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—Stereo 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 Stereo 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
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 warping. 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
evaluated the quality separately. Maalouf
availability of binocular depth perception in the regions, and
image into binocular and monocular regions based on the
quality pooling (3DPS) model, which divided the stereoscopic
the important region are more serious than other regions, as
geometric distortion in the edge region and the content loss in
are two major distortions of retargeted images, where the
stereopairs generated by content persistent cropping \cite{36}, geometrically consistent stereo seam carving \cite{1} and stereo scaling \cite{17}, respectively. The images in
the second row are the intermediate views of the associated stereopairs.

**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
\cite{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 \cite{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 \cite{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
geo...
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.

![Fig. 2. Example of (a) uniform grids and (b) non-uniform grids.](image)

### TABLE I

<table>
<thead>
<tr>
<th>Typical Artifacts in the Retargeted Stereopairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation cues</td>
</tr>
<tr>
<td>Image quality</td>
</tr>
<tr>
<td>Edge distortion</td>
</tr>
<tr>
<td>General geometric distortion</td>
</tr>
<tr>
<td>Important content loss</td>
</tr>
<tr>
<td>General content loss</td>
</tr>
<tr>
<td>Depth perception quality</td>
</tr>
<tr>
<td>Visual discomfort</td>
</tr>
<tr>
<td>Lack of depth sensation</td>
</tr>
</tbody>
</table>

### 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:

$$\delta_{f_{LR}} (x_L) = x_R - x_L - \hat{F}_{LR} (x_L) \quad (1)$$

$$\delta_{f_{RL}} (x_R) = x_L - x_R - \hat{F}_{RL} (x_R) \quad (2)$$

Then, the corrected pixel locations are written as:

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

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

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

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

where the blending weight $\alpha$ 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.
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 \( \nu(x^i_s, y^i_s) \), \( \nu(x^2_s, y^2_s) \), \( \nu(x^3_s, y^3_s) \) and \( \nu(x^4_s, y^4_s) \) be four intersection points of the axis and the boundary for the \( i \)-th superpixel in the source intermediate view, and let \( \nu(x^1_r, y^1_r) \), \( \nu(x^2_r, y^2_r) \), \( \nu(x^3_r, y^3_r) \) and \( \nu(x^4_r, y^4_r) \) 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:

\[
g_1 = \frac{2 \cdot d_1 \cdot d_2}{d_1^2 + d_2^2 + C} \exp\left(-\gamma \left(\cos(\theta'_1) - \cos(\theta^1_s)\right)\right) \tag{6}
\]

\[
g_2 = \frac{2 \cdot d_1 \cdot d_2}{d_1^2 + d_2^2 + C} \exp\left(-\gamma \left(\cos(\theta'_2) - \cos(\theta^2_s)\right)\right) \tag{7}
\]
where \( d_i = d_i' / d_i'' \) (i = 1, 2, ..., 5), \( \theta_i^1 (\theta_i^2) \) and \( \theta_i' (\theta_i'') \) are the angles of the first (second) source triangle and the first (second) retargeted triangle, respectively. \( \gamma \) 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:

\[
f_i = \frac{1}{2} \sum_{\ell} \alpha_{\ell} \cdot (g_i(\ell) + g_{i'}(\ell)) / \sum_{\ell} \alpha_{\ell}
\]

where \( \alpha_{\ell} \) denotes the average saliency value of the \( \ell \)-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 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:

\[
B_{\varphi}(x, y) = \begin{cases} 
1, & \sum_{(i, j) \in \Omega_k} S(x, y) > T, \; k = 1, 2, ..., L \\
0, & \text{Others}
\end{cases} 
\]

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

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:

\[
f_6 = \sum_{i=1}^{255} \| H_s(i) - \tilde{H}_s(i) \| \\
f_7 = \sum_{i=1}^{255} \| \tilde{H}_s(i) - H_s(i) \|
\]

where \( H_s \) and \( \tilde{H}_s \) are the normalized color histograms of the source and retargeted intermediate views, respectively, and \( \tilde{H}_s \) denote the normalized color histograms of the forward and backward rewarped images, respectively.
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:

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

where $\sigma_k$ denotes the standard deviation of the Gaussian model at the $k$-th scale. In this paper, $\sigma_0 = 0.3, \sigma_1 = 0.5$ and $\sigma_3 = 0.8$ are set empirically, and $k=0$ denotes the original scale. Then, the $k$-th DoG image is defined as:

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

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

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

where $\mu_k(i, j)$ and $\sigma_k(i, j)$ denote the local mean and standard deviations, and the local window size of $7 \times 7$ 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 $f_{s5}, f_{s6}$ and $f_{s10}$. As a result, the final content-aware feature component is represented as $F_{CA} = [f_{s5}, f_{s6}, f_{s7}, f_{s8}, f_{s9}, f_{s10}]$.

### 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., $\pm1^\circ, \pm2^\circ$ 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:

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

$$f_{12} = \frac{1}{A_4} \sum_{(x,y), \Omega_2} |D_p(x, y)| \quad (24)$$

$$f_{13} = \max \{D(x, y), D(y, x)\} \quad (25)$$

$$f_{14} = \frac{1}{A_3} \sum_{(x,y), \Omega_2} |D(x, y)| \quad (26)$$

where $D(x, y)$ and $D(y, x)$ 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, $\mu_p$ is the mean of the pixel disparity map, $\Omega_2$ and $\Omega_3$ 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 $\Omega_2$ and $\Omega_3$, 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:

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

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:

$$f_{16} = \frac{2 \cdot \kappa_m (\tilde{G}) \cdot \kappa_c (\tilde{G})}{[\kappa_m (\tilde{G})]^2 + [\kappa_c (\tilde{G})]^2 + C} \quad (28)$$

where $\tilde{G} = [G_m + G_c]/2$ is the gradient map of the left retargeted view, and $\kappa_m$ and $\kappa_c$ 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 $F_{DA} = [f_{11}, f_{12}, f_{13}, f_{14}, f_{15}, f_{16}]$.

### C. Quality Evaluation

#### TABLE II

**SUMMARY OF THE EXTRACTED 16-DIM FEATURES.**

<table>
<thead>
<tr>
<th>Feature Types</th>
<th>Symbol</th>
<th>Feature Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometric-aware features</td>
<td>$F_{GA}$</td>
<td>$f_{s1}, f_{s2}$</td>
</tr>
<tr>
<td>Content-aware features</td>
<td>$F_{CA}$</td>
<td>$f_{s5}, f_{s6}$</td>
</tr>
<tr>
<td>Depth-aware features</td>
<td>$F_{DA}$</td>
<td>$f_{s11}, f_{s12}, f_{s13}, f_{s14}$</td>
</tr>
</tbody>
</table>

With the extracted feature components $F_{GA}, F_{CA}$ and $F_{DA}$, 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...
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 3DVPD [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 SIRDQA. 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
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].

![Graphical representation of performance metrics](image-url)

**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**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Type</th>
<th>PLCC</th>
<th>SRCC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT-flow [37]</td>
<td>2D</td>
<td>0.2054</td>
<td>0.1625</td>
<td>1.0559</td>
</tr>
<tr>
<td>EMD [52]</td>
<td>2D</td>
<td>0.4892</td>
<td>0.4651</td>
<td>0.9384</td>
</tr>
<tr>
<td>HDPM [56]</td>
<td>2D</td>
<td>0.7757</td>
<td>0.7607</td>
<td>0.6767</td>
</tr>
<tr>
<td>3DAVM [54]</td>
<td>3D</td>
<td>0.5737</td>
<td>0.5699</td>
<td>0.8705</td>
</tr>
<tr>
<td>3DVDP [55]</td>
<td>3D</td>
<td>0.5768</td>
<td>0.5847</td>
<td>0.8781</td>
</tr>
<tr>
<td>Liu [15]</td>
<td>3D</td>
<td>0.3991</td>
<td>0.3960</td>
<td>0.8957</td>
</tr>
<tr>
<td>Zhou [16]</td>
<td>3D</td>
<td>0.8670</td>
<td>0.8494</td>
<td>0.5237</td>
</tr>
<tr>
<td>GDIL [17]</td>
<td>3D</td>
<td>0.8417</td>
<td>0.8041</td>
<td>0.5097</td>
</tr>
<tr>
<td>Proposed</td>
<td>3D</td>
<td>0.8827</td>
<td>0.8561</td>
<td>0.4435</td>
</tr>
</tbody>
</table>
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 $F_{GA}$ or content-aware component $F_{CI}$ delivers better performance than the scheme only using depth-aware component $F_{DI}$, which demonstrates that geometric distortion and information loss are the main factors leading to quality degradation of retargeted stereopairs.

V. Conclusion

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.
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


---

**TABLE VIII**

PERFORMANCE OF DIFFERENT COMBINATIONS ON THE TWO DATABASES

<table>
<thead>
<tr>
<th>Combination</th>
<th>NBU SIRQA</th>
<th>SIRD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLCC</td>
<td>SRCC</td>
</tr>
<tr>
<td>F_{G4} F_{E4} F_{D4}</td>
<td>0.7795 0.7600 9.3498</td>
<td>0.7081 0.685 0.6678</td>
</tr>
<tr>
<td>F_{G4} F_{E4}</td>
<td>0.7871 0.7791 9.2503</td>
<td>0.7838 0.7561 0.5877</td>
</tr>
<tr>
<td>F_{G4} F_{D4}</td>
<td>0.4418 0.4009 13.3760</td>
<td>0.6941 0.6879 0.6824</td>
</tr>
<tr>
<td>F_{E4} F_{D4}</td>
<td>0.8232 0.8121 8.4230</td>
<td>0.8433 0.8158 0.5007</td>
</tr>
<tr>
<td>F_{G4} F_{D4}</td>
<td>0.8504 0.7951 8.8386</td>
<td>0.8435 0.8202 0.5088</td>
</tr>
<tr>
<td>F_{E4} F_{D4}</td>
<td>0.8126 0.8117 8.6853</td>
<td>0.8393 0.8151 0.5134</td>
</tr>
</tbody>
</table>

**TABLE IX**

PERFORMANCE OF EACH COMPONENT ON THE TWO DATABASES

<table>
<thead>
<tr>
<th>Feature components</th>
<th>NBU SIRQA</th>
<th>SIRD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLCC</td>
<td>SRCC</td>
</tr>
<tr>
<td>f_1</td>
<td>0.1717</td>
<td>0.1536</td>
</tr>
<tr>
<td>f_2</td>
<td>0.6312</td>
<td>0.5826</td>
</tr>
<tr>
<td>f_3</td>
<td>0.6224</td>
<td>0.5844</td>
</tr>
<tr>
<td>f_4</td>
<td>0.3592</td>
<td>0.3329</td>
</tr>
<tr>
<td>f_5</td>
<td>0.0926</td>
<td>0.1234</td>
</tr>
<tr>
<td>f_6</td>
<td>0.2463</td>
<td>0.2398</td>
</tr>
<tr>
<td>f_7</td>
<td>0.7204</td>
<td>0.6964</td>
</tr>
<tr>
<td>f_8</td>
<td>0.1479</td>
<td>0.0717</td>
</tr>
<tr>
<td>f_9</td>
<td>0.2595</td>
<td>0.2214</td>
</tr>
<tr>
<td>f_10</td>
<td>0.2497</td>
<td>0.2171</td>
</tr>
<tr>
<td>f_11</td>
<td>0.1376</td>
<td>0.0836</td>
</tr>
<tr>
<td>f_12</td>
<td>0.1442</td>
<td>0.1337</td>
</tr>
<tr>
<td>f_13</td>
<td>0.2303</td>
<td>0.1975</td>
</tr>
<tr>
<td>f_14</td>
<td>0.1655</td>
<td>0.1125</td>
</tr>
<tr>
<td>f_15</td>
<td>0.2421</td>
<td>0.1708</td>
</tr>
<tr>
<td>f_16</td>
<td>0.2018</td>
<td>0.1672</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature components</th>
<th>NBU SIRQA</th>
<th>SIRD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLCC</td>
<td>SRCC</td>
</tr>
<tr>
<td>f_{G4}</td>
<td>0.8054</td>
<td>0.7951</td>
</tr>
<tr>
<td>f_{E4}</td>
<td>0.8126</td>
<td>0.8117</td>
</tr>
<tr>
<td>f_{D4}</td>
<td>0.8554</td>
<td>0.8502</td>
</tr>
</tbody>
</table>


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.

Yu-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), Daejeon, 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.

Authorized licensed use limited to: Xiamen University. Downloaded on May 25,2021 at 15:03:08 UTC from IEEE Xplore. Restrictions apply.