Content-Driven Retargeting of Stereoscopic Images

Jin Woo Yoo, Sehoon Yea, *Member, IEEE*, and In Kyu Park, *Member, IEEE*

***Abstract*—This letter proposes a novel warping-based method for the content-driven retargeting of stereoscopic images. Conventional algorithms in single image retargeting generally do not consider the scene depth saliency and disparity consistency when applied independently to left and right images . Therefore, the salient region and stereoscopic correlation of independently retargeted images can become corrupted. On the other hand, the proposed algorithm retains the stereo consistency of the retargeted images by matching the vertices of the grid and preserving the correspondence between them. Vertex disparity is propagated by a GPU interpolation to construct a sparse disparity map. To improve the capability of preserving visually important regions, the sparse disparity is used in conjunction with the image gradient in a saliency map computation. The experimental results show that the proposed method retains the correct disparity, while the salient objects remain undistorted in the retargeted stereoscopic image pairs.**

***Index Terms*—Disparity consistency, retargeting, sparse disparity, stereoscopic image.**

I. Introduction

RECENTLY, 3-D stereoscopic images and videos have become ubiquitous due to the rapid increase in contents as well as advances in hardware capability. Stereoscopic images and video contents are widely available in 3-D cinemas and 3-D broadcasting. They can be created and displayed using recently developed hardware such as 3-D camera rigs, 3-D televisions and 3-D smartphones.

To display a stereoscopic image on a range of devices with different resolutions and aspect ratios, naive methods simply resize or crop the original image to fit into the target display. However, they might cause a loss or unpleasant artifact of visually salient regions or objects. Therefore, the input images should be adjusted appropriately by considering image contents adaptively. Existing content-aware, image retargeting methods identify the important regions or objects, and minimize the distortion during resizing. Over the last few years, most existing methods of media retargeting deal with monocular images and videos only [1]–[6].

Manuscript received December 29, 2012; revised February 26, 2013; accepted March 06, 2013. Date of publication March 11, 2013; date of current version April 02, 2013. This work was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (2012R1A1A2009495), and also supported by Inha University. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ce Zhu.

J. W. Yoo and I. K. Park are with the School of Information and Communication Engineering, Inha University, Incheon 402-751, Korea (e-mail: jinwoo1028@gmail.com; pik@inha.ac.kr).

S. Yea is with LG Electronics, Seoul 137-130, Korea (e-mail: sehoon.yea@lge.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2013.2252165

Because a stereoscopic image pair consists of left and right images, it is essential for any retargeting technique to maintain the perceived depth while leaving the visually important regions undistorted. The natural extension of existing single image retargeting methods performs retargeting on the left and right images independently. However, the perceptual depth in the retargeting result can be distorted severely because single image retargeting algorithms do not consider the disparity consistency between a stereoscopic image pair.

Research in retargeting stereoscopic images and videos is still in its early stages and few studies have been reported. Basha *et al.* [7] and Utsugi *et al.* [8] extended the seam carving algorithm to stereoscopic images. In [7], the connected seam assumption is alleviated to piecewise-connected seams. A dense disparity map is used to identify the corresponding seams in the left and right images. As a result, the depth information in the retargeted image is preserved. However, similar to the original seam carving method for a single image, this method fails to preserve the important regions when the important regions have large featureless areas. Furthermore, an accurate disparity map should be calculated to find the seam correspondence. Chang *et al.* extended the warping-based retargeting algorithm [9] using the SIFT (scale-invariant feature transform) feature to retain the disparity of the corresponding SIFT feature points during retargeting to preserve the disparity consistency. However, the perceived depth in the featureless region is more likely to be distorted because it depends on the good distribution of feature points.

This letter presents a novel warping-based retargeting method for stereoscopic images. The proposed method reduces significantly the distortion of visually important regions as well as the distortion of the perceived depth in the retargeted images. Using the common knowledge of 3-D video authoring, where the object-of-interest is located closer to the camera, the sparse disparity map is constructed by vertex matching and interpolation on the mesh grid. Furthermore, a novel energy function is proposed to accommodate the sparse disparity map in the mesh warping procedure. As a consequence, the ability to preserve visually important regions with a salient depth is improved significantly.

II. Saliency Detection for Stereoscopic Images

The left image is subdivided using a uniform mesh grid $\mathbf{M} = (\mathbf{V}, \mathbf{E}, \mathbf{F})$, where $\mathbf{V}$, $\mathbf{E}$, and $\mathbf{F}$ represent the set of vertices, edges and quad faces, respectively. The corresponding grid mesh on the right image is generated by applying the proposed vertex matching. The typical resolution of the mesh grid is 20×20.

A. Vertex Matching and Sparse Disparity Map

In this letter, vertex matching is performed by modifying the conventional multi-resolution block matching. Block matching

on a single resolution would cause incorrect correspondence, particularly in a featureless region. Consequently, the sparse disparity map that interpolates the grid vertices becomes distorted too. In this approach, multi-resolution block matching is performed on three levels of the image pyramid (original, 1/4, and 1/16). The candidate vertex positions on the right image are the set of best positions at each level. The final vertex position is selected by measuring the window similarity again for the candidate vertex positions at the original level and taking the best with the maximum correlation.

A sparse disparity map is generated using the vertex matching result and hardware interpolation on the GPU. In 3-D stereoscopic images and videos, visually important objects are normally located close to the camera. Because conventional retargeting methods for a single image do not consider this, visually important objects close to the camera can be distorted easily when they have weak visual features or are textureless. In the proposed approach, a sparse disparity map is used to encode the geometric saliency evaluation.

It has been found that the sparse disparity map provides enough information to encode the geometric saliency. The retargeting result is almost equivalent even if a dense disparity map is used instead of the sparse disparity map. On the other hand, it takes significantly less time to obtain the sparse disparity map than the dense disparity map.

B. Saliency Map

The sparse disparity map can be used directly as a saliency map. On the other hand, when the entire region of the image has a similar depth, a sparse disparity map is almost uniform and does not provide any help. In this case, retargeting would work similar to simple scaling. Therefore, a desirable saliency map should consider both the visual and depth saliency. In this letter, an image gradient (magnitude) map G is used to encode the visual saliency, in which the high frequency region is assumed to be visually salient. Note that geometric saliency has positive correlation with the visual saliency. Finally, the saliency map S is defined as the weighted sum of the gradient map G and sparse disparity map D as follows.

$$S = wG + (1 - w)D \tag{1}$$

In (1), w controls the relative impact of the gradient map and the sparse disparity map. In our implementation, we assume the equal contribution from both maps, *i.e.*, $w = 0.5$. Rather than multiplying G and D, it is more effective to add them, since we don't use the quad significance as in [4] in which gradient magnitude is computed per-quad not per-pixel and consequently geometrically important but featureless regions get lost if we simply multiply them. Using the proposed saliency map, visually and geometrically important regions can be detected and preserved in further retargeting steps which are described in the next section. Fig. 1 gives an example of generating a saliency map.

III. Stereoscopic Image Retargeting

To effectively preserve the visually important regions during retargeting, the energy functions proposed by [4] are employed and improved to provided a better fit for stereoscopic images. In addition to the energy function, a novel

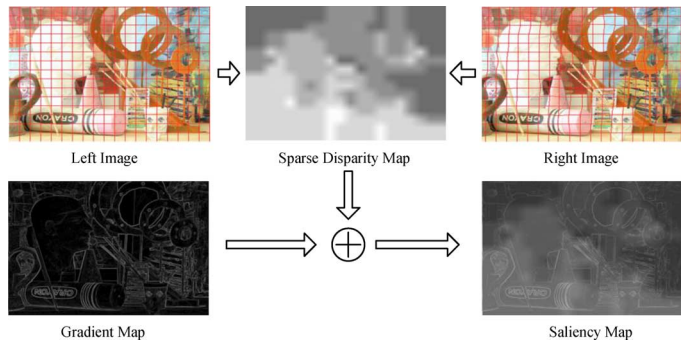


Fig. 1. Procedure of computing saliency map.

energy subfunction is proposed to consider the stereo consistency. In the retargeting procedure, the deformed vertex set $\mathbf{V}' = \{\mathbf{v}'_0, \mathbf{v}'_1, \ldots, \mathbf{v}'_{N-1}\}$ of the corresponding vertex set $\mathbf{V} = \{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{N-1}\}$ is obtained by minimizing the energy function.

A. Energy Function for Mesh Warping

The total energy function for stereoscopic image retargeting is defined as the sum of three different energy subfunctions as follows:

$$D = D_q + D_l + D_t \tag{2}$$

where D_q, D_l, and D_t are the distortion energy for quad deformation, grid line bending and stereo consistency, respectively. Given the initial mesh and target resolution of an image, the total energy function D is minimized and the deformed vertex set $\mathbf{V}'$ is found.

1) Quad Deformation and Line Bending Energy: The first two energy functions in (2) are employed from the existing work [4] and are generalized for use with stereoscopic images. The role of the quad deformation energy is to prefer uniform scaling of important quad cells and minimize cell deformation. Given the scale factor (s_f), the uniformly scaled edge vector ($\mathbf{v}_i - \mathbf{v}_j$), and deformed edge vector ($\mathbf{v}'_i - \mathbf{v}'_j$), the quad deformation energy function for each quad cell f can be defined as follows:

$$d_q(f) = \sum_{(i,j)\in\mathbf{E}(f)} \| (\mathbf{v}'_i - \mathbf{v}'_j) - s_f(\mathbf{v}_i - \mathbf{v}_j) \|^2 \tag{3}$$

In this letter, the quad deformation energy for a stereoscopic image is defined by summing (3) for all quads in the left and right images as follows:

$$D_q = \sum_{f\in\mathbf{F}} \left(w_f^L d_q^L(f) + w_f^R d_q^R(f) \right) \tag{4}$$

where the superscript L and R denote the left and right images, respectively. The weight term w_f is calculated by averaging the saliency values of the pixels in f.

On the other hand, to avoid the excessive deformation of quads, the line bending energy for stereoscopic images is defined as follows:

$$D_l = \sum_{(i,j)\in\mathbf{E}} \| \left(\mathbf{v}_i^{L\prime} - \mathbf{v}_j^{L\prime} \right) - l_{ij}^L \left(\mathbf{v}_i^L - \mathbf{v}_j^L \right) \|^2 + \sum_{(i,j)\in\mathbf{E}} \| \left(\mathbf{v}_i^{R\prime} - \mathbf{v}_j^{R\prime} \right) - l_{ij}^R \left(\mathbf{v}_i^R - \mathbf{v}_j^R \right) \|^2 \tag{5}$$

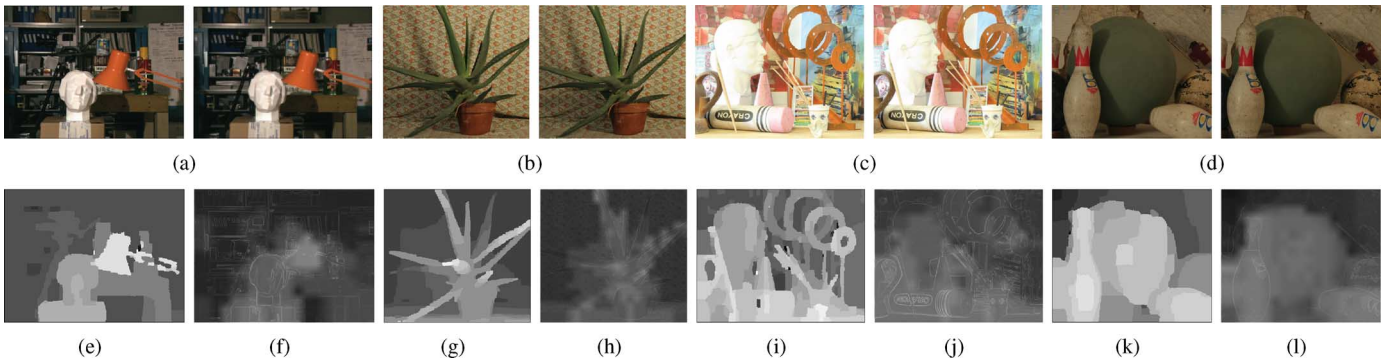


Fig. 2. Input images and the saliency maps. (a)–(d) The Middlebury test images (left and right) [10]. (e), (g), (i), (k) Reference disparity maps using belief propagation (BP) algorithm [11]. (f), (h), (j), (l) Estimated saliency maps.

Fig. 3. Retargeting result of standard stereoscopic image set. (a)–(h) Input images (left and right) (i) Estimated saliency map. (j) Disparity map of input. (k) Retargeting result (50% horizontal reduction). (l) Disparity map of (k). (m) Retargeting result (150% horizontal enlargement). (n) Disparity map of (m).

A low line bending energy means that the edge length ratio is maintained well after deformation. Together with the other energies, it prevents the excessive shearing of quads. The scale factors s_f and l_{ij} in (3) and (5) are computed optimally as described in [4]. However, they are computed independently for the left and right images at each iteration of optimization procedure.

2) Stereo Consistency Energy: The stereo disparity consistency between the left and right images can be preserved by applying the disparity in the sparse disparity map to each pair of the corresponding vertices. Therefore, using the initial and deformed vertices, the energy function for preserving the disparity consistency can be defined as follows:

$$D_t = \sum_{(i)\in\mathbf{V}} \| \left(\mathbf{v}_i^L - \mathbf{v}_i^{L'}\right) - \left(\mathbf{v}_i^R - \mathbf{v}_i^{R'}\right) \|^2 \qquad (6)$$

Using (6), the movement of a vertex on the right image is equivalent to the vertex movement on the left image, yielding preserved disparity consistency.

B. Energy Minimization

Once the energy function (2) is set up, it will be possible to apply any type of optimization method to minimize it. The present approach uses the conjugate gradient method. The initial guess of $\mathbf{V}'$ and the boundary constraint are the same as those used in [4].

IV. Experimental Result

In the experiment, we use a set of standard stereoscopic images [10] as shown in Fig. 2(a)–(d). Fig. 2(e), (g), (i), and (k) show the reference disparity map estimated by applying a well known stereo matching algorithm, *i.e.,* belief propagation (BP)

(a) (b) (c)

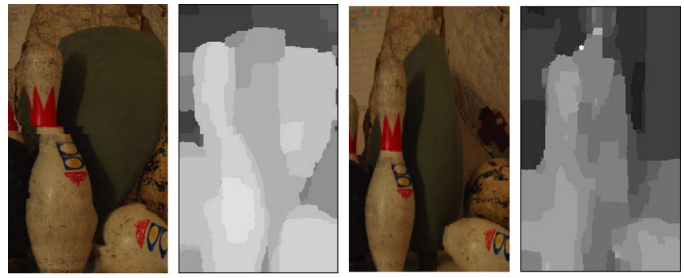
Fig. 4. Comparison with the existing single image retargeting algorithms. (a) 50% horizontal reduction using the seam carving [1] and (b) the scale-and-stretch [4] methods. (c) 150% horizontal enlargement using the scale-and-stretch method [4].

algorithm [11]. The estimated saliency maps are shown in Fig. 2(f), (h), (j), and (l). After retargeting the input images to horizontal 50% reduction and 150% enlargement, the resultant left and right images are shown in Fig. 3(a)–(c) and (d)–(f), respectively. The visually salient objects are well preserved without significant distortion. In particular, the objects with homogeneous textures but close to the camera (plaster cast and bowling ball) remain undistorted. Fig. 3(b) and (e) show the disparity map generated using the retargeting result by applying the BP algorithm. The disparity map is still correct compared to the reference disparity map (Fig. 2(e), (g), (i), and (k)). In this context, it is evaluated indirectly and qualitatively that the proposed method retains the stereo disparity consistency, *i.e.*, perceived depth consistency, after retargeting.

Compared to the conventional single image retargeting methods, the proposed algorithm can better preserve the visually salient regions. To show this, the seam carving [1] and scale-and-stretch [4] methods are used and retargeting is performed. The result is shown in Fig. 4. The featureless important objects (plaster cast and bowling ball) are distorted severely because seam carving and scale-and-stretch depend on the saliency map calculated on the visual features only, as shown in Fig. 4. On the other hand, the featureless important objects are preserved without distortion in the proposed algorithm because it fully considers the scene depth by the sparse disparity map, as shown in Fig. 3(a) and (d).

An additional experiment is carried out to compare the performance of the proposed method with the existing method for stereoscopic image retargeting [7], [9]. Fig. 5(a) and (b) show the result of [7]. Although the perceptual depth is preserved, serious distortion is observed, as shown in the bowling pin. This is because the assumption of seam connectivity is alleviated piecewise in [7]. On the other hand, as shown in Fig. 5(c) and (d), [9] reveals serious distortion in the important but featureless region (bowling ball). In addition, distortion of the perceptual depth occurs in some regions, which is illustrated in the disparity map calculated using the retargeting results. As shown in Fig. 3(a) and (b), the proposed method preserves the featureless important regions as well as the perceptual depth in the entire image.

The proposed algorithm has some limitations. Mismatching of the vertex can occur when the homogeneous region is excessively large in the image because this method relies on the result of vertex matching. This would cause an erroneous disparity map, yielding distorted retargeting results. Nevertheless, this is a limitation of all existing methods.



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

Fig. 5. Comparison with the existing stereoscopic image retargeting algorithms. 50% horizontal reduction. (a) Result using the seam carving based method [7]. (b) Corresponding disparity map of (a). (c) Result using the warping-based method [9]. (d) Corresponding disparity map of (c).

V. Conclusion

This letter proposed a novel warping-based method for retargeting stereoscopic images. The experimental results showed that retargeting is performed efficiently for stereoscopic images. Compared to other stereoscopic retargeting algorithms, the proposed method maintains the visually important regions and the perceived depth is significantly less distorted. Future work includes the natural extension of the proposed algorithm to 3-D stereoscopic video and multi-view video.

References

[1] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 10:1–10:9, Jul. 2007.
[2] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 23:1–23:11, Jul. 2009.
[3] A. Mansfield, P. Gehler, L. Van Gool, and C. Rother, "Scene carving: Scene consistent image retargeting," in *Proc. ECCV*, Sep. 2010.
[4] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 118:1–118:8, Dec. 2008.
[5] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 126:1–126:10, Dec. 2009.
[6] R. Gal, O. Sorkine, and D. Cohen-Or, "Feature-aware texturing," in *Proc. Eurographics Symp. on Rendering*, Jun. 2006, pp. 297–303.
[7] T. Basha, Y. Moses, and S. Avidan, "Geometrically consistent stereo seam carving," in *Proc. IEEE ICCV*, Nov. 2011, pp. 1816–1823.
[8] K. Utsugi, T. Shibahara, T. Koike, K. Takahashi, and T. Naemura, "Seam carving for stereo image," in *Proc. 3-DTV-Conf.*, Jun. 2010, pp. 1–4.
[9] 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.
[10] *The Middlebury Stereo Vision Page*, [Online]. Available: http://vision.middlebury.edu/stereo/
[11] Q. Yang, L. Wang, R. Yang, S. Wang, M. Liao, and D. Nistér, "Real-time global stereo matching using hierarchical belief propagation," in *Proc. BMVC*, Sep. 2006, pp. 989–998.