A Study of Perceptual Quality Assessment for Stereoscopic Image Retargeting

Zhenqi Fu[*], Yan Yang[*], Feng Shao[*] and Xinghao Ding[†]
[*]Ningbo University, Ningbo, China
E-mail: shaofeng@nbu.edu.cn
[†]Xiamen University, Xiamen, China
E-mail: dxh@xmu.edu.cn

***Abstract*— Subjective and objective perceptual quality assessment for stereoscopic retargeted images is a fundamentally important issue in stereoscopic image retargeting (SIR) which has not been deeply investigated. Here, a stereoscopic image retargeting quality assessment (SIRQA) database is proposed to study the perceptual quality of different stereoscopic retargeted images. To construct the database, we collect 720 stereoscopic retargeted images generated by eight representative SIR methods. The perceptual quality (mean opinion scores, MOS) of each stereoscopic retargeted image is subjectively rated by 30 viewers. For objective assessment, several publicly available quality evaluation metrics are tested on the database. Experimental results show that there is a large room for improving the accuracy of objective quality assessment in SIRQA by comprehensively considering geometric distortion, content loss and stereoscopic perceptual quality.**

I. INTRODUCTION

With the popularity of various 3-D display devices and applications, such as 3-D televisions and heterogeneous mobile phones, the requirements for displaying 3-D media on different terminal devices with different sizes and aspect ratios are much urgent [1]. To adapt stereoscopic images into different devices, the stereoscopic images should be modified appropriately to preserve original 3-D scene structures and avoid incorrect 3-D perception, and this process is described as stereoscopic image retargeting (SIR) [2].

Recently, many SIR operators have been proposed and can be broadly classified into two categories. The first category is discrete approach. Stereoscopic cropping (SCR) and stereoscopic seam carving (SSC) are two representative discrete algorithms. SCR generates a retargeted stereoscopic image via calculating the position of the clipping window, meanwhile SSC iteratively removes a pair of seams from stereoscopic image pair. For example, Wang *et al* [3] proposed a SCR algorithm based on stereoscopic saliency information. Basha *et al*. [4] proposed a geometrically consistent SSC algorithm by considering the visibility relations between pixels in the image pair. Lei *et al*. [5] proposed a pixel fusion-based SSC method that extends the single pixel fusion-based way to be applicable for stereoscopic image pair. The second category is continuous approach based on stereoscopic warping (SWARP). For example, Chang *et al*. [6] proposed a content-aware SWARP method for editing stereoscopic image to different displays based on sparse stereoscopic correspondences and disparity consistency constraints. Lee *et al*. [7] proposed a layer-based SWAPR method, in which each layer is warped by the mesh deformation. Li *et al*. [8] presented a method that can preserve the object shape and the depth simultaneously. Shao *et al*. [2] presented a QoE-guided SLWARP algorithm by taking user's Quality of Experience (QoE) into account.

However, different SIR methods have their advantages and disadvantages in shape preservation, content preservation and creating a comfortable and enjoyable 3-D experience. To prove the superiority of their retargeting results, most existing studies employ a small-scale subjective test. Although subjective evaluation is most reliable, it is expensive, time-consuming, and difficult to be embedded into the online optimization systems [9]. Therefore, objective stereoscopic image retargeting quality assessment (SIRQA) metrics should be presented to automatically select the best retargeting results for real applications.

Unfortunately, few works are presented in SIRQA [10-11]. To study the perceptual quality of stereoscopic retargeted images, we construct a SIRQA database, which contains 720 stereoscopic retargeted images generated by eight different SIR methods from 45 original stereoscopic images on two retargeting scales. The perceptual quality of each stereoscopic retargeted image is subjectively rated by 30 graduate students (20 male and 10 female) engaged in digital image processing. After processing the subjective ratings, the mean opinion scores (MOS) are obtained. Additionally, several publicly available quality metrics for stereoscopic retargeted images are evaluated on the built database.

The rest of this paper is organized as follows. Section II introduces the details of our database. In Section III, some objective quality metrics are introduced and evaluated on the built database. Finally, conclusions are presented in Section IV.

II. DATABASE

We build a SIRQA database for public research, which contains 720 stereoscopic retargeted images generated by eight representative SIR operators. In this section, we will introduce the database in detail.

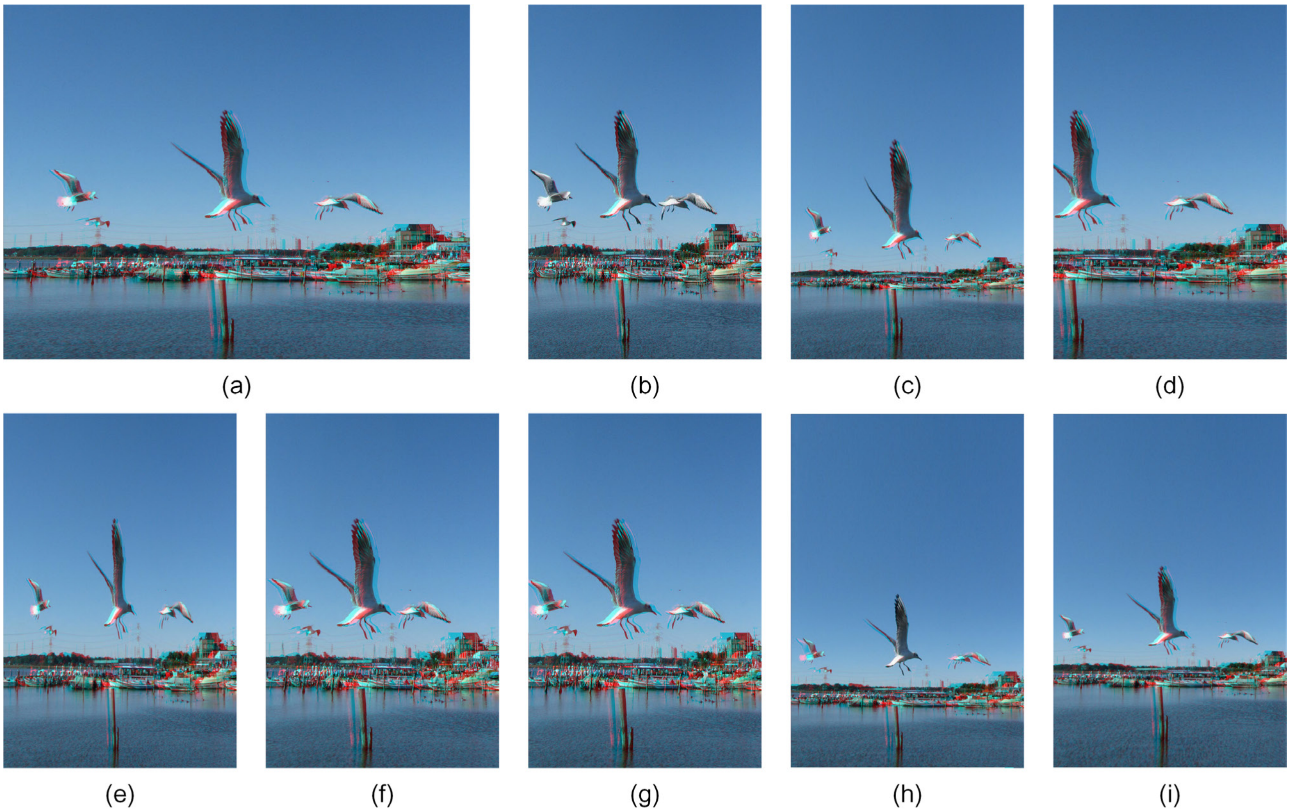

Fig. 1. Examples of stereoscopic retargeted images generated by eight SIR algorithms. (a) Original image. (b) MSC. (c) MSNS. (d) CPC. (e) SSCL. (f) GCSSC. (g) VASSC. (h) QOE. (i) SLWAP.

A. Original Images

Content-aware SIR methods generate retargeted images with high perceptual quality in which the background content can be directly removed or efficiently compacted, while the important foreground objects will be preserved. Furthermore, 3-D experiences such as visual comfort and depth perception have a crucial impact on SIR. In order to build a reasonable database and effectively reflect the performance of different SIR algorithms, we select 45 original stereoscopic images, including natural scenery, foreground object, geometric structure, face and people, and other indoor and outdoor scenes. Moreover, the original stereoscopic images selected in this paper have different disparity ranges, which can be used to clarify the influence of depth and visual comfort.

B. Retargeting Methods

Eight representative SIR methods including four discrete methods and four continuous methods are applied to build the database. In our database, the resolution of all original stereoscopic images is adjusted only in horizontal direction, and all original stereoscopic images are resized on two scales (shrinking the width to 75% and 50% of the original width) according to most of the existing SIR operators. The selected SIR algorithms are detailed in the following.

- Monocular seam carving (MSC) [12]: left and right images are resized by seam carving algorithm respectively.
- Monocular scale and stretch (MSNS) [13]: left and right images are resized by scale and stretch algorithm respectively.
- Content persistent cropping (CPC) [3]: using stereoscopic saliency information to calculate the position of the clipping window and automatically cropping the original stereoscopic image.
- Stereo scaling (SSCL): simply scaling the stereoscopic image into the target size.
- Geometrically consistent stereo seam carving (GCSSC) [4]: removing pixels that belong to the non-informative regions in left image and right image at the same time according to the principle of geometric consistency.
- Visual attention guided seam carving (VASSC) [14]: applying stereoscopic visual attention and binocular just-noticeable-difference (BJND) models to select seams in left and right images and seam replacement is performed for the occluded regions to prevent the geometry inconsistency.
- QoE-guided warping (QOE) [2]: taking shape preservation, visual comfort preservation, and depth perception preservation into account, and solving the optimization model in 3-D space. QOE aims to promote the user's quality of experience by preserving 3-D scene structure,

and balancing visual comfort and depth perception.

- Single-layer warping (SLWAP) [6]: an objective energy function is optimized based on feature matching and disparity consistency constraint. SLWAP can adapt depth to the comfort zone of the display while preserving the perceived shapes of prominent objects.

Among these eight methods, CPC, GCSSC, MSC and VASSC are discrete methods, while QOE, SSCL, SLWAP and MSNS are continuous algorithms. In particular, MSC and MSNS directly apply 2-D retargeting methods to 3-D images. It is known that some SIR methods aim to preserve the original disparity especially the SSC and SCR based algorithms and other methods aim to optimize the original disparity. In this paper, we focus on assessing perceptual quality of each stereoscopic retargeted image no matter how it is generated and what the resolution is. A set of retargeted images generated by eight SIR methods are shown in Fig. 1.

C. *Subjective Test*

A double stimulus continuous quality scale test methodology was used in the subjective test, in which the reference and the retargeted images are simultaneously presented on the screen. The subjective test environment and condition are described in ITU-R BT.500-11 and ITU-R 1438. A Samsung UA65F9000 65-inch Ultra HD 3D-LED TV with 3-D shutter glasses was used for the test. The default viewing distance was 3 times the screen height. Each stereoscopic retargeted image was randomly displayed on the screen lasting for 10s, and another 5s for voting. In order to avoid the contextual and memory effects on the subjective judgment, retargeted images generated from the same original image will not be presented consecutively. A total of 30 graduate students were participated in the subjective test. During the test, the participants were asked to rate the perceptual quality of stereoscopic retargeted images on a five-level scale: Excellent = 5, Good = 4, Fair =3, Poor =2 and Bad =1. After subjective voting, the raw subjective scores were converted to z-scores, and then scaled to the range of [0, 100] after removing outliers. Finally, the MOS value of each stereoscopic retargeted image is obtained as the mean of rescaled z-scores.

D. *Analysis and Discussion of the Subjective Score*

After subjective test, the SIRQA database is built, which comprises of the stereoscopic retargeted images and their corresponding MOS values. The MOS distribution is shown in Fig. 2. From Fig. 2, we can observe that the range of MOS is from 20 to 80, and most MOS distributions are concentrated upon the middle quality level, indicating the good separation of subjective quality.

In our database, the average MOS value under 50% retargeting scale (shrinking the width to 50% of the original width) is 39.47, and the average MOS value under 75% retargeting scale (shrinking the width to 75% of the original width) is 60.53, which indicate that it is difficult to preserve 3-D scene structure and avoid incorrect 3-D perception for a large retargeting scale.

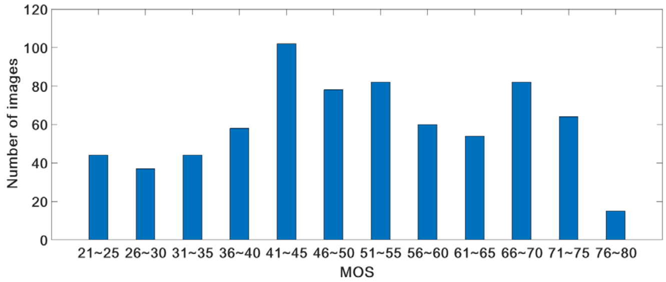

Fig. 2. The MOS distribution of the database

III. OBJECTIVE QUALITY METRIC

We adopt several publicly available quality evaluation metrics to test the database. The information about the metrics and the objective test results are detailed in the following.

A. *Quality Metric*

- SIFT-flow [15]: using SIFT descriptors to establish the relationship between the original and retargeted images, and the cost function based on the displacements of adjacent pixels is considered as the dissimilarity measure of two images.
- Bidirectional Similarity (BDS) [16]: computing a bidirectional mapping distances between the patches in the original and retargeted images.
- Earth-Mover's Distance (EMD) [17]: a dissimilarity index defined as the minimal cost in transforming the original distribution into target distribution.
- Hand-Craft and Deep Learned features (HCDL) [18]: using hand-crafted features and deep-learned features to evaluate the quality of retargeted image.
- Aspect Ratio Similarity (ARS) [19]: interpreting the geometric change by a backward registering in Markov random field and evaluating the aspect ratio similarity in local grids.
- Disparity Amplitude and Gradient (DAG) [20]: an attention model based visual discomfort assessment metric that extracted perceptually significant disparity amplitude and disparity gradient features.
- Image Quality and Stereoscopic Perception (IQSP): fusing ARS and DAG features to predict stereoscopic retargeted image quality via Support Vector Regression (SVR).

Since SIFT-flow, BDS, EMD, HCDL and ARS are 2-D image quality evaluation metrics, the mean score of two views is computed as the final perceptual quality for stereoscopic retargeted image. In the experiment, we employ Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SRCC) and root mean square error (RMSE) to evaluate the relationship between the objective quality metric scores and the provided MOSs. PLCC and RMSE are calculated after nonlinear regression [21]:

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

B. Performances

Table 1 gives the comparisons of PLCC, SRCC and RMSE. From the table, we can make the following observations: 1) SIFT-flow, BDS and EMD have poor correlation with subjective scores, due to ignoring image contents and stereoscopic perception; 2) HCDL and ARS achieve better evaluation performance compared with SIFT-flow, BDS and EMD, because these two metrics can capture the shape distortion and content loss more accurately; 3) Since DAG simply considers the stereoscopic perceptual quality, the evaluation results are not satisfactory; 4) IQSP achieves the best performance compared with other metrics by fusing image quality and stereoscopic perceptual quality.

Table I. Performance of different methods on the database.

Metric	PLCC	SRCC	RMSE
SIFT-flow	0.1019	0.0542	14.8731
BDS	0.2791	0.2751	14.3882
EMD	0.3480	0.3858	14.0167
HCDL	0.7261	0.7215	11.8422
ARS	0.7833	0.7745	9.2950
DAG	0.2841	0.2614	14.4157
IQSP	0.8203	0.8154	8.5934

IV. Conclusions

In this paper, we constructed a SIRQA database to study the visual quality of stereoscopic retargeted images. The database includes 720 stereoscopic retargeted images generated by eight typical SIR algorithms. We conducted subjective and objective tests on the database, and the experimental results show that comprehensively considering geometric distortion, content loss and stereoscopic perception can effectively improve the accuracy of SIRQA. For the future works, we will focus on designing special SIRQA measures via considering image semantic and stereoscopic perception.

References

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

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

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

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

[5] J. Lei, M. Wu, C. Zhang, F. Wu, N. Ling and C. Hou, "Depth-preserving stereo image retargeting based on pixel fusion," IEEE Transactions on Multimedia, vol. 19, no. 7, pp. 1442-1453, July 2017.

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

[7] K. Lee, C. Chung and Y. Chuang, "Scene warping: Layer-based stereoscopic image resizing," in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 49-5, 2012.

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

[9] Q. Jiang, F. Shao, W. Lin and G. Jiang, "Learning sparse representation for objective image retargeting quality assessment," IEEE Transactions on Cybernetics, vol. 48, no. 4, pp. 1276-1289, April 2018.

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

[11] Y. Zhou, W. Zhou, P. An and Z. Chen, "Visual comfort assessment for stereoscopic image retargeting," in Proc. of IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1-5, May 2018.

[12] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," ACM Transactions on Graph, vol. 26, no. 3, article no. 10, 2007.

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

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

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

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

[17] O. Pele and M. Werman, "Fast and robust earth mover's distances," in Proc. of IEEE International Conference on Computer Vision (ICCV), pp. 460-467, 2009.

[18] 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.

[19] Y. Zhang, Y. Fang, W. Lin, X. Zhang, L. Li, "Backward registration based aspect ratio similarity for image retargeting quality assessment," IEEE Transactions on Image Processing, vol. 25, no. 9, pp. 4286-4297, Sep. 2016.

[20] 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 Transactions on Circuits and Systems for Video Technology, vol. 23, no. 12, pp. 2077-2082, Dec. 2013.

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