Feedback Retargeting

Eitam Kav-Venaki and Shmuel Peleg

School of Computer Science and Engineering
The Hebrew University of Jerusalem
91904 Jerusalem, Israel

Abstract. Feedback retargeting combines the benefits of two previous retargeting methods: Bidirectional similarity [1] and Shift-Map [2]. The first method may have blurry areas due to patch averaging and the latter can remove entire objects. Feedback retargeting has the sharpness of shift-map and the completeness of bidirectional similarity, avoiding the removal of salient objects.

In Shift-Map retargeting the output image is made from segments of the input image, and this minimizes the forward direction of bidirectional similarity. An iterative feedback procedure is developed to take care of the backward direction, assuring that the input image can be reconstructed from the output image. This is done by using Shift-Map backwards, reconstructing the input image back from the output image. Areas in the input image that are difficult to reconstruct from the output image get a feedback priority score. A second Shift-Map retargeting is then performed, adding this feedback priority to the data term. These regions now have a higher priority to be included in the output.

After a few iterations of forward retargeting and backward feedback the retargeted image includes all salient features from the input image. Computational efficiency and image sharpness remain as high as in ordinary Shift-Map.

1 Image Retargeting Background

Image retargeting algorithms [1–7] take an input image A and generate an output image B having new dimensions, mostly having a new aspect ratio (e.g. reducing image width by two). These algorithms attempt to preserve some of the image's important qualities and features, that may be lost or distorted when using simple scaling or cropping.

The most recognized method for image retargeting is seam-carving [4, 8]. In this approach continuous seams of pixels are removed from the image in an iterative greedy manner, selecting in each step a seam whose removal will minimize the error measured by image gradients. Seam-carving methods and also methods that apply non-homogeneous warping to the image [5, 6, 9] may induce noticeable distortions even if objects with high gradients are unchanged. Obviously, some texture or details in the image will have to change their aspect ratio.

Some retargeting approaches [1, 2, 10] are based on algorithms that were used in the fields of texture synthesis and image completion (see [11–13]). Paper

K.N. Kutulakos (Ed.): ECCV 2010 Workshops, Part II, LNCS 6554, pp. 145–155, 2012.

[1] introduces a bidirectional similarity formulation. Every patch in the output image B should have a similar patch in the input image A ("coherence") and vice versa: every patch in the input image A should have a similar patch in the output image B ("completeness"). Shift-Map retargeting [2] has a high coherence but does not guarantee completeness. An epitome [14] is an example for visual summarization that imposes no coherence.

The minimization of the bidirectional dissimilarity score in [1] assumes retargeting with a small scale change. To obtain any significant scale change, multiple iterations of small image scaling are performed. Feedback retargeting as proposed in this paper can perform any scale change, with no need for repetitive small changes.

Another bidirectional-similarity score was developed by [7]. They combine several resizing operators (seam-carving, cropping, and scaling), and find an optimal combination that minimizes that score. Their work was followed by the work of [15], which use a similar technique to minimize a similarity score based on the formulation in [1] and on dominant color descriptors.

2 Shift-Map

Following [2, 16, 17], we define the relationship between the pixels in the output image $R(u, v)$ to pixels in the source image $I(x, y)$ by a shift-map $M(u, v) = (t_x, t_y)$. The pixel $R(u, v)$ in the output image will be derived from the source pixel $I(u + t_x, v + t_y)$. The optimal shift-map is defined using graph labeling, where the nodes are the pixels of the output image, and each output pixel is labeled by a shift $t = (t_x, t_y)$. The optimal shift-map M minimizes the following cost function:

$$E(M) = \alpha \sum_{p \in R} E_d(p, M(p)) + \sum_{(p,q) \in N} E_s(p, q, M(p), M(q)). \tag{1}$$

E_s is a smoothness term defined over neighboring pixels N in section 2.1. The data term E_d depends on whether the shift-map is forward or backward:

1. Forward direction, when retargeting the input image A to a smaller version B. In this direction we use $E_{d,ret}$ (Eq. 3 and Eq. 7). The role of α in that case, and the value that was used will be discussed in Sec. 4.
2. Backward feedback stage, when trying to reconstruct the input image A from the reduced image B. In this direction we use $E_{d,sim}$ (Eq. 4), and an α value of 1.

Once the graph is given, shift-map labeling is computed using the alpha-expansion algorithm [18–20]. As in [2] a hierarchical pyramid scheme is used to speed up the optimization.

Assignment of new locations to pixels using a similar energy minimization scheme was done in the texture-synthesis application of [11], and formulating image synthesis problems as a graph labeling problem was done by [2, 16, 17].

We follow [16] in using a heuristic adaptation of the alpha-expansion algorithm for smoothness costs where the triangle inequality is not guaranteed (Eq. 2). Video retargeting using shift-map is described in [21].

2.1 The Smoothness Term

The smoothness term E_s is identical to the one used in [2], and is based on the formulation in [22]. The smoothness term represents discontinuities added to the output image by discontinuities in the shift-map. The smoothness term should reflect, for each of the two neighboring output pixels, how different the value of its output neighbor is from the value of its input neighbor, as well as the difference of the output-neighbor gradient from the input-neighbor gradient. The smoothness term between two neighboring locations p and q in the output image R, having shift-maps $r = M(p)$ and $t = M(q)$, is defined as follows:

$$E_s(p,q,r,t) = \|I(p+r) - I(p+t)\|^2 + \|I(q+r) - I(q+t)\|^2 + \beta_s(\|\triangledown I(p+r) - \triangledown I(p+t)\|^2 + \|\triangledown I(q+r) - \triangledown I(q+t)\|^2). \tag{2}$$

we used a value of $\beta_s = 2$, and as in [2] the squared norm is used over RGB values.

2.2 Image Retargeting

In horizontal image retargeting, horizontal monotonicity is often required. I.e., if $M(u,v) = (t_x, t_y)$ and $M(u+1,v) = (t'_x, t'_y)$, than $t'_x \geq t_x$. This constraint assures that input objects are not duplicated in the output, and that the left-right relationship between objects will be maintained. In Shift-Map [2] this constraint is imposed through the smoothness term, giving a large penalty if $t'_x < t_x$.

The priority of input pixels to appear in the output image can be controlled using the data term as follows:

$$E_{d,ret}((u,v),(t_x,t_y)) = D(u+t_x, v+t_y), \tag{3}$$

where $D(x,y)$ is a value between 0 to 1 given to each input location (x,y). The higher the value of $D(x,y)$, the smaller are the chances to include the pixel at input location (x,y) in the output image.

2.3 A New Shift-Map Application: Similarity Guided Composition

Shift-Map can be used for the composition of an output image $R(x,y)$ from segments taken from a source image $I(x,y)$ while requiring that the resulting image $R(x,y)$ will be similar to a given target image $T(x,y)$. This is done by using the data term $E_{d,sim}$:

$$E_{d,sim}(p,t) = \|R(p) - T(p)\|^2 = \|I(p+t) - T(p)\|^2 \tag{4}$$

This data term is defined for every output pixel location $p = (u, v)$ and for every candidate shift-value $t = (t_x, t_y)$.

While similarity guided composition can be used to create very interesting visual effects, these will not be addressed here. In feedback retargeting only the following question is important: "how hard is it to reconstruct the target image T from pieces of an image I". The output image R itself is not used in retargeting. The next section will describe how to extract this information from the optimal shift-map.

3 Composition Score

After performing similarity guided composition and assigning shift-map labels to all output pixels, values of the data term (Eq. 4) and the smoothness term (Eq. 2) indicate how difficult it was to compose the target image T from pieces of the source image I. This indicates features in the target image that did not appear in the source image.

For an ordered pair of images $\langle T, I \rangle$, an optimal shift-map M will be computed to construct the target image T from the source image I. This will be done as described in Sec. 2 for similarity guided composition. Once the optimal shift-map has been computed, A composition score indicates for every image location $p \in T$ how hard it is to build its neighborhood from the source image I. This composition score is defined as follows:

$$E_{\langle T|I \rangle}(p) = \alpha E_d(p, M(p)) + E_{s,mean}(p, M), \tag{5}$$

where E_d is the data term as defined in Eq. 4, and $E_{s,mean}(p, M)$ is the average smoothness term E_s (Eq. 2) between a pixel p and its neighbors:

$$E_{s,mean}(p) = \frac{1}{4} \sum_{q \in N(p)} E_s(p, q, M(p), M(q)).$$

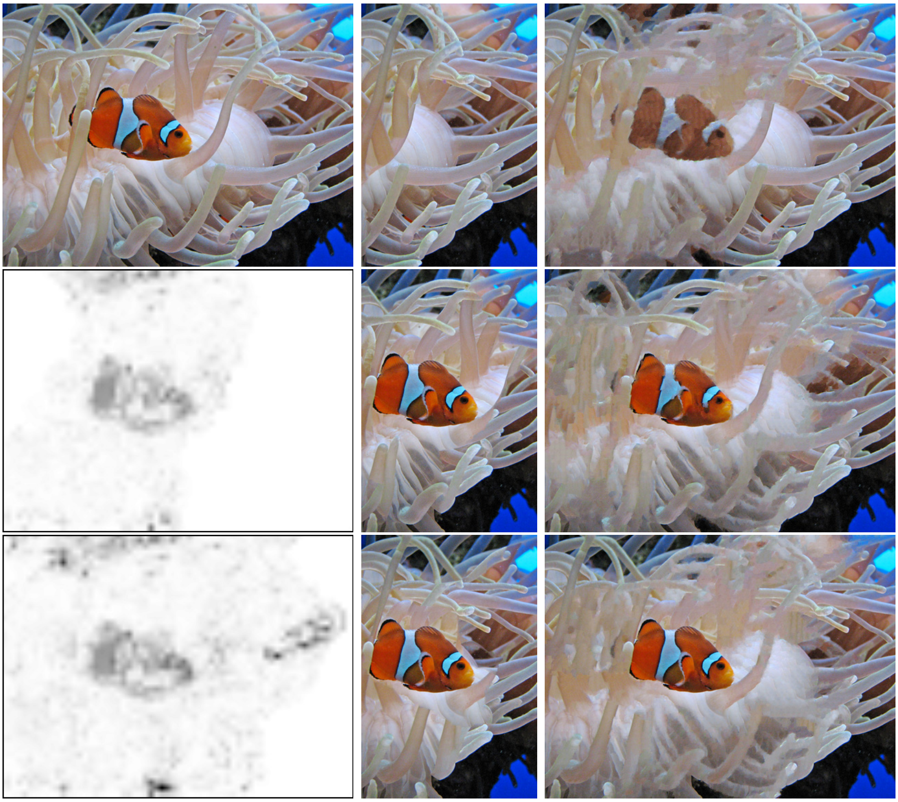
The composition-energy for the entire image is defined as

$$E(T|I) = \sum_{p \in T} E_{\langle T|I \rangle}(p), \tag{6}$$

and is equal to the value of $E(M)$ in Eq. 1.

It may be relevant to compare this score to the score used in [23], as both answer the abstract question: "How difficult is it to compose a signal S_1 from the signal S_2". The two algorithmic frameworks are designed for different tasks, and our score does not use SIFT descriptors nor an exhaustive search. In addition, our score is flexible enough to match patches of different sizes and shapes due to the use of similarity guided shift-map composition.

We get back to the retargeting of an input image A to a smaller output image B, and to the issue of similarity. Terms used in previous papers are also used here: "coherence" will indicate the ease of constructing the output B from the input A, and will be inversely proportional to the global composition energy $E(B|A)$.



(a) (b) Retargeted output (c) Construction of the source from the output using similarity guided composition.

Fig. 1. Iterations of feedback retargeting, reducing the width of an image by 2.
Top Row: (a) The source image A. (b) The initial output B_0 using Shift-Map retargeting, when data term is set to zero everywhere except the left and right columns. The fish disappeared. (c) Backward reconstruction of the input A from B_0 using similarity guided composition. The fish is poorly restored (best visible in color).
Middle Row: (a) The feedback score C_1 corresponding to regions in the input A that are poorly restored from B_0. High values are given to the fish area. (b) Shift-Map retargeting of the input A to B_1, this time using an updated data-term with the feedback values C_1. The fish returned! (c) Backward reconstruction of the input A from B_1 using similarity guided composition. This reconstruction is much better than the first attempt.
Bottom Row: The last iteration. (a) The cumulative feedback C_2 from the first two iterations. (b) Shift-Map retargeting to B_2 using the updated data-term. (c) The last and best Backward reconstruction of the input A from the retargeted image B_2.

"Completeness" will indicate the ease of constructing the input A from the output B, and will be inversely proportional to $E(A|B)$. Shift-Map retargeting [2] creates, by definition, a "coherent" output B.

In the bidirectional similarity work of [1], coherence and completeness are modeled slightly differently. In their work, the penalty for each rectangular patch (in different scales) equals to the SSD with its nearest neighbor in the other image. The global score adds up the scores of all patches. A similar formulation to [1] was used in [24] in the field of texture transfer.

In feedback retargeting we use Shift-Map [2] to reduce either the width or the height of the image. Shift-Map retargeting minimizes the value of $E(B|A)$ with respect to a constraint given by the data-term $E_{d,ret}$. In [2], the data term included maintaining the borders and perhaps a user-given saliency. Completeness, however, is not guaranteed by the original Shift-Map retargeting method. To improve completeness, feedback retargeting uses the data-term mechanism: input regions with lower value of $D(x, y)$ (Eq. 3) are more likely to appear in the output image. We propose to determine the value of $D(x, y)$ of an input pixel $p = (x, y)$ from the composition score for that pixel $E_{\langle A|B\rangle}(p)$.

4 Feedback Retargeting

Input regions with a high composition score $E_{\langle A|B\rangle}$ indicate that these input regions cannot be reconstructed accurately and easily from the output image B. If we decrease the value of $D(x, y)$ in Eq. 3 for pixels in these input regions, and recompute again the output B using Shift-Map, we increase the likelihood that components from these regions will appear in the output, increasing the completeness. The iterative feedback retargeting algorithm, creating a retargeted image having bidirectional similarity, is as follows:

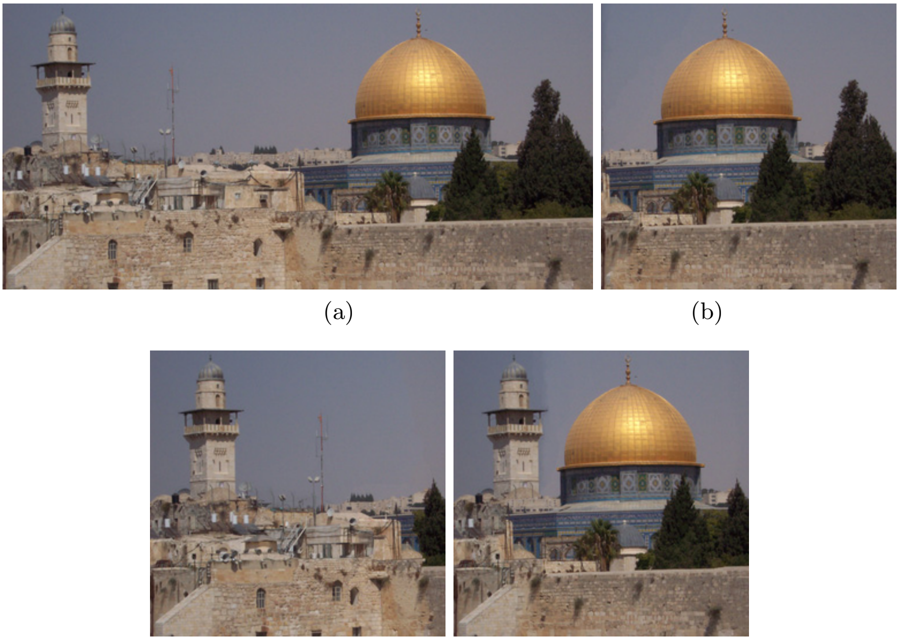
In each iteration we use Shift-Map retargeting to compute the retargeting result B from the input image A using the updated data-term. In the first iteration $D(x, y)$ is set to zero for all input pixels, and it is possible that important features of the input image will disappear in the initial retargeted image. Given the retargeted image B, the composition score $E_{\langle A|B\rangle}$ is computed using similarity guided composition as described in sections 2.3 and 3.

The composition score computed during the similarity guided composition is used as a feedback to improve the data term in the following way:

$$D(x, y) = e^{-C(x,y)} \tag{7}$$

where $C(x, y)$ is a cumulative feedback, summing the feedback $E_{\langle A|B\rangle}(x, y)$ from all previous iterations. We do not know which weight to use when adding $E_{\langle A|B\rangle}(x, y)$ to the cumulative feedback $C(x, y)$ in a way that will maximize completeness. We therefore search through a decreasing sequence of weights as follows, until we reach the first weight which reduces the global composition energy $E(A|B)$ therefore improves completeness.

$$C_{temp}(x, y) = C(x, y) + \frac{E_{\langle A|B\rangle}(x, y)}{\gamma 2^k}. \tag{8}$$



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

Fig. 2. Reducing width by 50%: (a) Source image. (b) Initial Shift-Map retargeting. (c) Intermediate result does not reduce the composition energy $E(A|B)$. (d) Final result with lower composition energy.

We use $\gamma = 0.05$, and the value of k is determined by searching $k = 0, 1, 2, \ldots$. Each k gives a candidate feedback C_{temp} computed by Eq. 8.

As mentioned before, we increase k until we reach the first value of k which reduces the global composition energy $E(A|B)$ (Eq. 6). In order to compute that value, retargeting of A to B, and reconstruction of B from A, are performed for each step of the search. An example with intermediate results of this process is shown in Fig. 2. Once an improved composition energy $E(A|B)$ is obtained, we update the value of C to this of C_{temp}, and iterations continue. This is a common way to perform gradient descent: A step in the direction of the gradient is tested; If the value of function is decreased the step is taken, and minimization continues, otherwise the step is not taken and a smaller step is tested.

Convergence is obtained after 2-3 iterations, once the improvement in the composition score is smaller than a given threshold. In each iteration we start with an initial value of k that was used in the previous iteration so we have about 6 calls for each of the shift-map optimization routines (retargeting or composition). The algorithm is demonstrated in Fig. 1.

To accelerate performance, the composition score $E_{\langle A|B\rangle}$ can first be computed on a downscaled version of the images. Then the composition map can be either magnified to the original resolution using bilinear interpolation, or the

Fig. 3. Feedback retargeting without border constraints: For every test image we show (from left to right): The input image, the result of [2] and our final output image.

hierarchical scheme of [2] can be used. Fig. 4 shows the difference between the two schemes. The synthesis itself is fine in both schemes, as the retargeting step is left the same, but the sensitivity to unique fine details increases. A compromise can obviously be made between the two schemes by performing optimization only on some of the levels of the pyramid. As for the Shift-Map retargeting step, one or two pyramid levels were used when the images were large. Horizontal shift-map is inherently faster due to a smaller number of allowed shifts. This makes the total running time only a few minutes.

Even though the value of C determines the priority of input pixels, it should not be confused with "saliency" or "importance" of the regions of the input image A. When an important feature is not removed by Shift-Map retargeting, it can have a low value of C.

As in the original shift-map algorithm, an infinite penalty was added if either the left-most or right-most column were not mapped to the new border location. Feedback retargeting works well also without this border constraint. Now that the values of $D(x, y)$ are not uniform, removing this constraint will not result in an under-determined problem. Practically, this means that pixels can be

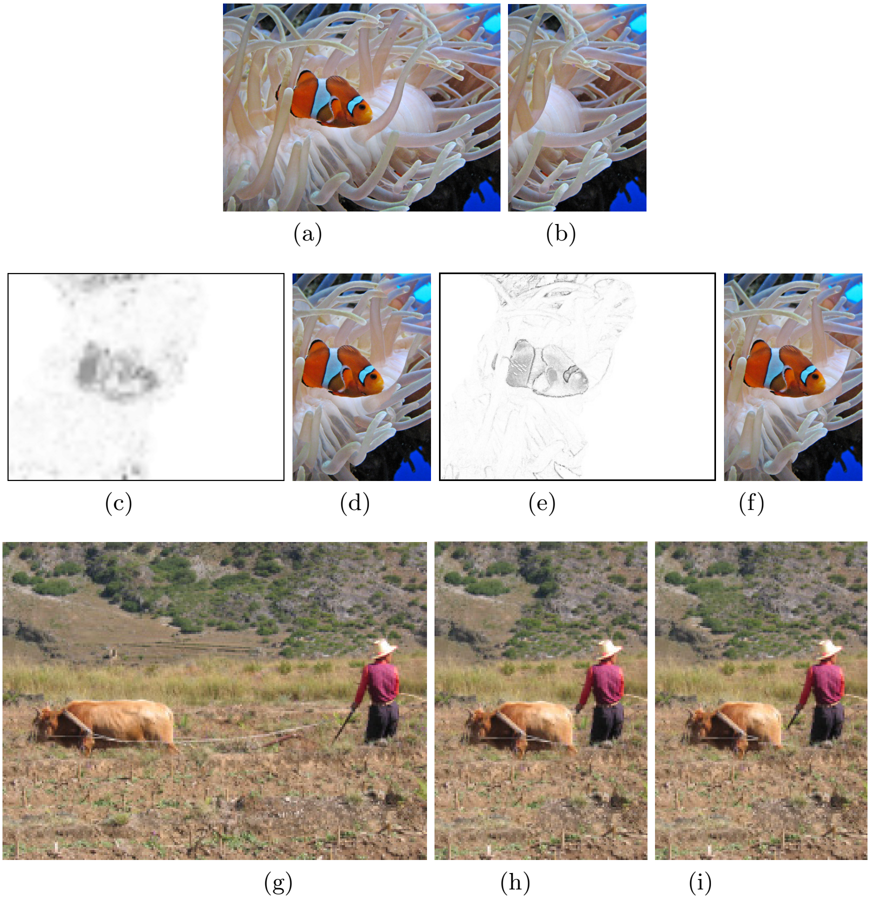


Fig. 4. Comparing coarse and fine feedbacks.
Fish: (a) Source image. (b) Initial retargeting (i.e., result of [2]). (c) Coarse Feedback. (d) Retargeted output using the coarse feedback (second iteration). (e) Feedback computed using the hierarchical scheme of [2]. (f) Retargeted output using this feedback. In this case a finer feedback does not improve the result.
Man and Cow: (g) Source image. (h) Final result using coarse feedback. (i) Final result using a finer feedback. The sensitivity to fine details is increased.

removed from the boundaries of the image as long as they does not increase the value of $E(A|B)$. But unlike the crop operator, we do not restrict pixels from the boundaries to be removed as entire columns. Fig. 3 shows results of this retargeting strategy. As it can be seen, in some cases most of the pixels removed were from the left and right boundaries, and in other cases most of the pixels removed were from the interior of the image.

The use of a negative exponent in Eq. 7 limits the data-term value to be between 0 and 1. This serves well our intentions, as we do not want the gradient

Simakov et. al [1] Shift-Map[2] Feedback Retargeting

Fig. 5. Comparison of Feedback retargeting with the Image-Summarization of [1]. Input image (same as in Fig. 4.g) and result of their method were taken from their article. Feedback retargeting does not suffer from blur caused by patch averaging.

descent technique to lead us to put too much weight on completeness. We also do not want to sacrifice coherence for completeness, but to make sure a certain amount of coherence will always be implied. The balance between the two can be controlled by the value of α in Eq. 1 in the forward retargeting stage. In our specific implementation we have used a value of $\alpha = 0.02$, and it worked well for most of our test images. But this value can be changed if the result should be more complete or coherent.

5 Comparison and Conclusion

Feedback retargeting combines the low level qualities of shift-map retargeting with the bidirectional similarity property [1]. As can be seen in Fig. 5, the result of [1] suffers from blurring caused by voting techniques, and possible distortions caused by repeated gradual resizing. Shift-Map retargeting does not capture the entire content of the input image. Feedback retargeting benefits from the best of both worlds, producing a good retargeting without blurring or distortions.

References

1. Simakov, D., Caspi, Y., Shechtman, E., Irani, M.: Summarizing visual data using bidirectional similarity. In: CVPR (2008)
2. Pritch, Y., Kav-Venaki, E., Peleg, S.: Shift-map image editing. In: ICCV, Kyoto (2009)
3. Setlur, V., Takagi, S., Raskar, R., Gleicher, M., Gooch, B.: Automatic image retargeting. In: Proceedings of the 4th International Conference on Mobile and Ubiquitous Multimedia (MUM 2005), pp. 59–68 (2005)
4. Avidan, S., Shamir, A.: Seam carving for content-aware image resizing. In: SIGGRAPH (2007)
5. Wolf, L., Guttmann, M., Cohen-Or, D.: Non-homogeneous content-driven video-retargeting. In: ICCV (2007)
6. Wang, Y., Tai, C., Sorkine, O., Lee, T.: Optimized scale-and-stretch for image resizing. In: SIGGRAPH (2008)

7. Rubinstein, M., Shamir, A., Avidan, S.: Multi-operator media retargeting. In: SIGGRAPH (2009)
8. Rubinstein, M., Shamir, A., Avidan, S.: Improved seam carving for video retargeting. In: SIGGRAPH (2008)
9. Krähenbühl, P., Lang, M., Hornung, A., Gross, M.: A system for retargeting of streaming video. In: SIGGRAPH Asia (2009)
10. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: Patchmatch: a randomized correspondence algorithm for structural image editing. In: SIGGRAPH (2009)
11. Kwatra, V., Schodl, A., Essa, I., Turk, G., Bobick, A.: Graphcut textures: image and video synthesis using graph cuts. In: SIGGRAPH, pp. 277–286 (2003)
12. Wexler, Y., Shechtman, E., Irani, M.: Space-time video completion. In: CVPR, vol. 1, pp. 120–127 (2004)
13. Kwatra, V., Essa, I., Bobick, A., Kwatra, N.: Texture optimization for example-based synthesis. In: SIGGRAPH, pp. 795–802 (2005)
14. Jojic, N., Frey, B.J., Kannan, A.: Epitomic analysis of appearance and shape. In: ICCV, vol. 1 (2003)
15. Dong, W., Zhou, N., Paul, J.C., Zhang, X.: Optimized image resizing using seam carving and scaling. In: SIGGRAPH (2009)
16. Rother, C., Kumar, S., Kolmogorov, V., Blake, A.: Digital tapestry. In: CVPR, pp. 589–596 (2005)
17. Rother, C., Bordeaux, L., Hamadi, Y., Blake, A.: Autocollage. In: SIGGRAPH, pp. 847–852 (2006)
18. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEET-PAMI 23, 1222–1239 (2001)
19. Kolmogorov, V., Zabih, R.: What Energy Functions Can Be Minimized via Graph Cuts? In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 65–81. Springer, Heidelberg (2002)
20. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEET-PAMI 26, 1124–1137 (2004)
21. Hu, Y., Rajan, D.: Hybrid shift-map for video retargeting. In: CVPR (2010)
22. Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S., Colburn, A., Curless, B., Salesin, D., Cohen, M.: Interactive digital photomontage. In: SIGGRAPH, pp. 294–302 (2004)
23. Boiman, O., Irani, M.: Similarity by composition. In: Schölkopf, B., Platt, J., Hoffman, T. (eds.) Advances in Neural Information Processing Systems 19, pp. 177–184. MIT Press, Cambridge (2007)
24. Wei, L.Y., Han, J., Zhou, K., Bao, H., Guo, B., Shum, H.Y.: Inverse texture synthesis. In: SIGGRAPH (2008)