

The paper proposes a new framework for horizontal and vertical stereo seam carving based on geometrically consistent approach of Basha [3-4]. The paper is well structured and written. My main concern about this paper is the experiments section. Major points are listed below:

1) In this paper the proposed method has been compared with only Basha [3-4] scheme. To have a fair comparison with this method, listed points must be considered.

- Quantitative evaluation must be done based on depth distortion score (B) which is defined in equation (18) of [4].




- Visual depth comparison in term of disparity and depth distortion map must be done to compare the obtained disparity maps with the ground truth and the result of [4].

- To perceive the depth of the obtained 3D scenes and to compare them with ground truth and results of [4], anaglyph images must be provided.

- Other methods which outperformed Basha's scheme must be considered for comparison. Some examples are as follows:

- Stereoscopic Visual Attention Guided Seam Carving for Stereoscopic Image Retargeting, *Journal of Display Technology*, 2016.
- Content- and disparity-adaptive stereoscopic image retargeting, *Journal of Modern Optics*, 2016
- Depth-Preserving Warping for Stereo Image Retargeting, *IEEE Transactions on Image Processing*, 2015.

2) For man dataset, the author reported the results of [3] with 10% and 25% shrinkage. In original work of Basha [3-4], results for the same image reported for 17% of shrinkage. These three results are shown in the following table. What is the justification of less distortions on vertical tracks in images with 17% of shrinkage?

		
<p>10% shrinkage (Figure 10 of the paper)</p>	<p>17% shrinkage (Figure 12 of [4] and figure 8 of [3])</p>	<p>25% shrinkage (Figure 11 of the paper)</p>

3) In the last subsection of Section V, the authors claim that “The experiment is done on 11 pairs of stereo images in Middlebury dataset.” Whereas images of Man, Car, People and Snowman belong to Flickr dataset, Diana belongs to Portrait dataset and the rest (6 of 11) belong to Middlebury dataset. Also in the first subsection of Section V the authors claim that “Fig. 10 shows results of the methods on five pairs of stereo images, namely, i) manU, ii) snowmanU, iii) couch, iv) lampshade1, and v) crusade. In the figure, the odd rows show the left view of a scene, and the even rows show the right view of a scene. To save the space, we here do not shown the input images that can be accessed publically¹

¹<http://vision.middlebury.edu/stereo/data/>”

Whereas the images of manU and snowmanU are not belong to Middlebury dataset and would not be found in the mentioned address.

4) By setting $\alpha = 0$ in equations shown in figure 15 (equations (4) in the text), the weight of E_{Gradient} will be 1 and the proposed method convert to Basha’s method. As the plots of figure 15 demonstrate, considering the scale of vertical axis, there is no meaning full improvement in the distortion measure in comparison to Basha’s method ($\alpha = 0$). Roughly speaking, the obtained result shows that the performance of the proposed method are independent to the value of α for each image pairs. Whereas one of the main contribution of the paper is to generalize the total energy model of Basha’s method using proposed appearance energy term which is based on α parameter.

Minor points are listed below:

- The references should be numbered and appear in the order they appear in the paper.
- The second sentence of Section III is too long (10 lines).
- What is the distortion measure used in figure 15, 16?
- How can one estimate α for a new image pairs?