Subjective and Objective Quality Assessment for Stereoscopic Image Retargeting

Zhenqi Fu , Feng Shao , *Member, IEEE,* Qiuping Jiang , *Senior Member, IEEE,*
Xiangchao Meng , *Member, IEEE,* and Yo-Sung Ho , *Fellow, IEEE*

Abstract—Binocular stereoscopic image retargeting (SIR) aims to adjust 3D images into target aspect ratios. In recent years, various SIR methods have been proposed, but there are few researches on visual quality assessment. As a consequence, we construct a benchmark stereoscopic image retargeting quality assessment database (NBU-SIRQA), which contains 720 stereoscopic retargeted images generated by eight representative SIR operators. Subjective test is conducted to obtain the mean opinion score (MOS) for each stereoscopic retargeted image. Additionally, we propose an objective SIRQA metric based on grid deformation and information loss (GDIL). The main idea of GDIL is to decompose the SIR operator into two transformations: monocular image retargeting transformation and viewpoint transformation. In each transformation, grid deformation and information loss are extracted simultaneously to represent image quality and 3D perception quality. Experimental results validated on our established NBU-SIRQA database show the superiority of our metric in measuring the quality of stereoscopic retargeted images over the existing approaches.

Index Terms—Stereoscopic image retargeting, quality assessment, monocular retargeting transformation, viewpoint transformation, grid deformation, information loss.

I. INTRODUCTION

WITH the popularity of stereoscopic 3D displays, stereoscopic image retargeting (SIR) brings great requirements for displaying images on different devices with specified aspect ratios and resolutions [1]. Especially, with the diversity of 3D display devices, such as phones, TVs and cinema screens, content adaptation via SIR should not only match the target aspect ratios and resolutions, but also adapt the different depth ranges. Although state-of-the-art SIR methods can handle the task for 3D images, it is still unclear whether a stereoscopic image can be successfully retargeted by a specific operator and what the mechanism the stereoscopic perception is generated during the retargeting. Thus, it is practical and necessary to evaluate the performances of different retargeting attributes objectively.

Conventional 2D image retargeting algorithms, such as manual cropping (CR), seam carving (SC) [2] and non-homogeneous warping (WARP) [3], aim to adjust image size while minimizing the geometric distortion and content loss in those visually important regions. However, different methods will demonstrate dissimilar results even for the same input. To address the issue of retargeting evaluation, 2D image retargeting quality assessment (IRQA) methods, such as bi-directional similarity (BDS) [4], SIFT-flow [5], earth mover's distance (EMD) [6] and aspect ratio similarity (ARS) [7], establish the inter-image correspondence to calculate the dissimilarity between the matched pixels/regions. For stereoscopic images, applying the same retargeting operation on the left and right images is often problematic, because the depth ranges should be also adjusted to maximize the visual experience (mainly from visual comfort and depth perception). Therefore, the perceptual quality of 3D retargeted images not only depends on image quality, but also the 3D perception properties [8], [9]. In 3D perception, depth information may help to better understand images while comfortable perception can improve the viewing experience. The existing SIR methods impose depth-preserving constraint [10] or visual comfort constraint [11] in optimization for generating the stereoscopic retargeted images, but these constraints are difficult to qualify by standard objective models. Therefore, stereoscopic image retargeting quality assessment (SIRQA) is a challenging task.

In recent years, many specific SIR methods have been developed. For most of these methods, subjective comparison is conducted to demonstrate the superiority of their retargeting results [10]–[13]. However, subjective comparison is expensive, time-consuming and difficult to be integrated into real-time optimization systems. To meet the requirement for high-quality image generation, the guideline of SIR process is still at an early stage of success, which makes quality evaluation of 3D retargeted images be a fundamental task. Therefore, objective SIRQA metric should be performed and used to maximize the

perceptual quality during the retargeting process, which has direct application in optimizing the retargeting operation with the supervision of quality.

In this paper, we carry out an in-depth investigation on quality assessment issue for SIR from both subjective and objective perspectives. For this purpose, we build a benchmark SIRQA database (NBU-SIRQA) to study the perceptual quality for SIR[1]. We select eight representative SIR methods and two retargeting scales to generate 720 stereoscopic retargeted images. To our knowledge, there is no large-scale SIRQA database on quality assessment for SIR. Based upon the database, we propose an objective SIRQA metric by measuring grid deformation and information loss (GDIL). Inspired by the work [14] which establishes a similarity transformation via rewarping to simulate different types of retargeting operators, and the work [7] which designs a unified framework of resampling grid generation to interpret the image retargeting, and also referred to the definition of 3D quality of experience (QoE) [12] as a combination of image quality and 3D perceptual quality (e.g., depth sensation and visual comfort), to simulate different types of SIR operators, the metric in this paper to measure image retargeting quality and 3D perceptual quality lies in decomposing each SIR operator into two complementary transformations: monocular retargeting transformation (MRT) and viewpoint transformation (VPT). We extract grid deformation and information loss features from the transformations to evaluate the stereoscopic retargeted image quality. Overall, the major contributions of our work are summarized as follows:

1) For the subjective study of SIRQA, we build a new SIRQA database named NBU-SIRQA, which includes 720 stereoscopic retargeted images generated from eight SIR operators on two retargeting scales. Subjective evaluation is implemented to obtain the subjective opinion score of each stereoscopic retargeted image.

2) We formulate each SIR operator into two independent transformations (i.e., MRT and VPT) to derive the stereoscopic retargeting modifications. As a result, the process to generate stereoscopic retargeted images using different SIR methods can be simulated in a straightforward way.

3) We apply grid deformation and information loss as feature representations extracted from MRT and VPT. Based upon the integrated features and the corresponding subjective opinion scores, a quality predictor can be learned to accurately predict the quality scores of stereoscopic retargeted images.

The rest of this paper is organized as follows. In Section II, we summarize the existing 2D/3D image retargeting methods and the previous works on IRQA and SIRQA. Section III introduces the new database. Section IV presents the motivations and feature representations of our GDIL metric. The experimental results are shown and discussed in Section V, and finally conclusions are drawn in Section VI.

II. RELATED WORK

A. 2D/3D Image Retargeting Methods

The existing 2D and 3D image retargeting techniques can be broadly classified into two categories: discrete and continuous approaches. For 2D image retargeting, CR, SC and shift-maps (SM) [15] are representative discrete algorithms which adjust image size by directly removing or inserting pixels in unimportant regions. In contrast, scaling (SCL), WARP, streaming video (SV) [16], scale-and-stretch (SNS) [17] formulate continuous solutions to preserve important image content. Particularly, multi-operators (MULTIOP) [18] combines several retargeting operators to iteratively adjust image size. For SIR, stereoscopic cropping and stereoscopic seam carving are two popular discrete algorithms. Wang *et al.* [13] adopted stereoscopic saliency information to guide stereoscopic cropping. Basha *et al.* [19] iteratively removed a pair of seams to adjust the aspect ratios of stereoscopic images. Chen *et al.* [20] proposed a stereoscopic seam carving algorithm based on stereoscopic gradient and saliency. Other discrete SIR methods can be found in [21]–[24]. Continuous SIR methods are based on stereoscopic warping. Chang *et al.* [11] presented a warping-based SIR framework by optimizing the sparse pixel correspondences and disparity consistency. Lee *et al.* [25] proposed a layer-based SIR method, in which warping operator is implemented on each layer. Li *et al.* [26] optimized stereoscopic warping energy functions preserving objects shapes and scene depth. Shao *et al.* [12] presented a QoE-guided warping algorithm to promote visual experience for stereoscopic retargeted images. Other continuous SIR algorithms can be found in [27]–[29].

B. 2D Image Retargeting Database and Its Perceptual Quality

To comparatively study the quality of different retargeting operators, Rubinstein *et al.* [30] constructed an IRQA database named RetargetMe. The database selects 37 source images and eight typical image retargeting methods to create 192 retargeted images. Pair-wise subjective test is conducted between two retargeted images to record the subjective rank score. Ma *et al.* [31] built a CUHK database for IRQA, which contains 171 retargeted images generated from 57 source images using three different retargeting operators. The perceptual quality of each retargeted image is subjectively rated on five quality scales, obtaining a mean opinion score (MOS) for each retargeted image. Hsu *et al.* [32] constructed a NRID database, which includes 175 retargeted images generated from 35 original images using five different retargeting methods. The subjective scores for NRID database are also obtained via pair-wise comparisons.

Benefited from these publicly available IRQA databases, many IRQA measurements have been proposed in recent years. Fang *et al.* [33] devised a structural similarity based IRQA metric (IR-SSIM) to evaluate how much structure information from the source image is preserved in the retargeted image at each spatial location. Hsu *et al.* [32] used scale-invariant feature transform flow (SIFT-flow) fields to represent local variance as geometric distortion, and computed saliency loss as content loss. Zhang

et al. [7] calculated aspect ratio similarity (ARS) in local blocks to elaborate geometric change. Liang *et al*. [34] predicted the quality of different retargeted images by considering the factors of salient content, artifact, global structure, aesthetic and symmetry. Jiang *et al*. [35] learned two over-complete dictionaries to evaluate the similarity of two images. Chen *et al*. [36] proposed a bi-directional natural salient scene distortion model for IRQA. Zhang *et al*. [37] improved the alignment algorithm and proposed a three-level representation model. Other relevant works can be found in [38]–[40].

C. 3D Image Retargeting Database and Its Perceptual Quality

As mentioned above, the existing SIR methods have advantages and disadvantages in handling content and depth optimization. For example, stereoscopic seam carving directly inserts or removes a pair of seams, so that it can well preserve the geometric consistency. However, high computational cost and poor object shape preservation are two main shortcomings for these types of methods. In contrast, stereoscopic warping distributes distortions on all spatial directions, which can better utilize available homogeneous regions to absorb distortions, but it may over-stretch or over-squeeze the salient objects. Therefore, it is very necessary to design efficient objective metrics to faithfully evaluate the quality of 3D retargeted images.

Currently, only few works are proposed for SIRQA. Liu *et al*. [41] built a SIRQA database and presented a learning-based objective SIRQA metric. The database contains 56 source images and 224 stereoscopic retargeted images produced by four SIR methods, including cropping, scaling, seam carving and warping. To assess the quality of 3D retargeted images, multiple features from disparity excessiveness, depth similarity, picture completeness, local and global shape distortions are extracted to learn a predictor. Zhou *et al*. [42] proposed a SIRQA database named SIRD, in which four stereoscopic image retargeting methods including cropping, seam carving, scaling and multi-operator are used to create the retargeted images. The database consists of 400 stereoscopic retargeted images and 100 source images. In addition, features from disparity range, disparity intensity distribution, boundary disparity, and image quality are used to evaluate the quality of 3D retargeted images. Despite of pioneering construction of these databases, the common limitations of these databases and quality evaluation methods are as follows: 1) Only few SIR methods are used to generate retargeted images, which may make the database cannot fully reflect the quality degradation factors; 2) Simply using image quality and stereoscopic perceptual quality for SIRQA may ignore the retargeting modifications between the original and stereoscopic retargeted images; 3) The performance of existing SIRQA metrics are not robust enough for real applications, thus, new SIRQA metrics are needed.

III. NBU-SIRQA DATABASE

To investigate quality assessment of stereoscopic retargeted images, we construct a new SIRQA database (NBU-SIRQA), which includes 720 stereoscopic retargeted images generated by eight typical SIR methods on two retargeting scales. Subjective quality evaluation of these stereoscopic retargeted images is formulated to capture the human option scores. Our database will be introduced in detail in the following parts.

A. Source Image

Since our framework is designed to evaluate the quality of different stereoscopic retargeted images, it is expected that the selected source images should have enough diversity in terms of image contents and depths. To distinguish different image retargeting methods, we select a set of attributes that can reflect major retargeting objectives (e.g., content preserving, shape preserving or artifacts preventing). The selected scenes include natural scenery, foreground object, geometric structure, face and people, and other indoor and outdoor scenes. Fig. 1 shows the selected 45 source images in the database. Also, in order to measure depth attributes on different SIR methods, the source images selected in the database have different disparity ranges to reflect comfortable or uncomfortable visual perception. An example of stereoscopic images with different disparity ranges is shown in Fig. 2.

B. Retargeting Method

To measure the diversity of collected stereoscopic images in the database, we select and employ eight typical SIR methods, including monocular seam carving, scale and stretch operators, stereoscopic cropping, stereoscopic scaling, stereoscopic seam carving and stereoscopic warping operators. We apply these operators to adjust the original image resolution on two retargeting scales: 25% and 50% (i.e., shrinking the width to 75% and 50% of the original width). The selected SIR methods are described below.

- Monocular seam carving (MSC) [2]: left and right images are resized via seam carving algorithm respectively.
- Monocular scale and stretch (MSNS) [17]: left and right images are resized by scale and stretch (SNS) algorithm respectively.
- Content persistent cropping (CPC) [13]: using stereoscopic saliency information to calculate the position of the clipping window and automatically cropping the image.
- Stereoscopic scaling (SSCL): simply scaling the original image into the target size.
- Geometrically consistent stereoscopic seam carving (GC-SSC) [19]: removing pixels that belong to the non-informative regions in left and right images according to the principle of geometric consistency.
- Visual attention guided seam carving (VASSC) [21]: applying binocular just noticeable difference (BJND) model and visual attention mechanism to select seams, and seam replacement is performed for the occluded regions to prevent the geometry inconsistency.
- QoE-guided warping (QOE) [12]: taking object shape, visual comfort and depth sensation into account and solving the optimization model in 3D space. QOE attempts to promote the visual experience for different stereoscopic retargeted images.
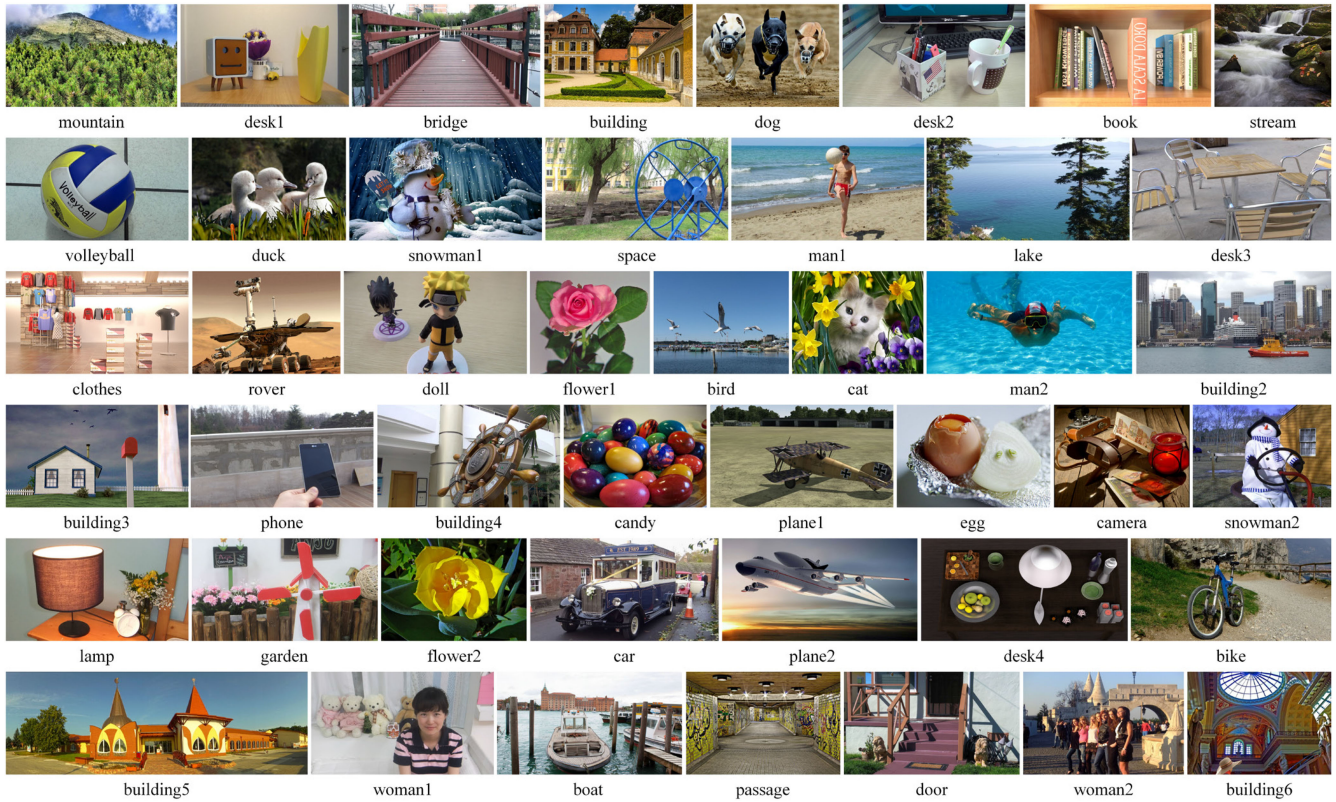
Fig. 1. The left images of 45 source stereoscopic images. The images demonstrate different resolutions, varying depth ranges and diverse contents. More detailed information of the source images can be found in our database.
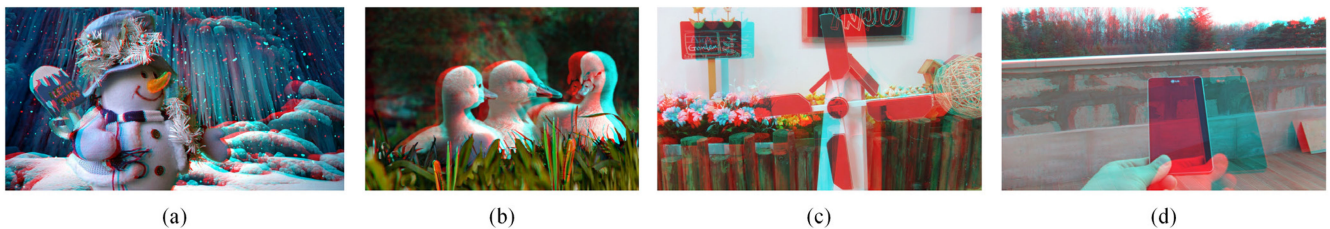

Fig. 2. Example of stereoscopic images with different disparity ranges. The mean absolute values of the disparity for (a), (b), (c), and (d) are 33.62, 44.02, 75.55 and 79.99 respectively. The mean top 5% absolute values of the disparity for (a), (b), (c), and (d) are 55.17, 89.09, 117.02, and 253.94, respectively.

- Single-layer warping (SLWAP) [11]: an objective energy function is optimized, based on feature matching and disparity consistency constraint. SLWAP can adapt depth to the comfort zone while preserving the shapes of prominent objects.

For the retargeting methods selected in the database, we have the following considerations: 1) Similar to the existing SIRQA databases [41], [42], four types of retargeting operators (e.g., cropping, scaling, seam carving and warping) are used to generate the retargeted images; 2) To well distinguish 2D and 3D properties, two 2D retargeting methods (e.g., MSC and MSNS) are included in database independently applied on the left and right images; 3) The existing databases [41], [42] omit uncomfortable stereoscopic images with large disparity ranges, while

visual comfort is one of the three attributes in 3D QoE. Therefore, two 3D methods (e.g., QOE and SLWAP) in optimizing depth ranges are included in our database. The main attributes of these methods are illustrated in Table I, where "√" denotes the attribute is considered while "×" indicates the attribute is not considered in the retargeting method. Since each method has its strength or weakness in shape preserving, object preserving, depth perception, visual comfort and geometry consistency, our database takes these attributes into account to comprehensively represent 2D and 3D retargeting objectives. Since it is difficult to evaluate each attribute independently in subjective perception, especially, strong depth perception means lower visual comfort under uncomfortable viewing, we only evaluate

TABLE I
MAIN ATTRIBUTES OF EIGHT SIR METHODS. ("√" DENOTES THE ATTRIBUTE IS CONSIDERED WHILE "×" DENOTES THE ATTRIBUTE IS NOT CONSIDERED IN THE RETARGETING METHOD)

Metric	MSC	MSNS	CPC	SSCL	GCSSC	VASSC	QOE	SLWAP
Shape preserving	√	√	√	×	√	√	√	√
Object preserving	√	√	√	×	√	√	√	√
Depth perception	×	×	√	×	√	√	√	√
Visual comfort	×	×	×	×	×	×	√	√
Geometry consistency	×	×	√	×	√	√	×	×

Fig. 3. Example of stereoscopic retargeted images generated by eight SIR algorithms. (a) Original image. (b) MSC. (c) MSNS. (d) CPC. (e) SSCL. (f) GCSSC. (g) VASSC. (h) QOE. (i) SLWAP.

the overall QoE of each stereoscopic retargeted image. An example of stereoscopic retargeted images generated by eight SIR methods is shown in Fig. 3.

C. Subjective Testing

Refer to [31], we adopt a double-stimulus absolute category rating (ACR) test methodology for the subjective testing, in which the reference and the stereoscopic retargeted images are simultaneously displayed on the screen. Subjective test environment and condition are designed according to the recommendations of ITU-R BT.500-11 and ITU-R 1438. A Samsung UA65F9000 65-inch Ultra HD 3D-LED TV with 3D shutter glasses was used for the testing. To avoid the possible contextual and memory effects on subjective judgment, the stereoscopic

retargeted images which is created from a same source image will be not displayed consecutively. The subjective ratings for stereoscopic retargeted images were rated on a five-level scale: 5 (excellent), 4 (good), 3 (fair), 2 (poor), and 1 (bad). Thirty graduate students (20 male and 10 female) participated in the subjective testing.

D. Processing of Subjective Scores

After subjective testing, human option ratings of all 720 stereoscopic retargeted images from 30 subjects are obtained, and the normalized z-scores are defined as [43]:

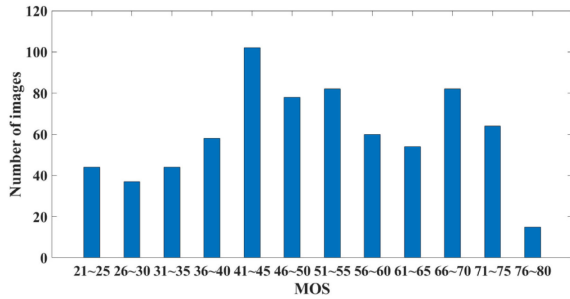$$z_{m,n} = \frac{s_{m,n} - \mu_m}{\sigma_m} \quad (1)$$

Fig. 4. The MOS distribution of NBU-SIRQA database.

where $s_{m,n}$ denotes the score assigned by subject m to the image n, and μ_m and σ_m denotes the mean and standard deviation calculated from all scores assigned by subject m. Then, after removing the outlier subjects (within 95% confidence interval), the normalized z-scores are converted to MOSs by the follow mapping functions:

$$z'_{m,n} = \frac{100\left(z_{m,n} + 3\right)}{6} \tag{2}$$

$$\text{MOS}_n = \frac{1}{K} \sum_m z'_{m,n} \tag{3}$$

where K is the number of subjects after subject rejection.

E. Analyses of the Subjective Scores

In Fig. 4, we demonstrate the histogram distribution of all the MOS values in the NBU-SIRQA database. We can observe that the MOS values span over a wide range of visual quality (from excellent to bad) and have a good spread at different visual levels. In addition, the distribution of MOS values indicates the diversity of images in the constructed NBU-SIRQA database. We further analyze the subjective scores from the perspective of source image, retargeting method and retargeting scale to better demonstrate the properties of the database.

Source Image: Since the essence of SIR is to preserve important shapes and contents while preventing artifacts, different source images will have different retargeting effects considering the texture/structure complexity. In addition, some SIR methods ignore the details of disparity optimization, leading to unsatisfactory retargeting results, especially when the source images have large disparity ranges. We illustrate the influence of source images with an example in Fig. 5. For the source image with a mid-size salient object and simple background shown in the first row, CPC can achieve better performance than QOE (with larger MOS value), while for the source image with multiple dispersive objects shown in the second row, CPC may destroy the completeness of the retargeted objects with unexpected content loss. However, for the source image with large disparity range shown in the last row, although CPC can well preserve the image content and shape, it ignores the optimization of disparity range, resulting in poor visual quality. Therefore, we collect the source images with diverse attributes (content, shape and disparity) to better demonstrate the influence of source images.
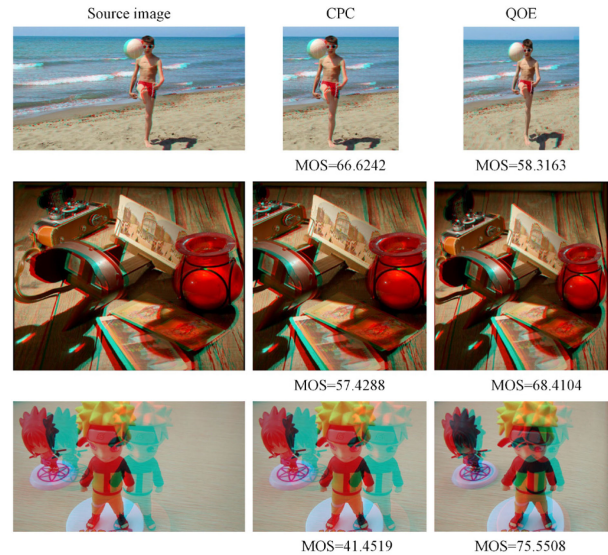
Fig. 5. Influence of the source image.
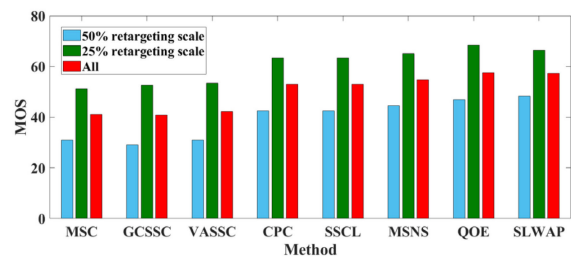
Fig. 6. The mean MOS values of different SIR methods.

Retargeting Method: The average MOS values of eight SIR methods are shown in Fig. 6. We can make the following observations from the figure: 1) Since discrete methods directly remove the pixels or blocks leading to noticeably jagged edges and artifacts in objects, especially when the image resolution is adjusted to 50% of the original width, the overall performance of the discrete methods (e.g., MSC, GCSSC, VASSC) is lower than the continuous methods; 2) Since the CPC creates a clipping window based on significance information to adjust image size, the overall performance of CPC is better than other three discrete methods; 3) Among these methods, QOE and SLWAP have the best retargeting results by considering visual comfort; 4) The continuous method MSNS resizes the left and right images respectively and its retargeting results are inferior to QOE and SLWAP.

Retargeting Scale: In fact, it is difficult to preserve object shape and important content on a large retargeting scale (i.e., large shrinking width for large scale). As illustrated in Fig. 6, the average MOS value on 50% retargeting scale (shrinking the width to 50% of the original width) is 39.47, and the average MOS value on 25% retargeting scale (shrinking the width to 75% of the original width) is 60.53. But for some source images, the subjects are more sensitive to content loss than structure deformation. Therefore, it is challenging and meaningful to evaluate

Fig. 7. MTR and VPT transformations.

the retargeted images on different retargeting scales to comprehensively investigate the performances of different retargeting methods.

IV. OBJECTIVE QUALITY METRIC FOR 3D IMAGE RETARGETING

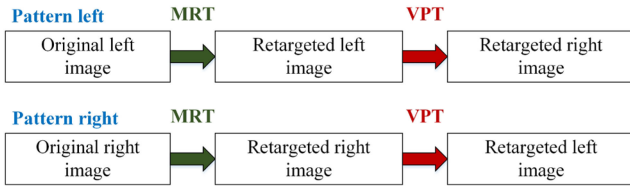Based upon the database, we propose an objective SIRQA metric via measuring grid deformation and information loss (GDIL). Differentiating with other methods that simply take image quality and stereoscopic perceptual quality into account, the novelty of our approach is to decompose each SIR operator into two transformations to reveal the artificial stereoscopic retargeting modifications. Next, we will introduce our approach in detail.

A. Motivation

Differentiating with SIR methods that the stereoscopic retargeted image is the algorithm output. In SIRQA, both the original and retargeted images are known in advance but the stereoscopic retargeting rule is unknown. As an inverse problem, it is important to find the artificial retargeting modifications and dig the relationship between the original and retargeted images. Most recently proposed IRQA methods attempt to find such modifications by using robust SIFT-flow [7], [33], [34], [36], [37], [44]. However, for SIRQA, the relationship between the original and retargeted images is more complex than 2D case due to the non-intuitive binocular visual characteristics, such as binocular fusion and binocular rivalry.

As mentioned earlier, inspired by the works [14] and [7] which employ a similarity transformation and a framework of resampling grid generation to simulate different types of retargeting operators respectively, and also referred to the definition of 3D quality of experience (QoE) [12], in this paper, we decompose each SIR operator into two transformations: monocular retargeting transformation (MRT) and viewpoint transformation (VPT) to simulate different types of SIR operators and reveal the artificial retargeting modifications. The MRT and VRT is performed on both left and right images (pattern left and pattern right), as shown in Fig. 7. MRT transformation transforms the original left/right image to its target size based on an estimated SIFT-flow map. Similar with MRT, VPT is applied to reconstruct the retargeted left and right images to clarify the 3D attributes via the disparity map [45]. As discussed in [7] and [30], approximately estimating retargeting modification is helpful to find the corresponding quality degradations between the original and stereoscopic retargeted images. The designed MRT is to evaluate the image quality of the retargeted left and right images while VPT

is to assess the 3D perceptual quality. As a result, by decomposing each SIR operator into two transformations, quantitative measurement of stereoscopic retargeting modifications will be more straightforward. The framework of GDIL is exhibited in Fig. 8.

B. MRT Feature Representation

MRT aims to resize the original monocular image (left or right) to a target aspect ratio. Thus, feature representation on MRT is also related to IRQA methods. In this paper, we use grid's aspect ratio similarity measurement to capture the 2D retargeting distortions.

First of all, the visual importance map is obtained by integrating a graph based visual saliency (GBVS) map [46] and a disparity map [45], calculated as:

$$V = \eta_1 \cdot \mathbb{N}(S_I) + \eta_2 \cdot \mathbb{N}(S_D) \qquad (4)$$

where S_I is the GBVS map of a monocular image and S_D is the disparity map, $\mathbb{N}(\cdot)$ is the normalization function, η_1 and η_2 are the weights for the GBVS map and disparity map respectively. In our experiment, we set $\eta_1 = \eta_2 = 0.5$.

After obtaining the visual importance map, pixel-to-pixel correspondence between the retargeted and original images is established by SIFT-flow algorithm. Then, the retargeted grids can be obtained by mapping all pixels within the original grids to the retargeted image according to the SIFT-flow map (a set of 16×16 grids are extracted in original image). Similar to [7], the gird deformation is estimated by using all pixels within a grid. We calculate the maximum width and height of each retargeted grid to measure the grid deformation:

$$f_1 = \sum_i \sum_j v_{i,j} \cdot 2 \cdot \left[\frac{w_{i,j} \cdot h_{i,j} + C}{w_{i,j}^2 + h_{i,j}^2 + C} \right]$$
$$\cdot \left[e^{-\alpha \cdot (0.5 \cdot (w_{i,j} + h_{i,j}) - 1)^2} \right] \qquad (5)$$

where $v_{i,j}$ represents the average visual importance value for the grid (i, j), $w_{i,j}$ and $h_{i,j}$ denote the maximum width and height of a grid respectively, C is a small positive constant to avoid the division by zero, and α is a parameter to balance content loss and shape distortion and. In the experiment, we set $C = 10^{-6}$ and $\alpha = 0.3$. Obviously, larger value of f_1 indicates better image structure preservation. However, when the original image structure is seriously distorted, the measured maximal width and height of a retargeted grid may be identical with the original grid, as shown by the example in Fig. 9. In such situation, f_1 cannot effectively measure the deformation for the retargeted grids. For this purpose, another feature measuring information loss is extracted to promote the accuracy of grid deformation measurement, defined as:

$$f_2 = \sum_i \sum_j v_{i,j} \cdot \tilde{s}_{i,j} / s_{i,j} \qquad (6)$$

where $\tilde{s}_{i,j}$ and $s_{i,j}$ denote the numbers of pixels in the matched grids of retargeted and original images respectively.
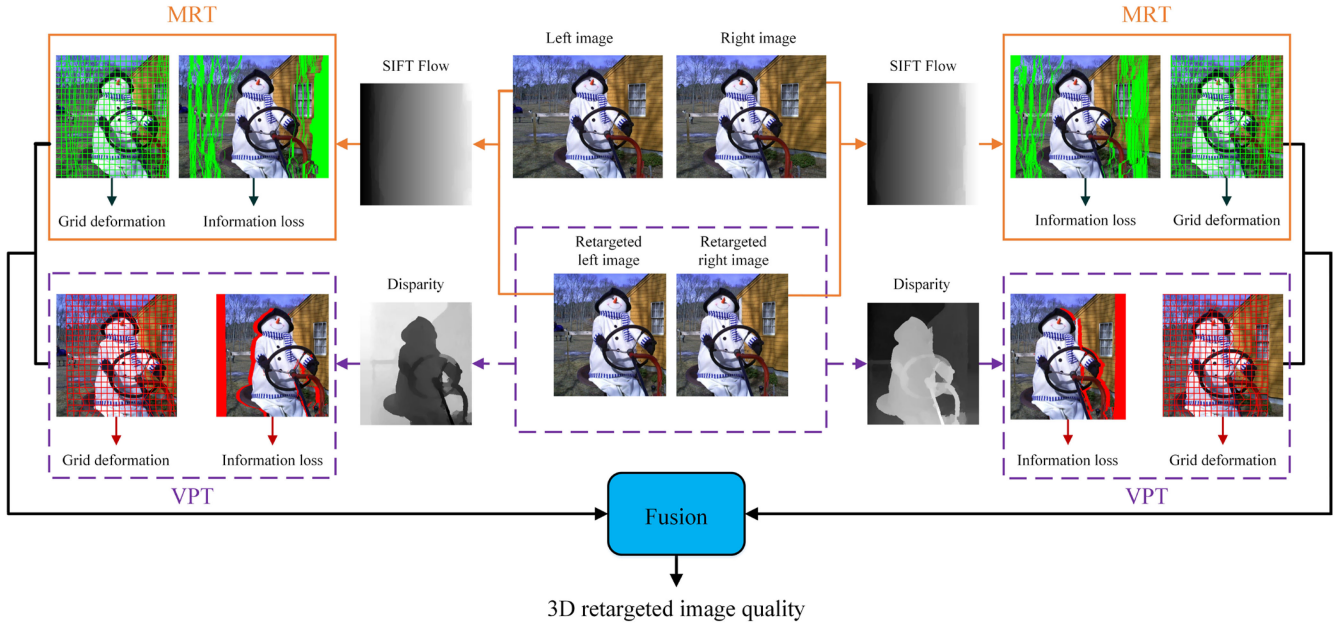
Fig. 8. The framework of the proposed GDIL method. The MRT (orange solid path) aims to evaluate the quality of left and right images. The VPT (purple dotted path) aims to evaluate the stereoscopic perceptual quality. GDIL metric (black solid path) obtains the objective quality score by fusing individual quality components.
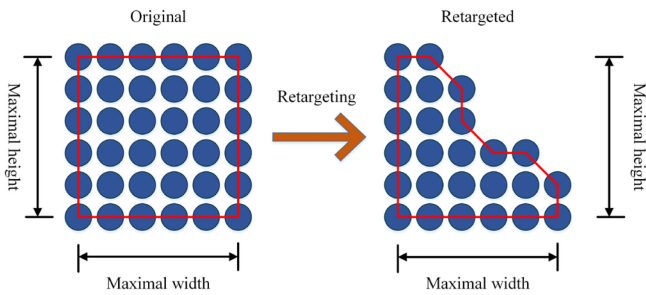

Fig. 9. An example of MRT gird deformation, in which the grid is seriously distorted but the maximal width and height are not changed.
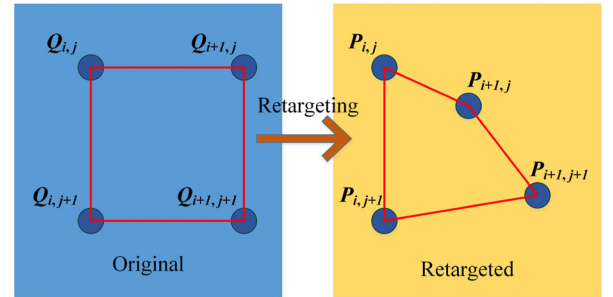

Fig. 10. An example of VPT grid deformation. Four grid vertices are used to estimate the grid deformation.

C. VPT Feature Representation

As discussed, visual comfort and depth sensation are two crucial factors influencing 3D perception in SIR. In this subsection, we propose two VPT driven stereoscopic perceptual properties measurements for SIRQA. These two measurements are named as VPT grid deformation and VPT information loss, similar to the feature representations on MRT. First, disparity map [45] is used to establish the relationship between the retargeted left and right images. Then, a set of grids (e.g., 16×16) in the retargeted left image and the corresponding grids in retargeted right image are extracted based on the disparity map. The grid deformation is calculated based on the average width and height of each grid to measure the possible changes of disparity gradient in affecting visual discomfort [47]–[49]. Instead of using all pixels within a grid, only four vertices of the grid are employed to estimate the grid deformation on VPT, since the fluctuation of disparity is relatively stable, compared with the object deformation on

MRT. The VPT grid deformation is defined as:

$$f_3 = \frac{1}{M \cdot N} \sum_i \sum_j 2 \cdot \left[\frac{\bar{w}_{i,j} \cdot \bar{h}_{i,j} + C}{\bar{w}_{i,j}^2 + \bar{h}_{i,j}^2 + C} \right] \quad (7)$$

$$\bar{w}_{i,j} = \frac{1}{2} \cdot \left[\left(p_{i+1,j}^x - p_{i,j}^x \right) + \left(p_{i+1,j+1}^x - p_{i,j+1}^x \right) \right] \quad (8)$$

$$\bar{h}_{i,j} = \frac{1}{2} \cdot \left[\left(p_{i,j+1}^y - p_{i,j}^y \right) + \left(p_{i+1,j+1}^y - p_{i+1,j}^y \right) \right] \quad (9)$$

where M and N are the number of girds in the horizontal and vertical directions, $\bar{w}_{i,j}$ and $\bar{h}_{i,j}$ denote the average width and height of the grid (i, j), p^x and p^y denote the x-coordinate and y-coordinate of a grid vertex P, respectively. An example of VPT grid deformation is illustrated in Fig. 10.

Since depth sensation is perceived from binocular disparity and monocular occlusion simultaneously, information loss is computed according to the area of occlusion and the area outside

the field-of-view (Out-FOV) and occlusion. Refer to [19], we define the occluded and Out-FOV regions as follows.

Let (l, k_1) and (l, k_2) be two pixels in left image, and D is the disparity map. It follows that the pixel (l, k_2) occludes (l, k_1):

$$k_1 + D(l, k_1) = k_2 + D(l, k_2) \, \exists k_1 < k_2 \qquad (10)$$

If a pixel (l, k) belongs to Out-FOV region, it needs to meet the following condition:

$$k + D(l, k) < 1 \, \text{or} \, k + D(l, k) > k_{\max} \qquad (11)$$

where $k_{\max}$ denotes the maximum width of retargeted image. After deriving the occluded and Out-FOV regions, we calculate information loss on VPT by:

$$f_4 = (s_1 + s_2) / (W \cdot H) \qquad (12)$$

where W and H denote the width and height of the retargeted image respectively, s_1 denote the area of Out-FOV, s_2 denote the area of occlusion.

D. Quality Evaluation

As illustrated in Fig. 7 and shown by the specific examples of MRT and VPT features in Fig. 8, the process for feature representations on MRT and VRT is performed on both left and right images (defined as pattern left and pattern right). To accurately capture the quality of each stereoscopic retargeted image, we opt to measure the similarity via combining these two aspects. By this way, we can acquire an eight-dimensional quality vector for each stereoscopic retargeted image:

$$f = \{f_1^L, f_1^R, f_2^L, f_2^R, f_3^L, f_3^R, f_4^L, f_4^R\} \qquad (13)$$

where the superscript L (R) denotes the feature is obtained from pattern left (right). With the estimated quality vector, we build a pre-trained quality predictor. In our implementation, we adopt random forest to train the quality predictor. Of course, other machine learning techniques can be employed here. We conduct experiments to analyze the impact of different pooling schemes, and the experiment results will be presented in the next section.

V. EXPERIMENTAL RESULTS AND ANALYSES

In this section, we quantify the performance of objective models on the proposed NBU-SIRQA database using three evaluation criteria: Pearson Linear Correlation Coefficient (PLCC), Spearman Rank order Correlation Coefficient (SRCC), and Root Mean Square Error (RMSE). SRCC is used to evaluate prediction monotonicity, while PLCC and RMSE are applied to assess prediction accuracy. In our experiment, PLCC and RMSE are calculated after a nonlinear mapping between the subjective and objective scores [50]:

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(x - \beta_3)}} \right) + \beta_4 x + \beta_5 \qquad (14)$$

where $\beta_1, \beta_2, \beta_3, \beta_4$ and β_5 are parameters determined by fitting between the subjective and objective scores.

Before training the quality predictor, it is necessary to construct a training and testing set. In the experiment, we randomly divide the whole dataset into two groups: 80% for training and

TABLE II
BASIC INFORMATION OF THE EXPERIMENTAL DATABASE

Database	NBU SIRQA
Retargeting ratio	0.5; 0.75
Original image No.	45
Retargeting image No.	720
Retargeting operator No.	8
Subject No.	30
Subjective score type	MOS

20% for testing. There will be no overlap between testing and training subsets. The average values of PLCC, SRCC and RMSE after 1000 iterations are reported as the final performance scores. The basic information of the experimental database is summarized in Table II.

A. Performance Comparisons with Other Methods

We compare our GDIL model with five state-of-the-art IRQA methods, including SIFT-flow [5], BDS [4], EMD [6], hand-craft and deep learned features (HCDL) [44] and ARS [7], and two recent SIRQA methods (Liu's method [41] and Zhou's method [42])[2]. For SIFT-flow, BDS, EMD, HCDL and ARS designed for 2D images, the quality of each view is separately estimated and weighted to generate a score. Table III gives the comparisons of PLCC, SRCC and RMSE scores between each competing measurement and GDIL metric on three datasets, including all stereoscopic images on 50% retargeting scale, 25% retargeting scales and all scales, respectively. From the table, we have the following observations: 1) Three traditional IRQA metrics (SIFT-flow, BDS and EMD) have poor consistency with the subjective rating, especially on single retargeting scale, due to ignoring important contents and the 3D perception; 2) Compared with HCDL and ARS methods which usually have high performance on 2D IRQA databases, the performance of our GDIL metric is significantly promoted by considering the 3D perceptual properties; 3) Compared with two 3D metrics (Liu's method and Zhou's method), our GDIL can achieve better performance. The reason is that MRT and VPT are effective to detect 2D and 3D retargeting semantics by revealing the artificial stereoscopic retargeting modifications. Overall, the evaluation results on all data are better than the results on 50% retargeting scale and 25% retargeting scale, and our GDIL measurement can achieve great performance promotion on predicting the quality of 3D retargeted images.

In addition, we conduct one-sided t-test at a significance level of 0.01 using 1000 SRCC values of all SIRQA metric pairs to demonstrate the statistical significance of the reported performance gains. The statistical test can be referred to [51]. The results of t-test are illustrated in Fig. 11. A value of "1", "0", or "-1" indicates that the metric in the row is statistical "better", "worse", or "indistinguishable" compared with the column

[2]Since these two measurements are not publicly available, the evaluation results are computed by our approximate implementations.

TABLE III
PERFORMANCE OF DIFFERENT METHODS ON NBU-SIRQA DATABASE

Database	Criteria	SIFT-flow	BDS	EMD	HCDL	ARS	Liu	Zhou	**GDIL**
25% retargeting scale	PLCC	0.1753	0.2112	0.1778	0.4157	0.4792	0.3887	0.6054	**0.6150**
	SRCC	-0.1321	-0.1519	-0.1619	0.3614	0.4491	0.3423	0.5616	**0.5685**
	RMSE	10.7355	10.6584	10.7306	9.9620	9.6887	9.7728	8.6625	**8.5868**
50% retargeting scale	PLCC	0.2553	0.0814	0.1828	0.3823	0.4993	0.3884	0.6202	**0.6555**
	SRCC	-0.2257	-0.1561	-0.1545	0.3599	0.4799	0.3522	0.5813	**0.6315**
	RMSE	9.9736	10.2812	10.1416	9.5425	8.9380	9.4414	8.1511	**7.7866**
All	PLCC	0.1019	0.2791	0.3480	0.7261	0.7833	0.6138	0.8146	**0.8371**
	SRCC	0.0542	0.2751	0.3858	0.7215	0.7745	0.5910	0.8054	**0.8315**
	RMSE	14.8731	14.3882	14.0167	11.8422	9.2950	12.0756	8.6338	**8.1083**

TABLE IV
PERFORMANCE OF EACH QUALITY COMPONENT

Criteria	f_1^L	f_1^R	f_2^L	f_2^R	f_3^L	f_3^R	f_4^L	f_4^R	**All**
PLCC	0.7851	0.7537	0.6349	0.6315	0.1305	0.1496	0.2233	0.2357	**0.8371**
SRCC	0.7687	0.7764	0.5629	0.5662	0.0972	0.1014	0.1866	0.1913	**0.8315**
RMSE	9.2597	9.8262	11.5509	11.5932	15.5449	14.7827	14.5959	14.5296	**8.1083**

	SIFT-flow	BDS	EMD	HCDL	ARS	Liu	Zhou	GDIL
SIFT-flow	0	-1	-1	-1	-1	-1	-1	-1
BDS	1	0	-1	-1	-1	-1	-1	-1
EMD	1	1	0	-1	-1	-1	-1	-1
HCDL	1	1	1	0	-1	1	-1	-1
ARS	1	1	1	1	0	1	-1	-1
Liu	1	1	1	1	1	0	-1	-1
Zhou	1	1	1	1	1	1	0	-1
GDIL	1	1	1	1	1	1	1	0

Fig. 11. Results of one-sided t-test conducted by using the SRCC scores of competing SIRQA models. A value of "1" indicates that the row model is statistically better than the column model, a value of "−1" indicates that the column model is statistically better, while a value of "0" indicates that the two models are statistically similar in performance.

mode respectively. From the result, we can observe that the proposed method is statistically better than all other metrics.

B. Validity of Each Quality Component

In this subsection, we conduct experiments to investigate the contribution of different quality components (features). First, we compare the performance of individual quality component. Comparison results are shown in Table IV, in which PLCC, SRCC and RMSE are directly computed between the individual quality scores and the MOSs. From the table, we have the following observations: 1) Among these measurements, grid deformation features on MRT (f_1^L and f_1^R) have the great impact on the overall performance; 2) Independently applying the quality components cannot obtain the satisfactory results,

especially for the quality components f_3^L, f_3^R, f_4^L and f_4^R, because each measurement is complementary in characterizing the quality degradations; 3) As expected, the performances on left and right patterns have some differences, because image contents, SIFT-flow map and disparity map of these two patterns are not exactly the same.

Furthermore, we investigate the evaluation results using different feature combinations, as shown in Table V. We can observe that the results of #1, #2 and #3 are similar and all better than #4, which indicates that grid deformation features on MRT (f_1^L and f_1^R) have the most significant contributions to the overall performance, in line with the result in Table IV. Only considering the left or right features, the results of #5 and #6 are very similar, but all do not exceed #9.

In addition, since we combine MRT (f_1^L, f_1^R, f_2^L, f_2^R) and VPT (f_3^L, f_3^R, f_4^L, f_4^R) features to evaluate the overall quality, the impact of two components is further analyzed. As shown in Table V, MRT component (#7) has a greater influence on the overall performance than VPT component (#8). The reason may be that, for most of source images selected in the database, image quality degradation induced by grid deformation and information loss is the primary factor under comfortable visual perception, while for those uncomfortable stereoscopic images, VPT (#8) must be considered to capture 3D perception quality and compensate the limitations of MRT. Overall, aggregated MRT and VPT results in the best performance.

C. Influence of Different Quality Pooling Methods

As discussed before, the final quality evaluation results of the proposed approach are highly dependent on the used quality pooling method. In this subsection, we conduct experiments to analyze the influence of different pooling schemes. We use six different pooling methods to fuse eight individual quality components, including linear regression, logistic regression,

TABLE V
PERFORMANCE OF DIFFERENT FEATURE COMBINATIONS

No.	f_1^L	f_1^R	f_2^L	f_2^R	f_3^L	f_3^R	f_4^L	f_4^R	PLCC	SRCC	RMSE
#1	✓	✓	✓	✓	✓	✓	×	×	0.8237	0.8167	8.4755
#2	✓	✓	✓	✓	×	×	✓	✓	0.8228	0.8174	8.4907
#3	✓	✓	×	×	✓	✓	✓	✓	0.8203	0.8127	8.6191
#4	×	×	✓	✓	✓	✓	✓	✓	0.7463	0.7295	9.9877
#5	×	✓	×	✓	×	✓	×	✓	0.8299	0.8229	8.4793
#6	✓	×	✓	×	✓	×	✓	×	0.8278	0.8206	8.3524
#7	✓	✓	✓	✓	×	×	×	×	0.8077	0.8043	8.7901
#8	×	×	×	×	✓	✓	✓	✓	0.3149	0.2648	14.4354
#9	✓	✓	✓	✓	✓	✓	✓	✓	**0.8371**	**0.8315**	**8.1083**

TABLE VI
PERFORMANCE COMPARISON OF DIFFERENT QUALITY POOLING METHODS

Criteria	Linear regression	Logistic regression	L-SVR	RBF-SVR	Poly-SVR	Random forest
PLCC	0.7852	0.8178	0.7895	0.8235	0.8353	0.8371
SRCC	0.7802	0.8198	0.7819	0.8214	0.8289	0.8315
RMSE	9.1344	8.7199	9.1376	8.4523	8.1928	8.1083

TABLE VII
INFLUENCE OF DIFFERENT RETARGETING METHODS

Test group	50% retargeting scale				25% retargeting scale				All			
	ARS		GDIL		ARS		GDIL		ARS		GDIL	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
Group-1	0.2598	0.2818	0.2171	0.0145	0.4732	0.4895	0.5232	0.4845	0.7344	0.7477	0.7447	0.7214
Group-2	0.4633	0.5070	0.5128	0.4821	-0.0153	-0.0155	0.4113	0.2318	0.8702	0.7910	0.8846	0.8221
Group-3	0.3865	0.3138	0.4823	0.1239	0.0124	0.1156	0.2845	0.2552	0.7583	0.7242	0.7797	0.7354
Group-4	0.0439	0.0379	0.2468	0.2109	0.0821	0.0606	0.4513	0.3917	0.8938	0.7576	0.9038	0.8136
Group-5	0.1248	0.0929	0.4290	0.2091	0.3201	0.2709	0.4854	0.4478	0.7985	0.7410	0.8137	0.7853
Group-6	0.1408	0.1743	0.3495	0.2599	0.3139	0.1603	0.5212	0.5295	0.7572	0.7300	0.8332	0.8109
Group-7	0.3116	0.2412	0.4837	0.3304	0.6528	0.3341	0.4493	0.4563	0.8446	0.7920	0.8659	0.8264
Group-8	0.3733	0.3071	0.5054	0.3079	0.2089	0.2563	0.3817	0.4318	0.8070	0.7710	0.8143	0.7950
Average	0.2630	0.2445	0.4033	0.2423	0.2560	0.2090	0.4455	0.4036	0.8080	0.7568	0.8300	0.7888

Support Vector Regression (SVR) with linear kernel (L-SVR), SVR with RBF kernel (RBF-SVR), SVR with polynomial kernel (Poly-SVR) and random forest. The same train-test process is used for the experiments. As observed from Table VI, non-linear models always perform better than the linear ones, because linear pooling assumes the contribution of each component on the overall quality is independent, but it is not always the fact if taking the interactions between each quality components into account. The best two results are obtained by using Poly-SVR and random forest. Compared with Poly-SVR, the used random forest strategy may not be the best choice for low-complexity evaluation as it needs large-scale trees in decision (the tree number is set to 50 in our experiments). Therefore, SVR with specific kernels or other quality pooling schemes can be selected by comprehensively taking accuracy, monotonicity and complexity into account.

D. Influence of Different Retargeting Methods and Scales

To investigate the influence of different SIR methods, we classify the whole database into eight groups (i.e., generated by the same SIR method): Group-1 (MSC), Group-2 (MSNS), Group-3 (CPC), Group-4 (SSCL), Group-5 (GCSSC), Group-6 (VACCS), Group-7 (QOE) and Group-8 (SLWAP). We apply leave-one-out cross-validation (LOOCV) on the database, i.e., seven groups for training and the rest one group for testing. Experimental results are shown in Table VII. From the table, we can make the following observations:

1) Our GDIL has low evaluation results on Group-1 and Group-3 (MSC and CPC), because the retargeted images obtained by the two methods are quite distinct from other methods (e.g., MSC always leads to binocular asymmetry

TABLE VIII
THE PREDICTED QUALITY VALUES FOR RETARGETED IMAGES IN FIG. 13

| Source image	Retargeted image	MOS	ARS		Liu		GDIL	
			Score	Difference	Score	Difference	Score	Difference
(a)	(b)	38.2612	45.5012	+7.2400	52.6296	+14.3684	40.6211	**+2.3599**
	(c)	51.3991	49.0100	-2.3891	59.6877	+8.2886	52.4048	**+1.0057**
(d)	(e)	42.2977	34.8186	-7.4791	46.3270	+4.0293	44.1692	**+1.8715**
	(f)	63.4577	63.3701	**-0.0876**	60.2649	-3.1928	65.7211	+2.2634
(g)	(h)	46.3395	63.6021	+17.2626	54.4822	+8.1427	53.0339	**+6.6944**
	(i)	67.4695	63.4667	-4.0028	59.9931	-7.4764	68.7381	**+1.2686**
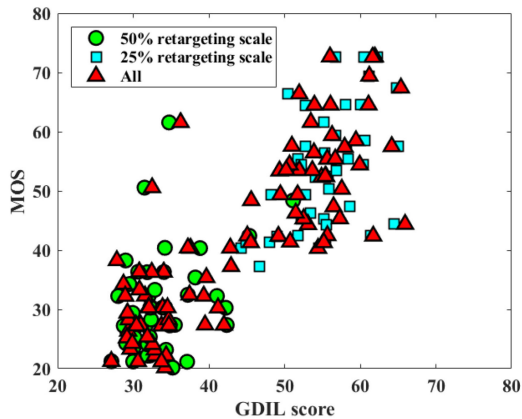
Fig. 12. Scatter plots of predicted quality scores against the subjective scores (MOS) on Group-6.

and geometric distortion, while CPC directly extracts important content using clipping window with information loss). Our GDIL has relatively high evaluation results on Group-4 (SSCL), because the squeezed distortion is normal in other retargeting methods. It also shows that the selected training samples should contain enough information for discriminating the images with different qualities.

2) Compared the results on two independent retargeting scales, the performance for a specific retargeting method on large scale is lower than the performance on small scale, due to the lower quality by removing more information, which also denotes retargeting scale is another factor affecting the ability of GDIL.

3) It should be recognized that in the view of the small number of samples and the very narrow diversity of quality ranges for each group on a single retargeting scale, it is impossible to achieve much high accuracy for quality prediction, while the aggregated data from all retargeting scales can achieve a relatively good consistency with subjective perception. The results from ARS metric (without training) and the scatter plots in Fig. 12 also clearly demonstrated the phenomenon.

E. Application for Selecting the Best Retargeting Results

The most direct application for our metric is to guide the selection of optimal retargeting operator performed on the stereoscopic images with the best perceptual quality, denoted as:

$$\hat{\Gamma} = \arg\max_{\Gamma} \text{GDIL}\left(\Gamma(I^L, I^R)\right) \tag{15}$$

where I^L and I^R are input left and right images, and $\Gamma(\cdot)$ indicates a specific SIR operator.

Fig. 13 and Table VIII provide an example to show the impact of different methods on selecting the best retargeting results. As shown in Fig. 13, the selected source image in (a) and (d) are visually comfortable, while the source image in (g) is visually uncomfortable (with large depth range beyond comfortable viewing zone). Each source image is retargeted using two SIR methods randomly selected from the eight SIR methods. We use ARS [7], Liu's method [41], and our GDIL metric to measure the quality of the stereoscopic retargeted images. Experimental results are reported in Table VIII. It is evident that our metric can obtain the closest results with the ground-truth subjective scores for almost all images. For ARS measurement, since it ignores the 3D perception properties, it is incapable to assess the quality of images, especially on (h). For Liu's method, it cannot accurately measure the object deformation and 3D perception properties for most of the stereoscopic retargeted images. Since our model measures grid deformation and information loss based on MRT and VRT, the adaptability to different SIR methods is obviously superior to others.

F. Discussions

In this paper, we decompose each SIR operator into two transformations, and use grid deformation and information loss as feature representations for SIRQA. Although the proposed model demonstrates its high performance compared with other state-of-the-art methods, it still has the following limitations: 1) Feature representations on MRT and VPT are highly dependent on the accuracy of the SIFT-flow and disparity estimation algorithms. However, the existing pixel-wise matching algorithms have the limitation in capturing structures, which may lead to incorrect correspondences; 2) In our method, the importance map is computed by directly combining GBVS map and disparity map. The influence of different visual importance maps should be further explored.
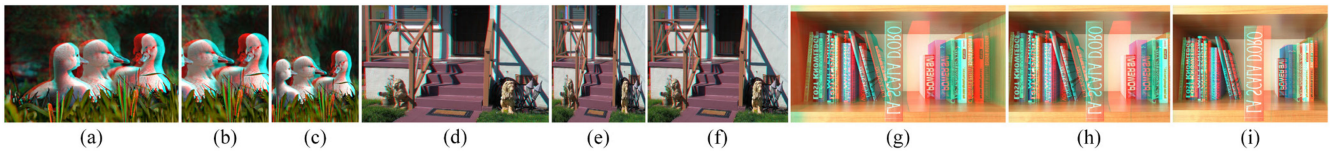
Fig. 13. Original images and corresponding retargeted images. (a), (d) and (g) are original images, (b), (c), (e), (f), (h), (i) are the retargeted images.
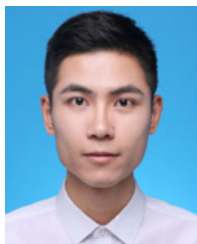
VI. CONCLUSION

In this paper, we propose a quality assessment model for stereoscopic image retargeting (SIR) based on grid deformation and information loss (GDIL). For assessment purpose, we construct a new stereoscopic image retargeting quality assessment database (NBU-SIRQA), which contains 45 original images and 720 stereoscopic retarged images. The stereoscopic retarged images are generated from eight SIR methods on two retargeting scales. By decomposing SIR operators into monocular retargeting transformation (MRT) and viewpoint transformation (VPT), grid deformation and information loss features are extracted to predict the quality of stereoscopic retarged images. The performance of our metric is extensively verified by the elaborately designed experiments. In the future work, we will focus on digging deep features for quality representation and concentrate on accurate quality pooling. In addition, we plan to design specific SIR algorithms in accordance with the stereoscopic image retargeting quality assessment methods.

REFERENCES

[1] G. Zhang, M. Cheng, S. Hu, and R. R. Martin, "A shape-preserving approach to image resizing," *Comput. Graph. Forum*, vol. 28, no. 7, pp. 1897–1906, 2010.

[2] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *Assoc. Comput. Machinery Trans. Graph*, vol. 26, no. 3, 2007, Art. no. 10.

[3] L. Wolf, M. Guttmann and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, 2007, pp. 1–6.

[4] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit.*, 2008, pp. 1–8.

[5] C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978–994, May 2011.

[6] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, 2009, pp. 460–467.

[7] Y. Zhang, Y. Fang, W. Lin, X. Zhang, and L. Li, "Backward registration based aspect ratio similarity for image retargeting quality assessment," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4286–4297, Sep. 2016.

[8] K. Lee, A. K. Moorthy, S. Lee, and A. C. Bovik, "3D visual activity assessment based on natural scene statistics," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 450–465, Jan. 2014.

[9] H. G. Kim, H. Jeong, H. Lim, and Y. M. Ro, "Binocular fusion net: Deep learning visual comfort assessment for stereoscopic 3D," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 4, pp. 956–967, Apr. 2019.

[10] J. Lei *et al.*, "Depth-preserving stereo image retargeting based on pixel fusion," *IEEE Trans. Multimedia*, vol. 19, no. 7, pp. 1442–1453, Jul. 2017.

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

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

[13] W. Wang, J. Shen, Y. Yu, and K. Ma, "Stereoscopic thumbnail creation via efficient stereo saliency detection," *IEEE Trans. Visualization Comput. Graph.*, vol. 23, no. 8, pp. 2014–2027, Aug. 2017.

[14] F. Shao, Z. Fu, Q. Jiang, G. Jiang, and Y. Ho, "Transformation-aware similarity measurement for image retargeting quality assessment via bidirectional rewarping," *IEEE Trans. Syst. Man Cybern.: Syst.* to be published, doi: 10.1109/TSMC.2019.2917496.

[15] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. IEEE Int. Conf. Comput. Vision*, 2009, pp. 151–158.

[16] P. Krähenbühl, M. Lang, A. Hornung, and M. H. Gross, "A system for retargeting of streaming video," *Assoc. Comput. Machinery Trans. Graph*, vol. 28, no. 5, 2009, Art. no. 126.

[17] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *Assoc. Comput. Machinery Trans. Graph*, vol. 27, no. 5, 2008, Art. no. 118.

[18] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *Assoc. Comput. Machinery Trans. Graph*, vol. 28, no. 3, 2009, Art. no. 23.

[19] T. Dekel Basha, Y. Moses, and S. Avidan, "Stereo seam carving: A geometrically consistent approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2513–2525, Oct. 2013.

[20] 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.

[21] F. Shao *et al.*, "Stereoscopic visual attention guided seam carving for stereoscopic image retargeting," *J. Display Technol.*, vol. 12, no. 1, pp. 22–30, Jan. 2016.

[22] B. Yan, K. Sun and L. Liu, "Matching-Area-Based seam carving for video retargeting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 302–310, Feb. 2013.

[23] B. Yue, C.-P. Hou, and Y. Zhou, "Improved seam carving for stereo image resizing," *EURASIP J. Wireless Commun. Netw.*, vol. 2013, Dec. 2013, Art. no. 116.

[24] J. Wang, Y. Fang, M. Narwaria, W. Lin, and P. Le Callet, "Stereoscopic image retargeting based on 3D saliency detection," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2014, pp. 669–673.

[25] K. Lee, C. Chung, and Y. Chuang, "Scene warping: Layer-based stereoscopic image resizing," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2012, pp. 49–5.

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

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

[28] S. Du, S. Hu, and R. R. Martin, "Changing perspective in stereoscopic images," *IEEE Trans. Visualization Comput. Graph.*, vol. 19, no. 8, pp. 1288–1297, Aug. 2013.

[29] J. W. Yoo, S. Yea, and I. K. Park, "Content-driven retargeting of stereoscopic images," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 519–522, May 2013.

[30] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *Assoc. Comput. Machinery Trans. Graph*, vol. 29, no. 6, 2010, Art. no. 160.

[31] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Image retargeting quality assessment: A study of subjective scores and objective metrics," *IEEE J. Sel. Top. Signal Process.*, vol. 6, no. 6, pp. 626–639, Oct. 2012.

[32] C. Hsu, C. Lin, Y. Fang, and W. Lin, "Objective quality assessment for image retargeting based on perceptual geometric distortion and information loss," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 3, pp. 377–389, Jun. 2014.

[33] Y. Fang *et al.*, "Objective quality assessment for image retargeting based on structural similarity," *IEEE J. Emer. Sel. Top. Circuits Syst.*, vol. 4, no. 1, pp. 95–105, Mar. 2014.

[34] Y. Liang, Y. Liu, and D. Gutierrez, "Objective quality prediction of image retargeting algorithms," *IEEE Trans. Visualization Comput. Graph.*, vol. 23, no. 2, pp. 1099–1110, Feb. 2017.

[35] Q. Jiang, F. Shao, W. Lin, and G. Jiang, "Learning sparse representation for objective image retargeting quality assessment," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1276–1289, Apr. 2018.

[36] Z. Chen, J. Lin, N. Liao, and C. W. Chen, "Full reference quality assessment for image retargeting based on natural scene statistics modeling and bi-directional saliency similarity," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5138–5148, Nov. 2017.

[37] Y. Zhang, K. N. Ngan, L. Ma, and H. Li, "Objective quality assessment of image retargeting by incorporating fidelity measures and inconsistency detection," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5980–5993, Dec. 2017.

[38] 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," *Comput. Vision Image Understanding*, vol. 168, pp. 172–181, Mar. 2018.

[39] Y. Guo, Y. Hao, and M. Yu, "Image retargeting quality assessment based on content deformation measurement," *Signal Process.: Image Commun.*, vol. 67, pp. 171–181, Sep. 2018.

[40] M. Karimi, S. Samavi, N. Karimi *et al.*, "Pyramidal modeling of geometric distortions for retargeted image quality evaluation" *Multimedia Tools Appl.*, vol. 77, no. 11, pp. 13799–13820, Jun. 2018.

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

[42] Y. Zhou, W. Zhou, P. An, and Z. Chen, "Visual comfort assessment for stereoscopic image retargeting," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2018, pp. 1–5.

[43] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.

[44] Z. Fu, F. Shao, Q. Jiang, R. Fu, and Y. Ho, "Quality assessment of retargeted images using hand-crafted and deep-learned features," *IEEE Access*, vol. 6, pp. 12008–12018, 2018.

[45] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, Jun. 2010, pp. 2432–2439.

[46] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2006, pp. 545–552.

[47] Y. J. Jung, H. Sohn, S. Lee, H. W. Park, and Y. M. Ro, "Predicting visual discomfort of stereoscopic images using human attention model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 12, pp. 2077–2082, Dec. 2013.

[48] C. Jung, H. Liu, and Y. Cui, "Visual comfort assessment for stereoscopic 3D images based on salient discomfort regions," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 4047–4051.

[49] Q. Jiang, F. Shao, G. Jiang, M. Yu, and Z. Peng, "Three-dimensional visual comfort assessment via preference learning," *J. Electron. Imag.*, vol. 24, no. 4, 2015, Art. no. 043002.

[50] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[51] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.

Feng Shao (Member, IEEE) received the B.S. and Ph.D degrees from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively, both in electronic science and technology. He is currently a professor with the 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.
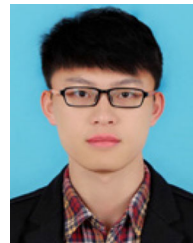
Qiuping Jiang (Senior Member, IEEE) received the Ph.D. degree from Ningbo University in June 2018. He is currently an Associate Professor with the School of Information Science and Engineering, Ningbo University, Ningbo, China. From January 2017 to June 2018, he was a Visiting Student with the School 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.

Xiangchao Meng (Member, IEEE) 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.

Yo-Sung Ho (Fellow, IEEE) 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.

Zhenqi Fu received the B.S. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2016, 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 image/video processing and quality assessment.