ADAPTIVE VIEWPOINT FEATURE ENHANCEMENT-BASED CONTRASTIVE STEREOSCOPIC IMAGE AND VIDEO QUALITY ASSESSMENT

[1]MALLEPALLI YUVAPRIYA, [2]K SUBRAHMANYAM

[1]PG Scholar, Dept. of ECE, Gokula Krishna College of Engineering, Sullurpeta, Tirupati (Dt),AP, India.

[2]Associate Professor, Dept. of ECE, Gokula Krishna College of Engineering, Sullurpeta, Tirupati (Dt),AP, India.

Abstract--The development of new 3D display technologies has drawn a lot of attention to modeling 3D visual saliency. Conventional techniques that rely on low-level features might not be effective when it comes to high-level semantic interpretation of 3D visual content. Exploring the stereoscopic image and video details is not being well served by the current methods for detecting and correcting stereoscopic image and video quality defects. For this, a better technical solution is required. Therefore, in this project, we proposed a Contrastive Stereoscopic Image and Video Quality Assessment method based on Adaptive View Point Feature Enhancement, which can effectively detect and correct quality defects using binocular vision. Through the use of a binocular stereoscopic saliency total module, which enables the creation of more standard saliency highlights for binocular vision, the relationship between left and right perspectives is examined. In order to truly incorporate the elements from top-level to down-even out, we plan a dynamic consideration based saliency include pyramid extraction module. Saliency maps are then created for stereoscopic images by evaluating the obtained saliency highlights.

Index Terms: 3D Visual Saliency,Stereoscopic Image and Video Quality Assessment,Binocular Vision Processing, Adaptive Viewpoint Feature Enhancement and Saliency Map Generation

I.INTRODUCTION

Many multimedia applications, such as video editing, compositing, and the creation of artificial images based on actual image content (also known as "image-based rendering"), require the 3D reconstruction of image and video scenes. In this paper, we concentrate on a non-photorealistic rendering (NPR) application that generates painterly rendering effects on stereo views, akin to hand-painted artwork, using an input stereo pair of images from a real scene. It is possible to view the computer-generated paintings in stereoscopic mode, which adds depth to the artistic rendering effects of the individual images. In contrast to traditional photorealistic rendering, computer graphics research has focused more on developing non-photorealistic rendering (NPR) techniques in recent years (Gooch and Gooch, 2001). (Litwinowicz, 1997) is one study on computer-generated painterly effects that focuses on the automated creation of impressionist paintings. Hertzmann (2003) presents a thorough analysis of stroke-based rendering techniques. Shiraishi and Yamaguchi (2000) and Gooch et al. (2002) have both addressed the combination of computer vision techniques and painterly rendering methods. The latter study's authors use the medial axis of segmented objects to direct the automated brush stroke generation.

Additionally, they propose adding depth information to their stroke-based rendering system, which can help with segmentation, for instance. The majority of depth map-based non-photorealistic rendering projects use artificial depth maps that are calculated from a 3D model. However, our method uses stereo analysis to process real-world scene images, which typically offer richer and more features.As 3D video content has become more widely available recently, primarily in 3D film, some stereoscopic effects that could confuse the human eye and negatively affect the 3D review experience have been investigated. Eye strain, headaches, and visual fatigue are just a few of the distressing side effects that can result from prolonged exposure to 3D video content that exhibits quality problems [1]. In order to address these problems, 3D cinematographers have established cinematographic guidelines that, if strictly adhered to throughout the production process, eliminate or mitigate these problems. When

Benoit et al. [3] suggested a method that combines two measurements in 2009, the need for SIQA improvement first gained traction.

Next, the contrast between the twisted and pure sound system image difference maps is processed. Another unexpected FR-SIQA [4] metric that outperforms the previous choice has been introduced. The authors used a straightforward definition of cyclopean view, which is impacted by the left-right binocular contest. Mama et al. [5] also presented NR-SIQA, which guides Human Visual System (HVS) reproduction by using binocular mix theory. SIQA specialists are increasingly relying on NR measures due to the general benefits they provide. Thus, in order to meet the needs of the majority of current applications, the majority of current SIQA approaches are dedicated to NRSIQA. As AI advances, Deep Learning (DL) in particular [6] enables the automated extraction of the best components, enabling them to outperform handwritten attributes.

Due to different display devices and visual experiences, image quality is an important metric for comparing how different image handling calculations and systems are presented. When it comes to image acquisition, pressure, transmission, storage, and display, Image Quality Assessment (IQA) is extremely important [16]. In addition, a range of SSIM in the gradient space was proposed by Liu et al. [17] to measure image distortion. Pixel-wise Gradient Magnitude Similarity (GMS) and its standard deviation were thought by Xue et al. [18] to be highly productive and accurate indicators of image quality. Liu et al. [8] examined the image-wise Simply Recognizable Contrast (JND) for packed images, which was predicted by a perceptually lossy/lossless predictor using a deep learning model, due to the limited visual awareness of HVS.

Stage congruency and image gradient magnitude were employed by Zhang et al. [9] to estimate element similarity. Furthermore, Sun et al. [19] correlated these highlights to measure the image quality after addressing element similarity with superpixel luminance/chrominance similarity and pixel gradient similarity. To determine the degree of image distortion, a regressor was created using help vector relapse after several strong correlative components were extracted [20]. However, because the remarkable omnidirectional and stereoscopic qualities of VR images were not taken into consideration, the exhibition obtained by simply applying these measurements to the VR contents was inadequate.

Since stereo images and videos have gained popularity, stereoscopic IQA (SIQA) has recently attracted the attention of experts. The stereo images were estimated using standard 2D-IQA models by Gorley and Holliman [21], who then applied the 2D-IQA models to the left- and right-view images to obtain a weighted normal score. In order to generate the difference map as a weight map in calculating the image quality scores, Wang et al. [22] added a layer of depth information. Rather than advancing SIQA in light of the existing 2D models, various analysts looked at the mind hypothesis and generative models to mimic the human cerebrum process. According to Hohwy et al. [23], prescient coding can help make sense of binocular contention while receiving visual boosts by analogizing the cerebrum neural registering system.

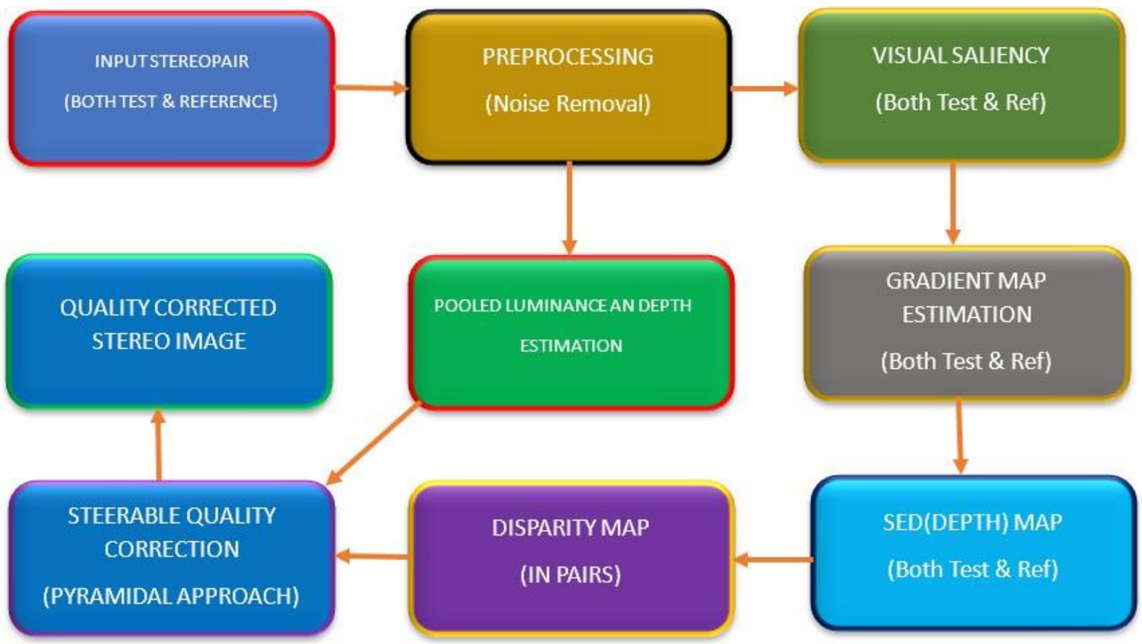
Zhu et al. suggested a SIQA that focused on the binocular combination and contention component in stereo vision after acknowledging that the "shock" from the image could explain the binocular competition. Fan et al. [24] discovered that the quality difference between the stereo images has a perceptual limit. Moreover, the omnidirectional property and projection distortion have not been considered by the SIQA. As a result, the current SIQA cannot be used to directly assess the quality of SOI. In order to produce stereoscopic images, Microsoft's Kinect-1 device for the XBox simultaneously records the depth map and the variety map (the depth map of the Kinect-1 may have openings that should be smoothed [25], which might cause noise).

These devices comprise a variety of 3D image or video applications, such as 3D video identification, 3D visual quality evaluation, and 3D delivery [26, 27]. This type of model takes variety information and depth factors into account. Although they didn't suggest a computational model in that review, Bruce and Tsotsos expanded the 2D model, which makes use of a visual pyramid handling design [16], by including neuronal units to illustrate the stereo vision. Because the systems of stereo vision actually represent a few examination challenges, such as how to fabricate and then apply the model for the stereoscopic vision, we only found two models in [27]. Based on our insight, planning the stereo-vision model is undoubtedly difficult.

II. PROPOSED WORK

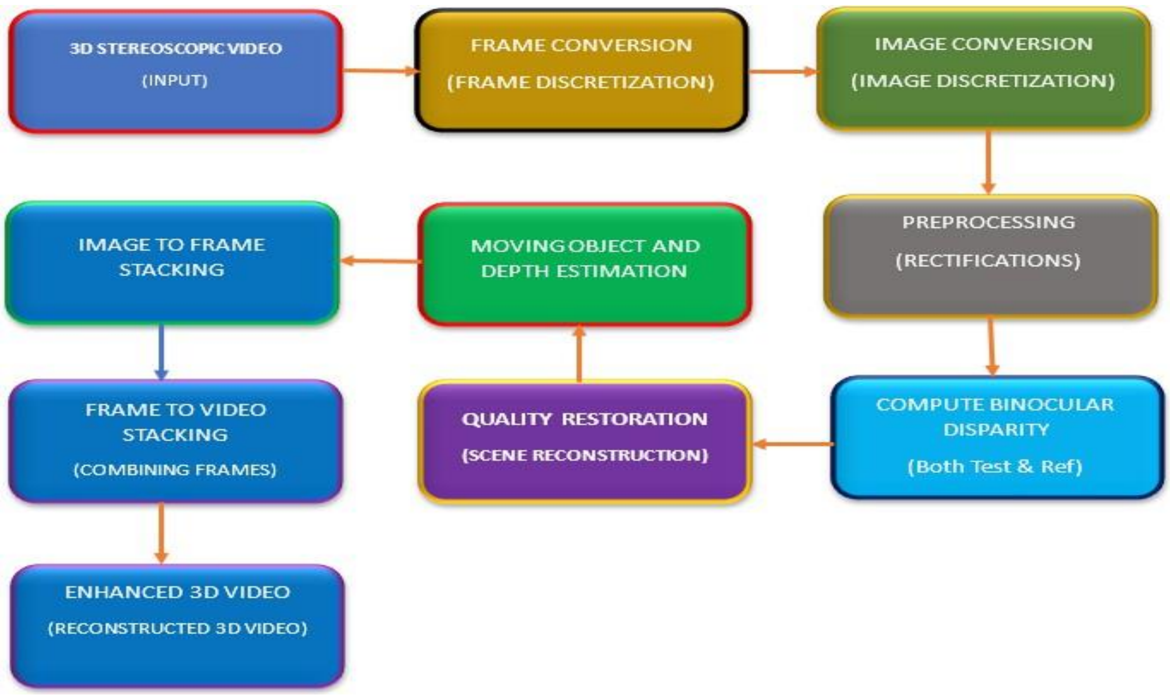
In this project, we suggested a model for evaluating the quality of stereoscopic omnidirectional images based on the visual saliency and the projection invariant feature. To address the discrepancy between the stretched projection formats and the viewports, the Scale-Invariant Feature Transform (SIFT) points are first used to derive the projection invariant monocular and binocular features of SOI. The expectation exactness is then worked with using the visual saliency model, which combines the chrominance and differentiation perceptual factors. Finally, we make the weight map based on the characteristics of the overall image and utilize it as an area beforehand, which can be modified to fit different projection designs.Delays in accessing 3D video content that exhibits quality problems can result in distressing symptoms like headaches, eye strain, and visual fatigue. As the availability of 3D video content increases, primarily through 3D film but also through 3D television, there are noticeable 3D video quality problems that could irritate the human visual system and negatively impact the 3D review experience. In order to address these problems, 3D cinematographers have established cinematographic guidelines that, if strictly adhered to throughout the creation process, eliminate or lessen these problems. The appropriate amount of defer was used in the 3D video's single frames in the current strategies. When it comes to image security, Image Quality Assessment (IQA) is crucial.

Delays in accessing 3D video content that exhibits quality problems can result in distressing symptoms like headaches, eye strain, and visual fatigue. As the availability of 3D video content increases, primarily through 3D film but also through 3D television, there are noticeable 3D video quality problems that could irritate the human visual system and negatively impact the 3D review experience. In order to address these problems, 3D cinematographers have established cinematographic guidelines that, if strictly adhered to throughout the creation process, eliminate or lessen these problems. The appropriate amount of defer was used in the 3D video's single frames in the current strategies. When it comes to image security, Image Quality Assessment (IQA) is crucial.



Fig(1): An overview of the suggested stereoscopic image enhancement technique in a schematic block.

The Saliency Map Estimation is then applied to the preprocessed Stereoscopic Image pairs (left and right) of the test and reference categories.The gradient maps for both types of stereoscopic image pairs are then created.For the test and reference categories of the stereoscopic image pairs, the ground truth and SED(Depth) maps are now created.To estimate the Disparity Maps, the estimated Saliency, Gradient, and SED maps are compared between the Test and reference Stereo pairs.



Fig(2): An overview of the proposed stereoscopic video enhancement system in schematic block form.

Along with the estimated disparity map, the entire 3D stereoscopic image was corrected for quality defects using the steerable pyramid approach, the pooled luminance vector, and depth information acquired using the Kinect depth normalization. Next, we use a variety of quality metrics, including SSIM, UQI, PLCC, and SROCC, to assess the enhanced stereoscopic image's quality. The process of identifying and fixing quality issues in 3D stereoscopic videos is then covered in detail. We start by taking the test stereoscopic video (left and right segments) into the workspace and preprocessing it to eliminate any noise. Before calculating the quality defects between the left and right stereoscopic video pairs, we first show the preprocessed frames. Next, we determine the quality gaps between the stereoscopic video pair by computing the Binocular Disparity map in both 2D and 3D. To calculate the 3D and 2D Binocular Disparity between the frames of the stereoscopic left and right videos, a separate module has been created. The primary parameter for restoring the 3D Stereoscopic video frame's quality will be the computed binocular disparity map. Until the entire video is improved, the same procedure is carried out for every frame in synchronization. The moving objects in the improved 3D stereoscopic videos are then identified, and the relative distance in meters between the camera and the moving objects is also assessed.

6.4 RESULTS AND DISCUSSION

In this work, we presented a Contrastive Stereoscopic Image and Video Quality Assessment method based on Adaptive View Point Feature Enhancement that can efficiently identify and fix quality issues using binocular vision.The following are the simulation results of the design, development, coding, implementation, and simulation of the suggested stereoscopic image and video quality enhancement technique in the Matlab environment. We use both stereoscopic and all-encompassing characteristics to model the quality expectation cycle of SOI. By imposing them on left-and right-view images, the monocular and binocular elements are meant to assess the apparent quality degradation. In this work, we try to capture invariant elements that are consistently present in 2D projection designs and viewports.

A brittle estimation of depth prompts is suggested by the fact that the number of matched matches is not precisely equal to the number of SIFT highlight focuses in a single view. Additionally, it maintains a decreasing pattern as the distortion caused by the two perspectives increases. Furthermore, in contrast to the ERP data set, these SIFT highlight focuses can also be identified in viewports and other projection organizers. The performance of the suggested PINTS-SOIQA on different projection designs is also evaluated. Fig. 3 displays the stereoscopic Left Reference Image along with its Saliency Map.

Fig. 3: Reference 3D Stereo Left Image with Saliency Map.

These images were taken from six excellent reference images. ERP design is used to store all images. Additionally, Figs. 4 and 5 show the Edge Maps of Stereoscopic Left, Ground Truth, and Disparity Map images.

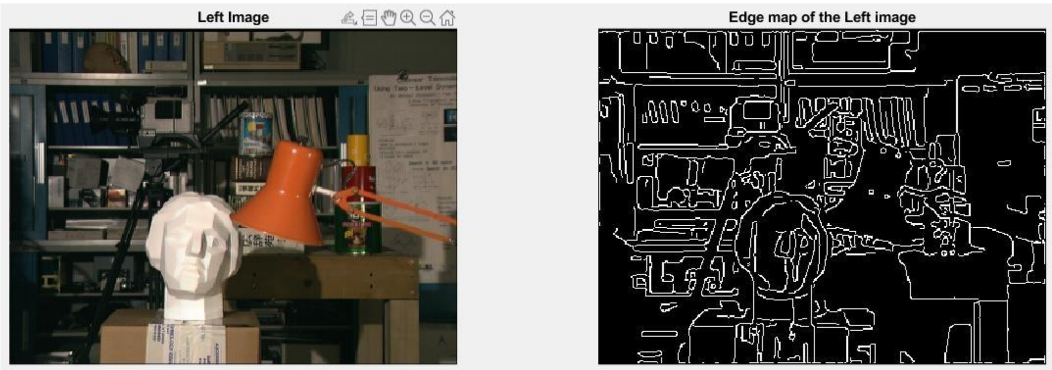


Fig.4: An edge map of the stereoscopic left image.

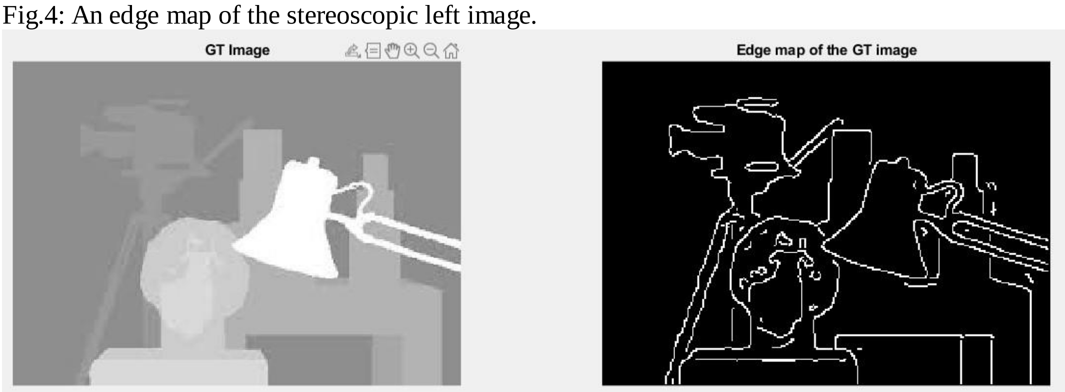


Fig. 5 : An edge map of a stereoscopic ground truth image.

Additionally, the Test and Reference Saliency and Gradient Maps The following figures also display stereoscopic LEFT and RIGH images along with their Salient Edges and Depth (SED) maps. Fig. 6 displays the quality score of the 3D stereoscopic image quality enhancement.

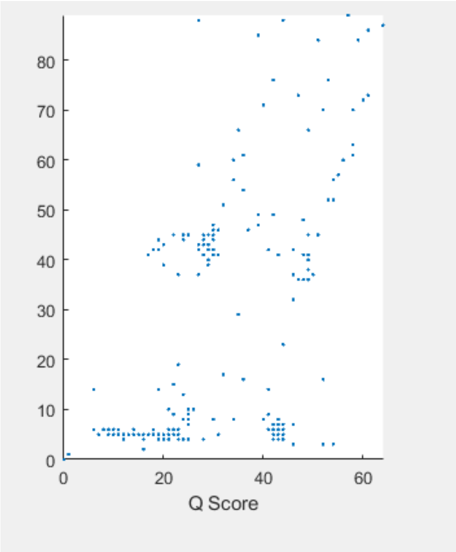


Fig. 6 : Improvement score for stereoscopic images.

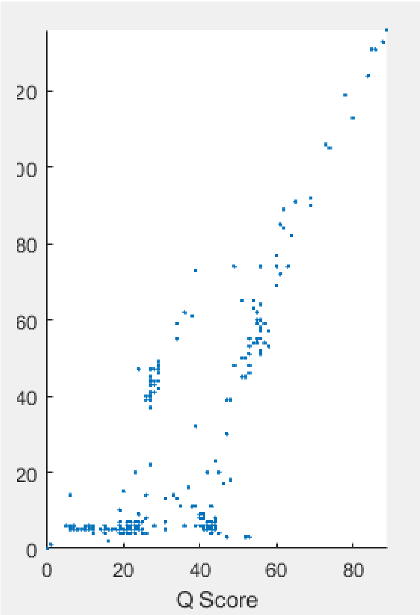


Fig.7: Video quality enhancement score for stereoscopic images.

The reference image coordinates and distorted image matches are used to accommodate the corresponding Mean Assessment Scores (Greenery). JPEG, JPEG2000, and HEVC intra are the three distinct codecs used to compress reference images. It contains 180 uniformly distorted images and 216 unevenly distorted images with two objectives, 4096 x 2048 and 2560 x 1280 pixels. The results of the detection and improvement of stereoscopic video quality defects are shown below.

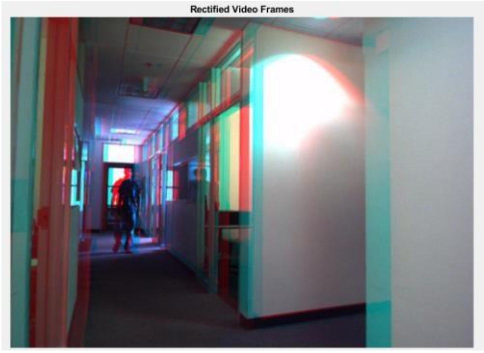


Fig.8: Stereoscopic 3D video frames that have been rectified.

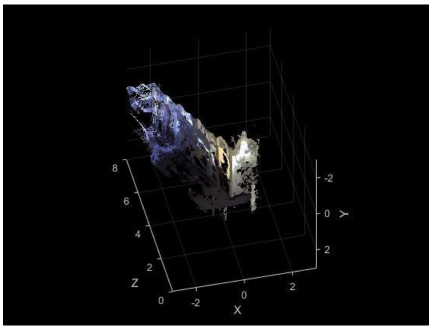


Fig.9: 3D Stereoscopic Binocular Disparity Map Estimate.

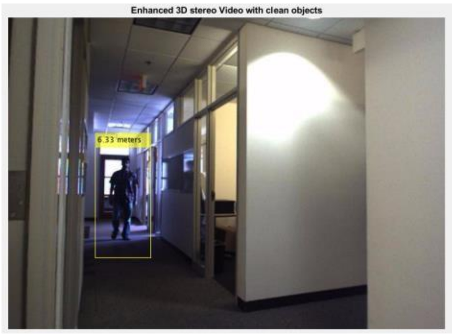


Fig.10: 3D Stereoscopic video frames with enhancements.

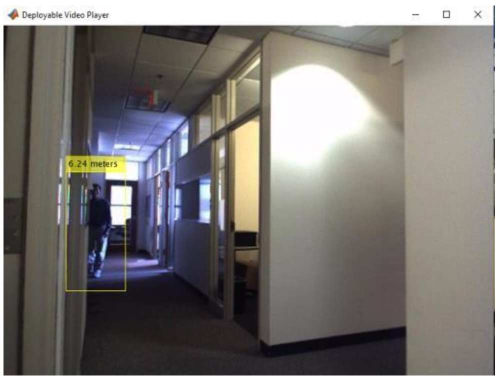


Fig.11: Final 3D stereoscopic video frames with enhancements.

On each data set, we compared the suggested PINTS-SOIQA with a few best-in-class objective quality assessment metrics, such as one No Reference (NR) IQA metric and eleven Full Reference (FR) IQA measurements. Four categories were established for these measurements. We immediately adopted five 2D IQA measurements from the standard IQA for correlation, namely PSNR, SSIM, RFSIM [52], VIF, and GSM [6]. One rearranged image with left-view at the top and right-view at the bottom was used to process quality scores. S-PSNR, WS-PSNR, and CPP-PSNR are three OIQA measurements that were used to estimate each view's exhibition. Additionally, SIQA measurements CHEN [55] and FI-PSNR [54], which were developed to predict the quality of stereo images, were known to provide a more thorough presentation correlation.

The round structure used to produce and address SOI makes it difficult to store or transmit and cannot be encoded using current coding standards. Thus, we should do reverse projection on the disentangling side and plane-to-circle mapping on the encoding side. Because of its flexibility and ease of use, ERP has become the most widely used projection design. The client behavior is taken into account by the Shortened Square Pyramid (TSP), which embraces the first goal in equatorial locales and downsamples in other districts, causing significant distortion in low goal regions.

CONCLUSIONS

The perceptual evaluation of stereoscopic content is significantly improved by the suggested Adaptive Viewpoint Feature Enhancement-Based Contrastive Stereoscopic Image and Video Quality Assessment method. This technique efficiently improves depth perception, reduces artifacts, and elevates the visual experience by utilizing contrastive analysis and adaptive viewpoint feature enhancement. By capturing the subtle differences between left and right perspectives, this technique ensures precise quality assessment, unlike traditional approaches that struggle with subtle distortions and compression artifacts. Furthermore, the incorporation of a dynamic attention-based saliency pyramid enhances the extraction of crucial features, resulting in assessments of stereoscopic quality that are more precise and trustworthy.

By providing a solid foundation for automated stereoscopic image and video evaluation, this study lessens the need for arbitrary human judgments. It can be used in a variety of domains where depth accuracy is essential, such as virtual reality, medical imaging, and entertainment. The approach guarantees adaptability across a variety of datasets by refining contrastive learning techniques and optimizing feature extraction, which makes it appropriate for real-world applications. To further develop its use in immersive media technologies, future studies can concentrate on expanding this methodology to dynamic and real-time processing scenarios.

FUTURE SCOPE:

The suggested Adaptive Viewpoint Feature Enhancement-Based Contrastive Stereoscopic Image and Video Quality Assessment method's future scope includes extending its use to dynamic and real-time scene analysis, allowing for a smooth transition into immersive technologies like 3D gaming, virtual reality, and augmented reality. Optimizing computational efficiency to enable real-time processing without sacrificing accuracy can be the subject of future research. Deep learning techniques can also improve the model's adaptability to different stereoscopic datasets, guaranteeing reliable performance in a variety of settings. The technique can also be used in remote sensing, autonomous navigation, and medical imaging, where accurate depth perception and quality evaluation are essential. In order to improve assessments based on human visual preferences, future developments might incorporate user perception models, which would increase the efficacy of evaluating stereoscopic content.

REFERENCES

1. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
2. A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, May 2010.
3. L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Image retargeting quality assessment: A study of subjective scores and objective metrics," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 626–639, Oct. 2012.
4. K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.
5. Z. Tu, Y. Wang, N. Birkbeck, B. Adsumilli, and A. C. Bovik, "UGC-VQA: Benchmarking blind video quality assessment for user generated content," *IEEE Transactions on Image Processing*, vol. 30, pp. 4449–4464, 2021.
6. A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality

analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
7. A. K. Moorthy, L. K. Choi, A. C. Bovik, and G. de Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 652–671, Oct. 2012.
8. K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.
9. Z. Sinno and A. C. Bovik, "Large-scale study of perceptual video quality," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 612–627, Feb. 2019.
10. W. Sun, T. Wang, X. Min, F. Yi, and G. Zhai, "Deep learning based full-reference and no-reference quality assessment models for compressed UGC videos," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2021, pp. 1–6.
11. D. Tran, H. Wang, L. Torresani, and M. Feiszli, "Video classification with channel-separated convolutional networks," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, Oct. 2019, pp. 5552–5561.
12. Y. Wang, J. Ke, H. Talebi, J. G. Yim, N. Birkbeck, B. Adsumilli, P. Milanfar, and F. Yang, "Rich features for perceptual quality assessment of UGC videos," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 13435–13444.
13. W. Wen, M. Li, Y. Zhang, Y. Liao, J. Li, L. Zhang, and K. Ma, "Modular blind video quality assessment," *arXiv preprint arXiv:2402.19276*, Feb. 2024.
14. H. Wu, C. Chen, J. Hou, L. Liao, A. Wang, W. Sun, Q. Yan, and W. Lin, "FAST-VQA: Efficient end-to-end video quality assessment with fragment sampling," in *Proc. Eur. Conf. Computer Vision (ECCV)*, Oct. 2022, pp. 538–554.
15. H. Wu, E. Zhang, L. Liao, C. Chen, J. Hou, A. Wang, W. Sun, Q. Yan, and W. Lin, "Exploring video quality assessment on user generated contents from aesthetic and technical perspectives," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, Oct. 2023, pp. 20144–20154.
16. W. Sun, H. Wu, Z. Zhang, J. Jia, Z. Zhang, L. Cao, Q. Chen, X. Min, W. Lin, and G. Zhai, "Enhancing blind video quality assessment with rich quality-aware features," *arXiv preprint arXiv:2405.08745*, May 2024.
17. A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
18. L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
19. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
20. Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, May 2011.
21. M. Liu, J. Wang, J. Zhang, and J. Zhang, "Stereoscopic image quality assessment based on binocular vision system and visual saliency," *IEEE Access*, vol. 7, pp. 153982–153992, 2019.
22. K. Ma, H. Yong, Q. Zhao, and D. Tao, "Perceptual quality assessment of smartphone photography," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2017, pp. 3255–3263.
23. S. A. Golestaneh and L. J. Karam, "Spatially-varying SSIM-based contrastive quality index for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1232–1244, Mar. 2016.
24. J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 206–220, Feb. 2017.
25. S. Y. Kim, S. H. Bae, and M. Kim, "Deep video quality assessor: From spatio-temporal visual sensitivity to a convolutional neural aggregation network," in *Proc. Eur. Conf. Computer Vision (ECCV)*, Sep. 2018, pp. 219–234.