International Journal of Information Technologies and Systems Approach
Volume 16 • Issue 3

Information Visualization Based on Visual Transmission and Multimedia Data Fusion

Lei Jiang, Hunan University of Arts and Science, China*

ABSTRACT

With the rapid development of information technology, the application media of visual identity design has been greatly broadened, and the requirements of dynamic variability and interactivity of vision have become higher and higher. The introduction of data as an element has brought new semantic endowments to visual design. To solve the problems of noisy image reconstruction and many outliers in traditional art design, this paper uses an improved phase correlation algorithm to reconstruct multimedia video images. In addition, based on the visual characteristics of human eyes, a multi-feature fusion viewpoint image quality evaluation algorithm is proposed. The simulation results show that the method adopted in this paper improves the uneven image texture in art design, and its real effect is greatly enhanced.

KEYWORDS

Image Quality Evaluation, Improved Phase Correlation Algorithm, Multimedia Data, Visual Recognition

INTRODUCTION

With the rapid development of the multimedia field, art design is organically combined with multimedia, and the advent of a new multimedia era also brings many opportunities for art design (Coral et al., 2022). Digging into the constant changes of people's aesthetic views, the visual impact of art design in the multimedia field should also be constantly developed and changed, and richer content should be given to make the visual effect of art design appear more stylish. Usually, multimedia is a type of information carrier, and its main presentation form is binary. It is mostly used to obtain, disseminate, process, and record information, such as digital animation, video images, sounds, images, graphics, and culture. Based on the multimedia context, the content of information dissemination is more and more diverse; the audience and the scope of dissemination are more and more extensive; the channels, media, and platforms of information dissemination are increasingly rich; and the timeliness and interactivity are getting higher and higher (Jarodzka, 2021; Bayraktar et al., 2019). Virtual reality (VR) technology is the most concerned image processing method in the era of big data. It simulates the surrounding environment and uses sensor devices to project and map virtual things, actions, or environments. If people watch video images, they can switch any viewpoint. This feature satisfies the

DOI: 10.4018/IJITSA.320229 *Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

visual pursuit of realism and immersion, and it is the development direction of the new generation multimedia video system (Cheng et al., 2021). At the sender, a small amount of viewpoint information is transmitted, and at the receiver, more viewing viewpoints are obtained through virtual viewpoint rendering technology (Shen et al., 2021; Shin et al., 2021).

VR is an effective way to use deep learning technology to process and visually evaluate video images. Zhang et al. (2018) proposed a no-reference video quality evaluation method based on a convolutional neural network (CNN) and resampling weakly supervised learning. This method mainly transmits the frequency histograms of distorted images and videos to resample the training set. Yan et al. (2019) proposed a no-reference image quality evaluation method based on CNN multitask learning that includes two tasks: prediction of statistical features of natural scenes and prediction of quality scores. However, virtual viewpoint rendering distortion is a variety of geometric distortions, such as artifacts, holes, and distortions. These distortions are different from coding and compression distortion all over the whole image and are mainly reflected in the texture edge region. Therefore, the existing 2D video image quality evaluation methods are directly used to evaluate the virtual viewpoint distortion, and the results obtained do not conform to the subjective perception of human eyes.

Therefore, in this research, I artistically process video images based on multimedia data fusion and reconstruct multimedia video images with an improved phase correlation algorithm to help artists complete their artistic design faster and better. In addition, I propose the use of a viewpoint image quality evaluation algorithm based on multifeature fusion and the visual characteristics of human eyes. This algorithm provides a reference for the visual evaluation of artistic image information.

RELATED RESEARCH WORKS

Visual Image Reconstruction

As one of the key steps in 3D reconstruction, image matching can be used to detect the translation distance of each pixel between the target image and the reference image. In 3D reconstruction, an image-matching algorithm is usually used to extract dense disparity maps of stereo image pairs. Jiang et al. (2021) put forward an image reconstruction model under 3D laser scanning that uses LOD and laser scanning to divide the corresponding visually highlighted areas according to different resolutions, extract the features of the target individuals, build binocular viewpoints, and obtain the detailed change values and parameter distribution in the image. This model effectively improves the simulation accuracy of the virtual target individuals and completes the visual reconstruction. However, the anti-noise and reconstruction effects of the visual images are not ideal. To meet the users' demand for visual realism, Wang (2018) used VR and VRGIS technology to describe unknown things. This technology uses a windowing algorithm and multi-level resolution structure to complete the digital image model and construct the visual image of the virtual landscape. However, it needs a lot of storage space in the system for calculation, requires high-running equipment, and is not suitable for practical application and popularization.

A feature-based stereo matching algorithm mainly extracts feature descriptions (such as edges or straight lines) from the image for stereo matching. Therefore, the matching accuracy of images with less texture, low signal-to-noise ratio, and unstable feature descriptions will be greatly reduced. Besides, to pursue more stable features, the feature extraction process usually has great computational complexity. Among the image registration algorithms based on the frequency domain, the most typical one is the matching algorithm based on phase correlation (Rasmy et al., 2021; Zheng et al., 2019). This algorithm mainly uses the translation characteristic of the Fourier transform to estimate the motion vector between two images. Among them, the translation characteristic refers to the corresponding relationship between the displacement in the spatial domain and the linear phase change in the frequency domain. The matching algorithm based on the correlation function not only has sub-pixel accuracy but also is not affected by global linear illumination changes in image contrast

and image intensity (Tzimiropoulos et al., 2010). Therefore, phase correlation has high robustness to noise, making it suitable for noisy scenes. Compared with the traditional feature extraction algorithm (Matthies et al., 2007), it is independent of feature extraction and has a faster update speed.

Goldberg et al. (2012) used Internet visual information to manipulate and edit the similarity features of objects in images more finely, thus improving the synthesis effect of image editing. Chia et al. (2011) used data-driven features to color images. Users provide text labels and reliable and similar outlines for important foreground objects in images. The system performs color conversion by searching for the best matching images in Internet visual resources, and the system will return multiple results for users to screen the best images that meet their expectations.

Video Image Quality Evaluation

The quality evaluation method of synthetic viewpoint perceived by human eyes is of great significance to improve MVD coding performance. Yang et al. (2016) considered the influence of binocular one-eyed characteristics and visual saliency on human subjective perception and proposed a stereo image quality evaluation method based on torus saliency. Karimi et al. (2017) put forward a general no-reference stereo image quality evaluation method for many kinds of distortions. By extracting the contrast and phase features of the image and using an unsupervised feature learning regression model, the predicted objective quality score is obtained. Khan et al. (2018) considered that the edges in the salient region of the image are more in line with the depth perception of human eyes, so the edges in the salient region of the original image and the distorted image were extracted, and the predicted objective quality score was obtained by calculating the structural similarity between them. Yang et al. (2018) simulated the human visual cortex's perception of depth edge information and color information when viewing images and proposed a no-reference stereo image quality evaluation method based on a segmented stacked automatic encoder. Bosc et al. (2011) extracted the distortion area by analyzing the distortion characteristics of the virtual viewpoint image and calculated the structural similarity index (SSIM) value in the distortion area to get the final quality score. Gu et al. (2017) offered an image quality evaluation method without a reference virtual viewpoint and calculated the proportion of distortion area in the whole image to get the final quality score. Tian et al. (2017) proposed a no-reference virtual viewpoint image quality evaluation method by extracting features from virtual viewpoint images with different distortion types. Yong et al. (2015) considered the binocular matching characteristics of human eyes when viewing stereoscopic images, calculated SSIM values of distorted images and parallax-mapped images in distorted areas, and then weighted the average of the left and right viewpoints to get the final quality score.

Most of the existing multimedia video image quality methods directly average the left and right viewpoint features to obtain the stereoscopic viewpoint features, but they are not suitable for the virtual viewpoint with asymmetric distortion of the left and right viewpoints because the 3D virtual viewpoint may have the situation that the left (right) viewpoint is the original viewpoint and the right (left) viewpoint is the virtual viewpoint. For stereoscopic video with asymmetric distortion of left and right views, synthesizing stereoscopic views according to the amount of information presented by a single view is necessary.

MULTIMEDIA VIDEO IMAGE RECONSTRUCTION BASED ON IMPROVED PHASE CORRELATION ALGORITHM

Visual Perception at Different Visual Distances

Figure 1 shows that watching stereoscopic video from a virtual viewpoint from different distances forms the video content of irregular scale in the user's visual system, resulting in users having different visual experience effects (Gu et al., 2017). When the user is close to the viewing screen, the size of the picture projected to both eyes is larger, and the user is sensitive to the local details in the video. Therefore, ignoring the effect of the whole video content is easy. If the user is far away from the viewing screen, the size

of the picture projected to both eyes is small, so the user can watch only the whole content of the video and cannot capture the specific details in the video. At the same time, users have different subjective perceptions of virtual viewpoint distortion when watching virtual viewpoint videos at different scales. When the distance from the screen is close, the video on the retina is larger, and viewers can see more distortion of local texture edges in the video. However, when the viewer is far away from the screen, users can notice only the overall serious distortion area in the video, and they ignore the distortion of many details in the video. Thus, predicting the quality of virtual viewpoint stereo video more accurately and making it consistent with the subjective perception of users is a challenge.

Phase Consistency Feature Extraction

During the algorithm's calculation of the image feature map, the processed image transmits its own information to the feature image, so there is a certain relationship between the processed image and the generated feature map. In addition, the distortion degree of the image is closely related to the change of its entropy.

Consider two images Y and S and remember that the entropies of the two images are H(Y) and H(S), respectively, and the joint entropy is H(Y, S) as shown in equations (1)–(3).

$$H\left(Y\right) = -\sum_{h} P_{Y}\left(h\right)\log P_{Y}\left(h\right) \tag{1}$$

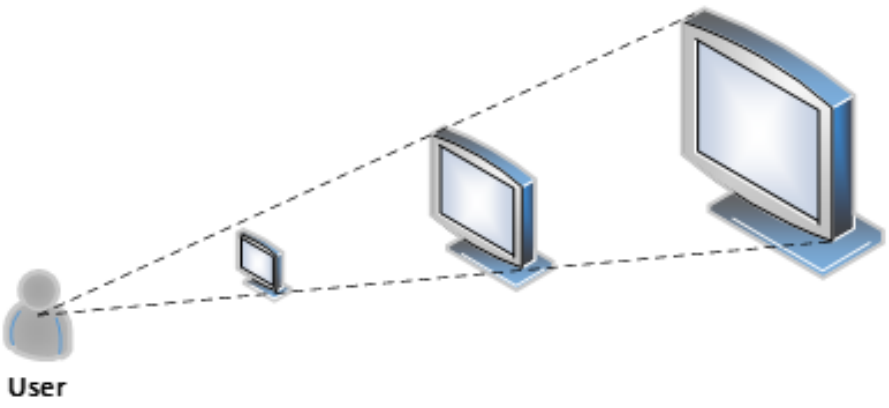
$$H\left(S\right) = -\sum_{k} P_{S}\left(k\right)\log P_{S}\left(k\right) \tag{2}$$

$$H\left(Y,S\right) = -\sum_{h,k} P_{YS}\left(h,k\right)\log P_{YS}\left(h,k\right) \tag{3}$$

Note that $P_{Y}\left(h\right)$ and $P_{S}\left(k\right)$ are the probability distributions of image Y and S, respectively, and $P_{YS}\left(h,k\right)$ represents the joint distribution of the two images. The mutual information of two images can be expressed using the formula in equation (4).

$$M\left(Y,S\right) = H\left(Y\right) + H\left(S\right) - H\left(Y,S\right) \tag{4}$$

Figure 1.
Schematic Diagram of Multiscale Visual Perception



Cross-entropy measures the transfer of image information by reflecting the difference in image gray distribution. Suppose two images Y and S, whose normalized gray histograms are $E=\left\{e_0,e_1,\ldots e_{L-1}\right\}$ and $Q=\left\{q_0,q_1,\ldots q_{L-1}\right\}$, respectively, then the cross-entropy of Y and S can be calculated as shown in equation (5).

$$D\left(E,Q\right)=\sum_{i=0}^{L-1}e_i\ln\frac{e_i}{q_i} \tag{5}$$

If the same image is placed at different distances from the human eye, the content of the image received by the human eye will be obviously different. If the image with a closer distance is observed, the human visual system is more sensitive to the detailed information of the image. If the observation distance is large, the human visual system will capture the whole image. In the field of computer vision research, images of different scales can be obtained by down-sampling and low-pass filtering, both of which can be used to simulate the influence of the change of human eye observation distance on visual perception.

Phase Correlation Algorithm

Algorithm description. The phase correlation method can effectively suppress the influence of low-frequency noise on the image and can balance and optimize the simulation scene constructed by multimedia data. Its algorithm is described in detail as follows.

Assume that there are two image sequences of the same type, $f\left(n\right)$ and $g\left(n\right)$, and both of their sizes are N. The discrete Fourier function of them is shown in equations (6) and (7).

$$F\left(k\right)=\sum_{n=-M}^{M}f\left(n\right)W_N^{kn}=A_F\left(k\right)e^{j\theta\xi^{jk}} \tag{6}$$

$$G\left(k\right)=\sum_{n=-M}^{M}g\left(n\right)W_N^{kn}=A_G\left(k\right)e^{j\theta G\left(k\right)} \tag{7}$$

Among these images, $n=-M,\cdots,M$ and $N=2M+1,W_N=e^{-\frac{2n}{N}}$. $A_F\left(k\right)$ and $A_G\left(k\right)$ represents the image amplitude. $e^{j\theta F\left(k\right)}$ and $e^{j\theta G\left(k\right)}$ represent the phase area. Therefore, the normalized image phase difference after conversion is as shown in equation (8).

$$\hat{R}\left(k\right)=\frac{F\left(k\right)\overline{G\left(k\right)}}{\left|F\left(k\right)\overline{G\left(k\right)}\right|}=e^{j\theta\left(k\right)} \tag{8}$$

Among these images, $\overline{G\left(k\right)}$ is used to describe the conjugate complex $G\left(k\right)$, $\theta\left(k\right)=\theta_F\left(k\right)-\theta_G$.

Then the inverse discrete Fourier transformation function of $\hat{R}$ is as shown in equation (9).

$$\hat{r}\left(n\right)=\frac{1}{N}\sum_{n=-M}^{M}\hat{R}\left(k\right)W_N^{-kn} \tag{9}$$

When image sequences $f(n)$ and $g(n)$ are similar, $\hat{r}$ in equation (9) will have a peak α, whose coordinate orientation δ is the relative deviation of $f(n)$ and $g(n)$. According to this peak feature, the visual deviation between two eyes can be effectively removed, and 3D stereo matching in video image reconstruction can be effectively and accurately completed.

Algorithm improvement. When the algorithm is determining the azimuth coordinates $\hat{\mathrm{r}}(\mathrm{n}_1, \mathrm{n}_2)$ to find the peak value, if the global retrieval method is used, the reconstruction error will be generated. Therefore, reducing the retrieval range and the occurrence of reconstruction error through the correlation between the reconstruction levels and the limiting rules is necessary. According to the transverse binocular limit rule, if the corrected longitudinal deviation is equal to 0, the longitudinal retrieval interval can be reduced. It can also reduce the horizontal retrieval interval by inferring the correlation of any reconstruction level.

Adding a low pass filtering step to $\hat{r}(n_1, n_2)$ and then calculating its peak value, can reduce the peak range, as shown in equation (10).

$$h(k_1, k_2) = \begin{cases} 1, |k_1| \leqslant U_1, |k_2| \leqslant U_2 \\ 0, \text{ otherwise} \end{cases} \tag{10}$$

If the relative offset interval between transverse and longitudinal is too small, the algorithm will not ensure the accuracy of the offset value. If the limited interval is too large, it will not have the ability to reduce the error reconstruction. Therefore, U_1 and U_2 are the most ideal retrieval range.

In the original algorithm, the error reconstruction is evaluated and corrected only once after the reconstruction is completed. However, in practical application, there is an intermediate level of reconstruction process, and the generated error reconstruction will extend the error to the next resolution. At the same time, the standard of its evaluation and correction is the comparison between the fixed threshold α_{th} and the peak α, but the overall peak of each resolution has a great change. Therefore, more accurate dynamic means are used to determine the threshold α_{th}.

The peaks α in any reconstruction level are statistically sorted and arranged into sets. If the edge of the restricted region of level l is assumed to be the threshold α_{th}, then $q_l(m_l)$ can be calculated. If the peak α of phase correlation is greater than or equal to the threshold α_{th}, it can be proved that this point is a cluster point, as shown in equation (11).

$$\begin{aligned} q_{lc}(m_l) &= q_l(m_l) \\ d_{lc}(m_l) &= m_l - q_l(m_l) \end{aligned} \tag{11}$$

Suppose $\alpha < \alpha_{th}$, and it can be proved that this point is an outlier. First, select all points within the 5×5 interval of the adjacent region excluding this point, construct the sequence according to its size, and take the median value, which is defined as the d value of this point, as shown in equation (12).

$$\begin{aligned} d_l'(m_l) &= \left(d_1^{\text{med}}, d_2^{\text{med}}\right) \\ q_l'(m_l) &= m_l - d_l'(m_l) \end{aligned} \tag{12}$$

Second, the phase correlation calculation shown in equation (13) is used to obtain a new peak value if $\alpha > \alpha_{th}$.

$$\begin{aligned} q_{lc}\left(m_l\right) &= q_l^{'}\left(m_l\right) \\ d_{lc}\left(m_l\right) &= m_l - d_l^{'}\left(m_l\right) \end{aligned} \tag{13}$$

Otherwise $d_{lc}\left(m_l\right) = d_l^{'}\left(m_l\right)$.

IMAGE QUALITY EVALUATION BASED ON VISUAL TRANSMISSION

Visual Information Matching

Because of different viewing angles, there are subtle differences and parallax between the same scene viewed by the left eye and that seen by the right eye. The brain vision system will get the accurate positioning of objects in the 3D space according to these viewpoints and form a stereoscopic image with depth in the brain from the surface. As shown in Figure 2, because of the differences between the left viewpoint and the right viewpoint, the video image can be divided into a stereo matching area and a monocular area. The stereo matching area is a public area that can be seen by both the left and right viewpoints, whereas the monocular area is an occlusion exposed area that exists only in the left or right viewpoint.

Assumptions about will point to receive the monocular visual information to $C_l\left(x\right) = \rho_l\left(x\right)e^{i\phi_l\left(x\right)}$ and $C_r\left(x\right) = \rho_r\left(x\right)e^{i\rho_r\left(x\right)}$,respectively . Then the response after binocular view information fusion is denoted as BE(x), as shown in equation (14).

$$\begin{aligned} BE\left(x\right) &= \left|C_l\left(x\right) + C_r\left(x\right)\right|^2 \\ &= \rho_l^2\left(x\right) + \rho_r^2\left(x\right) + \rho_l\left(x\right)\rho_r\left(x\right)\cos\left(\text{”}\,\phi\left(x\right)\right) \end{aligned} \tag{14}$$

Among these variables, $\rho_l\left(x\right)$ and $\rho_r\left(x\right)$ respectively represent the extent of visual information, and $\phi_l\left(x\right)$ and $\phi_r\left(x\right)$ respectively represent the phase of the visual information. The formula $\text{”}\,\phi\left(x\right) = \phi_l\left(x\right) - \phi_r\left(x\right)$ represents the difference between the phase information.

Image Quality Evaluation Based on Multifeature Fusion

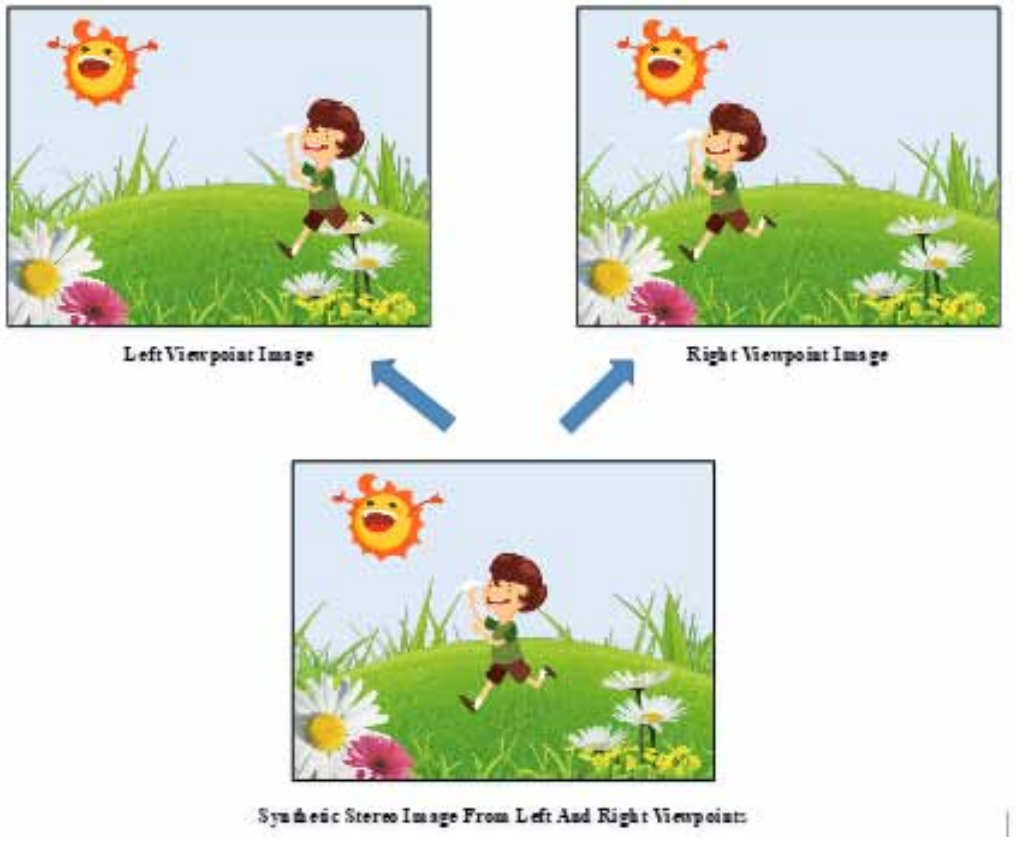
To solve the problems mentioned in the previous section, this paper proposes a multifeature fusion method for stereo image quality evaluation of a synthetic viewpoint that includes three parts:

- quality evaluation based on distorted area
- quality evaluation based on distorted edge area
- quality evaluation based on singular value

By assigning different weights to the three mass scores, the final objective mass score is obtained by linear weighted summation.

The overall flow of the algorithm is shown in Figure 3. First, the quality scores of the left and right viewpoints based on distortion region, distortion edge, and singular value features are extracted. The left and right viewpoints are then linearly weighted with 0.5 to get the three-part stereo quality

Figure 2.
Synthesis of Images from Different Viewpoints



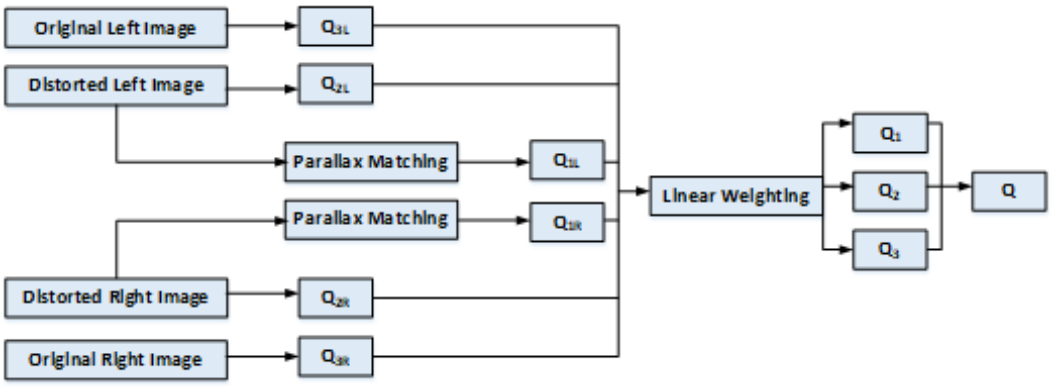
scores. Finally, different weights are assigned to each part of the 3D mass fraction, and the final objective mass fraction is obtained by linear weighted summation of three parts of mass fraction according to different weights.

In Figure 3, $Q_{1\,\mathrm{L}}\,?\,Q_{2\,\mathrm{L}}$ and $Q_{3\,\mathrm{L}}$ represent the quality scores of the left viewpoint based on the distortion region, the distortion edge region, and the singular value. $Q_{1\mathrm{R}}\,?\,Q_{2\mathrm{R}}$ and $Q_{3\mathrm{R}}$ represent the quality scores of the right viewpoint based on distortion regions, distortion edge regions, and singular values. $Q_1\,?\,Q_2$ and Q_3 represent the stereo quality scores based on the distorted region, the distorted edge region, and the singular value. Q is the final objective quality score.

Based on the distortion area part, first, the mapped left viewpoint image is obtained by using the right parallax and the right viewpoint distortion image. The difference between the left viewpoint distortion image and the mapped left viewpoint image is then calculated, and the distortion area that accords with human perception is obtained by using the threshold value. Finally, the mean SSIM (MSSIM) values of the left viewpoint distortion image and the mapped left viewpoint image are calculated to obtain Q_{IL}.

In the part based on distortion edges, the morphological edge of the image brightness channel is extracted. The difference between the morphological edge of the original image and the distorted image is then calculated, and the distorted edge area that accords with the human eye perception is

Figure 3.
Feature Fusion Process



obtained by using the threshold value. Finally, the MSSIM value of the original image and the distorted image in the distorted area perceived by the human eye is calculated to obtain Q_{2L}.

The singular value-based part is obtained by performing singular value decomposition on the original image and the distorted image and calculating the difference between the original image and the distorted image, Q_{3L}.

In this paper, three features of the image–distortion area, distortion edge area, and singular value–are extracted, and the final objective quality score is obtained by linearly weighting the quality scores of the three parts, as shown in equation (15).

$$Q = w_1 \times Q_1 + w_2 \times Q_2 + w_3 \times Q_3 \tag{15}$$

In this equation, $w_1 ? w_2 ? w_3$ represent the weight coefficients of each part, and the value range is [0 1]. The weights are w_1 is 0.3, w_2 is 0.45 and w_3 is 0.25.

EXPERIMENT AND ANALYSIS

Data Sources

In this paper, the virtual viewpoint stereo image database is used as the sample dataset of synthetic viewpoint stereo images (Lin et al., 2014). This database contains seven sequences, and each sequence contains left and right viewpoint images. At the same time, this database also provides the difference mean opinion scores (DMOs) of each stereo image pair. The larger the DMOs, the better the halo of the stereo image pair.

Analysis of Simulation Results

To avoid the influence of nonlinear factors in the process of subjective quality evaluation on the objective evaluation model, the quality scores obtained from the objective evaluation model need to be fitted nonlinearly. In this paper, the four-parameter fitting method is adopted to fit the objective mass fraction with the subjective mass fraction. The specific fitting method is shown in equation (16).

$$\mathrm{DMOS}_{\mathrm{p}} = \frac{a-b}{1+\exp\left[-\dfrac{Q-c}{abs(d)}\right]} + b \tag{16}$$

Among the values in equation (16), a b c d are constant, after nonlinear fitting, and the predicted value of the objective model is DMOSp.

Figure 4 shows the curve fitting diagram of DMOSp and subjective score DMOS. It can be seen that the quality evaluation algorithm of synthetic viewpoint stereo image based on multifeature fusion proposed in this paper is in good agreement with subjective scores and can better evaluate the quality of synthetic viewpoint stereo image.

To further verify that the objective quality score obtained by this algorithm is consistent with the subjective quality score, three commonly used evaluation indexes—SROCC, PLCC and RMSE—are used to judge the performance of this algorithm. Among them, SROCC is used to reflect the monotonicity of the objective evaluation model, whereas PLCC and RMSE are used to reflect the accuracy of the objective evaluation model. The closer the value of PLCC and SROCC is to 1, the smaller the value of RMSE is, indicating that the correlation between the objective score and the subjective score is better. The objective algorithm in this paper is compared with five algorithms, such as SSIM, MSSIM, and Bosc, on the database, and the results are shown in Figures 5 and 6.

It can be seen from the experimental results that the proposed algorithm fits the subjective score well, and its PLCC and SROCC are above 0.86, which is better than other algorithms, and the RMSE also decreases. The distortion region and distortion edge of the algorithm proposed in this paper are both based on structural similarity to evaluate the distorted image, which is sensitive to the brightness distortion region. At the same time, people often have a texture-masking effect in the artistic design and cannot well identify the distortion of the image details. Therefore, the algorithm

Figure 4.
Results of Objective and Subjective Fitting

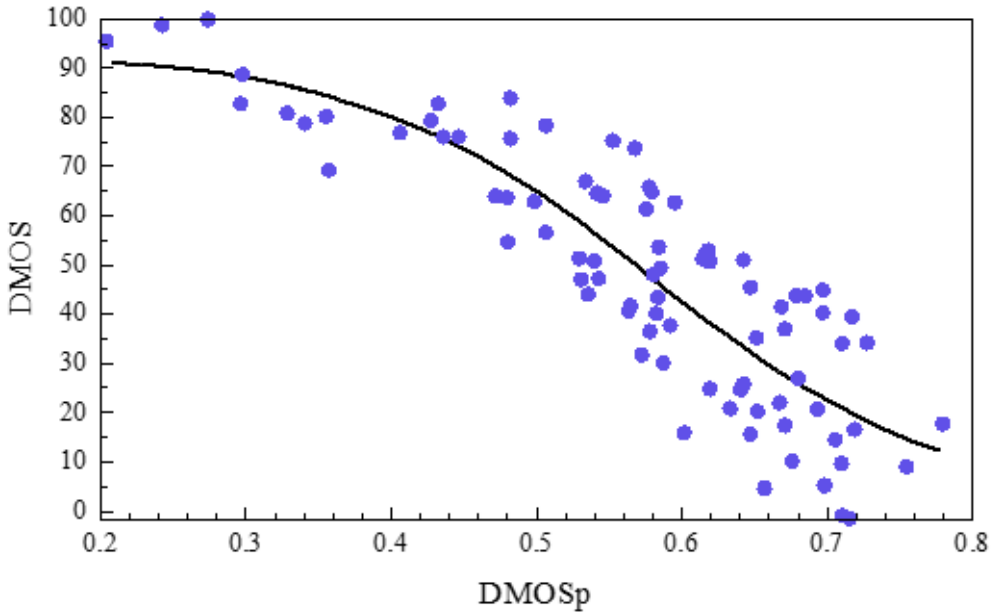


Figure 5.
Comparison of PLCC and SPOCC Algorithms

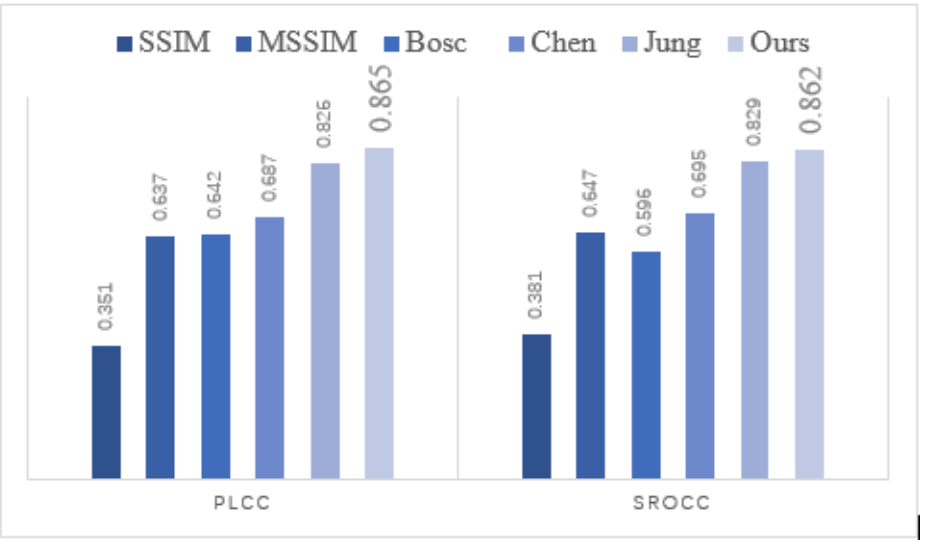
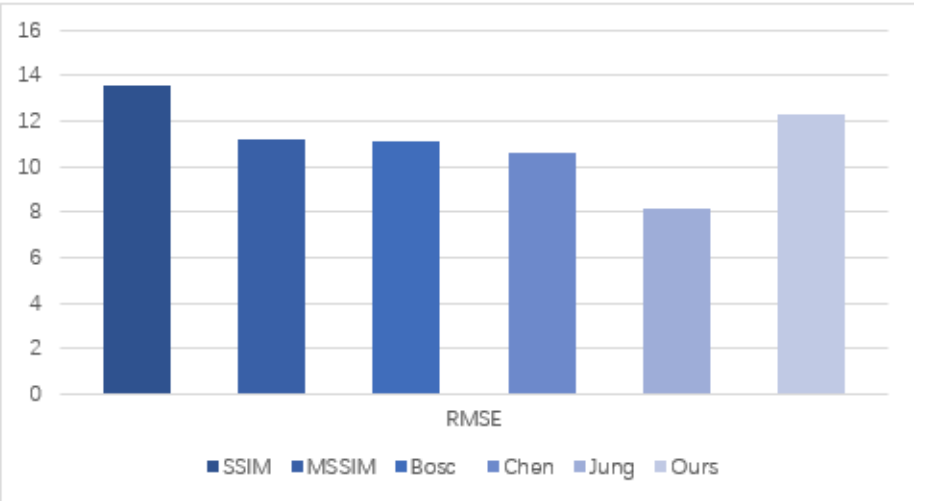


Figure 6. Comparison of Different Algorithms in Terms of RMSE



in this paper is more consistent with the visual characteristics of human eyes, making it conducive to the overall innovation of art design.

CONCLUSION

This paper describes the artistic processing of video images based on multimedia data fusion and explains how to use an improved phase correlation algorithm to reconstruct multimedia video images, thereby helping artists to complete their artistic design faster and better. In addition, this paper proposes a multifeature fusion viewpoint image quality evaluation algorithm based on the

visual characteristics of human eyes. This algorithm provides a reference for visual evaluation of art image information. Three commonly used evaluation indexes (SROCC, PLCC, and RMSE) are used to judge the performance of the proposed algorithm. The results show that the proposed multifeature fusion algorithm is in good agreement with the subjective score and can better evaluate the quality of the synthesized view stereo image. This research solves the problem of high noise and more outliers in the traditional art design and comprehensively innovates the art design more in line with the characteristics of human vision.

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers who have provided valuable comments on this paper.

CONFLICT OF INTEREST

The authors of this publication declare there is no conflict of interest.

FUNDING AGENCY

This research received no specific grant from any funding agency in the public, commercial, or nonprofit sectors.

REFERENCES

Bosc, E., Pepion, R., Le Callet, P., Koppel, 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*(7), 1332–1343. doi:10.1109/JSTSP.2011.2166245

Cheng, Q., Shan, H., Zhuang, W., Yu, L., Zhang, Z., & Quek, T. Q. (2021). Design and analysis of MEC-and proactive caching-based 360° Mobile VR Video Streaming. *IEEE Transactions on Multimedia*, *24*, 1529–1544. doi:10.1109/TMM.2021.3067205

Chia, A. Y.-S., Zhuo, S., Gupta, R. K., Tai, Y.-W., Cho, S.-Y., Tan, P., & Lin, S. (2011). Semantic colorization with internet images. *ACM Transactions on Graphics*, *30*(6), 1–8. doi:10.1145/2070781.2024190

Coral, M. A., & Bernuy, A. E. (2022). Challenges in the digital transformation processes in higher education institutions and universities. *International Journal of Information Technologies and Systems Approach*, *15*(1), 1–14. doi:10.4018/IJITSA.290002

Goldberg, C., Chen, T., Zhang, F.-L., Shamir, A., & Hu, S.-M. (2012). Data-driven object manipulation in images. *Computer Graphics Forum*, *31*(2), 265–274. doi:10.1111/j.1467-8659.2012.03005.x

Gu, K., Jakhetiya, V., Qiao, J.-F., Li, X., Lin, W., & Thalmann, D. (2017). Model-based referenceless quality metric of 3D synthesized images using local image description. *IEEE Transactions on Image Processing*, *27*(1), 394–405. doi:10.1109/TIP.2017.2733164 PMID:28767368

Gu, K., Qiao, J. F., Le Callet, P., Xia, Z., & Lin, W. (2017). Using multiscale analysis for blind quality assessment of DIBR-synthesized images. In *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)*. IEEE. doi:10.1109/ICIP.2017.8296380

Jarodzka, H. (2021). Research methods in multimedia learning. In R. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (pp. 41–54). Cambridge University Press. doi:10.1017/9781108894333.006

Jiang, Q., Shi, Y.-E., Yan, F., Zheng, H., Kou, Y.-Y., & He, B.-G. (2021). Reconstitution method for tunnel spatiotemporal deformation based on 3D laser scanning technology and corresponding instability warning. *Engineering Failure Analysis*, *125*, 105391. doi:10.1016/j.engfailanal.2021.105391

Jung, Y. J., Kim, H. G., & Ro, Y. M. (2015). Critical binocular asymmetry measure for the perceptual quality assessment of synthesized stereo 3D images in view synthesis. *IEEE Transactions on Circuits and Systems for Video Technology*, *26*(7), 1201–1214. doi:10.1109/TCSVT.2015.2430632

Karimi, M., Nejati, M., Reza Soroushmehr, S. M., Samavi, S., Karimi, N., & Najarian, K. (2017). Blind stereo quality assessment based on learned features from binocular combined images. *IEEE Transactions on Multimedia*, *19*(11), 2475–2489. doi:10.1109/TMM.2017.2699082

Khan, S., & Channappayya, S. S. (2018). Estimating depth-salient edges and its application to stereoscopic image quality assessment. *IEEE Transactions on Image Processing*, *27*(12), 5892–5903. doi:10.1109/TIP.2018.2860279 PMID:30059303

Lin, Y.-H., & Wu, J.-L. (2014). Quality assessment of stereoscopic 3D image compression by binocular integration behaviors. *IEEE Transactions on Image Processing*, *23*(4), 1527–1542. doi:10.1109/TIP.2014.2302686 PMID:24569441

Matthies, L., Maimone, M., Johnson, A., Cheng, Y., Willson, R., Villalpando, C., Goldberg, S., Huertas, A., Stein, A., & Angelova, A. (2007). Computer vision on Mars. *International Journal of Computer Vision*, *75*(1), 67–92. doi:10.1007/s11263-007-0046-z

Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review. *Computers & Education*, *141*, 103618. doi:10.1016/j.compedu.2019.103618

Rasmy, L., Sebari, I., & Ettarid, M. (2021). Automatic sub-pixel co-registration of remote sensing images using phase correlation and Harris detector. *Remote Sensing*, *13*(12), 2314. doi:10.3390/rs13122314

Shen, K., Tu, R., Yao, R., Wang, S., & Luhach, A. K. (2021). Decorative art pattern mining and discovery based on group user intelligence. *Journal of Organizational and End User Computing*, *33*(6), 1–12. doi:10.4018/JOEUC.20211101.oa20

Shin, S. I., Hall, D. J., Han, S., Paradice, D., & Lang, T. (2021). Do social networking fan page posts matter for corporate image?: Modified elaboration likelihood model perspective. *Journal of Organizational and End User Computing*, *33*(6), 1–23. doi:10.4018/JOEUC.20211101.oa17

Tian, S., Zhang, L., Morin, L., & Deforges, O. (2017). NIQSV+: A no-reference synthesized view quality assessment metric. *IEEE Transactions on Image Processing*, *27*(4), 1652–1664. doi:10.1109/TIP.2017.2781420 PMID:29324418

Tzimiropoulos, G., Argyriou, V., Zafeiriou, S., & Stathaki, T. (2010). Robust FFT-based scale-invariant image registration with image gradients. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(10), 1899–1906. doi:10.1109/TPAMI.2010.107 PMID:20479492

Wang, G. (2018). Indoor landscape reconstruction technology based on virtual reality. *Modern Electronics Technique*, *41*(4), 147–149.

Yan, B., Bare, B., & Tan, W. (2019). Naturalness-aware deep no-reference image quality assessment. *IEEE Transactions on Multimedia*, *21*(10), 2603–2615. doi:10.1109/TMM.2019.2904879

Yang, J., Sim, K., Gao, X., Lu, W., Meng, Q., & Li, B. (2018). A blind stereoscopic image quality evaluator with segmented stacked autoencoders considering the whole visual perception route. *IEEE Transactions on Image Processing*, *28*(3), 1314–1328. doi:10.1109/TIP.2018.2878283 PMID:30371364

Yang, J., Wang, Y., Li, B., Lu, W., Meng, Q., Lv, Z., Zhao, D., & Gao, Z. (2016). Quality assessment metric of stereo images considering cyclopean integration and visual saliency. *Information Sciences*, *373*, 251–268. doi:10.1016/j.ins.2016.09.004

Zhang, Y., Gao, X., He, L., Lu, W., & He, R. (2018). Blind video quality assessment with weakly supervised learning and resampling strategy. *IEEE Transactions on Circuits and Systems for Video Technology*, *29*(8), 2244–2255. doi:10.1109/TCSVT.2018.2868063

Zheng, Y., & Zheng, P. (2019). Image matching based on Harris-Affine detectors and translation parameter estimation by phase correlation. In *Proceedings of the 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*. IEEE. doi:10.1109/SIPROCESS.2019.8868783