Geometrically Consistent Stereo Seam Carving

Tali Basha
Tel Aviv University
Tel Aviv 69978, Israel
talib@eng.tau.ac.il

Yael Moses
The Interdisciplinary Center
Herzliya 46150, Israel
yael@idc.ac.il

Shai Avidan
Tel Aviv University
Tel Aviv 69978, Israel
avidan@eng.tau.ac.il

Abstract

Image retargeting algorithms attempt to adapt the image content to the screen without distorting the important objects in the scene. Existing methods address retargeting of a single image. In this paper we propose a novel method for retargeting a pair of stereo images. Naively retargeting each image independently will distort the geometric structure and make it impossible to perceive the 3D structure of the scene. We show how to extend a single image seam carving to work on a pair of images. Our method minimizes the visual distortion in each of the images as well as the depth distortion. A key property of the proposed method is that it takes into account the visibility relations between pixels in the image pair (occluded and occluding pixels). As a result, our method guarantees, as we formally prove, that the retargeted pair is geometrically consistent with a feasible 3D scene, similar to the original one. Hence, the retargeted stereo pair can be viewed on a stereoscopic display or processed by any computer vision algorithm. We demonstrate our method on a number of challenging indoor and outdoor stereo images.

1. Introduction

Digital images are displayed on a variety of digital devices, each of which might require a different aspect ratio. The core idea of image retargeting algorithms is to adapt the image content to the screen without distorting the important objects in the scene. The rapid pace of technology makes it possible to view 3D content on a large range of devices, from cellphones to large TV screens. In addition, stereophotography is becoming increasingly popular, with a large number of stereo images appearing online. As a result, image retargeting algorithms need to be adapted to work on stereo image pairs.

We propose a novel method for retargeting stereo image pairs. The input to our method is assumed to be a rectified stereo image pair and a disparity map. The input disparity map may be computed from the pair of images by an available stereo algorithm, or be given by any other algorithm or device.

The 3D information provides valuable cues for retargeting, as previously demonstrated by retargeting algorithms for a single image [8]. Clearly, stereo image retargeting can also benefit from the 3D information provided by the other image; however, it also poses new challenges. In addition to retargeting both images without distorting the important objects in each, it should also obtain a stereo pair consistent with 3D geometry. Otherwise, the user will not be able to view and enjoy the retargeted stereo pair.

Our method retargets the input pair in the horizontal domain while minimizing the distortion of each image as well as the distortion in depth. A key property of our method is that the retargeted stereo pair has a feasible 3D interpretation that is similar to the original one. Thanks to this geometric consistency, our retargeted pair can be viewed on a stereoscopic display or processed by any computer vision algorithm that makes use of a stereo pair (e.g., cosegmentation or tracking).

1.1. The General Idea

We generalize the single image seam carving algorithm [1, 10] to work on a stereo pair. Instead of removing a seam from a single image, our algorithm iteratively removes a pair of seams from the stereo image pair.

A naive extension of the single image seam carving algorithm is to independently apply it to each of the images (see the blue box in Fig 1**b**). It disregards the geometry and damages the 3D structure of the scene. For example by removing a pixel from one image while keeping its corresponding pixel in the other one. To overcome this problem, a joint retargeting of both images must be considered. In particular, the selection of seams in both images should be coupled. A straightforward approach is to compute the seam in one of the images, say the left one, and then map it to the right image via the disparity map. This is clearly sub-optimal as it does not utilize the information available in the right image or the depth map. Fig. 2 demonstrates the results using this approach (for more details see Sec. 5).

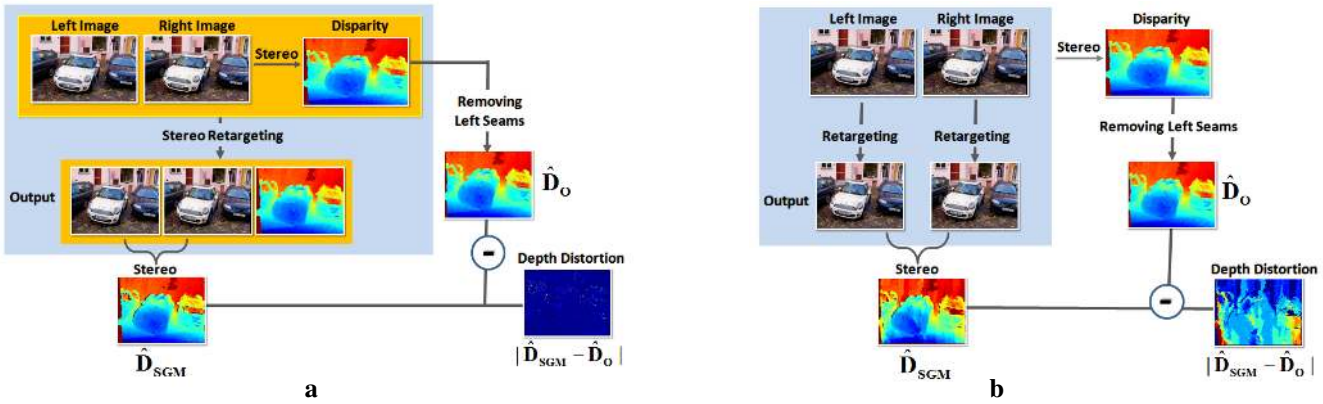


Figure 1. **Geometric Evaluation. (a):** the results of applying our stereo retargeting algorithm. **(b):** the results of applying single image SC [10], on each of the input images. On both left and right sides: $\widehat{D}_{SGM}$ is computed by applying stereo [4] on the retargeted pair of images; the original disparity values of the remaining pixels are stored in $\widehat{D}_o$. Finally, the depth distortion is measured by $|\widehat{D}_{SGM} - \widehat{D}_o|$. The color code is blue for low values and red for high ones; red indicates a difference of at least six pixels.

In fact, the problems run even deeper, mainly due to occlusions; there is no guarantee that seam pixels in the left image have matching pixels in the right. And the change in 3D shape must be carefully considered to avoid an inconsistent change in the visibility relation of the scene points. In particular, pixels that are visible only in one of the views should not be revealed. Thus, the benefits provided by the additional information in the right image and the disparity map may create new challenges.

The proposed method overcomes these challenges by generalizing seam carving to simultaneously carve a pair of seams in both images, while minimizing distortion in appearance and depth. Seam selection and disparity map modification are subject to geometric constraints that take into account the visibility relations between pixels in the images (occluded and occluding pixels). These geometric constraints *guarantee* consistency of the target images with a feasible 3D scene, as formally proven in Appendix A, and empirically demonstrated in Sec. 5.

2. Background

Image and video retargeting algorithms have been extensively investigated in recent years. These algorithms try to change the aspect ratio of an image in a way that does not hurt the proportions of the important objects in the image. The various algorithms differ in how they determine the importance of different pixels in the image and in how they use this information. An excellent overview and comprehensive study of the topic is given in [9].

Here, we extend the Seam Carving algorithm. The algorithm was first introduced in [1] and extended in [10]. The algorithm works by iteratively computing a seam with minimal visual distortion in the image and removing it. These algorithms assumed that seams are connected paths in the image, yet this is not a necessary assumption and Grundmann *et al.* [3] recently showed that piece-wise connected paths are more flexible for video retargeting, where the goal is to retarget frames sequentially and rely on piece-wise connected seams to better fit the retargeting to previous frames. However, they do not consider stereo data.

Most work reported thus far in the literature focused on retargeting a single image or video. Yet, the rise of 3D content makes it necessary to extend image retargeting algorithms to work with 3D content. This was addressed, to some extent, by the following authors. In Lang *et al.* [7] the authors work with stereo images (and video); however, their goal is to adjust the disparity map according to various stylistic considerations and not retarget the stereo images. Mansfield *et al.* [8] assume the input is an image and a relative depth map (provided by the user) and the output is an image. They extend seam carving to *scene carving* and show that scene carving is indeed scene consistent, can introduce occlusions, and can also handle pixel reappearance, say when one layer moves behind another layer and reappears on the other side.

There are a number of important distinctions between our work and that of [8]. First, we assume that the input is a pair of stereo images. They, on the other hand, assume the input to be an image with a depth map. As a result, they cannot produce a retargeted stereo pair without resorting to image synthesis techniques to fill in gaps, as they have to deal with the occlusions caused by the different points of view of the stereo image pair. Second, we assume a per-pixel stereo map, as opposed to representing the scene as a collection of well-defined fronto-parallel planes.

The only attempt to consider a stereo pair was proposed by Utsugi *et al.*[11]. They also use the disparity map to retarget the stereo image pair. However, the primary goal of preserving the geometric consistency of the output image pair is not defined and not obtained by [11]. Moreover,

a b c d e f g

Figure 2. **The straightforward approach for seam coupling. (a-b)**: The pair of input images, and **(c)**, their ground truth disparity map. **(d):** The single image retargeted left image using [10], and **(e)** the retargeted right image computed by carving corresponding seams. **(f):** The disparity map computed by applying the SGM stereo algorithm on **(d-e)**. **(g):** The updated disparity map computed by our method.

the appearance of their retargeted images is impaired (see example in Fig. 6).

3. Geometric Consistency

A primary goal of our method is to preserve a 3D interpretation of the scene from the retargeted image pair, while minimizing the 3D distortion of the original scene. We prove that our algorithm preserves the geometric consistency of the input pair. Namely, it is possible to define correspondence between pixels that is consistent with the epipolar geometry as well as with visibility relations between 3D points in the retargeted pair.

Since all pixels in both images are either left-shifted or remain in their original location, the epipolar geometry is preserved (as well as the rectification). In addition, our method maintains the original input correspondence and the visibility relations between 3D points in the scene, by imposing the following constraints:

$\mathbf{C_1}$: *Corresponding pixels in the original images are either both removed or remain corresponding in the output images.*

$\mathbf{C_2}$: *3D points that are visible in the reference view but occluded in the other are not revealed.*

Constraint $\mathbf{C_1}$ is directly satisfied by our method since the disparity map is used to couple the seams (see Sec. 4.1). Constraint $\mathbf{C_2}$ is satisfied by using occlusion to restrict the set of candidate seam pixels (see Sec. 4.3). In Appendix *A* we formally define constraint $\mathbf{C_2}$ and prove that occluding and occluded pixels do not change their role in the retargeted pair.

It is important to note that these constraints can be proven to be unsatisfiable for horizontal seams.

4. The Method

The input to our method is a pair of $m \times n$ rectified stereo images, $\{I_L, I_R\}$, and a disparity map, D, where the disparity map can be computed by any stereo algorithm (we use [4]). Without loss of generality, we consider the disparity with respect to the left image, which is taken to be the reference image. The output of our algorithm is a pair of retargeted images, $\{\widehat{I}_L, \widehat{I}_R\}$ and an updated disparity map, $\widehat{D}$. The retargeted images are geometrically consistent with a feasible 3D scene.

4.1. Seam Coupling

The geometric coupling of the two seams, $S_L = \{s_L^i\}_{i=1}^m$ and $S_R = \{s_R^i\}_{i=1}^m$, is obtained by using the correspondence defined by D. Formally, each of the seam's pixels in the left image at row i, $s_L^i = (i, j_L(i)) \in S_L$, is matched to a seam pixel in the right image, $s_R^i = (i, j_R(i)) \in S_R$, as follows:

$$s_R^i = (i, j_R(i)) = (i, j_L(i) + D(s_L^i)), \quad (1)$$

where $j_L, j_R : [m] \rightarrow [n]$, and $[m] = [1, ..., m]$. The estimated disparity map, $D : [n] \times [m] \rightarrow \mathbb{Z} \cup \perp$, maps pixels of I_L to their corresponding pixels in I_R, if the correspondence is known, and to $\perp$ otherwise. It follows that the seams contain only pixels for which the disparity is defined.

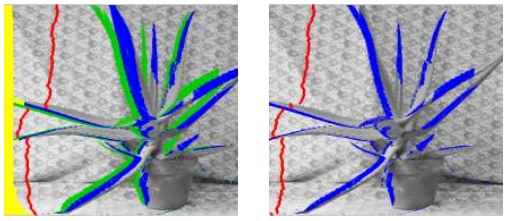
Note that a continuous seam in the left image generally corresponds to a piecewise continuous seam in the right image since the seam may cross depth discontinuities. Therefore, we drop the assumption that a seam (in either I_L or I_R) is continuous and consider piecewise seams from now on, which we refer to as *generalized seams* (See Fig. 3).

4.2. The Energy Function

The energy function of the stereo seam carving method consists of an intensity term and a 3D geometry term. Removing a seam's pixel from each of the stereo image pair has the local effect of creating new adjacent pixels in the target image. 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 $j_R^{\pm}$, respectively. Since the left and right image seams are coupled, $j_R^{\pm}$, is uniquely defined by $j_L^{\pm}$ and the disparity map, D. Therefore, we define 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,

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

where α controls the relative impact of each of the terms. Since we use generalized seams, $j^{\pm} \in [m]$ can be any pixel in row $i-1$ (unlike the continuous case in which $j^{\pm} \in \{j-1, j, j+1\}$).



a **b**

Figure 3. (**a**): The left image masked with the computed occluded pixels in green, and occluding pixels in blue; out of field of view pixels are colored in yellow. (**b**): The right image masked with the corresponding occluding pixels in blue. In this example both left and right seams (in red) are discontinuous.

4.2.1 Appearance Energy

We generalize the original *forward energy* criterion [10] that aims at minimizing the resulting distortion in the retargeted image caused by the intensity differences between new adjacent pixels. The appearance distortion $E_{intensity}(i,j,j^{\pm})$ is taken to be the sum of the energy terms, E_L and E_R, for removing a pair of coupled pixels from the left and right images. That is,

$$E_{intensity}(i,j,j^{\pm}) = E_L(i,j,j^{\pm}) + E_R(i,j_R(i),j_R^{\pm}(i-1)), \quad (3)$$

where the coupling of the left and right seams is captured via the disparity map as defined in Sec.4.1.

The energy of removing a specific pixel, (i,j) from image I, left or right, is given by:

$$E(i,j,j^{\pm}) = E^v(i,j,j^{\pm}) + E^h(i,j) \quad (4)$$

where E^h and E^v are the forward energy terms due to the new gradients in the horizontal and vertical directions, respectively. In particular, E^h is given by:

$$E^h(i,j) = |I(i,j+1) - I(i,j-1)|. \quad (5)$$

In the vertical direction, the new gradients depend on the position of the seam in row $i-1$, $j^{\pm}$. Accordingly, the vertical forward energy is given by:

$$E^v(i,j,j^{\pm}) = \begin{cases} V_1 & j^{\pm} < j \\ 0 & j^{\pm} = j \\ V_2 & j^{\pm} > j \end{cases} \quad (6)$$

where

$$\begin{aligned} V_1 &= \textstyle\sum_{k=j^{\pm}+1}^{j} |I(i-1,k) - I(i,k-1))| \\ V_2 &= \textstyle\sum_{k=j+1}^{j^{\pm}} |I(i-1,k-1) - I(i,k))| \end{aligned} \quad (7)$$

4.2.2 Depth Energy

The computed depth map provides valuable cues for seam selection, and a 3D forward energy term, E_D, is used to minimize the disparity distortion. It is defined similarly to the forward energy of the intensity values, by replacing the intensity function, I, with the disparity map D in Eq. 4-7. In practice, in order to compensate for the differences in range between the intensity and the disparity values, we normalize both I and D in the range of zero to one.

In addition, the object's distance from the camera often correlates with its saliency. Hence, we increase the energy of pixels that are the projections of nearby 3D points. Moreover, our method is strongly based on the disparity map, which is computed by a stereo algorithm that is regarded as a black box. Errors in the estimated map may result in incorrect coupling of seam pixels. We prefer removing pixels for which we have high confidence of disparity values, measured by the difference in the intensities of corresponding pixels. That is,

$$G(i,j) = |I_L(i,j) - I_R(i,j+D(i,j))|. \quad (8)$$

The total forward 3D energy is a weighted sum between three components:

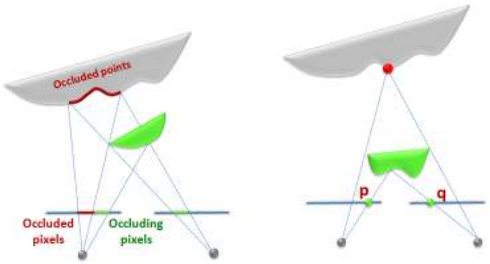
$$E_{3D}(i,j,j^{\pm}) = E_D(i,j,j^{\pm}) + \beta|D_n(i,j)| + \gamma G(i,j), \quad (9)$$

where D_n is the normalized disparity map.

4.3. Maintaining Pixel Visibility

An *occluded* pixel in the reference image is defined as the projection of a 3D point that is not visible in the right view due to another 3D point that occludes it (red in Fig. 4a). Occluded pixels do not have corresponding pixels in the right image; our method does not remove them from the image.

Furthermore, in order to satisfy the geometric constraint, $\mathbf{C}_2$, occluded pixels must not be revealed. Otherwise, no coherent 3D interpretation can justify the visibility of the revealed pixel only in one image and not in the other. To this end, we ensure that occluded pixels in the original right image remain occluded in the retargeted right image, by



a **b**

Figure 4. **(a):** Occluded pixels, red, have no corresponding pixels in the right image. The occluding pixels, green, are visible in both views. **(b):** The ordering constraint does not hold: removing the red point causes the point p to shift left while the point q remains in its original location.

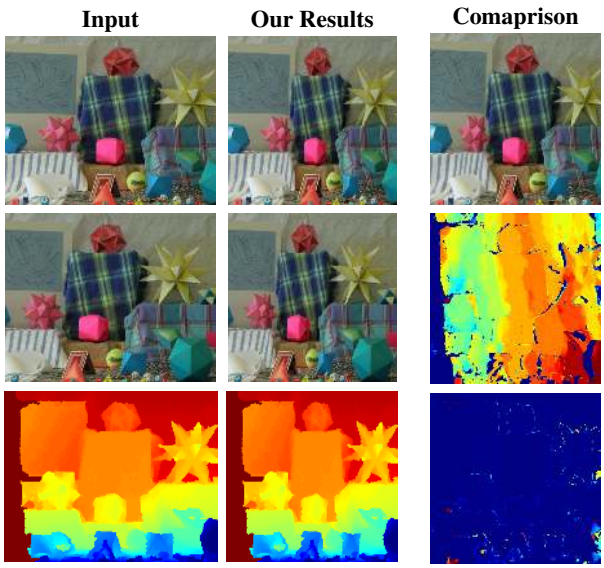


Figure 5. **Moebius Dataset.** In the first column (top to bottom), the input left and right images and the input disparity map. In the second column, our results, with respect to the first column. The third column shows the results of applying single image SC [10] to the left input image (top); the distortion in depth caused by independent single image retargeting (middle); the distortion in depth caused by our stereo retargeting method. Depth distortion scores: single image SC=85%; stereo pair SC=3.2%.

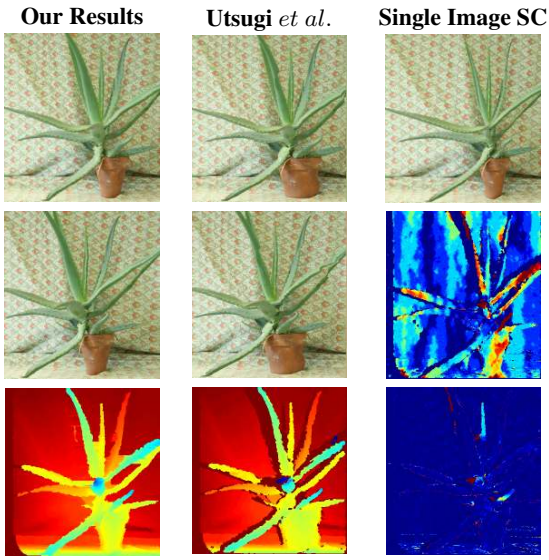


Figure 6. **Aloe Dataset.** The input stereo pair is shown in Fig. 2(**a-b**). In the first column (top to bottom), our results for the left and right images and the disparity map. In the second column, the results of Utsugi *et al.*[11] method. In the third column, the results of applying single image SC [10] to the left input image (top); see caption of Fig.5. The depth distortion scores are: single image SC=47%; stereo pair SC=2.9%.

avoiding removing pixels that may reveal them, namely *occluding* pixels. An occluding pixel is defined to be the projection of a visible 3D point in both views that accounts for the occlusion of one or more 3D points in the right view (see green in Fig. 4a). Our choice of removing only pixels that are neither occluded nor occluding, is proved to guarantee that the original visibility relation (i.e., occluded-occluding pairs) is preserved (see Appendix A).

The set of occluding and occluded pixels is computed once from the input disparity map, D and represented by a binary map, $O(i,j)$ where $O(i,j) = 1$ if pixel (i,j) is an occluded or occluding pixel. This map is computed using a simplified Z-buffer approach.

In the examples we considered, the number of occluded and occluding pixels is typically 20%. An example of both the occluded and occluding maps is given in Fig. 3.

4.4. Stereo Seam Selection and Carving

The energy term defined in Eq. 2 is now accumulated in a cost matrix M to select a pair of seams using dynamic programming. The seams are coupled as defined in Sec. 4.1. We set $M(i,j) = \infty$ for pixels that do not satisfy the visibility constraints, namely if $O(i,j) = 1$ (see Sec 4.3).

By default, we prefer continuous seams (where $j^{\pm} \in j-1, j, j+1$). However, if a continuous path is blocked at pixel (i,j) by occluded/occluding pixels, we allow discontinuous seams. Formally, we consider two cases, according to whether it is necessary to switch at the pixel (i,j) from a continuous to discontinuous seam:

$$M(i,j) = \begin{cases} \min\limits_{j^{\pm} \in \{j-1,j,j+1\}} E_{total}(i,j,j^{\pm}); & T(i,j) = 0 \\ \min\limits_{j^{\pm} \in [m]} E_{total}(i,j,j^{\pm}); & T(i,j) = 1, \end{cases} \tag{10}$$

where, T is the binary map of size $n \times m$. $T(i,j)$ indicates whether a continuous path is blocked in row $i-1$ by occluding/occluded pixels. That is, $T(i,j) = 1$ if $O(i-1, j^{\pm}) = 1$ for $j^{\pm} \in \{j-1, j, j+1\}$.

As in [10], removing a seam pixel from a row results in shifting pixels in that row. Specifically, all pixels to the right of the removed pixels are shifted left by one pixel. The remaining pixels are unchanged. Formally, the shifting function $f_L(i,j) : [m] \times [n] \rightarrow [m] \times [n-1]$ maps the i^{th} input row to the i^{th} output row. Let $s_L^i = (i, j_L(i))$ be the pixel to be removed from the left image. Then, the shifting mapping is defined by:

$$f_L(i,j) = \begin{cases} j & \text{if } j < j_L(i) \\ j-1 & \text{if } j > j_L(i) \\ \perp & \text{if } j = j_L(i) \end{cases} \tag{11}$$

Likewise, $f_R(i,j)$ is the corresponding mapping function in the right image, where $j_L(i)$ is replaced by $j_R(i)$ (as defined in Eq. 1).

After carving a seam, the new disparity map, $\widehat{D}$, is obtained by removing the left seam S_L from the previous D and updating the disparity values of the remaining pixels. In particular, the updated disparity value, $\widehat{D}$, of a pixel (i,j) is

a **b** **c** **d**

Figure 7. **Diana Dataset. (a):** The input pair of images. **(b):** The retargeted pair of images produced by our method. **(c):** The input disparity map, our result, and the evaluated depth distortion. **(d):** The result of single image SC on the left image.

given by:

$$\hat{D}(i, f_L(i,j)) = f_R(i, j + D(i,j)) - f_L(i,j). \qquad (12)$$

Geometric Interpretation: We next describe the geometric interpretation of the carving. From Eq. 11 it follows that each pixel may either be shifted one pixel to the left, or remain in its original location. If a pair of corresponding pixels remains in its original location, the associated 3D point remains the same as in the original scene. When pixels are shifted, the position of the associated 3D points change. If the two pixels in a corresponding pair are both shifted left, the original depth is preserved, namely $\widehat{D} = D$ (see Eq. 12). The associated 3D point changes its location accordingly by a left translation, parallel to the image plane. Most pixels will either remain in their original location or be shifted together. However, when the ordering constraint does not hold (see Fig. 4**b**), a pixel may be shifted in one of the images, while its corresponding pixel remains in its original location. In this case, the disparity is changed by one pixel, which corresponds to a small change in depth. For example, if the left pixel remains in its location, then the 3D point moves one step along the left pixel ray.

5. Results

We tested our method on challenging indoor and outdoor scenes. In all experiments, we used the SGM stereo algorithm [4] to compute the input disparity; hole filling was performed on regions for which the disparity was not computed. The algorithm was implemented in MATLAB. The following datasets were considered:

Middlebury: The *Middlebury* stereo datasets [5], *Moebius* (Fig. 5) and *Aloe* (Fig. 6), are challenging because the scene is highly textured and contains objects at different depths. About 20% of the pixels in the original reference image cannot be removed, since they are either occluding or occluded (see Fig. 3).

Portrait: A pair of images (Fig. 7), provided by [6]. The main challenge in this pair is that the salient object, which covers most of the image, should not be distorted. Also, a significant part of the left image is out of the field of view of the right camera, and hence cannot be removed by our algorithm.

Flicker: A set of stereo images, with large depth range, downloaded from *Flicker* (Fig. 8-10). The images were manually rectified using [2].

A main contribution of our method is the production of a geometrically consistent retargeted image pair that preserves the original depth values of the remaining points. We next describe the evaluation of these properties.

Geometric Evaluation: We evaluate depth distortion by measuring the deviation of the updated disparity values from their original values. Our evaluation scheme is described in Fig.1: a disparity map, $\widehat{D}_{SGM}$, is computed on the retargeted pair of stereo images. The computed map, $\widehat{D}_{SGM}$, reflects the geometry contained in the pair of retargeted images, regardless of the method used to produce them.

The depth distortion is measured by comparing the disparity value of each pixel in $\widehat{D}_{SGM}$ with its original value. In particular, we compute $\widehat{D}_o$, which consists of the original disparity values, D, after removing the relevant seams with respect to the reference view. The absolute difference, $|\widehat{D}_o - \widehat{D}_{SGM}|$, is shown for all our experiments. For comparison, we evaluate the depth distortion caused by independent single image retargeting (see Fig. 1**b**). For quantitative evaluation, we define the *depth distortion* score to be the percentage of pixels whose depth, $\widehat{D}_{SGM}$, has been changed by more than one pixel.

Test 1: We tested our algorithm on all the abovementioned datasets using a fixed set of parameters for the 3D weight: $\beta = 0.08$ and $\gamma = 0.5$ (see Eq. 9). The image width was reduced by 20% for the Middlebury datasets and by 17% for the rest. The results are presented in Fig. 5-9.

Our experiments show that the output pair is geometrically consistent and the original depth values are preserved. It is also clearly seen that a significant depth distortion is caused when naive independent retargeting of each image is considered. (See right columns in each of the figures.)

To evaluate the appearance distortion, we show the single image seam carving result of the left image [10]. The large number of geometric constraints that our method must satisfy limits the number of candidate seams; the constraints are thus expected to impair the results compared to the single image retargeting. Still, the 3D information and the use of generalized seams compensate for this problem. Our results are similar (e.g., Fig. 5-6) to those of single image seam carving and in some cases much better. For example, our method successfully preserves the face appearance

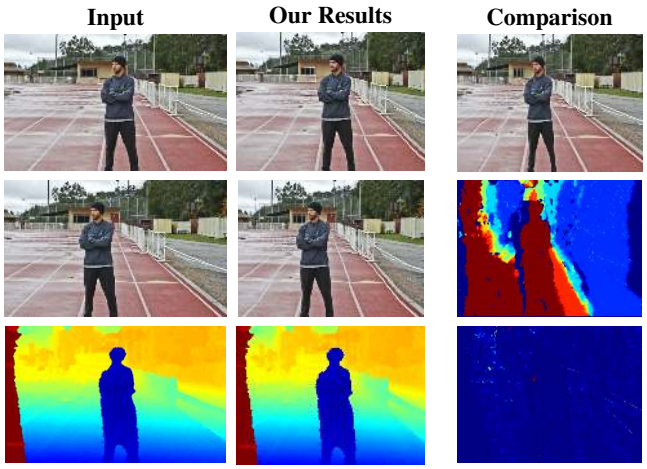


Figure 8. **Man Dataset from Flicker.** see caption of Fig. 5. Depth distortion scores: single image SC=49%; stereo pair SC=0.43%.

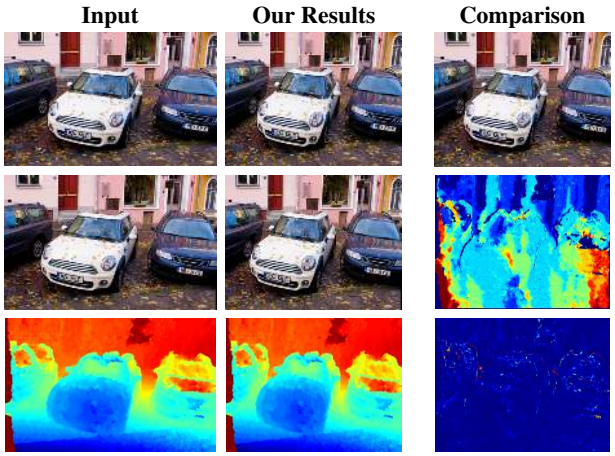


Figure 9. **Cars Dataset from Ficker.** see caption of Fig. 5.Depth distortion scores: single image SC=70%; stereo pair SC=1.3%.

(Fig. 7) as well as the face depth (Fig. 7(c)), without prior knowledge, such as face location, used by [12]. Fig. 8 shows another example in which the perspective of the running track is nicely preserved and the man is not deformed, in contrast to the single image seam carving.

Test 2: We test a straightforward approach to using the disparity map. That is, a single image retargeting is computed for the reference image, and the corresponding seams in the right image are computed directly via the disparity map. Fig. 2 presents the retargeted images (**d-e**). The ground truth disparity map of the original pair, **c**, should be compared to the disparity map computed using the SGM on the naive retargeted pair, **f**, and to our results, **g**. It is clear from these results that the straightforward coupling is insufficient for obtaining geometric consistency. (The retargeted images using our results are presented in Fig. 6.)

Test 3: So far we have used fixed parameters for the 3D weight. However, as can be seen in the second column of Fig. 10, the man on the left almost "lost" his leg. Allowing user interaction for setting the weight of objects according to their depth (the parameter β in Eq. 9) improves the results (third column in Fig. 10). The head of the person on the right was not distorted regardless of this parameter, in contrast to the single image seam carving which distorts it dramatically. We note that the geometric consistency of the retargeted images is obtained regardless of the choice of parameters, and depth distortion remains negligible.

Finally, we compare the disparity computed by our method, $\hat{D}$, with that computed by the SGM algorithm on the retargeted images, $\widehat{D}_{SGM}$. For all the tests described above, the differences is less than 2%.

6. Conclusions

We extended seam carving to work on a stereo image pair. Retargeting each image independently will distort the geometric structure, making it impossible to perceive the 3D structure of the scene. We have shown how to extend the single image seam carving to work on a pair of images, and proved that the proposed algorithm is guaranteed to give a geometrically consistent result. The retargeted images can thus be viewed on a stereoscopic display or processed by any computer vision algorithm. As far as limitations go, we depend on the quality of the stereo matching, and when the stereo matching fails, we might fail as well. On the positive side, our method takes advantage of both appearance and depth cues and can deal with scenes that are more difficult to deal with using single image seam carving. We demonstrated the advantages of our method on several challenging stereo images.

Appendix A

We formally define the *occluded* and *occluding* pixels and prove that the geometric constraint, $\mathbf{C_2}$, is satisfied by our method. In particular, we prove the operation of removing a pair of seams pixels and the following update of the disparity map guarantee that the original visibility of 3D points in the interpreted scene is preserved. (The consistency with the epipolar geometry was discussed in Sec. 3.)

Definition: Occluded and Occluding Pixels
Let (i, j_b) and (i, j_f) be two pixels in the left image. The pixel (i, j_f) occludes (i, j_b) *iff* $j_b < j_f$ and the two pixels are mapped to the same pixel in the right image. That is,

$$j_f + D(i, j_f) = j_b + D(i, j_b). \tag{13}$$

It follows that a pixel (i, j) is *not* an occluding\occluded *iff*

$$j + D(i, j) \neq j' + D(i, j') \quad \forall j' \neq j. \tag{14}$$

Lemma #1:

The operation of removing a seam point, $p_L = (i, j_L)$ preserves the ordering between the remaining pixels on the row i. Formally, given two pixels, $p_1 = (i, j_1)$ and $p_2 = (i, j_2)$: $j_1 < j_2 \Leftrightarrow f_L(i, j_1) < f_L(i, j_2)$, where f is defined in Eq. 11.

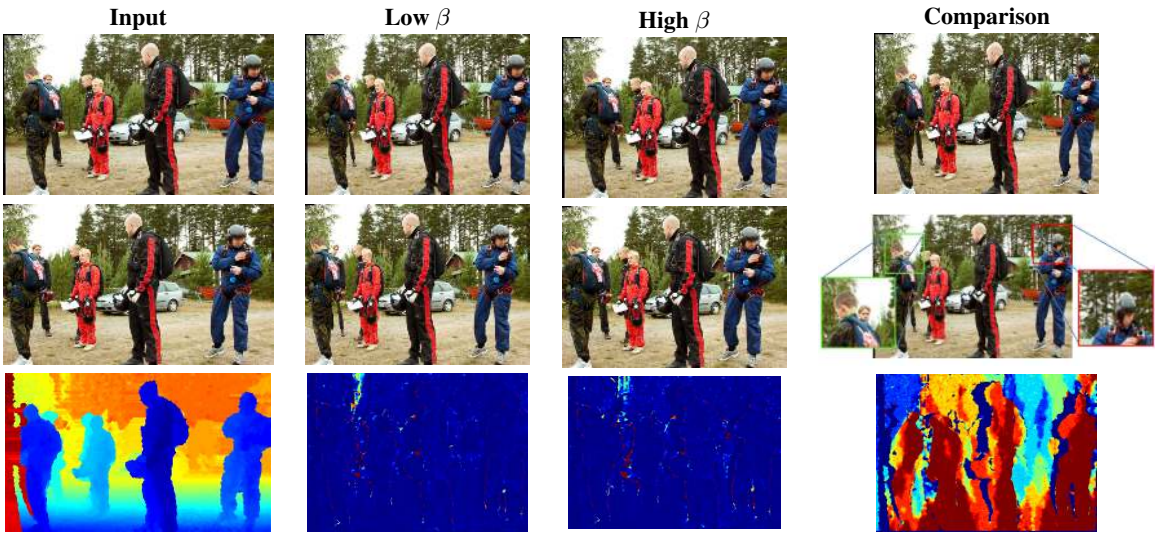


Figure 10. **People Dataset.** The first column shows the input pair and the computed disparity map. The second and third columns show our results using low and high weights, respectively; the bottom figures show the depth distortion, respectively. The third column shows the results of applying single image SC to the left input image; the bottom figure shows the depth distortion caused by single image SC. Depth distortion scores: single image SC=79%; stereo pair SC, low β=2.1%; stereo pair SC, high β=2.6%.

It follows directly from this Lemma that: $j_1 = j_2 \Leftrightarrow f_L(i, j_1) = f_L(i, j_2)$.

Proof: If both p_1 and p_2 are on the same side of the seam, then by Eq. 11 the order is preserved. Therefore, the only case to consider is when the seam pixel, (i, j_L), is in between the two pixels: without loss of generality, $j_1 < j_L < j_2$. In this case, $f_L(i, j_1) = j_1$ and $f_L(i, j_2) = j_2 - 1$. Since this scenario is possible only if the gap between j_1 and j_2 is at least one pixel, it follows that $j_1 < j_2 - 1$. In particular, we obtain that $j_1 < j_2$ and $f_L(i, j_1) < f_L(i, j_2)$.

Claim: Let $p_f(i, j_f)$ and $p_b = (i, j_b)$ be two pixels in the reference view. Pixel p_f occludes p_b in the original image pair *iff* $(i, f_L(i, j_f))$ occludes $(i, f_L(i, j_b))$ after removing the seam pixels.

Proof: We have to show that: $j_b < j_f$ and $j_f + D(i, j_f) = j_b + D(i, j_b)$ *iff* $f_L(i, j_b) < f_L(i, j_f)$ and $f_L(i, j_b) + \hat{D}(i, f_L(i, j_b)) = f_L(i, j_f) + \hat{D}(i, f_L(i, j_f))$. Using the definition of $\hat{D}$ (see Eq. 12), it follows that:

$$\begin{aligned} f_L(i, j_b) \quad + \quad & \hat{D}(i, f_L(i, j_b)) = \\ & f_L(i, j_b) + f_R(i, j_b + D(i, j_b)) - f_L(i, j_b) = \\ & f_R(i, j_b + D(i, j_b)) \end{aligned}$$

$$\begin{aligned} f_L(i, j_f) \quad + \quad & \hat{D}(i, f_L(i, j_f)) = \\ & f_L(i, j_f) + f_R(i, j_f + D(i, j_f)) - f_L(i, j_f) = \\ & f_R(i, j_f + D(i, j_f)), \end{aligned} \tag{15}$$

Now, using Lemma 1 we obtain that

$$j_b < j_f \Leftrightarrow f_L(i, j_b) < f_L(i, j_f), \text{ and,} \tag{16}$$

$$\begin{aligned} f_R(i, j_b + D(i, j_b)) = f_R(i, j_f + D(i, j_f)) \Leftrightarrow \\ j_b + d(j_b) = j_f + d(j_f). \end{aligned} \tag{17}$$

To complete the proof, the above equations are put together:

$$\begin{aligned} f_L(i, j_b) + \hat{D}(f_L(i, j_b)) = f_L(i, j_f) + \hat{D}(i, f_L(i, j_f)) \Leftrightarrow \\ j_b + D(i, j_b) = j_f + D(i, j_f), \end{aligned}$$

Acknowledgments.

This work was supported in part by an Israel Science Foundation grant 1556/10 and a grant from the Israeli Ministry of Science and Technology.

References

[1] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM Trans. Graph.*, 26(3):10, 2007.

[2] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000.

[3] M. Grundmann, V. Kwatra, M. Han, and I. Essa. Discontinuous seam-carving for video retargeting. In *CVPR*, 2010.

[4] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. *PAMI*, 30(2):328–341, 2008.

[5] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. In *CVPR*, pages 1–8, 2007.

[6] F. Huguet and F. Devernay. A variational method for scene flow estimation from stereo sequences. In *ICCV*, 2007.

[7] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross. Nonlinear disparity mapping for stereoscopic 3d. *ACM Trans. Graph.*, 29(3):10, 2010.

[8] A. Mansfield, P. V. Gehler, L. J. V. Gool, and C. Rother. Scene carving: Scene consistent image retargeting. In *ECCV*, pages 143–156, 2010.

[9] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir. A comparative study of image retargeting. *ACM Trans. Graph.*, 29(5), 2010.

[10] M. Rubinstein, A. Shamir, and S. Avidan. Improved seam carving for video retargeting. *ACM Trans. Graph.*, 27(3):16–16, 2008.

[11] K. Utsugi, T. Shibahara, T. Koike, K. Takahashi, and T. Naemura. Seam carving for stereo images. In *3DTV-Conference*.

[12] L. Wolf, M. Guttmann, and D. Cohen-Or. Non-homogeneous content-driven video-retargeting. In *ICCV*, pages 1–6, 2007.