This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3081259, IEEE Transactions on Multimedia

Combining Retargeting Quality and Depth Perception Measures for Quality Evaluation of Retargeted Stereopairs

Xuejin Wang, Feng Shao, *Member*, *IEEE*, Qiuping Jiang, Zhenqi Fu, Xiangchao Meng, Ke Gu, *Member*, *IEEE*, Yo-Sung Ho, *Fellow*, *IEEE*

Abstract—Stereoscopic Image Retargeting (SIR) aims to adapt stereoscopic images and videos to 3D display devices with various aspect ratios by emphasizing the important content while retaining surrounding context with minimal visual distortion. To address the issue of SIR evaluation, this paper presents a new objective quality assessment method for retargeted stereopairs by combining image quality and depth perception measures. Specifically, the image quality measure is conducted between the source and retargeted intermediate views generated by the view synthesis method to characterize the geometric distortion and content loss of the retargeted stereopair, while several depth-aware features are extracted to measure the visual comfort/discomfort and depth sensation when human views a 3D scene. Then, the extracted features are integrated into an overall perceptual quality prediction. Experiment results on NBU SIRQA and SIRD databases verify the superiority of our method.

***Index Terms*—Quality assessment; Stereoscopic image retargeting; Superpixel-based method; Depth perception.**

I. Introduction

WITH the rapid development of 3D display devices, various 3D displays can be used for stereoscopic image visualization, ranging from high-resolution cinema screens to low-resolution mobile devices. For optimal display or use in different applications, a variety of Stereoscopic Image Retargeting (SIR) techniques [1-4] have been developed to adapt stereoscopic images/videos to screens with various aspect ratios and resolutions. However, it is still challenging to generate a perfect retargeted stereopair for arbitrary scenes without producing any noticeable artifacts, such as shape twisting, visually important content loss and visual discomfort. Therefore, an effective stereoscopic image retargeting quality assessment (SIRQA) metric is urgently needed for promoting SIR techniques.

In recent years, extensive researches [5-8] on image retargeting quality assessment (IRQA) for 2D retargeted images indicate that geometric distortion and content loss are two major factors leading to quality degradation of retargeted images, and the existing 2D IRQA metrics [9-11] explore the quality-aware features to characterize the two types of distortion. Among these metrics, the grid-based methods [12-14] have been demonstrated to be relatively effective in measuring the geometric distortion of retargeted images.

By contrast with 2D retargeted images, it is a more challenging issue to evaluate the perceptual quality of retargeted stereopairs. So far, only a few works have been proposed for SIRQA [15-17], especially lack of large-scale databases for SIRQA purpose. Among these works, Liu *et al*. [15] combined five features, including picture completeness, local distortion, global distortion, depth similarity and disparity excessiveness to predict the quality of retargeted stereopairs. Zhou *et al*. [16] proposed a visual comfort assessment metric for SIR by estimating the disparity range, disparity intensity distribution, boundary disparity as well as image quality. Fu *et al*. [17] evaluated the perceptual quality of retargeted stereopairs by introducing monocular image retargeting transformation and viewpoint transformation to reveal the artificial retargeting modifications. Although the abovementioned metrics have positive effects on 2D IRQA or SIRQA, they still have the following limitations: 1) These SIRQA metrics performed feature extraction on left and right views separately without utilizing the disparity information, and then combined the extracted features to represent the 3D image quality. However, studies on stereoscopic images [18] provide the evidence that the quality perception of stereoscopic images is a more complex process, which cannot be expressed as a simple combination of monocular features. 2) Most grid-based methods adopted uniform grids to represent the deformation of retargeted images without considering the structure characteristic of the image. Although the non-uniform grid, e.g., superpixel, was applied in [19], which measured the

This work was supported by the Natural Science Foundation of China (grant 62071261, 61901236, 41801252, 62076013, 62021003), and Natural Science Foundation of China (grant R18F010008). It was also sponsored by K. C. Wong Magna Fund in Ningbo University. *Corresponding author*: *Feng Shao*.)

Xuejin Wang, Feng Shao, Qiuping Jiang, and Xiangchao Meng are with the Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China (e-mail: 1020468620@qq.com; shaofeng@nbu.edu.cn; jiangqiuping@nbu.edu.cn, mengxiangchao@nbu.edu.cn).

Zhenqi Fu is with the School of Information Science and Engineering, Xiamen University, Xiamen 361005, China. (e-mail: 920012597@qq.com)

K. Gu is with the Beijing Advanced Innovation Center for Future Internet Technology, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke@bjut.edu.cn).

Yo-Sung Ho is with the School of Information and Communications, Gwangju Institute of Science and Technology (GIST), Gwangju 500-712, Korea (e-mail: hoyo@gist.ac.kr).

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3081259, IEEE Transactions on Multimedia

content loss of retargeted images by utilizing the statistic information of the superpixel, the shape modification of the superpixel was not involved. To overcome the limitations and further explore more effective features for SIRQA, we propose a new quality assessment method for stereoscopic image retargeting by performing image quality measure on the generated intermediate view and depth perception measure from two aspects (i.e., visual comfort/discomfort and depth sensation). In summary, the main novelties of the proposed method are three-folds:

1) We generate an intermediate view for quality assessment of retargeted stereopair based on the experimental evidence that stereoscopic quality assessment on the intermediate view yields higher correlations with human subjective judgments than that on the monocular views separately [18].

2) We adopt superpixel-based method instead of the uniform grid-based method to measure the geometric distortion of the retargeted stereopair based on the fact that the shape of the superpixel can effectively represent the structure characteristic of the image.

3) Inspired by the visual physiology evidence that HVS perception tends to be more stable when monocular regions are more visually similar to the binocular background [20], we develop a new statistical similarity to measure the perceptual stability of the generated retargeted stereopair.

The rest of this paper is organized as follows. Section II gives a brief review of related works and presents the motivation of this work. Section III details the proposed method. The experimental results on NBU SIRQA and SIRD databases are presented in Section IV. Finally, the conclusions are derived in Section V.

II. Related Work and Motivation

A. Image Retargeting Operators

The image retargeting techniques can be roughly divided into two categories, i.e., discrete approaches and continuous approaches. Seam carving [21] is a typical discrete approach, which resizes an image by removing or inserting pixels (seams) in low-importance regions. However, the jagged edges and content loss may appear in the visually important objects, which is the common weakness of the discrete approaches. In contrast, the performance of continuous approaches relies on the designed energy functions, such as warping [22], which adjusts the image resolution by redistributing density without discarding image contents, but the geometric distortion may degrade the visual quality of retargeted images created by continuous approaches.

For stereoscopic image retargeting, besides object and shape preservation, depth perception is another important factor affecting the performance of SIR approaches. For instance, Chen *et al*. [23] combined seam carving with depth-aware saliency for SIR. Shao *et al*. [24] incorporated stereoscopic visual attention and binocular just-noticeable difference models for the energy optimization of SIR. Lin *et al*. [25] utilized the object correspondences between the left and right views to implement the object-coherence warping. Chang *et al*. [2] adapted image depth to the comfort zone while preserving the shapes of visually important objects. Shao *et al*. [4] developed a QoE-guided warping method by jointly taking the perception factors into account, including image quality, depth perception and visual comfort.

B. Quality Assessment for 2D Image Retargeting

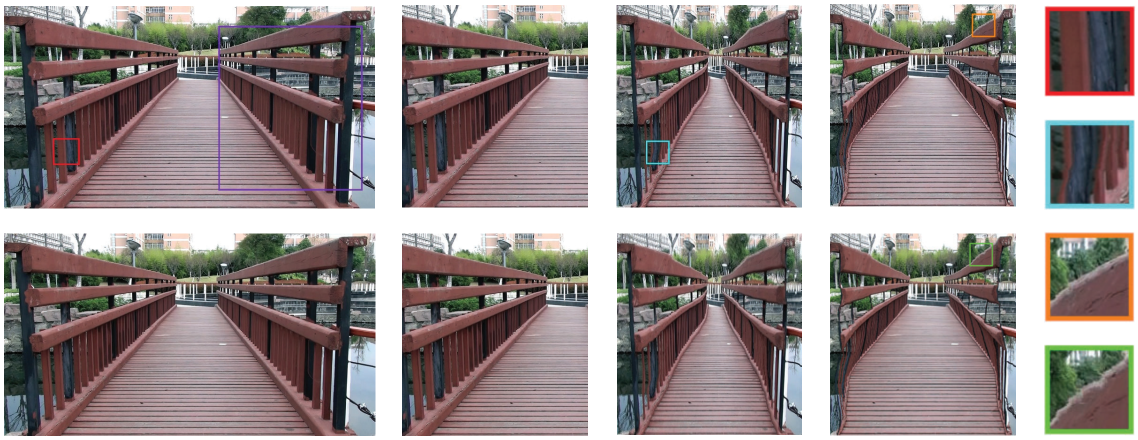
Since source images and retargeted images do not have the same resolution, the traditional full reference IQA metrics are not suitable for evaluating the quality of retargeted images. In recent years, numerous works [9-13] have been done to develop effective IQA metrics for 2D image retargeting. To overcome the resolution gap between the source and retargeted images, the commonly-adopted preprocessing method is to utilize the dense correspondence to attain the alignment between the source and retargeted images, followed by the specific feature extraction and quality prediction operations. For instance, Fang *et al.* [9] designed an IR-SSIM metric which extended SSIM to IRQA by measuring the local quality between the matched patches of source and retargeted images. Unlike the traditional distortion types such as blur, compression artifacts and white noise, geometric distortion and content loss are two major factors leading to quality degradation of retargeted images. Hsu *et al.* [10] designed an IRQA metric by measuring the geometric distortion based on the local variance of SIFT-flow vector and estimating the content loss based on the saliency map. Liang *et al*. [11] evaluated the quality of retargeted images by jointly considering salient region, artifact, global structure, aesthetic and symmetry. Zhang *et al*. [12] proposed an aspect ratio similarity (ARS) metric for image retargeting by exploiting the local block changes to evaluate the visual quality of retargeted images. Karimi *et al*. [13] combined shape features, area features and aspect ratio features to estimate the geometric distortion and content loss of retargeted images. Shao *et al*. [21] [26] presented a transform-aware similarity measurement metric for IRQA to estimate geometric distortion and content loss of retargeted images via bidirectional rewarping. Li *et al*. [19] proposed a quality evaluation model for image retargeting by extracting instance-level semantic features including shape twisting, size similarity, content loss and location movement. Although these metrics have delivered moderate performance in evaluating the quality of retargeted images, there is still large room for improvement in effectively and accurately measuring geometric distortion and content loss of retargeted images.

C. Quality Assessment of Stereoscopic Images

Different from the perceptual quality of independent 2D images in a stereopair, the perceptual quality of the stereopair is a comprehensive result of multiple factors, such as image quality, depth quality and visual comfort [27]. The existing 3D IQA can be roughly divided into two categories based on

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3081259, IEEE Transactions on Multimedia



(a) (b) (c) (d)

Fig. 1. Example of the generated intermediate view. The images in the first row are the left views of (a) the source stereopair and (b)-(d) the retargeted stereopairs generated by content persistent cropping [36], geometrically consistent stereo seam carving [1] and stereo scaling [17], respectively. The images in the second row are the intermediate views of the associated stereopairs.

whether the 3D visual perceptual property is considered [28]. The first category [29] directly applied 2D IQA algorithms on the left and right views of the stereopair separately, and then combined the two scores into an overall quality. The second category takes various 3D perceptual properties into account, such as stereopsis, binocular suppression and binocular fusion. Lee *et al*. [30] presented a 3D Perception-based Stereo image quality pooling (3DPS) model, which divided the stereoscopic image into binocular and monocular regions based on the availability of binocular depth perception in the regions, and evaluated the quality separately. Maalouf *et al*. [31] performed 3D IQA by calculating the sensitivity coefficients of the cyclopean views and the coherence between the disparity maps of the reference and distorted stereopairs. Chen *et al*. [28] utilized the disparity information and Gabor filter response to construct a cyclopean view for the stereopair, and applied full reference 2D IQA models on the reference and test cyclopean views to predict the 3D quality scores. Shao *et al*. [32] learned binocular receptive field properties to simulate simple and complex cells in the primary visual cortex, and characterized their impacts on quality estimation of stereoscopic images. In addition, visual comfort/discomfort is another important factor in affecting the overall perceptual quality of stereopairs. Most researches predicted visual comfort/discomfort scores based on statistical characteristics of excessive disparities, including mean of disparity and range of disparity [33], horizontal disparity and vertical disparity [34], zone of comfort and depth of focus (DoF) [35]. However, the existing 3D quality metrics evaluate the 3D image quality or visual comfort/discomfort separately, without fully taking both aspects into account.

D. Motivation of This Work

From the analyses of the above related works, we have the following summaries: 1) Content loss and geometric distortion are two major distortions of retargeted images, where the geometric distortion in the edge region and the content loss in the important region are more serious than other regions, as shown by an example in Fig. 1(b) associated with the purple rectangle of Fig. 1(a) and the blue rectangle of Fig. 1(c) associated with the red rectangle of Fig. 1(a), respectively; 2) The grid-based methods are demonstrated to be effective in characterizing the geometric distortion of retargeted images [12-14]; 3) The 3D image quality method that considers 3D perceptual properties is more reasonable than other methods; 4) Most existing methods adopted disparity statistical characteristics to predict visual comfort/discomfort of stereoscopic images [33-35]; 5) The existing 3D perceptual quality metrics separately evaluate the 3D image quality or visual comfort/discomfort. Although these quality metrics deliver moderate performances in evaluating the quality of 2D retargeted images or stereoscopic images, we target at proposing an effective quality metric for stereoscopic image retargeting motivated by the following considerations: 1) Different from the perception of 2D image in the stereopair, the perception of stereoscopic vision is actually the interaction of left and right views, which can be approximatively modeled by the perception of the intermediate view generated from the stereopair (similar to the virtual cyclopean vision). As shown by an example in Fig. 1(d), the distortion in the green rectangle of the intermediate view is more serious than the distortion in the orange rectangle of the left view. However, the existing SIRQA metrics [15-17] separately computed the quality scores of the left and right views without considering the 3D visual perceptual property. Thus, we attempt to evaluate the 3D image quality of retargeted stereopairs by investigating the quality of the intermediate view instead of the left or right view. 2) The grid-based methods use the uniform grids to measure the geometric distortion without considering the image structure, which may assign different textures on the same grid, as shown by an example in the orange rectangle of Fig. 2(a), where the background and foreground belong to the same grid. In this paper, we attempt to adopt the non-uniform grids (e.g., superpixel, as shown in Fig. 2(b)) as the bases for measuring the geometric distortion of retargeted stereopairs. 3) Besides

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3081259, IEEE Transactions on Multimedia

the statistical characteristics of disparities, we aim at exploring more effective features to predict visual comfort/discomfort of retargeted stereopairs based on the research of visual psychology, e.g., the perceptual conclusion drawn in [20] indicates that 3D perception on stereoscopic images tends to be more stable when monocular regions are more visually similar to the binocular background. To summarize the perception factors considered in the paper, the typical distortion types and evaluation cues in the retargeted stereopairs are listed in Table I.

(a) (b)

Fig. 2. Example of (a) uniform grids and (b) non-uniform grids.

TABLE I
TYPICAL ARTIFACTS IN THE RETARGETED STEREOPAIRS.

Evaluation cues	Distortion type
Image quality	Edge distortion
	General geometric distortion
	Important content loss
	General content loss
Depth perception quality	Visual discomfort
	Lack of depth sensation

III. PROPOSED METHOD

In this paper, we propose a new SIRQA method for the retargeted stereopairs based on image quality and depth quality measures. Fig. 3 shows the framework of the proposed method. First, the intermediate views for the source and retargeted stereopairs are generated from the respective left and right views. Then, the geometric-aware quality and content-aware quality are measured on the intermediate views to evaluate the geometric distortion and content loss induced by imperfect retargeting operators, respectively. Additionally, depth-aware quality measure is studied from two aspects, i.e., visual comfort/discomfort and depth sensation, to characterize the 3D perception quality of a retargeted stereopair. Finally, the extracted feature components are fused to obtain the overall quality score.

A. Image Quality Measure

1) Virtual view synthesis: As claimed in [18], stereoscopic image quality assessment aims to evaluate the quality of the true cyclopean view when a stereopair is stereoscopically presented. However, it is difficult to simulate the true cyclopean view as it requires to consider the display geometry, the presumed fixation, vergence and accommodation. Thus, we attempt to generate an intermediate view close to the quality of the true cyclopean view.

Since the camera parameters of the benchmark SIRQA databases are unknown, an optical flow-based view synthesis method is developed to generate a virtual view with high quality from an input stereopair. Specifically, the bidirectional optical flows $\hat{F}_{LR}$ from the left view to right view and $\hat{F}_{RL}$ from the right view to left view are first estimated using the SIFT-flow algorithm [37]. With the bidirectional optical flows, we pre-warp the left view I_L to the target location using the 'halfway' strategy [38] based on the optical flow $\hat{F}_{LR}$, denoted as forward warping, obtaining a pre-warped left view $\hat{V}_L$. Similarly, the pre-warped right view $\hat{V}_R$ can be obtained based on the optical flow $\hat{F}_{RL}$ by backward warping. To compensate for parallax in the two pre-warped views before blending, we calculate a parallax correction based on the bidirectional optical flows F_{LR} and F_{RL} between the two views [39]. Let x_L and x_R be matched pixels in the pre-warped left and right views, respectively, the local flow displacements can be calculated as:

$$\delta_{F_{LR}}(x_L) = x_R - x_L - F_{LR}(x_L) \tag{1}$$

$$\delta_{F_{RL}}(x_R) = x_L - x_R - F_{RL}(x_R) \tag{2}$$

Then, the corrected pixel locations are written as:

$$\hat{x}_L = x_L + \alpha \cdot \hat{F}_{LR}(x_L) \tag{3}$$

$$\hat{x}_R = x_L + (1-\alpha) \cdot \hat{F}_{RL}(x_R) \tag{4}$$

Finally, the virtual view I_V is synthesized using the linear blending as:

$$I_V(x_V) = (1-\alpha) \cdot I_L(\hat{x}_L) + \alpha \cdot I_R(\hat{x}_R) \tag{5}$$

where the blending weight α is set as a default value of 0.5 for generating the intermediate view. The generated intermediate view is shown by an example in Fig. 1(a). It can be seen that our view synthesis method is able to generate the intermediate view with high quality. Unlike the cyclopean view generated by fusing the stereopair, the disparity map and energy responses simulating the binocular rivalry property [18], our intermediate view is synthesized to primarily reflect the geometric inconsistency between the left and right retargeted views induced by the imperfect stereoscopic image retargeting process.

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3081259, IEEE Transactions on Multimedia

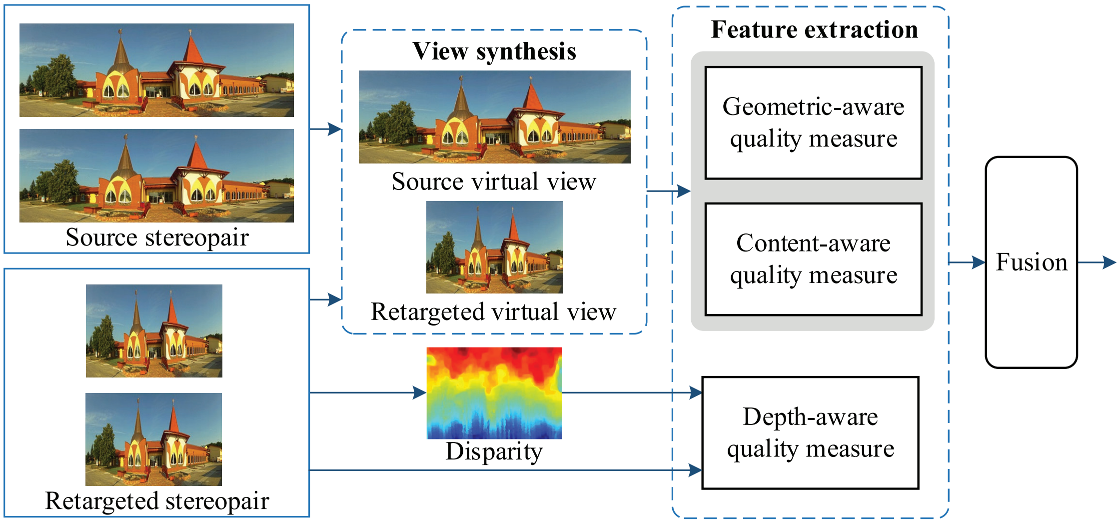


Fig. 3. Framework of the proposed method.

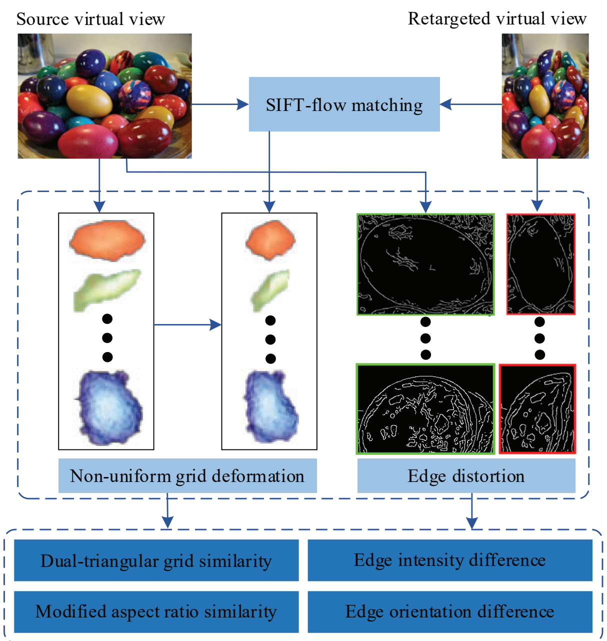


Fig. 4. Framework of the geometric-aware quality measure.

2) Geometric-aware quality measure: As shown in Fig. 1(c) and (d), geometric distortion is one of the major factors leading to the quality degradation of the retargeted stereopairs. In order to effectively measure the geometric distortion, as illustrated in Fig. 4, four geometric-aware features, including Dual-Triangular Grid Similarity (DTGS), Modified Aspect Ratio Similarity (MARS), Edge Intensity Difference (EID) and Edge Orientation Difference (EOD) are extracted from the source and retargeted intermediate views.

In contrast to the existing metrics [12-14] that used the uniform grid-based method to estimate the geometric distortion, we employ the superpixel-based approach to establish the non-uniform grid distribution, which is more coincident with the perception of image structure. To be specific, we adopt the Simple Linear Iterative Clustering (SLIC) [40] to detect the superpixels of the source intermediate view, and the corresponding results of the retargeted intermediate view are obtained based on the optical flow between the source and retargeted intermediate views. For the extraction of DTGS, the rectangular coordinate with the center of a superpixel as origin is first set up for the superpixel. As shown in Fig. 5, let $\upsilon(x_\ell^1, y_\ell^1)$, $\upsilon(x_\ell^2, y_\ell^2)$, $\upsilon(x_\ell^3, y_\ell^3)$ and $\upsilon(x_\ell^4, y_\ell^4)$ be four intersection points of the axis and the boundary for the ℓ-th superpixel in the source intermediate view, and let $\upsilon(\tilde{x}_\ell^1, \tilde{y}_\ell^1)$, $\upsilon(\tilde{x}_\ell^2, \tilde{y}_\ell^2)$, $\upsilon(\tilde{x}_\ell^3, \tilde{y}_\ell^3)$ and $\upsilon(\tilde{x}_\ell^4, \tilde{y}_\ell^4)$ be the matched intersection points for the corresponding superpixel in the retargeted intermediate view, where the first and second triangles are obtained by dividing the quadrilateral along the ordinate axis, the similarity between the first (second) source triangle and the first (second) retargeted triangle is calculated as:

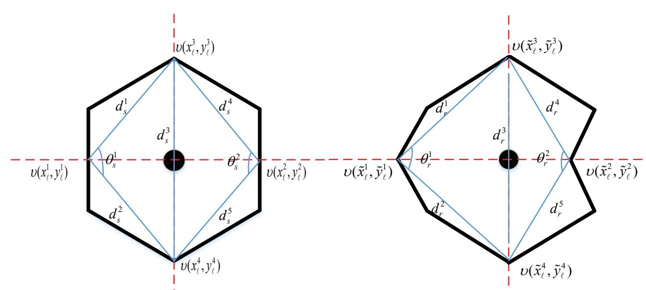


Fig. 5. Dual-triangular grid in the superpixel.

$$g_1 = \frac{2 \cdot d_1 \cdot d_2}{d_1^{\,2} + d_2^{\,2} + C} \exp\left(-\gamma\left(\cos\left(\theta_s^1\right) - \cos\left(\theta_r^1\right)\right)\right) \tag{6}$$

$$g_2 = \frac{2 \cdot d_4 \cdot d_5}{d_4^{\,2} + d_5^{\,2} + C} \exp\left(-\gamma\left(\cos\left(\theta_s^2\right) - \cos\left(\theta_r^2\right)\right)\right) \tag{7}$$

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3081259, IEEE Transactions on Multimedia

where $d_i = d_r^i / d_s^i$ $(i = 1,2,\ldots,5)$, θ_s^1 (θ_s^2) and θ_r^1 (θ_r^2) are the angles of the first (second) source triangle and the first (second) retargeted triangle, respectively, γ is empirically set as 0.3 to balance the translation (scaling) and rotation distortions of a superpixel, and C is a small constant to avoid the division by zero. We set $C = 10^{-6}$ in the experiment. Then, the DTGS is defined as:

$$f_1 = \frac{1}{2}\sum_{\ell}\omega_\ell \cdot \left(g_1(\ell) + g_2(\ell)\right) \Big/ \sum_{\ell}\omega_\ell \tag{8}$$

where ω_ℓ denotes the average saliency value of the ℓ-th superpixel obtained by GBVS algorithm [41]. As revealed in our previous work [26], the vertex-based method and ARS are complementary to each other in detecting the change of grids induced by translation, scaling and rotation. Thus, we compute the MARS of the superpixel as a supplement to the DTGS:

$$g = \frac{2 \cdot w \cdot h}{w^2 + h^2 + C}\exp\left(-\gamma(u-1)^2\right) \tag{9}$$

$$f_2 = \sum_{\ell}\omega_\ell \cdot g(\ell) \Big/ \sum_{\ell}\omega_\ell \tag{10}$$

where $w = w_r / w_s$, $h = h_r / h_s$, in which w_φ and h_φ ($\varphi \in \{r,s\}$) are the maximum width and height of the superpixel, respectively. In comparison to the homogeneous regions, distortions in the edges are more conspicuous to HVS. Therefore, the edge intensity and orientation differences between the source and retargeted intermediate views are calculated to measure the degree of edge distortion [42]. Specifically, for an input grayscale I, the edge intensity map is defined as:

$$M = \frac{\left|G_x + G_y\right| \cdot E}{2} \tag{11}$$

$$J = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{12}$$

where $G_x = I \otimes J$ and $G_y = I \otimes J^T$ are the horizontal and vertical gradients, respectively. $\otimes$ and T denote the convolution and transpose operators, respectively. The edge map E is detected using the Canny operator [43]. The edge orientation map is computed as:

$$O = \arctan\frac{180° \cdot G_y}{\pi \cdot G_x} \tag{13}$$

Then, the EID and EOD between the source and retargeted intermediate views are calculated as:

$$f_3 = \sum_{i=1}^{10}\left|\mathcal{H}_r^M(i) - \mathcal{H}_s^M(i)\right| \tag{14}$$

$$f_4 = \sum_{i=1}^{10}\left|\mathcal{H}_r^O(i) - \mathcal{H}_s^O(i)\right| \tag{15}$$

where $\mathcal{H}_\varphi^M$ and $\mathcal{H}_\varphi^O$ ($\varphi \in \{r,s\}$) are the edge intensity and orientation histograms, respectively. The range of edge intensity or edge orientation is quantized into 10 equally spanned bins. As a result, the final geometric-aware feature component is represented as $\mathbf{F}_{GA} = [f_1, f_2, f_3, f_4]$.

3) Content-aware quality measure: During the SIR process, image resolution of the source stereopair is adjusted to a target resolution by unavoidably discarding partial image contents. Thus, it is necessary to evaluate the information loss/preservation in the SIR results. In this paper, we estimate the information loss/preservation by calculating the statistical distance/similarity between the source and retargeted intermediate views from the global and local perspectives, respectively. Fig. 6 shows the framework of the content-aware quality measure. Let S be the normalized saliency map detected from the intermediate view using GBVS algorithm [41], the important content preservation feature can be calculated as:

$$B_\varphi(x,y) = \begin{cases} 1, & \sum_{(x,y)\in\Omega_k} S(x,y) > T,\ k = 1,2,\ldots,L \\ 0, & \text{Others} \end{cases} \tag{16}$$

$$f_5 = \frac{2 \cdot A_s \cdot A_r}{A_s^2 + A_r^2 + C},\quad A_\varphi = \sum B_\varphi \tag{17}$$

where Ω_k denotes the set of pixels in the k-th superpixel, L is the number of superpixels in the intermediate view, and V_{th} is a threshold to divide the image into important and non-important regions. Here, V_{th} is empirically set to 0.25.

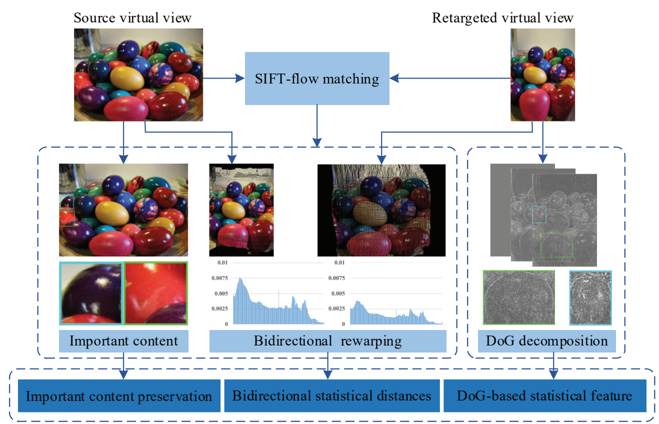


Fig. 6. Framework of the content-aware quality measure.

For another global perspective, bidirectional statistical distances between the source and retargeted intermediate views are calculated according to the forward rewarping and backward rewarping [26], respectively, as:

$$f_6 = \sum_{i=1}^{255}\left|\mathcal{H}_r(i) - \vec{\mathcal{H}}_s(i)\right| \tag{18}$$

$$f_7 = \sum_{i=1}^{255}\left|\overleftarrow{\mathcal{H}}_r(i) - \mathcal{H}_s(i)\right| \tag{19}$$

where $\mathcal{H}_s$ and $\mathcal{H}_r$ are the normalized color histograms of the source and retargeted intermediate views, respectively, and $\vec{\mathcal{H}}_s$ and $\overleftarrow{\mathcal{H}}_r$ denote the normalized color histograms of the forward and backward rewarped images, respectively.

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3081259, IEEE Transactions on Multimedia

Additionally, the above quality-aware features extracted from the intermediate views are the relative distance/similarity between two intermediate views without involving the intermediate view itself, but the quality of the retargeted intermediate view may indirectly reflect the consistency/ inconsistency between the left and right retargeted images. Since the Difference-of-Gaussian (DoG) decomposition is a validated effective way to capture the edge and texture characteristics of the scenes [44], the DoG-based statistical features are computed to evaluate the naturalness of the retargeted intermediate view. Specifically, for an input grayscale I associated with the retargeted intermediate view, the Gaussian low-pass filter is adopted for the scale space representation as:

$$F_k(x,y) = I(x,y) \otimes G(x,y,\sigma_k),\ k \in [0,3] \tag{20}$$

where σ_k denotes the standard deviation of the Gaussian model at the k-th scale. In this paper, $\sigma_1 = 0.3$, $\sigma_2 = 0.5$ and $\sigma_3 = 0.8$ are set empirically, and k=0 denotes the original scale. Then, the k-th DoG image is defined as:

$$D_k(x,y) = F_k(x,y) - F_{k+1}(x,y),\ k \in [0,2] \tag{21}$$

For each DoG image, the MSCN coefficients can be calculated by the local mean subtraction and divisive normalization [45]:

$$\hat{D}_k(i,j) = \frac{D_k(i,j) - \mu_k(i,j)}{\sigma_k(i,j)+1} \tag{22}$$

where $\mu_k(i, j)$ and $\sigma_k(i, j)$ denote the local mean and standard deviations, and the local window size of 7×7 is set in the calculation. Finally, the standard deviation of the achieved MSCN coefficients associated with each DoG image is computed as the statistical feature, denoted as f_8, f_9 and f_{10}. As a result, the final content-aware feature component is represented as $\mathbf{F}_{CA} = [f_5, f_6, f_7, f_8, f_9, f_{10}]$.

B. Depth-aware Quality Measure

Visual comfort/discomfort and depth sensation are another two main aspects in the 3D perception of a retargeted stereopair [17]. Firstly, we characterize the visual comfort/discomfort based on the following factors: 1) The binocular disparity that exceeds the tolerated value may lead to the failure in binocular fusion; 2) Since the object which produces the crossed disparity is perceived in front of the screen while the object which produces the uncrossed disparity is perceived behind the screen, the crossed disparity plays an important role in affecting the visual comfort; 3) The viewer will feel visual discomfort if the disparity does not lie within the visual comfort zone [46], e.g., [-1°, 1°] of disparity angle. Considering the above aspects, the corresponding statistical features, including variance of pixel disparity (f_{11}), mean of pixel disparity in the visual discomfort zone (f_{12}), range of angular disparity (f_{13}) and mean of crossed angular disparity (f_{14}), are calculated as:

$$f_{11} = \frac{1}{W \cdot H} \sum_{(x,y)} \left[D_p(x,y) - u_p \right]^2 \tag{23}$$

$$f_{12} = \frac{1}{A_1} \sum_{(x,y)\in\Omega_d} \left| D_p(x,y) \right| \tag{24}$$

$$f_{13} = \max\{D_a(x,y)\} - \min\{D_a(x,y)\} \tag{25}$$

$$f_{14} = \frac{1}{A_2} \sum_{(x,y)\in\Omega_c} \left| D_a(x,y) \right| \tag{26}$$

where $D_p(x, y)$ and $D_a(x, y)$ denote the pixel disparity value and angular disparity value at the pixel position (x, y), respectively, W and H are the width and height of the disparity map, respectively, μ_p is the mean of the pixel disparity map, Ω_d and Ω_c denote the index sets corresponding to the visual discomfort zone and the crossed (negative) disparities, respectively, A_1 and A_2 denote the number of pixels in Ω_d and Ω_c, respectively.

Secondly, as revealed in [17], monocular regions (e.g, the occluded region and the region out of field-of-view (Out-FOV)), can reflect depth sensation. However, the SIR process may change the size of the two regions, which will destroy the depth perception of the retargeted stereopair. To reflect the distortion, the statistical feature is calculated as:

$$f_{15} = \frac{1}{W \cdot H}(A_3 + A_4) \tag{27}$$

where A_3 and A_4 denote the number of pixels in the occluded regions and the regions Out-FOV, respectively. The occluded regions and the regions Out-FOV are detected using the method in [47]. Moreover, the perceptual experiments in [20] show that perception tends to be more stable when monocular regions are more visually similar to the binocular background. Thus, the statistical gradient similarity between the monocular regions and binocular background in the retargeted stereopair is calculated to measure the perception stability:

$$f_{16} = \frac{2 \cdot \kappa_m(\bar{G}) \cdot \kappa_g(\bar{G})}{\left|\kappa_m(\bar{G})\right|^2 + \left|\kappa_g(\bar{G})\right|^2 + C} \tag{28}$$

where $\bar{G} = \left|G_x + G_y\right|/2$ is the gradient map of the left retargeted view, and κ_m and κ_g denote the skewness statistics associated with the monocular regions and the binocular background, respectively. As a result, the final depth-aware feature component is represented as $\mathbf{F}_{DA} = [f_{11}, f_{12}, f_{13}, f_{14}, f_{15}, f_{16}]$.

C. Quality Evaluation

TABLE II
SUMMARY OF THE EXTRACTED 16-DIM FEATURES.

Feature Types	Symbol	Feature Components
Geometric-aware features	$\mathbf{F}_{GA}$	f_1 - f_4
Content-aware features	$\mathbf{F}_{CA}$	f_5 - f_{10}
Depth-aware features	$\mathbf{F}_{DA}$	f_{11} - f_{16}

With the extracted feature components $\mathbf{F}_{GA}$, $\mathbf{F}_{CA}$ and $\mathbf{F}_{DA}$, as shown in Table II, we first map these feature components from feature space to quality space using support vector regression (SVR) with the radial basis function (RBF) kernel to learn a quality predictor. Then, the quality predictor is utilized to

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3081259, IEEE Transactions on Multimedia

TABLE III
BASIC INTRODUCTION OF THE USED BENCHMARK DATABASES

Databases	Source stereopairs	Retargeted stereopairs	Retargeting scale	SIR operators
NBU-SIRQA [17]	45	720	0.25, 0.5	Monocular Seam Carving [21], Monocular Scale and Stretch [48], Content Persistent Cropping [36], Stereo Scaling, Geometrically Consistent Stereo Seam Carving [1], Visual Attention Guided Seam Carving [24], QoE-guided Warping [4], Single-layer Warping [2]
SIRD [16]	100	400	0.3	Stereo Cropping, Stereo Seam Carving [1], Stereo Scaling, Stereo Multi-operator [49]

predict the quality score of the retargeted stereopair. Note that other machine learning methods can also be used for quality pooling, we conduct the related experiments to investigate the impact of different pooling methods in the section that follows.

IV. EXPERIMENTS

A. Databases and Experimental Protocols

In the experiment, two benchmark databases, NBU SIRQA [17] and SIRD [16], are used to test the performance of the proposed method. The former database contains 45 source stereopairs and 720 retargeted stereopairs created by eight SIR operators, and two retargeting scales (25% and 50%) are included. The latter database consists of 100 source stereopairs and 400 SIR results produced using four SIR operators, and the retargeting scale is 30%. Both databases provide the mean opinion score (MOS) for each retargeted stereopair as ground truth of the image quality, where a larger MOS means the better image quality. The basic introduction of the databases is summarized in Table III. As suggested in [50], the performance of an objective IQA metric can be evaluated from two aspects, i.e., prediction accuracy and prediction monotonicity. In this work, we adopt three common performance criteria, i.e. Pearson Linear Correlation Coefficient (PLCC), Spearman Rank order Correlation Coefficient (SRCC) and Root Mean Square Error (RMSE), to benchmark the IQA metrics. PLCC and RMSE are used to measure the prediction accuracy, while SRCC is used to measure the prediction monotonicity. A metric with higher PLCC and SRCC, and lower RMSE is deemed to have good performance. For each database, we randomly divide it into two non-overlapping image subsets: 80% of the data for training while the rest 20% for testing. The average result after 1000 iterations is reported.

B. Performance Comparison

We compare the proposed method with eight IRQA metrics on NBU SIRQA database, including: five metrics designed for 2D retargeted images (SIFT-flow [37], BDS [51], EMD [52], HCDL [53] and ARS [12]), and three metrics designed for retargeted stereopairs (Liu's method [15], Zhou's method [16] and GDIL [17]). The results of the whole database and the subsets associated with the individual retargeting scale are presented in Table IV. From the table, we have the following observations: 1) Most 3D metrics (i.e., Zhou's method [16] and GDIL [17]) perform better than 2D metrics own to considering the 3D perceptual properties. 2) The best performance of these metrics is achieved by GDIL [17], which is still lower than our method owing to adopting the uniform grid-based method which ignores the structure attribute of the image. 3) The results on the whole database show higher performance than those on the two subsets due to the more obvious difference between the retargeted stereopairs associated with different retargeting scales than that associated with the same retargeting scale. Moreover, we have plotted in Fig. 7 the SRCC values of the competitive methods for retargeted stereopairs associated with each of 45 source stereopairs on NBU SIRQA database. It can be seen that our method delivers higher performance than most competitive methods, or it is comparable to GDIL for some groups. Overall, our method achieves the best performance on the whole database and the two subsets, which validates the effectiveness of the newly adopted intermediate view-based method and the superpixel-based approach.

Table V further summarizes the experimental results on SIRD database. It is observed from the table that both 2D IRQA and 3D IQA design for traditional images (i.e. 3DAVM [54] and 3DVDP [55]) cannot well evaluate the quality of retargeted stereopairs due to the fact that the former fails to handle the complicated 3D perceptual task while the latter is unable to capture the geometric distortion and content loss in retargeted stereopairs. The proposed method delivers the best performance in terms of both prediction accuracy and prediction monotonicity against other competing metrics, including 2D IRQA, 3D IQA design for traditional images and SIRQA. It indicates that the combination of geometric-aware, content-aware, and depth-aware quality measures can work well on retargeted stereopairs.

C. Impact of Training Set

To explore whether the quality prediction performance of the proposed method relies heavily on the training data, we conduct the experiments by changing the percentage of the training set size which varies from 20% to 80% and the remaining images are used for testing, and the average performance indices across 1000 random trials are presented in Table VI. Observations show that the proposed method still achieves relatively good performance on NBU SIRQA and

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3081259, IEEE Transactions on Multimedia

TABLE IV
PERFORMANCE OF DIFFERENT METRICS ON NBU SIRQA DATABASE

Metric	25% retargeting scale			50% retargeting scale			All		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
SIFT-flow [37]	0.1753	-0.1321	10.7355	0.2553	-0.2257	9.9736	0.1019	0.0542	14.8731
BDS [51]	0.2112	-0.1519	10.6584	0.0814	-0.1561	10.2812	0.2791	0.2751	14.3882
EMD [52]	0.1778	-0.1619	10.7306	0.1828	-0.1545	10.1416	0.3480	0.3858	14.0167
HCDL [53]	0.4157	0.3614	9.9620	0.3823	0.3599	9.5425	0.7261	0.7215	11.8422
ARS [12]	0.4792	0.4492	9.6887	0.4993	0.4799	8.9380	0.7833	0.7745	9.2950
Liu [15]	0.3887	0.3423	9.7728	0.3884	0.3522	9.4414	0.6138	0.5910	12.0756
Zhou [16]	0.6054	0.5616	8.6625	0.6202	0.5813	8.1511	0.8146	0.8054	8.6338
GDIL [17]	0.6150	0.5685	8.5868	0.6555	0.6315	7.7866	0.8371	0.8315	8.1083
Proposed	**0.6638**	**0.6459**	**8.2885**	**0.7367**	**0.7165**	**6.9256**	**0.8554**	**0.8502**	**7.7562**

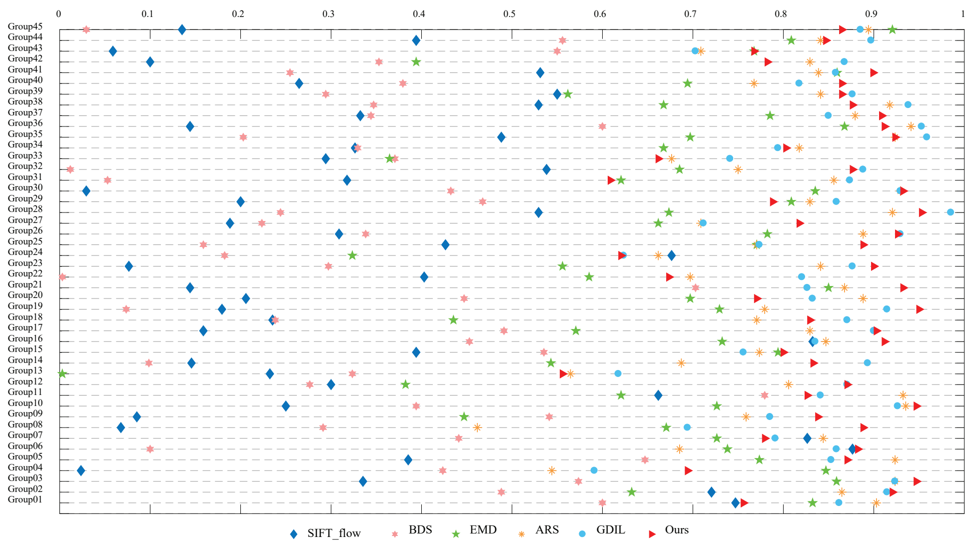


Fig. 7. Comparisons of the SRCC values of the competitive methods for retargeted stereopairs associated with each of 45 source stereopairs on NBU SIRQA database.

TABLE V
PERFORMANCE OF DIFFERENT METHODS ON SIRD DATABASE

Metric	Type	PLCC	SRCC	RMSE
SIFT-flow [37]	2D	0.2054	0.1625	1.0559
EMD [52]	2D	0.4892	0.4651	0.9384
HDPM [56]	2D	0.7757	0.7607	0.6767
3DAVM [54]	3D	0.5737	0.5699	0.8705
3DVDP [55]	3D	0.5768	0.5847	0.8781
Liu [15]	3D	0.3991	0.3960	0.8957
Zhou [16]	3D	0.8670	0.8494	0.5237
GDIL [17]	3D	0.8417	0.8041	0.5097
Proposed	**3D**	**0.8827**	**0.8561**	**0.4435**

SIRD databases even only 50% of the images are used for training, and the performance of the proposed method does not change drastically with the reduction of training data, which coincides with the conclusions drawn from the learning-based approaches [57,58].

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3081259, IEEE Transactions on Multimedia

TABLE VI
PERFORMANCE RESULTS WITH DIFFERENT TRAINING SET SIZES

Train-Test	NBU SIRQA			SIRD		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
20%-80%	0.7749	0.7748	9.4410	0.8282	0.8090	0.5360
30%-70%	0.7996	0.7992	8.9628	0.8491	0.8277	0.5053
40%-60%	0.8152	0.8139	8.6394	0.8615	0.8395	0.4858
50%-50%	0.8281	0.8260	8.3672	0.8693	0.8465	0.4724
60%-40%	0.8388	0.8357	8.1254	0.8749	0.8514	0.4613
70%-30%	0.8488	0.8444	7.9019	0.8800	0.8550	0.4516
80%-20%	0.8554	0.8502	7.7562	0.8827	0.8561	0.4435

TABLE VII
PERFORMANCE COMPARISON OF DIFFERENT POOLING METHODS ON NBU SIRQA DATABASE

Criteria	RF	ELM	L-SVR	Poly-SVR	RBF-SVR
PLCC	0.8585	0.8481	0.8215	0.7729	0.8554
SRCC	0.8499	0.8442	0.8159	0.7861	0.8502
RMSE	7.6054	7.9361	8.5196	9.3754	7.7562

D. Impact of Different Pooling Methods

To investigate the impact of different quality pooling methods on the performance of the proposed method, we use five different pooling schemes to fuse into the extracted features, including Random Forest (RF), Extreme Learning Machine (ELM), SVR with the linear kernel (L-SVR), SVR with the polynomial kernel (Poly-SVR), and SVR with the RBF kernel (RBF-SVR), and the results are shown in Table VII. The table shows that RF, ELM and RBF-SVR deliver better performance than L-SVR and Poly-SVR. Although RF and RBF-SVR have the similar performance, the two schemes cost 124.092s and 63.058s during 1000 iterations train-test process on a personal computer with Intel Core i5-9400 CPU @2.9 GHz and an 8 GB RAM, respectively. Therefore, RBF-SVR can be adopted by considering prediction accuracy, prediction monotonicity and computational complexity comprehensively.

E. Ablation Study

To separately investigate the contribution of each individual feature component as well as their mutual effects, we conduct the ablation study on the two databases. The experimental results are summarized in Table VIII, and the detailed results for each feature are listed in Table IX. From the table, we have two observations: Firstly, each component makes positive contribution to the overall performance, which indicates that these components have complementary information. Secondly, the contribution of the components is not the same, where the scheme only using geometric-aware component $\mathbf{F}_{GA}$ or content-aware component $\mathbf{F}_{CA}$ delivers better performance than the scheme only using depth-aware component $\mathbf{F}_{DA}$, which demonstrates that geometric distortion and information loss are the main factors leading to quality degradation of retargeted stereopairs.

E. Further Discussion

In this paper, we propose a quality evaluation method for retargeted stereopairs by combining image quality and depth perception measures. Although the proposed method has delivered better performance than the existing metrics designed for 2D retargeted images and the state-of-the-art quality metrics specially designed for retargeted stereopairs, the following issues still need to be considered in the future work:

1) The quality of synthesized intermediate views largely depends on the performance of virtual view synthesis method, which further affects the accuracy of image quality measures. Therefore, the more effective virtual view synthesis method is expected to promote the performance.

2) In addition to image quality and depth perception quality, image aesthetics is another potential factor affecting the overall perceptual quality of retargeted stereopairs. Towards a more powerful quality metric for retargeted stereopairs, subjective or objective aesthetics evaluation of retargeted stereopairs may be considered in the future research.

V. CONCLUSION

In this paper, we present a quality assessment method for retargeted stereopairs by jointly measuring image quality on the synthesized intermediate view and the depth perception quality from two aspects, i.e., visual comfort/discomfort and depth sensation. Specifically, the superpixel-based and edge-based features between the source and retargeted intermediate views are extracted to evaluate the geometric distortion, while the

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3081259, IEEE Transactions on Multimedia

TABLE VIII
PERFORMANCE OF DIFFERENT COMPONENTS ON THE TWO DATABASES

Combination			NBU SIRQA			SIRD		
$\mathbf{F}_{GA}$	$\mathbf{F}_{CA}$	$\mathbf{F}_{DA}$	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
√			0.7795	0.7600	9.3498	0.7081	0.685	0.6678
	√		0.7871	0.7791	9.2503	0.7838	0.7561	0.5877
		√	0.4418	0.4009	13.3760	0.6941	0.6879	0.6824
√	√		0.8232	0.8121	8.4230	0.8433	0.8158	0.5007
√		√	0.8054	0.7951	8.8386	0.8435	0.8202	0.5088
	√	√	0.8126	0.8117	8.6853	0.8393	0.8151	0.5134
√	√	√	0.8554	0.8502	7.7562	0.8827	0.8561	0.4435

TABLE IX
PERFORMANCE OF EACH COMPONENT ON THE TWO DATABASES

Feature components		NBU SIRQA			SIRD		
		PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
$\mathbf{F}_{GA}$	f_1	0.1717	0.1536	14.6832	0.4000	0.2846	0.8591
	f_2	0.6312	0.5826	11.5668	0.4291	0.3401	0.8551
	f_3	0.6224	0.5844	11.6615	0.6227	0.5864	0.7404
	f_4	0.3592	0.3329	13.9511	0.2443	0.2336	0.9198
$\mathbf{F}_{CA}$	f_5	0.0926	0.1234	14.8034	0.3678	0.1039	0.8812
	f_6	0.2463	0.2398	14.4167	0.6529	0.6407	0.7182
	f_7	0.7240	0.6964	10.2629	0.6947	0.7483	0.6685
	f_8	0.1479	0.0717	14.7036	0.3812	0.3646	0.8764
	f_9	0.2595	0.2214	14.3889	0.3014	0.2687	0.9051
	f_{10}	0.2497	0.2171	14.4352	0.2789	0.2542	0.9132
$\mathbf{F}_{DA}$	f_{11}	0.1376	0.0836	14.7726	0.2640	0.3740	0.9175
	f_{12}	0.1442	0.1337	14.7134	0.2898	0.2960	0.9096
	f_{13}	0.2303	0.1975	14.4632	0.3689	0.3564	0.8826
	f_{14}	0.1655	0.1125	14.6746	0.1128	0.0197	0.9458
	f_{15}	0.2421	0.1708	14.4254	0.4744	0.4561	0.8355
	f_{16}	0.2018	0.1672	14.6462	0.1610	0.1050	0.9376

local statistical similarity and global statistical distance are calculated to reveal the content loss. Meanwhile, several depth-aware features are combined to characterize the visual comfort/discomfort and depth sensation. Experimental results verify the superiority of the proposed method against the competing metrics on the NBU SIRQA and SIRD databases.

REFERENCES

[1] T. Dekel Basha, Y. Moses, and S. Avidan, "Stereo seam carving a geometrically consistent approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 10, pp. 2513–2525, Oct. 2013.

[2] C.-H. Chang, C.-K. Liang, and Y.-Y. Chuang, "Content-aware display adaptation and interactive editing for stereoscopic images," *IEEE Transactions on Multimedia*, vol. 13, no. 4, pp. 589–601, Aug. 2011.

[3] B. Li, L. Duan, C. Lin, T. Huang, and W. Gao, "Depth-preserving warping for stereo image retargeting," *IEEE Transactions on Image Processing*, vol. 24, no. 9, pp. 2811–2826, Sep. 2015.

[4] F. Shao, W. Lin, W. Lin, Q. Jiang, and G. Jiang, "QoE-guided warping for stereoscopic image retargeting," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4790–4805, Oct. 2017.

[5] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Transactions on Graphics*, vol. 29, no. 6, article no. 160, 2010.

[6] Y. Liu, X. Luo, Y. Xuan, W. Chen, and X. Fu, "Image retargeting quality assessment," *Computer Graphics Forum*, vol. 30, no. 2, pp. 583–592, 2011.

[7] L. Ma, L. Xu, H. Zeng, K. N. Ngan, and C. Deng, "How does the shape descriptor measure the perceptual quality of the retargeting image?" *in Proc. of IEEE International Conference on Multimedia and Expo Workshops*, 2014, pp. 1–6.

[8] S. A. F. Oliveira, S. S. A. Alves, J. P. P. Gomes, and A. R. Rocha Neto, "A bi-directional evaluation-based approach for image retargeting quality assessment," *Computer Vision and Image Understanding*, vol. 168, pp. 172–181, Mar. 2018.

[9] Y. Fang, K. Zeng, Z. Wang, W. Lin, Z. Fang, and C. Lin, "Objective quality assessment for image retargeting based on structural similarity," *IEEE Journal of Emerging and Selected Topics in Circuits and Systems*, vol. 4, no. 1, pp. 95–105, Mar. 2014.

[10] C. Hsu, C. Lin, Y. Fang, and W. Lin, "Objective quality assessment for image retargeting based on perceptual geometric distortion and information loss," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 3, pp. 377–389, Jun. 2014.

[11] Y. Liang, Y. Liu, and D. Gutierrez, "Objective quality prediction of image retargeting algorithms," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 2, pp. 1099–1110, Feb. 2017.

[12] Y. Zhang, Y. Fang, W. Lin, X. Zhang, and L. Li, "Backward registration based aspect ratio similarity for image retargeting quality assessment,"

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3081259, IEEE Transactions on Multimedia

IEEE Transactions on Image Processing, vol. 25, no. 9, pp. 4286–4297, Sep. 2016.

[13] M. Karimi, S. Samavi, N. Karimi, S. M. R. Soroushmehr, W. Lin, and K. Najarian, "Quality assessment of retargeted images by salient region deformity analysis," *Journal of Visual Communication and Image Representation*, vol. 43, pp. 108–118, Feb. 2017.

[14] Y. Niu, S. Zhang, Z. Wu, T. Zhao, and W. Chen, "Image retargeting quality assessment based on registration confidence measure and noticeability-based pooling," *IEEE Transactions on Circuits and Systems for Video Technology*, doi: 10.1109/TCSVT.2020.2998087.

[15] Y. Liu, L. Sun, and S. Yang, "Learning-based quality assessment of retargeted stereoscopic images," *in Proc. of IEEE International Conference on Multimedia and Expo*, 2016, pp. 1–6.

[16] Y. Zhou, W. Zhou, P. An, and Z. Chen, "Visual comfort assessment for stereoscopic image retargeting," *in Proc. of IEEE International Symposium on Circuits and Systems*, 2018, pp. 1–5.

[17] Z. Fu, F. Shao, Q. Jiang, M. Chao, and Y. Ho, "Subjective and objective quality assessment for stereoscopic 3D image retargeting," *IEEE Transactions on Multimedia*, doi: 10.1109/TMM.2020.3008054.

[18] M.-J. Chen, C.-C. Su, D.-L. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for binocular rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, Oct. 2013.

[19] L. Li, Y. Li, J. Wu, L. Ma, and Y. Fang, "Quality evaluation for image retargeting with instance semantics," *IEEE Transactions on Multimedia*, doi: 10.1109/TMM.2020.3016124.

[20] Z. Başgöze, D. N. White, J. Burge, and E. A. Cooper, "Natural statistics of depth edges modulate perceptual stability," *Journal of Vision*, vol. 20, no. 8, pp. 1–21, Aug. 2020.

[21] S. Avida and A. Shamir, "Seam carving for content-aware image resizing," *ACM Transactions on Graphics*, vol. 26, no. 3, article no. 10, 2007.

[22] S.-S. Lin, C.-H. Lin, I.-C. Yeh, S.-H. Chang, C.-K. Yeh, and T.-Y. Lee, "Content-aware video retargeting using object-preserving warping," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 10, pp. 1677–1686, Oct. 2013.

[23] Y. Chen, Y. Pan, M. Song, and M. Wang, "Improved seam carving combining with 3D saliency for image retargeting," *Neurocomputing*, vol. 151, pp. 645–653, Mar. 2015.

[24] F. Shao, W. Lin, W. Lin, G. Jiang, M. Yu, and R. Fu, "Stereoscopic visual attention guided seam carving for stereoscopic image retargeting," *Journal of Display Technology*, vol. 12, no. 1, pp. 22–30, Jan. 2016.

[25] S.-S. Lin, C.-H. Lin, S.-H. Chang, and T.-Y. Lee, "Object-coherence warping for stereoscopic image retargeting," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 5, pp. 759–768, May 2014.

[26] F. Shao, Z. Fu, Q. Jiang, G. Jiang, and Y. Ho, "Transformation-aware similarity measurement for image retargeting quality assessment via bidirectional rewarping," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, doi: 10.1109/TSMC.2019.2917496.

[27] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "Quality of experience model for 3DTV," *in Proc. of SPIE*, 2012, pp. 1–9.

[28] M-J Chen, C-C Su, D-K Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, Oct. 2013.

[29] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," *in Proc. of International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2010, pp. 1–6.

[30] K. Lee and S. Lee, "3D perception based quality pooling: Stereopsis, binocular rivalry, and binocular suppression," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 3, pp. 533–545, Apr. 2015.

[31] A. Maalouf and M.-C. Larabi, "CYCLOP: A stereo color image quality assessment metric," *in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 1161–1164.

[32] F. Shao, K. Li, W. Lin, G. Jiang, M. Yu, and Q. Dai, "Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 2971–2983, Oct. 2015.

[33] M. Lambooij, M. Fortuin, I. Heynderickx, and W. Ijsselsteijn, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *Journal of Imaging Science and Technology*, vol. 53, no. 3, pp. 1–14, May 2009.

[34] D. Kim and K. Sohn, "Visual fatigue prediction for stereoscopic image," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 2, pp. 231–236, Feb. 2011.

[35] Q. Jiang, F. Shao, G. Jiang, M. Yu, and Z. Peng, "Three-dimensional visual comfort assessment via preference learning," *Journal of Electronic Imaging*, vol. 24, no. 4, 2015, pp. 1–12.

[36] W. Wang, J. Shen, Y. Yu, and K. Ma, "Stereoscopic thumbnail creation via efficient stereo saliency detection," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 8, pp. 2014–2027, Aug. 2017.

[37] C. Liu, J. Yuen, and A. Torralba, "SIFT Flow: Dense correspondence across scenes and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978–994, May 2011.

[38] S. Niklaus and F. Liu, "Context-aware synthesis for video frame interpolation," *in Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1701–1710.

[39] L. Baker, S. Mills, S. Zollmann, and J. Ventura, "CasualStereo: Casual capture of stereo panoramas with spherical structure-from-motion," *in Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces*, 2020, pp. 782–790.

[40] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2281, Nov. 2012.

[41] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *in Proc. of Advances in Neural Information Processing Systems*, pp. 545–552, Dec. 2006.

[42] Y. Zhou, L. Li, K. Gu, Z. Lu, B. Chen, and L. Tang, "DIBR-synthesized image quality assessment via statistics of edge intensity and orientation," *IEICE Transactions on Information and Systems*, vol. 100, no. 8, pp. 1929–1933, Aug. 2017.

[43] J. F. Canny, "A computation approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 670–700, Dec. 1986.

[44] Y. Zhou, L. Li, S. Wang, J. Wu, Y. Fang, and X. Gao, "No-reference quality assessment for view synthesis using DoG-based edge statistics and texture naturalness," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4566–4579, Sep. 2019.

[45] Q. Wang, J. Chu, L. Xu, and Q. Chen, "A new blind image quality framework based on natural color statistic," *Neurocomputing*, vol. 173, pp. 1798–1810, Jan. 2016.

[46] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *Journal of Vision*, vol. 11, no. 8, pp. 1–29, 2011.

[47] T. Dekel Basha, Y. Moses, and S. Avidan, "Stereo seam carving: A geometrically consistent approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 10, pp. 2513–2525, Oct. 2013.

[48] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Transactions on Graphics*, vol. 27, no. 5, article no. 118, 2008.

[49] L. Zhu and Z. Chen, "Multi-operator stereoscopic image retargeting based on human visual comfort," accepted by *China Multimedia*, 2017.

[50] (Aug. 2003). VQEG, Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment—Phase II. [Online]. Available: http://www.vqeg.org/

[51] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," *in Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[52] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proc. of IEEE International Conference on Computer Vision*, 2009, pp. 460–467.

[53] Z. Karni, D. Freedman, and C. Gotsman, "Energy-based image deformation," *Computer Graphics Forum*, vol. 28, no. 5, pp. 1257–1268, 2009.

[54] J. Park, S. Lee, and A. C. Bovik, "3D visual discomfort prediction: Vergence, foveation, and the physiological optics of accommodation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 3, pp. 415–427, Jun. 2014.

[55] J. Park, H. Oh, S. Lee, and A. C. Bovik, "3D visual discomfort predictor: analysis of horizontal disparity and neural activity statistics," *IEEE*

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.

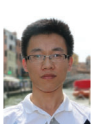
Transactions on Image Processing, vol. 24, no. 3, pp. 1101–1114, Mar. 2015.
[56] J. Lin, L. Zhu, Z. Chen, and X. Chen, "Objective quality assessment for image retargeting based on hybrid distortion pooled model," *in Proc. of International Workshop on Quality of Multimedia Experience*, 2015, pp. 1–6.
[57] Q. Jiang, F. Shao, W. Lin, and G. Jiang, "BLIQUE-TMI: Blind quality evaluator for tone-mapped images based on local and global feature analyses," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 2, pp. 323–335, Feb. 2019.
[58] G. Yue, C. Hou, and T. Zhou, "Blind quality assessment of tone-mapped images considering colorfulness, naturalness and structure," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3784–3793, Jul. 2018.

Xuejin Wang received the B.S. and M.S. degrees from Fuzhou University, Fuzhou, China, in 2012 and 2015, respectively. She is currently pursuing the Ph.D. degree at Ningbo University, Ningbo, China. Her current research interests include image/video processing and quality assessment.

Feng Shao (M'16) received his B.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively, all in Electronic Science and Technology. He is currently a professor in Faculty of Information Science and Engineering, Ningbo University, China. He was a visiting Fellow with the School of Computer Engineering, Nanyang Technological University, Singapore, from February 2012 to August 2012. He received 'Excellent Young Scholar' Award by NSF of China (NSFC) in 2016. He has published over 100 technical articles in refereed journals and proceedings in the areas of 3D video coding, 3D quality assessment, and image perception, etc.

Qiuping Jiang is currently an Associate Professor with the School of Information Science and Engineering, Ningbo University, Ningbo, China. He received the Ph.D. degree from Ningbo University in June 2018. From January 2017 to June 2018, he was a Visiting Student with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include image processing, visual perception modeling, and computer vision. He was a recipient of the JVCI 2017 Best Paper Award Honorable Mention as the first author. He is a reviewer for several prestigious journals and conferences, such as the IEEE T-NNLS, IEEE T-IP, IEEE T-CSVT, IEEE T-MM, IEEE T-SIPN, ICME, ICIP, etc.

Zhenqi Fu received the B.S. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2016. He received the M.S. degree from Ningbo University, Ningbo, China, in 2019. He is currently pursuing the Ph.D. degree with Xiamen University, Xiamen, China. His current research interests include machine learning, image processing and quality assessment.



Xiangchao Meng (M'18) received the B.S. degree in geographic information system from the Shandong University of Science and Technology, Qingdao, China, in 2012, and the Ph.D. degree in cartography and geography information system from Wuhan University, Wuhan, China, in 2017. He is currently a Lecturer with the Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, China. His research interests include variational methods and remote sensing image fusion.

Ke Gu (M'19) received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. He is currently a Professor with the Beijing University of Technology, Beijing, China. His research interests include environmental perception, image processing, quality assessment, and machine learning. He received the Best Paper Award from the IEEE Transactions on Multimedia (T-MM), the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo (ICME) in 2016, and the Excellent Ph.D. Thesis Award from the Chinese Institute of Electronics in 2016. He was the Leading Special Session Organizer in the VCIP 2016 and the ICIP 2017, and serves as a Guest Editor for the Digital Signal Processing (DSP). He is currently an Associate Editor for the IEEE ACCESS and IET Image Processing (IET-IPR), and an Area Editor for the Signal Processing Image Communication (SPIC). He is a Reviewer for 20 top SCI journals.

Yo-Sung Ho (SM'06–F'16) received the B.S. and M.S. degrees in electronic engineering from Seoul National University, Seoul, Korea, in 1981 and 1983, respectively, and the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 1990. He joined Electronics and Telecommunications Research Institute (ETRI), Daejon, Korea, in 1983. From 1990 to 1993, he was with Philips Laboratories, Briarcliff Manor, NY, where he was involved in development of the Advanced Digital High-Definition Television (AD-HDTV) system. In 1993, he rejoined the technical staff of ETRI and was involved in development of the Korean DBS digital television and high-definition television systems. Since 1995, he has been with Gwangju Institute of Science and Technology (GIST), Gwangju, Korea, where he is currently Professor of Information and Communications Department. His research interests include digital image and video coding, image analysis and image restoration, advanced video coding techniques, digital video and audio broadcasting, three-dimensional video processing, and content-based signal representation and processing. He is a fellow of IEEE.

1520-9210 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.