0.7305 | 0.5381 |
| SSIM                 | 0.8895 | 0.8823 | 0.7016 |
| ADD-SSIM             | 0.9112 | 0.9104 | 0.7412 |
| **Haar+Db4**         | **0.9511** | **0.9535** | **0.8169** |

As shown in Table 3, comparing the SROCC, the performance of the proposed algorithm is 2.26% higher than that of the FSIM algorithm [10], 10.34% better than the PSIM algorithm [11], 23.59% higher than the PSNR algorithm performance, 30.53% higher than the SQMS algorithm [12] performance, 8.07% higher than the SSIM algorithm [13] performance, and 4.73% higher than the ADD-SSIM algorithm [14] performance. The six algorithms are all full reference algorithms. The full reference algorithm needs to access all the information of the original image. But the proposed algorithm as a reduced-reference algorithm has better performance than the full reference algorithm, and only needs to access part of the original image, which proves that the proposed algorithm has certain superiority.

5. Conclusion

By analyzing the characteristics of common wavelets and finding their complementarity in feature extraction, the integrated wavelet decomposition accords with the characteristics of human visual system. This paper proposes a 360-degree panoramic image reduced-reference quality assessment method based on 2-layer Haar wavelet Db4 wavelet feature extraction and fusion. In this experiment,
170 360-degree panoramic images in CVIQD are used to verify the prediction accuracy of this algorithm. The experimental results show that the proposed algorithm outperforms the results of only one wavelet decomposition and feature extraction. As 360-degree video and images are increasingly used, research on quality evaluation algorithms for 360-degree images will have more practical implications.

Acknowledgment
This paper was financially supported by the Major Science and Technology Program for Water Pollution Control and Treatment of China (2018ZX07111005).

References
[1] Li, L., Zhou, Y., Gu, K., Lin, W., Wang, S., (2018) Quality assessment of DIBR-synthesized images by measuring local geometric distortions and global sharpness. IEEE Trans. Multimedia. 20: 914-926.
[2] Gu, K., Jakhetiya, V., Qiao, J., Li, X., Lin, W., Thalmann, D., (2018) Model-based referenceless quality metric of 3D synthesized images using local image description. IEEE Trans. Image Process., 27: 394-405.
[3] Yue, G., Hou, C., Gu, K., Zhou, T., Zhai, G., (2018) Combining local and global measures for DIBR-synthesized image quality evaluation,” IEEE Trans. Image Process. 28: 2075-2088.
[4] Jakhetiya, V., Gu, K., Singhal, T., Guntuku, S. C., Xia, Z., Lin, W., (2018) A highly efficient blind image quality assessment metric of 3D-synthesized images using outlier detection. IEEE Trans. Ind. Informat., pp. 1-1
[5] Jakhetiya, V., Gu, K., Lin, W., Li, Q., Jaiswal, S. P., (2018) A prediction backed model for quality assessment of screen content and 3D synthesized images. IEEE Transactions on Industrial Informatics, 14: 652-660.
[6] Ling, S., Gutiérrez, J., Gu, K., Callet, P. L., (2019) Prediction of the influence of navigation scan-path on perceived quality of free-viewpoint videos. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 9: 204-216.
[7] Wang, G., Wang, Z., Gu, K., Xia, Z., (2019) Blind quality assessment for 3D-synthesized images by measuring geometric distortions and image complexity. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton.
[8] Gu, K., Qiao, J., Callet, P. L., Xia, Z., (2017) Using multiscale analysis for blind quality assessment of DIBR-synthesized images. In 2017 IEEE International Conference on Image Processing (ICIP). Beijing.
[9] He, L., Tao, D., Li, X., Gao, X., (2012) Sparse representation for blind image quality assessment. In: IEEE Conference on Computer Vision and Pattern Recognition. Providence, pp. 1146–1153.
[10] Zhang, L., Zhang, L., Mou, X., Zhang, D., (2011) FSIM: A Feature Similarity Index for Image Quality Assessment. IEEE Trans. Image Process., 20: 2378-2386.
[11] Gu, K., Li, L., Lu, H., Min, X., Lin, W., (2017) A fast reliable image quality predictor by fusing micro- and macro-structures. IEEE Transactions on Industrial Electronics, 64: 3903-3912.
[12] Gu, K., Wang, S., Yang H., Lin, W., Zhai, G., Yang, X., Zhang, W., (2016) Saliency-guided quality assessment of screen content images. IEEE Trans. Multimedia., 18: 1098-1110.
[13] Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process, 13: 600-612.
[14] Gu, K., Wang, S., Zhai, G., Lin, W., Yang, X., Zhang, W., (2016) Analysis of distortion distribution for pooling in image quality prediction. IEEE Transactions on Broadcasting, 62: 446-456.
[15] Gu, K., Li, L., Lu, H., Min, X., Lin, W., (2017) A fast reliable image quality predictor by fusing micro- and macro-structures. IEEE Transactions on Industrial Electronics, 64: 3903-3912.

[16] Gu, K., Zhou, J., Qiao, J., Zhai, G., Lin, W., Bovik, A. C., (2017) No-reference quality assessment of screen content pictures. IEEE Trans. Image Process, 26: 4005-4018.

[17] Sandic-Stankovic, D., Kukolj, D., Callet, P. L., (2016) DIBR-synthesized image quality assessment based on morphological multi-scale approach. EURASIP Journal on Image and Video Processing.

[18] Sun, W., Gu, K., Zhai, G., Lin, S., Calle, P. L., (2017) CVIQD: Subjective quality evaluation of compressed virtual reality images. In: IEEE International Conference on Image Processing (ICIP). Beijing.

[19] Bosc, E., Pépion, R., Callet, P. L., Köppel, M., Ndjiki-Nya, P., Pressigout, M., Morin, L., (2011) Towards a new quality metric for 3-D synthesized view assessment. IEEE Journal of Selected Topics in Signal Processing, 5: 1332-1343.

[20] Liu, Y., Zhai, G., Gu, K., Liu, X., Zhao, D., Gao, W., (2018) Reduced-reference image quality assessment in free-energy principle and sparse representation. IEEE Trans. Multimedia, 20: 379–391.

[21] Wang, S., Gu, K., Zhang, X., Lin, W., Zhang, L., Ma, S., Gao, W., (2016) Subjective and objective quality assessment of compressed screen content images. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 6: 532 – 543.

[22] Jung, S., Whangbo. T., (2017) Study on inspecting VR motion sickness inducing factors. In: 4th International Conference on Computer Applications and Information Processing Technology (CAIPT). Kuta Bali.

[23] Battisti, F., Bosc, E., Carli, M., Callet, P. L., Perugia, S., (2015) Objective image quality assessment of 3D synthesized views. Signal Processing: Image Communication, 30: 78-88.

[24] Deng, N., Jiang, C., (2012) Selection of optimal wavelet basis for signal denoising. In: 9th International Conference on Fuzzy Systems and Knowledge Discovery. Sichuan.

[25] Stankovic, D. S., Kukolj, D., P. L. Callet, (2015) DIBR-synthesized image quality assessment based on morphological pyramids. In: 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON) . Lisbon.

[26] Wang, S., Zhang, X., Ma, S., Gao, W., (2013) Reduced reference image quality assessment using entropy of primitives. In: Proc. Picture Coding Symp. (PCS). pp. 193-196. San Jose.

[27] Gu, K., Zhai, G., Yang, X., Zhang, W., (2015) Using free energy principle for blind image quality assessment. IEEE Trans. Multimedia., 17: 50-63.

[28] Gu, K., Qiao, J., and Li, X., (2019) Highly efficient picture-based prediction of PM2.5 concentration. IEEE trans. Industrial Electronics, 66: 3176-3184.