PIXEL FUSION BASED STEREO IMAGE RETARGETING

Bahetiyaer Bare, Ke Li, Bo Yan, Xiaoyu Qi*

School of Computer Science
Shanghai Key Laboratory of Intelligent Information Processing,
Fudan University, Shanghai, China
byan@fudan.edu.cn

Hamid Gharavi

Advanced Network Technologies Division
National Institute of Standards and Technology (NIST), 100 Bureau Drive
Gaithersburg, MD 20899
gharavi@nist.gov

ABSTRACT

Image retargeting attempts to adapt images to different devices while preserving the salient contents. Most existing methods address retargeting of a single image. In this paper, we propose a novel image retargeting method for resizing a pair of stereo images. Naively retargeting each image independently will distort the geometric structure and will impair the perception of the 3-D structure of the scene. We introduce an extension to the 2-D image retargeting method that works on a pair of stereo images. We demonstrate the performance of our method on a number of challenging indoor and outdoor stereo images. Experimental results show that our method is able to provide visually comfortable resized images when the resizing ratio is relatively high.

Index Terms— Stereo images, image retargeting, seam searching, pixel fusion

1. INTRODUCTION

Due to recent advances in image processing technology, 3-D stereoscopic images and videos have become increasingly popular. As stereoscopic images have a fixed resolution and aspect ratio, they do not always fit the devices. So we need to adapt them to fit different screens. This process is called image retargeting. Here, we first introduce two main types of image retargeting methods and then discuss the extensions which apply to 3-D stereoscopic images.

With extensive investigation in image retargeting algorithms, two main kinds of methods emerged. One is discrete, including seam carving [1] and shift map [2], which removes and shifts pixels in the images. The other is continuous method, including scale-and-stretch [3] and warping [4], which are processed in a quad mesh based on image content.

Shai Avidan [1] developed seam carving, which has been considered as a basic image retargeting algorithm. It finds the least important seams in each row of an image, namely, finds the path of minimum cost from one end to another and then automatically removes them to reduce the image size or insert seams to extend it. Thus, it adapts the size of the image to the display terminals without distorting important objects.

Previous works mainly focus on a single image or video. With 3-D contents coming into the consumer market, it is necessary to extend these methods to 3-D areas. Basha $et\ al.$ [5] extended the seam carving algorithm [1] [6] to work on stereo images. The input of their method is a rectified stereo image pair and a disparity map. Based on seam carving, the algorithm iteratively removes a pair of seams in both the left and right images rather than removing seams independently. It processes seam coupling to correlate the right image to the left using the disparity map. Then, in order to avoid depth distortion, it considers both appearance and depth energy. As it removes a pair of pixels which are not occluded or occluding ones at one time, it does not generate geometrical inconsistency. So the method can be applied to make the retargeted pair suitable for stereoscopic display.

However, there exist drawbacks since the algorithm uses seam carving when it processes image retargeting. It has the same disadvantage as seam carving, namely, distorting important objects when an image is overly shrunken. It is also difficult to find unnoticeable areas if the significant objects have large featureless areas. So it is not sufficiently content aware.

As to continuous method, image warping digitally manipulates an image to make any shapes portrayed in it significantly distorted. It is used to correct image distortion or for creative purposes. This technique is also applicable to video.

Chang $et\ al.$ [7] extended the warping-based retargeting method. In order to keep disparity consistency, it uses the SIFT (scale-invariant feature transform) feature to retain the disparity of the corresponding SIFT feature points. Since it depends on the distribution of feature points, the perceived depth of the featureless region is more likely to be distorted if the distribution is not so good.

In this paper, we firstly review the related work in section 2. Then in section 3, based on the seam searching based pixel fusion method [8], we propose an improved 3-D image retar-

*This work is supported by NSFC (Grant No.: 61370158).

geting method. Experimental results demonstrated in section 4 prove that our method is able to provide visually comfortable resized images when the resizing ratio is relatively high. Finally, we draw our conclusions in section 5.

2. RELATED WORK

Basha *et al.* [5] have proposed a geometrically consistent approach to retarget 3-D stereo images based on seam carving [1] [6]. The input of the method is a $W \times H$ rectified stereo image pair and a disparity map D, where the disparity map is computed using the SGM (semiglobal matching) method [9] with the input image pair.

First, they do seam coupling to match each seam in the right image to that of the left. The geometric coupling of the two seams, $S_L = \{s_L^i\}_{i=1}^W$ and $S_R = \{s_R^i\}_{i=1}^W$, is obtained by using the correspondence defined by disparity map D, as follows [5]:

$$s_R^i = (i, j_R(i)) = (i, j_L(i) + D(s_L^i)), \tag{1}$$

where $s_L^i = (i, j_L(i)) \in S_L$ and $s_R^i = (i, j_R(i)) \in S_R$ are the seam's pixels in the left and right image at row i, $j_L, j_R : [W] \rightarrow [H]$, and $[W] = [1, ..., W]$. Without loss of generality, the method then considers piecewise seams since the seams may be piecewise continuous.

In order not to lose geometrical consistency, the energy function of this method includes both appearance energy and depth energy. Since the resulting gradients in the retargeted left and right images depend on the seam pixel in the previous row, denoted by $j_L^{\pm}$ and $j_R^{\pm}$, the method defines the energy function (w.r.t. the left image) in accordance with the seam pixel in the previous row, $j^{\pm}$ (which is short for $j_L^{\pm}$). That is [5],

$$E_{total}(i, j, j^{\pm}) = E_{\text{intensity}}(i, j, j^{\pm}) + \alpha E_{3D}(i, j, j^{\pm}), \tag{2}$$

where α controls the relative impact of each of the terms, $E_{intensity}$ is the appearance energy and E_{3D} is the depth energy. Since they use generalized seams, $j^{\pm} \in [W]$ can be any pixel in row $i-1$ (unlike the continuous case in which $j \in \{j-1, j, j+1\}$).

The appearance energy is the sum of the energy of the left and right image which is generalized from [6]. That is [5],

$$E_{\text{intensity}}(i, j, j^{\pm}) = E_L(i, j, j^{\pm}) + E_R(i, j_R, j_R{}^{\pm}), \tag{3}$$

where E_L and E_R are the energy of the left and right image. The depth energy is a weighted sum between three components [5]:

$$E_{3D}(i, j, j^{\pm}) = E_D(i, j, j^{\pm}) + \beta |D_n(i, j)| + \gamma G(i, j), \tag{4}$$

where E_D is the 3-D forward energy term, D_n is the normalized disparity map D, $G(i, j)$ is the difference in the intensities of corresponding pixels, β and γ are weights set by the user.

Seam carving is then processed with the constraints that corresponding pixels are either both removed or remain corresponding in the output images and that 3-D points are not revealed if they are visible in the reference view but occluded in the other. The method defines a cost matrix M to select seam pairs by dynamic programming. To maintain the constraints, it sets $M(i, j) = \infty$ for pixels that not meeting the constraints. Other pixels are divided into continuous and discontinuous. The calculation of $M(i, j)$ is as follows [5]:

$$M(i, j) = \left\{ \begin{array}{c} \min\limits_{j^{\pm} \in \{j-1, j+1\}} E_{total}(i, j, j^{\pm}); T(i, j) = 0 \\ \min\limits_{j^{\pm} \in \{W\}} E_{total}(i, j, j^{\pm}); T(i, j) = 1 \end{array} \right\}, \tag{5}$$

where T is the binary map of size $H \times W$. $T(i, j)$ indicates whether a continuous path is blocked in row $i-1$ by occluding/occluded pixels [5].

Removing a seam pixel from a row results in shifting pixels in that row. It defines the shifting function by [5]

$$f_L(i, j) = \left\{ \begin{array}{c} j, j < j_L(i) \\ j-1, j > j_L(i) \\ \infty, j = j_L(i) \end{array} \right\}, \tag{6}$$

where $f_L(i, j) : [W] \times [H] \rightarrow [W] \times [H-1]$ maps the i^{th} input row to the i^{th} output row, $(i, j_L(i))$ is the pixel removed from the left image. Likewise, $f_R(i, j)$ is the corresponding mapping function in the right image, where $j_L(i)$ is replaced by $j_R(i)$ [5].

The disparity map D is updated after carving each seam for the next process until the image pair meets the need. Finally, a retargeted stereo image pair and an updated disparity map can be obtained from this method.

3. THE METHOD

The input to our method is a pair of $W \times H$ rectified stereo image pairs $\{I_L, I_R\}$ and a disparity map D. We compute the disparity map D by the OpenCV implementation of the SGM [9] stereo algorithm. We use our 3-D retargeting method which is based on a seam searching based pixel fusion method [8] and specially designed energy for 3-D scene, to obtain retargeted stereo image pair $\{I'_L, I'_R\}$. The goal of our method is to obtain retargeted images that are visually comfortable and geometrically consistent in 3-D scene.

The total energy should include depth energy and appearance energy. Different from [5], since our method does not remove seam pixels from original images, our energy function does not consider the seam pixel in the previous row, which is denoted as $j^{\pm}$ (see section 2). The computed depth map provides valuable cues for the retargeting method. Depth energy term E_D is used to minimize the disparity distortion. We define the depth energy term E_D by the gradient of the input disparity map D, as follows:

$$E_D = |\nabla D| \,. \tag{7}$$

Our method strongly relies on the disparity map, which is computed by SGM [9] stereo algorithm that is regarded as a black box. Estimated map errors may result in incorrect coupling of pixels. We minimize pixels for which we have high confidence of disparity values, measured by the $Diff$. $Diff$ is the difference in the intensities of corresponding pixels. That is,

$$Diff(i,j) = |I_L(i,j) - I_R(i,j+D(i,j))| \,, \tag{8}$$

where I_L and I_R are the left and right images, (i,j) is the coordinate of pixels and D is the disparity map.

The total depth energy E_{3D} is a weighted sum between three components:

$$E_{3D}(i,j) = E_D(i,j) + \beta|D_n(i,j)| + \gamma Diff(i,j), \tag{9}$$

where β and γ are parameters set by the user, D_n is the normalized disparity map D.

Now we consider the appearance energy. Appearance energy should include the pair of stereo images' importance map. In our method, importance map is represented by the saliency of every pixel. The saliency is computed by Itti's algorithm [10] (We use a MATLAB[1] implementation from http://www.saliencytoolbox.net). Accordingly, the appearance energy is given by:

$$E_{intensity}(i,j) = SM_L(i,j) + SM_R(i,j), \tag{10}$$

where SM_R and SM_L are the saliency map of the right image and the left one respectively.

According to the definition of depth energy E_{3D} and the appearance energy $E_{intensity}$, the total energy is given by:

$$E_{total}(i,j) = \alpha E_{3D}(i,j) + E_{instensity}(i,j), \tag{11}$$

where α is a parameter set by the user to change the weight between each component. Then, we normalize the result in the range of 0 to 1. Particularly, since we do not remove the occluding or occluded pixels, we set $E_{total}(i,j) = 1$ for these pixels.

[1]The use of this product is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products identified are necessarily the best available for the purpose.

We use seam searching based pixel fusion method [8] to resize 2-D images. First, this method generates the importance map of the original image. Then, it calculates the number of searched seams. Next, based on this number, seam carving is applied in the original image and all the pixels can be divided into W groups. Furthermore, each group's scaling factor is assigned based on whether it reduces the image width or not. Finally, pixel fusion which is proposed by [11] is used to generate the retargeted image based on the assigned scaling factors [8].

We also use seam searching based pixel fusion algorithm [8] and the computed energy E_{total} to retarget the left image I_L and get the result I'_L. To get the retargeted right image, we use the allocated scaling factors of left image and the disparity map D. That is,

$$SF_R\,(i,j) = SF_L[(i,j) + D(i,j)], \tag{12}$$

where SF_R and SF_L are the allocated scaling factors of the right image and the left one, (i,j) is the coordinate of the pixels. Then, we use pixel fusion [11] to generate the retargeted right image with SF_R. Finally, we can obtain the retargeted stereo image pair $\{I'_L, I'_R\}$.

4. EXPERIMENTS AND RESULTS

We tested our method on very challenging outdoor and indoor scenes. These scenes are challenging because the scenes are highly textured and contain objects at different depths. In all experiments, we used the OpenCV implementation of the SGM stereo algorithm [9] to compute disparity. We filled the regions with its original values for which the disparity was not computed. Our algorithm was implemented in MATLAB[1] and the datasets are publicly available.

We tested our method on the following datasets:

a) Flickr[1]: A set of stereo images (Figure 1), with large depth range, are downloaded from Flickr[1]. The images were manually rectified using [12].

b) Portrait: A pair of images (Figure 2), provided by [13]. The main challenge in this pair is that the salient objects covers the most of the image and these salient objects should not be distorted.

At the beginning, we use our method to retarget the input stereo image pair. Then, we generate the disparity map D_{SGM} by using the retargeted image pair. As defined by Basha *et al.* [5], for maintaining geometric consistency, the resized image should have a similar disparity map to the original disparity map. For comparison, we take the left image to be the reference image and use its scaling factors to retarget the disparity map D. D_0 is the retargeted disparity map, which is taken to be the reference disparity map. Finally, the depth distortion is measured by $|D_{SGM} - D_0|$.

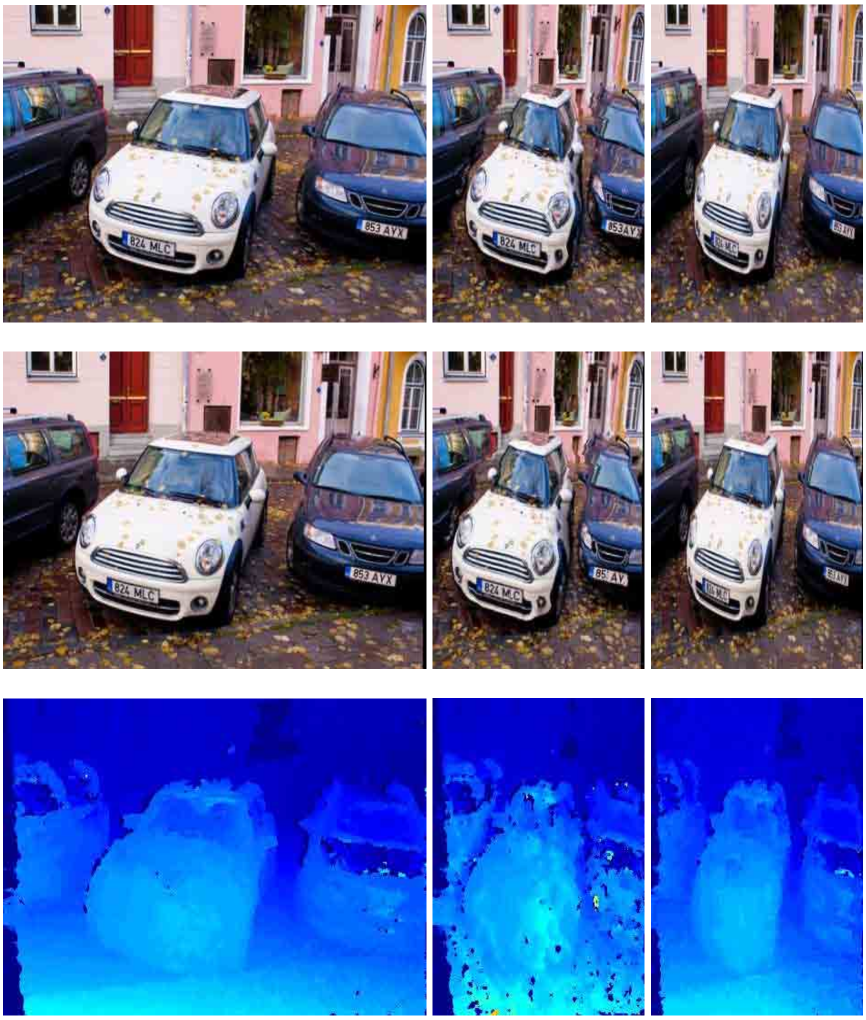

Fig. 1. Car dataset. In the first column (from left to right): the input left image, the input right image and the input disparity map (top to bottom). In the second column, the results of the method of Basha $et\ al.$ [5]. In the third column, the results of our method. The depth distortion scores: Basha $et\ al.$ [5], $B = 52.7\%$; Our method, $B = 2.8\%$.

For quantitative evaluation, we employ B value, which is proposed in [5], to be the depth distortion score. The B value is calculated for all our experiments. B value is computed by the percentage of pixels whose depth, D_{SGM}, has been changed by more than one pixel. Namely [5],

$$B = \frac{1}{N} \sum_{(i,j)} (|D_0(i,j) - D_{SGM}(i,j)| > 1), \tag{13}$$

where D_0 is the retargeted disparity map D, (i,j) is the coordinate of the pixels and N is the total pixel number.

We tested our algorithm on the challenging 3-D scenes using a fixed set of parameters for the 3-D weight: $\beta = 0.68$, $\gamma = 0.5$ and $\alpha = 0.5$. The image width was reduced by 50% for the all datasets. The results are shown in Figures 1 to 2. For comparison, we took the method of Basha $et\ al.$ [5] as reference method.

Car dataset (Figure 1) includes three main objects (the three cars) and how to preserve the three main objects become a challenge. Figure 1 shows that our method well preserves the three main objects and avoids the possible distortions. Basha $et\ al.$'s [5] method distorts the main objects obviously.

The main challenge of the Diana dataset (Figure 2) is that salient object, which covers most of the image, should not be distorted. Moreover, a significant part of the left image is beyond the view of the right camera, this part cannot be removed by retargeting algorithm. Figure 2 shows that our method well preserves the face appearance as well as face depth. In comparison, Basha $et\ al.$'s [5] method distorts the face appearance obviously and removes the significant part of the left image which is beyond the view of the right camera.

The depth of the retargeted 3-D image should be perceived same as that of the input 3-D image. In Figure 3,

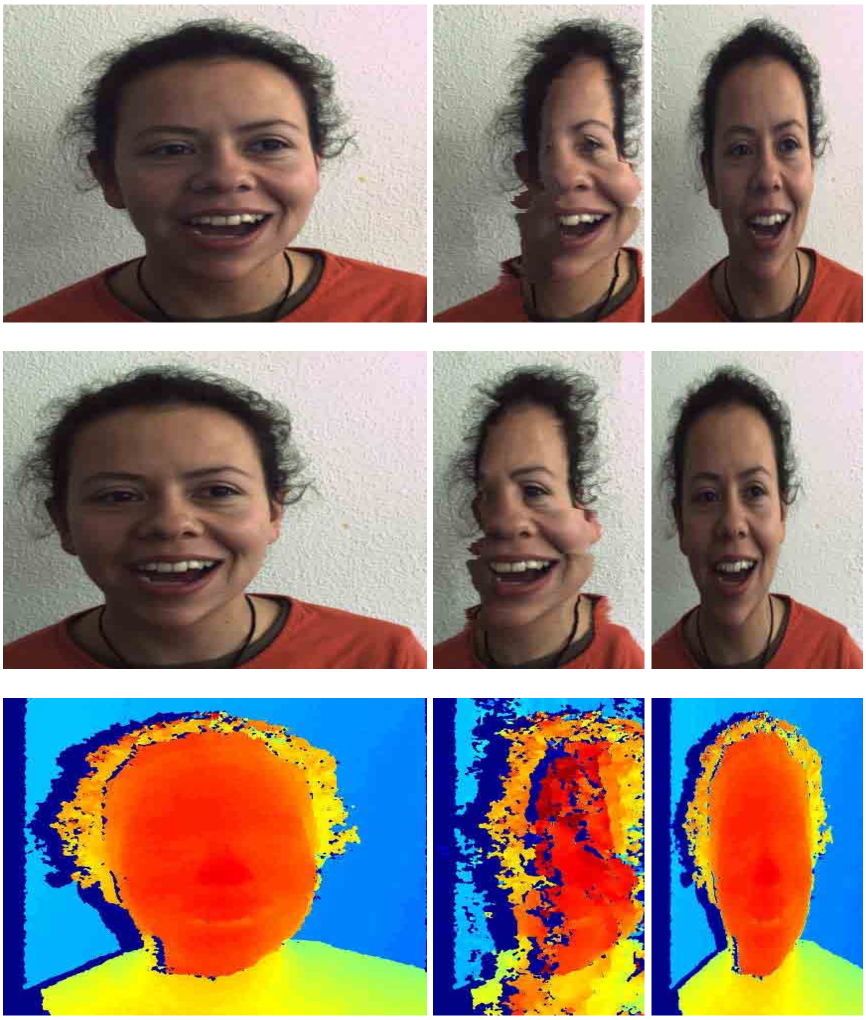

Fig. 2. Diana dataset. In the first column (from left to right): the input left image, the input right image and the input disparity map (top to bottom). In the second column, the results of the method of Basha $et\ al.$ [5]. In the third column, the results of our method. The depth distortion scores: Basha $et\ al.$ [5], $B = 43.8\%$; Our method, $B = 8.6\%$.

we use anaglyph (red-green) images and depth distortions ($|D_{SGM} - D_0|$) to demonstrate our method's performance. As we can observe from Figure 3, our method can retarget the input 3-D image naturally and can preserve its depths as well. Method of Basha $et\ al.$ [5] distorts the main object and has obvious depth distortions.

The calculated depth distortion scores which are given in Table 1 proves that our method achieves good performance than Basha $et\ al.$'s [5] method. For example, our method generates 2.8% depth distortion in Car dataset, which is better than that of Basha $et\ al.$'s method (52.7% depth distortion).

Table 1. Depth distortion scores in terms of B value in (13).

	Basha's method [5]	our method
Car dataset	52.7%	2.8%
Diana dataset	43.8%	8.6%

5. CONCLUSIONS

In this paper, we presented a novel 3-D image retargeting method for retargeting a pair of stereo images. We extended the seam searching based pixel fusion method to work on a pair of stereo images and achieved good performance. The most remarkable feature of our method is preserving the salient content as well as its depth without possible distortions, which may be generated by other methods. Experimental results proved that our method is able to provide visually comfortable resized images when the resizing ratio is relatively high.

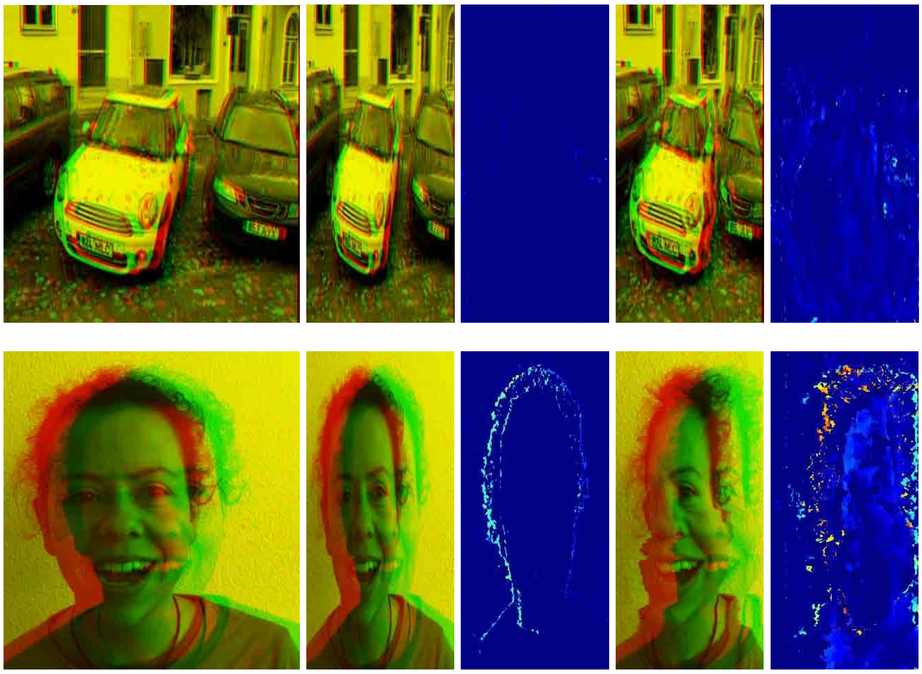

Fig. 3. Anaglyph images and depth distortions. First column (from left to right) is the input stereo pair of each of the input image pairs: Car and Diana (from top to bottom). Second column is our result. Third column is the depth distortion of our result. Fourth column is the result of Basha $et\ al.$ [5]. Fifth column is the depth distortion of Basha $et\ al.$ [5]. The anaglyph images should be viewed with red-green glasses.

6. REFERENCES

[1] Shai Avidan and Ariel Shamir, "Seam carving for content-aware image resizing," in *ACM Transactions on graphics (TOG)*, 2007, p. 10.

[2] Yael Pritch, Eitam Kav-Venaki, and Shmuel Peleg, "Shift-map image editing," in *IEEE 12th International Conference on Computer Vision*, 2009, pp. 151–158.

[3] Yu-Shuen Wang, Chiew-Lan Tai, Olga Sorkine, and Tong-Yee Lee, "Optimized scale-and-stretch for image resizing," in *ACM Transactions on Graphics (TOG)*, 2008, p. 118.

[4] Lior Wolf, Moshe Guttmann, and Daniel Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *IEEE 11th International Conference on Computer VisionICCV'2007*, 2007, pp. 1–6.

[5] Tali Basha, Yael Moses, and Shai 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.

[6] Michael Rubinstein, Ariel Shamir, and Shai Avidan, "Improved seam carving for video retargeting," in *ACM Transactions on Graphics (TOG)*, 2008, p. 16.

[7] Che-Han Chang, Chia-Kai Liang, and Yung-Yu Chuang, "Content-aware display adaptation and interactive editing for stereoscopic images," *IEEE Transactions on Multimedia*, vol. 13, no. 4, pp. 589–601, August 2011.

[8] Bo Yan, Ke Li, Xiaochu Yang, and Tingxiao Hu, "Seam searching based pixel fusion for image retargeting," *IEEE Transactions on Circuits and Systems for Video Technology*, 2014.

[9] Heiko Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 2, pp. 328–341, 2008.

[10] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[11] Tzu-Chieh Yen, Chia-Ming Tsai, and Chia-Wen Lin, "Maintaining temporal coherence in video retargeting using mosaic-guided scaling," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2339–2351, 2011.

[12] Andrea Fusiello, Emanuele Trucco, and Alessandro Verri, "A compact algorithm for rectification of stereo pairs," *Machine Vision and Applications*, vol. 12, no. 1, pp. 16–22, 2000.

[13] Frederik Huguet and Frédéric Devernay, "A variational method for scene flow estimation from stereo sequences," in *IEEE 11th International Conference on Computer Vision (ICCV'2007)*, 2007, pp. 1–7.