A No-reference Image Quality Assessment Algorithm Based on Human Visual Perception Process Reconstruction

Shuo Zhang
Tsinghua Shenzhen International Graduate School, Tsinghua University
zszhangshuo4@163.com

Ping Su
Tsinghua Shenzhen International Graduate School, Tsinghua University
su.ping@sz.tsinghua.edu.cn

Jianshe Ma
Tsinghua Shenzhen International Graduate School, Tsinghua University
ma.jianshe@sz.tsinghua.edu.cn

ABSTRACT

Traditional no-reference image quality assessment algorithms mostly focus on the objective indicators of the image, such as the IE (information entropy) and clarity of the image. These indicators can reflect the quality of the image at the objective level to a certain extent, but it cannot well reflect the human visual evaluation of the image. In fact, it's still a relatively difficult problem to obtain the human visual evaluation of the image through a no-reference image quality assessment algorithm, considering the complexity of the imaging mechanism of the HVS (Human Visual System), for example, the connection mode of photoreceptor cells and ganglion cells which still cannot be well expressed by any specific model, along with the individual differences of human visual perception. This paper starts with the perception process reconstruction of the HVS, uses the similarity between the reconstructed image and the original image to reflect whether the image conforms to the request of human visual perception, carries out parameter optimization and correlation verification. Finally, it is found that the proposed algorithm has better relevance, lightness and universality than the existing algorithms, and can more effectively restore the scoring result of human visual perception of the test image.

CCS CONCEPTS

• **Computing methodologies**; • **Computer graphics**; • **Image manipulation**;

KEYWORDS

HVS, perception process reconstruction, Fourier transform, no-reference image quality assessment algorithm

ACM Reference Format:
Shuo Zhang, Ping Su, and Jianshe Ma. 2022. A No-reference Image Quality Assessment Algorithm Based on Human Visual Perception Process Reconstruction. In *2022 4th International Conference on Video, Signal and Image Processing (VSIP 2022), November 25–27, 2022, Shanghai, China.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3577164.3577179

VSIP 2022, November 25–27, 2022, Shanghai, China

ACM ISBN 978-1-4503-9781-0/22/11…$15.00
https://doi.org/10.1145/3577164.3577179

1 INTRODUCTION

The evaluation of images by human eyes is a very subjective process. Technically, we need to gather a certain number of subjects and ask them to give a score on the test image. Then we can obtain the final visual perception score of the image by statistical means. If we can calculate the human eye rating of an image by algorithm, we can simplify the above process, which will directly benefit the research aimed at optimizing human perception. Because we can believe that the research has improved the quality of the image (observed from the perspective of human eyes) by the improvement of the score given by the algorithm, and there is no need to recruit subjects to score the experimental results subjectively one by one. However, it is still of great significance to reconstruct the process of information receiving and processing by human eyes and predict the results of human visual perception towards an image by building mathematical models or algorithm design.

There are mainly two existing image quality evaluation algorithms for human visual perception. One is to use the full-reference method, to consider part of the mechanism of human visual imaging. By focusing on a specific feature, it can magnify and quantify the distortion compared with the original image, and then evaluate the human visual perception result towards it. Wang et al. [1] used Gabor filter to simulate human visual characteristics, extracted features of the original image and the distorted image in the same way and then compared them. The disadvantage of this method is that it requires the existence of the original image and only focuses on certain features and parameters, which cannot fully reflect the perception process of the HVS. This limitation may also magnify individual differences. The other is to use the no-reference method to train the neural network using some datasets. For example, Mittal et al. [2] proposed BRISQUE (Blind/referenceless image spatial quality evaluator) algorithm which transformed the image into MSCN (Mean Subtracted Contrast Normalized coefficients) domain, extracted 36 features, and used LIVE IQA dataset to train an SVM (Support Vector Machine), so that it can recognize different types of distortion and give scores. The disadvantage of this method is that the images in the existing datasets are mostly made up of results of original images through different degrees of distortion (white noise, blur, Mosaic), and the images that need to be evaluated in the actual application scene may not be quantized by identifying the distortion, but are more differentiated in image content. This means that the images to be evaluated may have the same sharpness, and the surface is not covered with noise, in which scenario, the algorithm trained by establishing a relationship between the distorted image and the human visual rating may not be able to make a judgment consistent with the real person. On this basis, Mittal et al. [3] proposed the NIQE (Natural Image Quality Evaluator) algorithm,

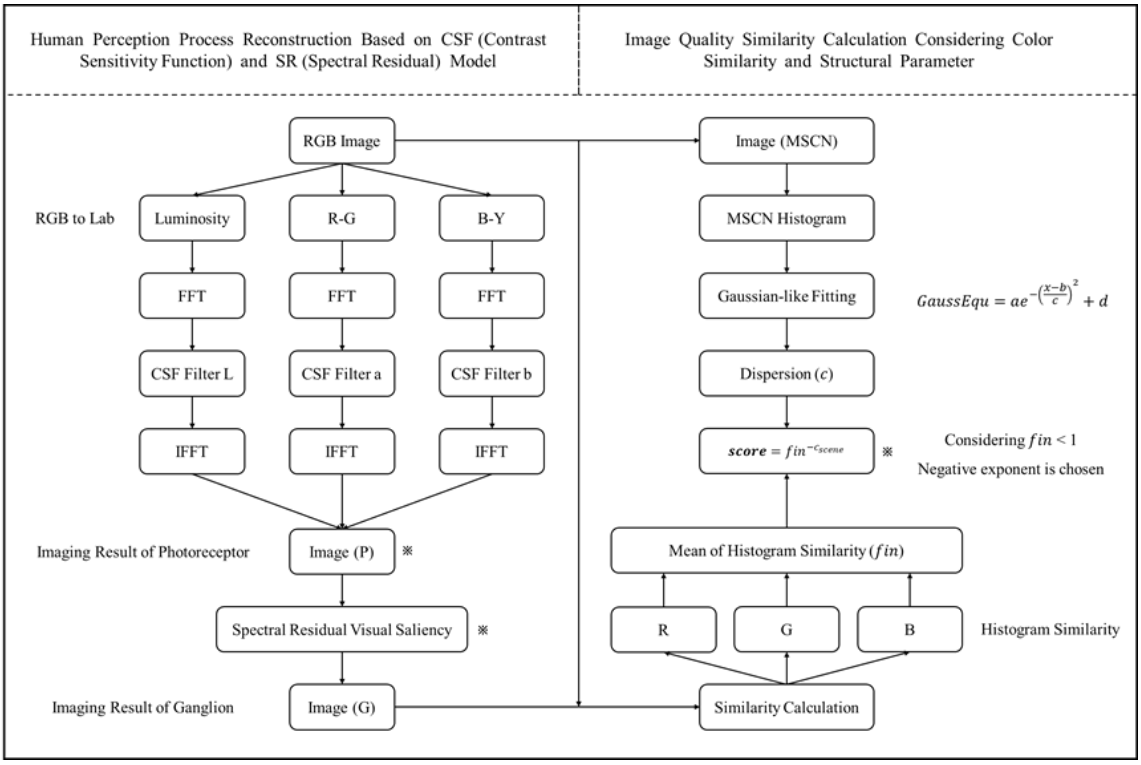

Figure 1: Schematic diagram of the algorithm (※indicates that the process requires additional parameter design).

which, unlike BRISQUE, does not identify distortion types, but scores the image by calculating the difference between the features of the image to be tested and 125 pre-selected natural images of different sizes. The problem with NIQE algorithm is that the selection of pre-selected natural images directly determines whether the algorithm can well conform to the image rating of real human eyes. There are also no-reference image quality assessment algorithms using the method of regional calculation to obtain the score of the human eye on images, Venkatanath et al. [4] assume that the human eye vision focuses on prominent area, and the quality of the various regions can be synthesized and represents the quality of the whole image. Then they proposed PIQE (Perception-based Image Quality Evaluator) algorithm which calculates distortion for each block and sums them up to obtain distortion score of the whole image which is believed to be able to conform to the evaluation result of HVS towards this image. The disadvantage of PIQE algorithm is that it does not have any parameter evaluate the image from the overall perspective, which also has an important impact on the human eye's rating of the image.

Unlike previous research, we started from the perspective of the image sampling process of HVS, rather than training based on a pre-collected database. We believe that although it's about HVS, which can be subjective, it still exists the possibility of a mathematical model which can well give the same results as most humans. Therefore, we start with the perception process reconstruction of the HVS, and use the similarity between the reconstructed image and the original image to reflect whether the image conforms to the request of human visual perception. The closer the image is to the original image after the HVS sampling, the less distortion the image will go through in actual human visual perception process, which we believe is more consistent with the human visual perception.

2 ALGORITHM

The schematic diagram of the algorithm is shown in Fig. 1, which is divided into the process of human visual perception process reconstruction and the process of similarity calculation and scoring.

2.1 Human Visual Perception Process Reconstruction

Firstly, we use three channels in Lab color space to simulate the effects of three kinds of photoreceptor cells (rods and two kinds of cones) in human eyes, and filter the three channels in frequency domain according to the CSF (contrast sensitivity function) measured by Yao's [5] experiment, to simulate the image sampling process of three kinds of photoreceptor cells in human eyes. Existing biology research shows that 1 million ganglion cells are responsible for receiving the optical signal collected by 130 million photoreceptor cells, but its internal coupling mode is not clear yet, we use the SR (spectral residual visual saliency) [6] model to simulate the subjective choice of the nerves towards each region of the image. We obtain the spatial saliency map through inverse Fourier transform of the calculation result of SR model, and use it to modulate the imaging result of photoreceptor cells in the spatial domain to complete the reconstruction of human visual perception process.

CSF of L component is as follows:

$$CSF\ Filter_L:\ C_L(f_\theta) = a \cdot f_\theta \cdot \exp(-b \cdot f_\theta)\left[1 + c \cdot \exp(b \cdot f_\theta)\right]^{0.5} \tag{1}$$

The coefficients are calculated as follows:

$$a = \frac{540 \cdot \left(1 + \frac{0.7}{L}\right)^{-0.2}}{1 + \frac{12}{\omega \cdot \left(1 + \frac{f_\theta}{3}\right)^2}},\ b = 0.3 \cdot \left(1 + \frac{100}{L}\right)^{0.15},\ c = 0.06,\ f_\theta = \frac{N}{\theta} \tag{2}$$

Where, L stands for the average brightness of the grating, f_θ stands for the angular frequency of the grating observed from human eye, ω stands for the size of the simulated pupil corresponding to the spatial view of the human eye, N stands for the number of periods of the target grating within the view range, and θ stands for the viewing angle.

CSF of a, b component is as follows:

$$CSF\ Filter_a:\ C_{rg}(f_\theta) = a \cdot exp\ b(f_\theta)^c, a = 1, b = -0.152, c = 0.893 \tag{3}$$

$$CSF\ Filter_b:\ C_{by}(f_\theta) = a \cdot exp\ b(f_\theta)^c, a = 1, b = -0.2041, c = 0.9 \tag{4}$$

Since this study focuses on the image, f_θ stands for the angular frequency of the grating. Considering the spectral coordinates (f_x, f_y) of any point on the spectral map obtained by Fourier transform is the spatial frequency of the point, and the spatial map can be obtained by inverse transformation of any point of the spectral map, we can think of an image as a combination of gratings. So f_θ of images can be calculated as follows:

$$f_\theta = \frac{\sqrt{f_x^2 + f_y^2}}{\theta} \tag{5}$$

Other parameters also need to be set according to the specific scene determined by this study. The left and right range of the static visual field of the human eye can reach 160°, shrinks to about 90-100° at 40km/h, 75° at 60km/h. In this study, θ is assigned to be 75°. The size of the pupil of human eye is generally 2-3mm, so the value of ω is set to 2mm.

As shown in Fig. 2, we use CSF Filters L, a and b to filter the spectrum of image components of three corresponding channels, and then synthesize them to obtain the perceptual image results of photoreceptor cells.

$$Image(P) = \begin{bmatrix} \boldsymbol{F}^{-1}\{\boldsymbol{F}\{L^*\} \cdot CSF\ Filter_L\} \\ \boldsymbol{F}^{-1}\{\boldsymbol{F}\{a^*\} \cdot CSF\ Filter_a\} \\ \boldsymbol{F}^{-1}\{\boldsymbol{F}\{b^*\} \cdot CSF\ Filter_b\} \end{bmatrix} \tag{6}$$

As shown in Fig. 3, we use SR model to simulate the post-processing process of ganglion cells on images collected by photoreceptor cells, the detailed operations are as follows:

$$M(u,v) = abs\left(\boldsymbol{F}\{f(x,y)\}(u,v)\right) \tag{7}$$

$$A(u,v) = angle\left(\boldsymbol{F}\{f(x,y)\}(u,v)\right) \tag{8}$$

$$L(u,v) = \log\left(M(u,v)\right) \tag{9}$$

$$R(u,v) = L(u,v) - h_n(u,v) * L(u,v) \tag{10}$$

$$SR(x,y) = g(x,y) * \left\|\boldsymbol{F}^{-1}\{\exp(R + jA)\}(x,y)\right\|^2 \tag{11}$$

Where, $M(u,v)$ and $A(u,v)$ stands for the amplitude and phase of the image after Fourier transform, $L(u,v)$ stands for the logarithmic spectrum of the image, $R(u,v)$ stands for the spectrum residual of the image, and $h_n(u,v)$ stands for a mean (smoothing) filter of size 3*3. $SR(x,y)$ stands for the spatial saliency map obtained by inverse Fourier transform of the spectral residual $R(u,v)$, and $g(x,y)$ is a Gaussian filter with size 10 and standard deviation 3.8.

We use the visual saliency map to perform special filtering on $Image(P)$ which is collected by photoreceptor cells, to obtain the perception results of the ganglion cells.

$$Image(G) = SR(x,y) \cdot Image(P) \tag{12}$$

We ignore the complex function of subsequent nerves and use $Image(G)$ perceived by ganglion cells as the result of human visual perception process reconstruction. Even if this means that only photoreceptor cells and ganglion cells are of consideration during the distortion process of the original image, as well as during the imaging process in the subsequent similarity test, we still believe that it is sufficient to distinguish different image quality.

2.2 Scoring through Similarity Calculation

As shown in Fig.4, we calculate the similarity between the image after distortion of human visual perception process and the original image to evaluate the degree that the image conforms to the request of human visual perception. Firstly, we use the mean of histogram matching degree of three channels (RGB) to represent the similarity degree of two images.

$$\Delta = \frac{\sum_{i=1}^{n} \min(H1(i), H2(i))}{min(\sum_{i=1}^{n} H1(i), \sum_{i=1}^{n} H2(i))} \tag{13}$$

$$fin = mean(\Delta k_R, \Delta k_G, \Delta k_B) \tag{14}$$

However, the histogram matching degree of three channels can only reflect the similarity of the two images in three colors (RGB), without considering the relationship between each pixel and the surrounding pixels, which is, the structural characteristics of the image. Therefore, the current method is one-sided in extracting features and comparing them to reflect the similarity before and after processing, which cannot fully evaluate the similarity of two images.

MSCN (Mean Subtracted Normalized coefficients), which is proposed by Mittal [4], is generally considered to establish the relationship between each pixel and the surrounding pixels. It has also been found that MSCN histogram always has Gaussian-like characteristics in images and different kinds of images differ in the fitting parameters. We believe that the result of mapping an image to the MSCN space can reflect the relationship between each pixel on the image and its surrounding pixels, or the overall structural characteristics, which can be quantified by Gaussian-like fitting parameters.

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \tag{15}$$

Where, $I(i,j)$ stands for the intensity of the center pixel, $\mu(i,j)$ stands for the mean value, $\sigma(i,j)$ stands for the variance, and C is a constant, which prevents the denominator from being 0 when a large area of continuous pixels with the same value (sky, grassland) occurs in the calculation area. Figure 5 shows the calculation process of MSCN and its histogram.

Taking the driver's perception during driving as an example, we selected 30 foreground images with wide and clear driver's vision

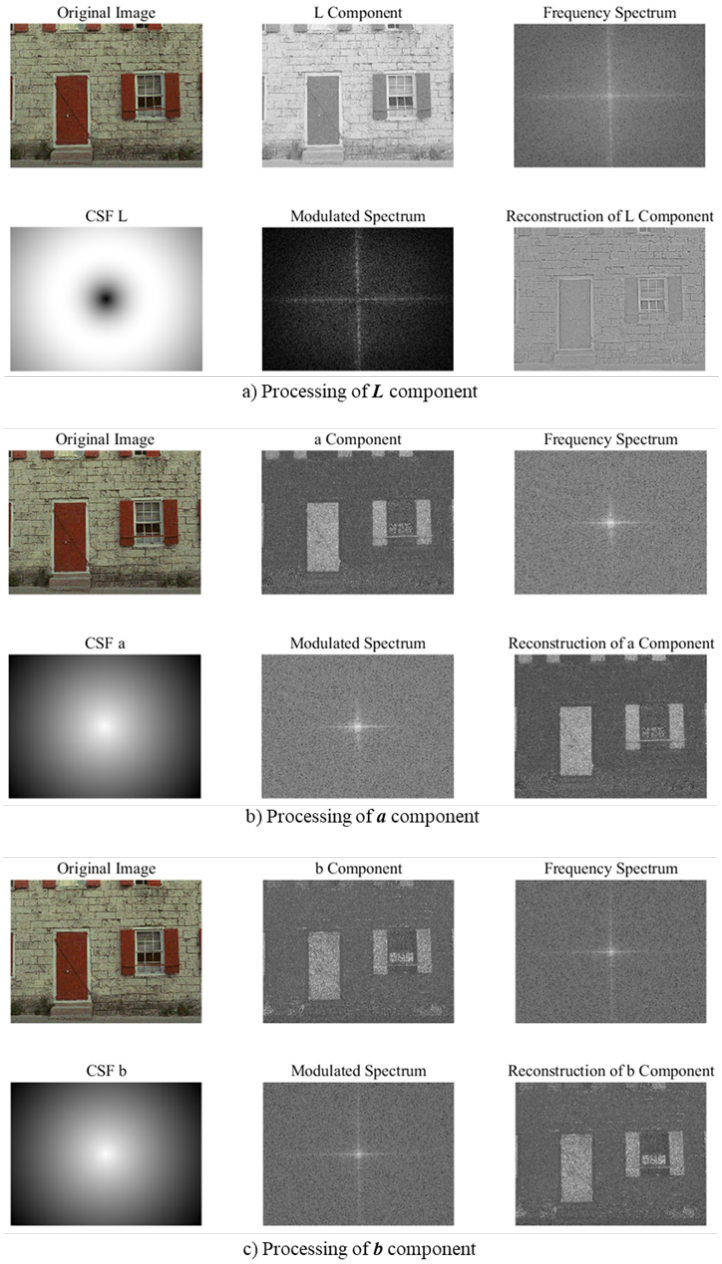

Figure 2: Image modulation process of CSF model on three channels (Lab).

and 30 images with poor overall perception quality caused by interference such as glare and dark areas. It can be seen from the results of Gaussian-like fitting of their MSCN histogram. Dispersion c_{scene} is the key to classify image through their structural characteristics. The higher c_{scene} is, the closer the image structure is to the natural image.

We run correlation test on TID2008 dataset for multiple coupling modes of color similarity parameter fin and structural parameter c_{scene}. And finally, we determine that the fitting parameter c_{scene} modulates the color similarity fin in an exponential way, making the features expressed by the two parameters complementary to each other, so that the similarity degree of the two images can be more fully reflected. It is worth noting that since fin is always no more than 1, the inverse is taken to ensure that the final score is a monotonically increasing function with respect to both fin and c_{scene}. Figure 7 shows the performance of the proposed algorithm on TID2008 dataset.

$$score = fin^{-c_{scene}} \tag{16}$$

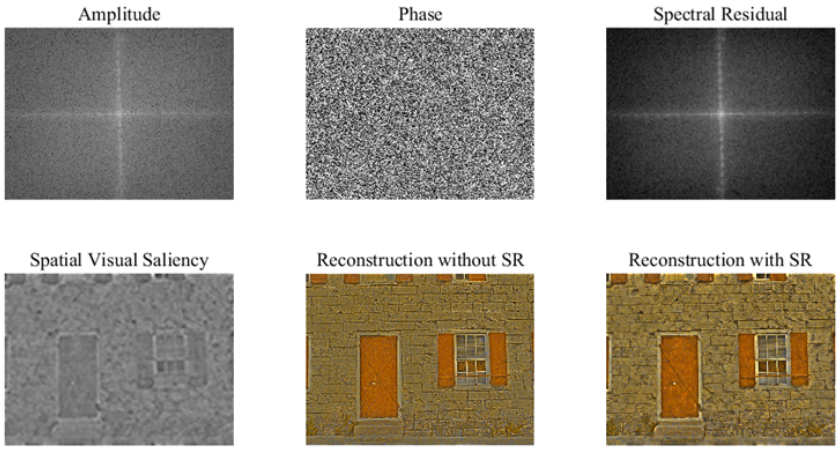

Figure 3: Image modulation process of SR model in special domain.

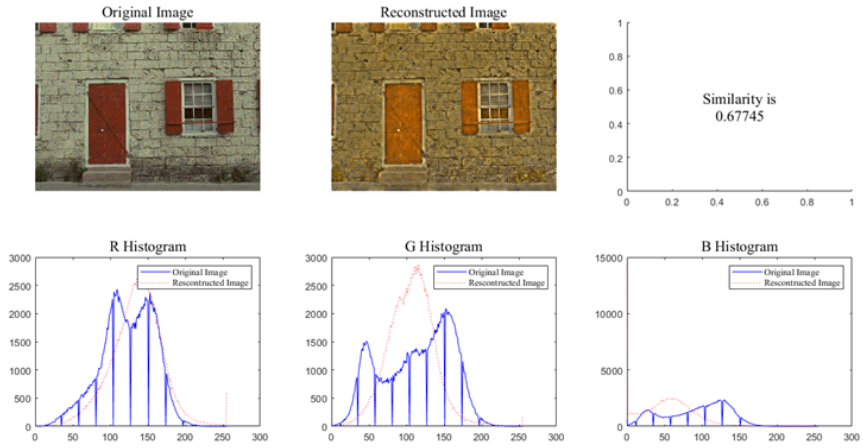

Figure 4: Calculation process of color similarity by the mean of histogram matching degree of three channels (RGB).

3 PARAMETER OPTIMIZATION

3.1 SR Modulation Coefficient

The result of logarithmic spectral residual calculation is a spatial saliency image, which is regarded as a weight to modulate the imaging results of photoreceptor cells. Since it is the weighted calculation of multiplication, it will involve the coefficient problem, the selection of coefficient will directly affect the overall brightness of the image, and then affect the calculation result of color similarity. Therefore, based on the principle that the average brightness of images before and after eye sampling should be consistent, we calculated that the SR modulation coefficient should be 2.1.

3.2 Parameters Coupling Coefficient

Adjusting the coupling coefficient between the color similarity and the fitting parameter of MSCN can improve the correlation between the scoring results and TID2008 dataset. The default coefficients of fin and c_{scene} are both 1. We respectively set the two coefficients to k and j, and traverse the optimization in steps of 0.01 on a scale from 0 to 1, as shown in Fig. 8

$$fin_{new} = (k \cdot fin)^{-(j \cdot c_{scene})} \tag{17}$$

Thus, the optimal coupling coefficient of color similarity parameter fin and structural parameter c_{scene} is obtained as follows, and the R-square of linear fitting of TID2008 dataset is raised to 0.1439.

$$fin_{new} = (0.47 \cdot fin)^{-(0.44 \cdot c_{scene})} \tag{18}$$

3.3 Lab Color Space Reconstruction Coefficient

In the reconstruction process, adjusting the proportion of L, a and b channels in the model can also improve the correlation between the scoring results and TID2008 dataset. In the previous article, we combine the three channels equally, and the results obtained are not only not high enough correlation, but also not consistent with the characteristics of human eyes. We conduct ergodic optimization for channels a and b in the interval from 0 to 2.5 with steps of 0.1. As shown in Fig. 9 (a), no extreme point is found in this traversal range, so according to its current trend, the traversal range is expanded to the upper left. As shown in Fig. 9 (b), we can finally determine the

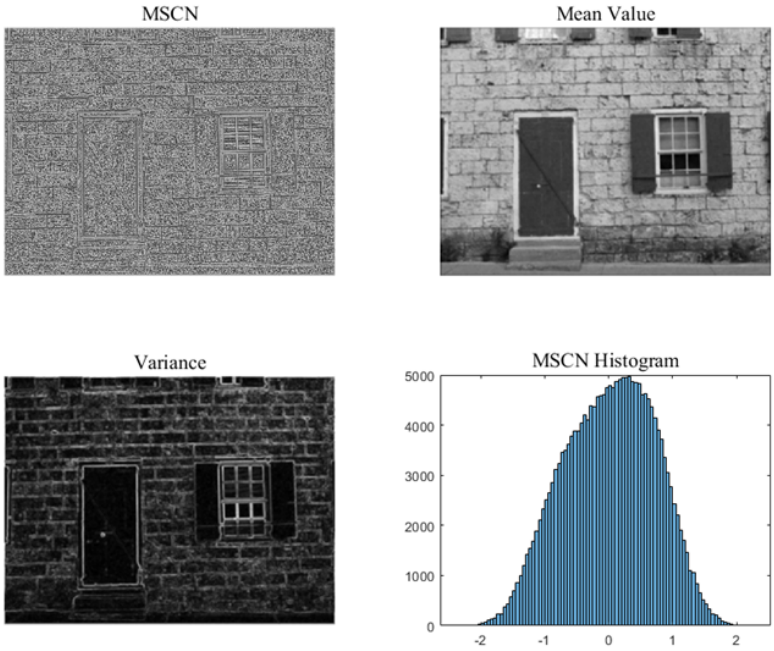

Figure 5: Calculation process of MSCN and its histogram.

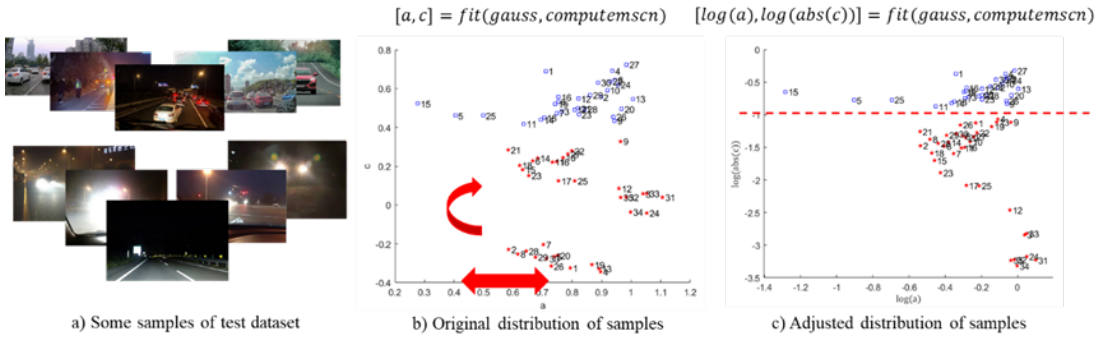

Figure 6: The process of finding fitting parameters that can distinguish the structural characteristics of an image.

synthetic ratio of three channels (Lab) is 1:0.4:3, which increases the R-square of linear fitting of TID2008 dataset to 0.1488, an increase of 36% from the original value.

4 CONCLUSION

We perform Kendall and Spearman correlation tests between the results obtained by the algorithm proposed in this paper as well as other mainstream no-reference image quality assessment algorithm and the image scores in TID2008 dataset, while Spearman correlation coefficient best reflects the monotony relationship between the two groups of data. Other algorithms selected in this comparison test include image information entropy, image structural sharpness, PIQE and NIQE, among which PIQE and NIQE are often used to calculate the human eye rating of image quality. PIQE goes through the training process of several natural images, while NIQE is calculation-based. It is worth noting that we did not carry out correlation test between BRISQUE's results and TID2008 dataset, because BRISQUE's training set and TID2008 dataset largely overlap, making the correlation test meaningless. And the reason why the values of NIQE and PIQE are negative is that their scores are negatively correlated with the ratings of human eyes.

As shown in the following table, we can see that the algorithm proposed in this paper has obvious advantages in correlation, and the performance of the fitting parameters is also better than other algorithms after linear fitting of the results. This shows that the algorithm proposed can more effectively restore and evaluate the result of human visual perception of the test image. The algorithm proposed is completely based on model and mathematical calculation, rather than training, therefore, we believe that it has better universality while being lighter.

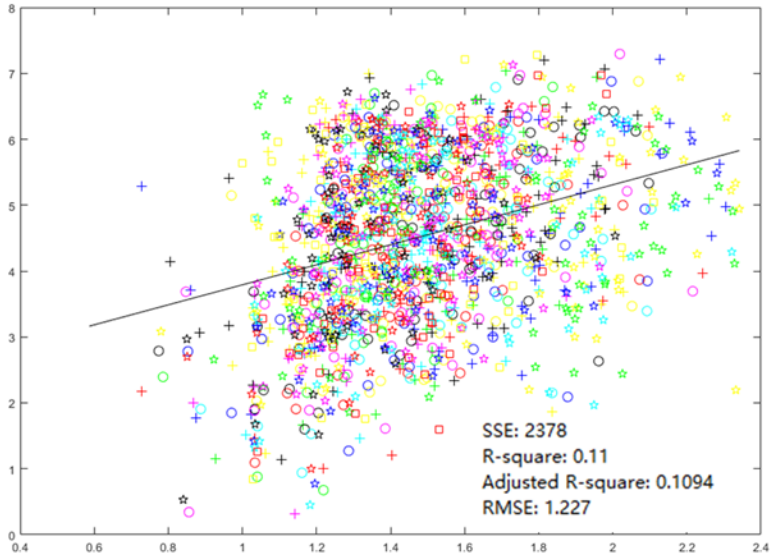

Figure 7: The performance of the proposed algorithm on TID2008 dataset. Each different marker represents the image with different processing based on the same original image, and the point groups of each different marker are evenly distributed along the line without clustering, which indicates that the proposed algorithm can distinguish image quality rather than image itself.

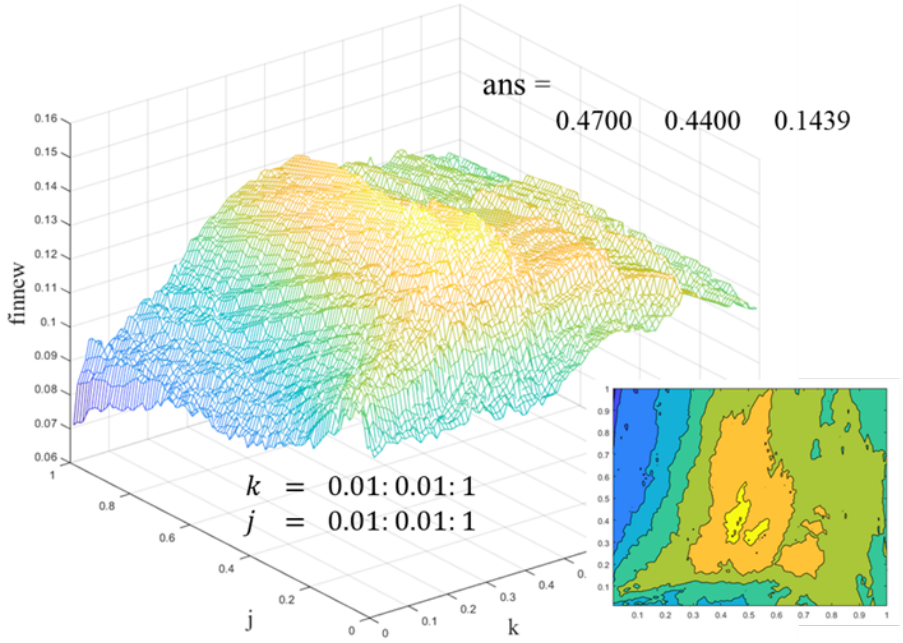

Figure 8: Optimization results of parameter coupling coefficient.

5 DISCUSSION

This paper starts with human visual perception process reconstruction, and determines whether the image conforms to the request of human visual perception by comparing the similarity between the reconstructed image and the original image. After parameter optimization, it can well reflect the human eye's rating of image quality. Some shortcomings of this study are given here, which can be used as some directions for future research:

1. We do not reconstruct the human visual perception process completely.

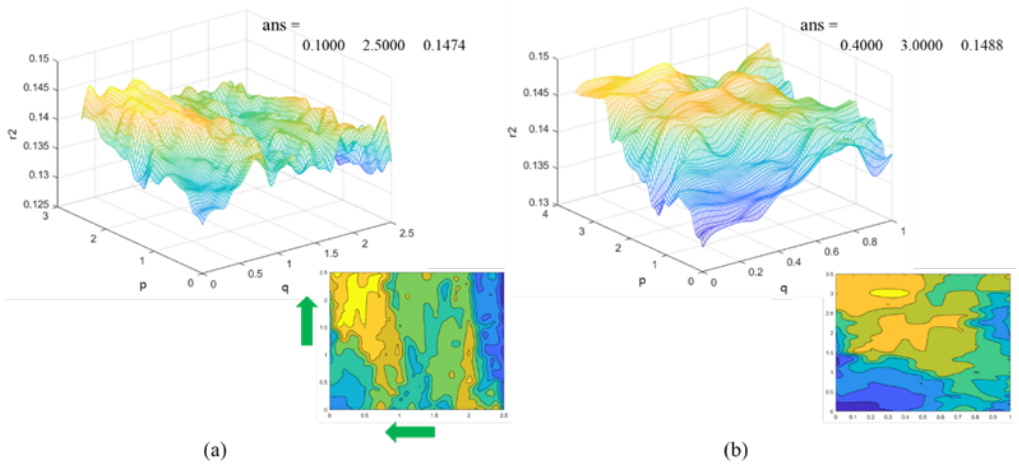

Figure 9: (a) Traversal optimization with a range of 0-2.5 (p), 0-2.5 (q) and a step size of 0.1; (b) Traversal optimization with a range of 0-4 (p), 0-1 (q) and a step size of 0.1; p stands for the ratio of channel *a* to channel *L* and q represents the ratio of channel *b* to channel *L*.

Table 1: Results of correlation test and linear fitting on TID2008 dataset

Algorithm	Kendall	Spearman	SSE	R-square	RMSE
IE	0.0579	0.0906	2403	0.000271	1.272
NRSS	0.0708	0.1061	2386	0.0000002	1.229
NIQE	-0.1251	-0.1863	2628	0.03271	1.293
PIQE	-0.2350	-0.3265	2504	0.1822	1.214
Proposed	0.2374	0.3552	2316	0.1488	1.205

2. Large step size of parameter optimization (limited by computing power).

3. The selected method cannot maximally restore the similarity of two images.

With the deepening of the cognition of the structure and mechanism of HVS, the modeling of the human visual perception process will be closer to the reality. With more perfect parameter optimization and image similarity evaluation method, the calculation results can be more accurate to the evaluation of the human eye to the measured image.

ACKNOWLEDGMENTS

This research is supported by National Key Research and Development Program of China (2021YFB2802004).

REFERENCES

[1] Wang X , Ding Y . Full reference image quality assessment based on Gabor filter[J]. Zhejiang Daxue Xuebao (Gongxue Ban)/Journal of Zhejiang University, 2013, 47(3):422-430.
[2] Mittal A , Moorthy A K , Bovik A C . No-Reference Image Quality Assessment in the Spatial Domain[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2012, 21(12):4695.
[3] Mittal A , Fellow, IEEE, *et al.* Making a 'Completely Blind' Image Quality Analyzer[J]. IEEE Signal Processing Letters, 2013, 20(3):209-212.
[4] Venkatanath N , Praneeth D , Chandrasekhar B , *et al.* Blind image quality evaluation using perception based features[C]// 2015 Twenty First National Conference on Communications (NCC). 0.
[5] Yao J C , Liu G Z . Objective assessment method of image quality based on visual perception of image content[J]. Wuli Xuebao/Acta Physica Sinica.
[6] Hou X , Zhang L . Saliency Detection: A Spectral Residual Approach [C] // IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2007.