

Importance Filtering for Image Retargeting

Yuanyuan Ding
Epson R&D, Inc.
yding@erd.epson.com

Jing Xiao
Epson R&D, Inc.
xiaoj@erd.epson.com

Jingyi Yu
University of Delaware
yu@eecis.udel.edu

Abstract

Content-aware image retargeting has attracted a lot of interests recently. The key and most challenging issue for this task is how to balance the tradeoff between preserving the important contents and minimizing the visual distortions on the consistency of the image structure. In this paper we present a novel filtering-based technique to tackle this issue, called "importance filtering". Specifically, we first filter the image saliency guided by the image itself to achieve a structure-consistent importance map. We then use the pixel importance as the key constraint to compute the gradient map of pixel shifts from the original resolution to the target. Finally, we integrate the shift gradient across the image using a weighted filter to construct a smooth shift map and render the target image. The weight is again controlled by the pixel importance. The two filtering processes enforce to maintain the structural consistency and yet preserve the important contents in the target image. Furthermore, the simple nature of filter operations allows highly efficient implementation for real-time applications and easy extension to video retargeting, as the structural constraints from the original image naturally convey the temporal coherence between frames. The effectiveness and efficiency of our importance filtering algorithm are confirmed in extensive experiments.

1. Introduction

Due to the fast growing diversity of display devices, an image often needs to be displayed across various settings such as aspect ratios. To maintain desirable visual quality across all conditions, a proper way of resizing is required. As a result, content-aware image retargeting has been an active research topic recently.

An image records the visual information of the covered scene viewed from a certain angle. When retargeted to a different setting such as aspect ratio, the original visual contents will be inevitably altered. To preserve the important contents close to the original, the other "un-important" pixels have to take more sacrifice. This generally changes

the overall image structure and often leads to visual distortion in the target image. A straight line may become badly curved in the target image, if different parts of it happen to have different importance. Such structural distortion often causes the target image look hardly natural. Therefore careful and proper treatment is required to minimize such distortion while preserving the important contents. How to balance this tradeoff is the key and most challenging issue for content-aware retargeting.

Many approaches have been successfully developed to address this problem. A comprehensive introduction of the recent development in this area is presented in [14], where the existing methods are categorized into two types. The first is discrete methods, including seam carving [1, 15] and shift-maps [13, 8]. This type of methods try to remove or copy unimportant pixels while keeping the important ones rigid. The second is continuous methods, including feature-aware texture mapping [4, 23], scale-and-stretch [22], and energy-based deformation [10]. These methods try to compute a non-uniform warping function from the original to the target image, which is designed to retain the important contents and warp the unimportant regions. To reduce the distortion on overall image structure, both types of methods use constraints from the image features to optimize the retargeting manipulation with local smoothness. Since the image features are usually computed at individual pixels or in a local patch, the global image structure can still be distorted in many occasions, as shown in Fig. 1. In [16], Rubinstein et al. propose a multi-operator approach that optimizes the combination of several methods and greatly reduces the visual distortion on image structure, though the essential problems for its individual components still remain.

To further minimize the visual distortion on image structure while preserving important contents, we propose to constrain the retargeting process directly using the original image itself, such that all the information, global and local, can be used together to enable the optimal overall quality. Based on this thinking, we develop the *importance filtering* algorithm for content-aware image retargeting. The algorithm consists of three major steps. First, we compute the image saliency using [12] and construct an importance map



Figure 1. Image retargeting by improved seam carving [15], scale-and-stretch [22], stream video [11], shift map [13], multi-operator [16], and our method. Note the distorted flags and chairs by [15, 11], unnatural stretch by [22, 11], and cut on humans and misplacement on roof and chairs by [13]. The results by [16] are close to ours, yet ours tend to better retain prominent areas while minimizing the distortion.

based on it. The image saliency by [12] was developed to measure visual attractiveness and used for human detection, but not designed to be consistent with the image structure, e.g. the pixels from the same object can have very different saliency. Therefore we define an importance map that represents the saliency as well as has the consistent structure with the original image, i.e. the pixels on the same object should be likely assigned the similar importance. Such a map is achieved using the guided filters [7, 6], i.e. filtering the image saliency under the guidance of the original image.

The resulting structure-consistent importance map provides the key constraint to determine how much a pixel is allowed to shift from the original to the target image. Ideally we want the neighboring pixels with similar importance to shift together so that their structure as a group will not be distorted. Meanwhile the pixels with high importance should not shift much with respect to the neighboring pixels with similar importance, such that their shape remains close to the original. On contrary the pixels with low importance should be allowed to move more relative to the neighbors. Note that these constraints are mainly on the relative shift of neighboring pixels, i.e. the gradient of the pixel shift. To satisfy these criterions, in the second step we develop a mapping function to compute the gradient map of pixel shifts based on the importance map.

The final step of our algorithm integrates the shift gradients across the image to construct a smooth shift map and render the target image. Since the shift for retargeting is usually one-dimensional, horizontal or vertical, the direct integration can still be inconsistent along the other dimension. This will again cause visual distortion. We propose an importance-weighted filtering method to address this issue. The method forces the integrated pixel shifts to be smooth along both dimensions and consistent across the image. The important pixels are weighted more so that the filtering process favors more on preserving their associated contents.

The combination of the two filtering processes in our algorithm ensures the consistency of overall image structure and yet preserves the important contents in the target image.

Our method significantly differs from existing content-aware retargeting approaches [1, 15, 22, 23, 16, 11, 24]. First and most importantly, we for the first time use the original image directly to constrain the retargeting manipulation. It is the key for our algorithm to minimize the distortion on the overall image structure. Second, we enforce the constraints from pixel importance to estimate the shift gradient, unlike the other methods where the saliency is used to compute the shift directly. Integrating the gradients to construct the pixel shifts avoids the undesired distortion such as pixel swap along the scan lines, which often occurs to direct mapping from saliency to pixel shift. Third, both the filtering operations in our algorithm are simple and allow highly efficient implementation. Since the original image frame is directly used as guidance in the filtering processes, the temporal coherence across frames is naturally conveyed. Combining with the efficiency advantage, our method can be easily extended for real-time video retargeting with little additional effort. We have conducted extensive experiments and comparisons based on the *RetargetMe* benchmark [14]. The results confirm the effectiveness and efficiency of our importance filtering algorithm.

2. Related Work

Numerous algorithms have been proposed for media retargeting across various settings such as aspect ratios. Traditionally this has been achieved by uniformly warping the contents to the target setting or cropping a single important region and discarding the rest [17, 18]. Though maintaining the overall structural consistency, such methods often either distort or discard partially the prominent image contents. To better present the important contents, content-aware meth-

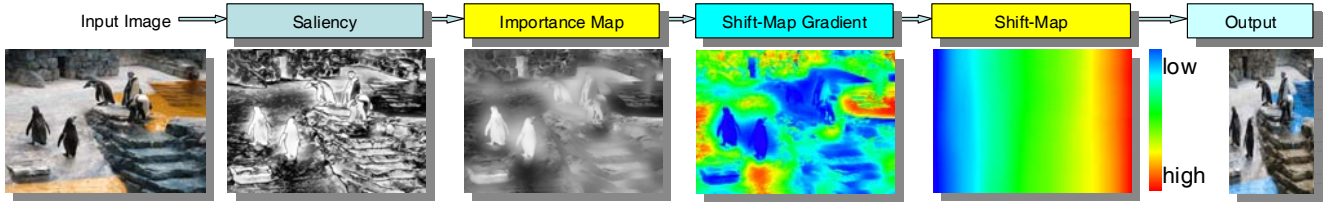


Figure 2. The pipeline of our algorithm.

ods has become the main stream for media retargeting, pioneered by the seam carving method [1].

Seam carving [1] resizes an image by reducing or adding one seam at each iteration. Each seam consists of a continuous chain of the least important pixel from each row or column so that the carving operation would not alter the important contents. This method has been extended for video retargeting [15, 3] and to allowing discontinuous seams for improving the quality of the target video [5]. In [16], a multi-operator approach is proposed to optimize the combination of seam carving with cropping and uniform scaling methods. It reduces the visual distortion on image structure from individual operators and improves the target image quality. Instead of manipulating one seam at a time, the shift-map method, another smart idea, optimizes the cropping and blending of the important image regions to construct the target image [13]. It thus better preserves the important image contents, though at the risk of significant change on the image structure. This method is also extended for video retargeting in [8].

Another category of methods try to compute a continuous warping function from the original image to the target [4, 23, 22, 10]. The warping is non-uniform in such a way that the important contents receive little changes while the un-important areas, e.g. homogeneous regions, suffer the most distortion. For this purpose [4] applies similarity constraint when warping user-specified important regions. Another nice method in [23] proposes a saliency-weighted linear system to compute the non-uniform mapping for individual pixels. The mapping does not enforce constraints to maintain the image structure and thus can lead to visual distortion. To reduce the distortion, [20] presents a method that applies joint bilateral filters on pixel shifts to rectify the image structure. In [10] a novel energy optimization scheme is proposed to constrain the distortions. The method in [22] divides an image into uniform grids and computes non-uniform warping that is small on important grids and big on un-important ones. The local warping functions are iteratively optimized by enforcing smoothness constraints on neighboring grids. This greatly reduces the distortion on the overall image structure.

The existing methods normally use image saliency directly to retain important pixels and local smoothness to restrain undesired visual distortion. They have achieved great success respectively. However, since the image features are

usually computed from individual pixels or a local area, the global image structure can still be distorted in many occasions, as shown in Fig. 1. A more comprehensive review and comparison is presented in [14]. For further improvement, we propose to use the original image directly as a global constraint to guide the retargeting manipulation, which preserves prominent contents as well as minimizes the visual distortions. We call the method importance filtering.

3. Importance Filtering

Our method resizes an image I of width X and height Y to a target I' of resolution $[X', Y']$. Without loss of generality, our illustration focuses on the example of fixing the height and resizing the width, i.e. $X' \neq X$ and $Y' = Y$. Application on the other cases is straightforward. Fig. 2 shows the basic pipeline of our importance filtering algorithm. We first compute the image saliency S^0 from I , and filters it under the guidance of I to obtain the pixel importance map S . We then estimate the gradient field G of pixel shifts from I to I' , by applying a non-linear mapping function on S . Finally, we compute the pixel shift-map M by integrating the gradient G with an importance-weighted filter and use it to render the target image. Below we explain the details of these steps.

3.1. Image Saliency and Importance Map

Image saliency is usually computed from local image features to measure the significance of pixels. Various saliency measurements have achieved success for image retargeting, e.g. gradient magnitudes [1, 23], neighborhood discontinuity [15], and patch based visual attention [9, 22]. We use the visual attention-based method in [12] together with the face detector in [19] to compute image saliency. Fig. 3(M) shows the saliency map computed from 3(L). This saliency map nicely captures the visual attractiveness of local regions. However it is not designed to be consistent with the geometric structure in the original image. As shown in Fig. 3(M), the saliency varies a lot on the same object such as the legs.

For content-aware retargeting, our goal is to preserve the important contents as well as minimize the distortions on image structure. For this purpose the pixels on the same object need to shift in nearly the same way. Since the pixel shift will be later decided by its importance, the desired importance should be close in the same object. The saliency by



Figure 3. Importance map construction: (L) Original, (M) Visual attention-based saliency, (R) Importance map by guided filtering.

[12] is thus not good enough for our purpose. To construct an importance map that matches the image structure, we use the guided filtering method [7] to enhance the saliency under the guidance of the original image. The guided filter is an excellent structure-preserving filter recently proposed [7]. A conceptually similar filter is presented in [6]. The filter considers the target as a linear transform of the guidance to constrain the smoothing process. The target thus nicely resembles the structure of the guidance after filtering. This is exactly what we need for pixel importance. Fig.3(R) shows the importance map by guided filtering on 3(M) under the guidance of 3(L). Clearly the importance inside the same object is more consistent and the image structure is better retained.

3.2. Shift-Map Gradient

In our approach, the retargeting is achieved by shifting the pixel coordinates and warping their colors from the original to the target image. A pixel (x, y) is retargeted to (x', y) in the target image with the shift of $M(x, y) = x' - x$. Such shifts across the image form the shift-map. Their signs show the shift direction, shrinking or enlarging, and the absolute values mean the amount of shifting. Fig.4(b) shows an example of the absolute shift-maps using uniform scaling and our importance filtering method respectively. The resulting target images are shown in Fig.4(a). We can see that the absolute shift values generally increase monotonically with respect to x . The shift-map is also desired to be smooth along the y dimension to avoid distortion, about which more details will be discussed in section 3.4.

The importance map in section 3.1 provides the key constraint to create the shift-map. Ideally the neighboring pixels with similar importance should shift together to maintain their structure as a group. To retain the prominent object shapes close to the original, pixels with high importance should not shift much relative to the neighbors with similar importance. On contrary pixels with low importance should contribute more on the overall shift across the image. Clearly these constraints are mainly on the relative shift of neighboring pixels, i.e. the gradient of the shift-map. Constant shift gradients refer to a uniform scaling of the local neighborhood. Zero gradient means rigid translation of the associated area and big gradient corresponds to large deformation. For simplicity of illustration, we define the shift gradient along the width dimension (x),

$$G(x, y) = \nabla_x M(x, y) \quad (1)$$

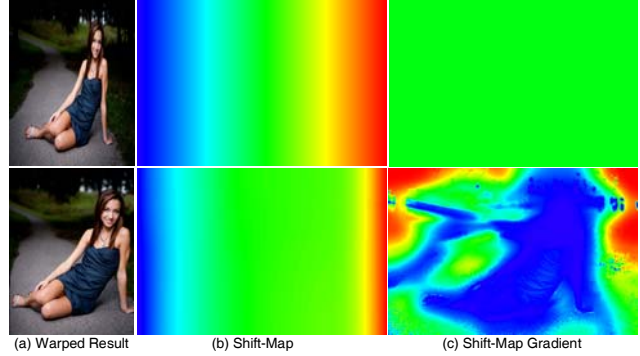


Figure 4. Shift-map and gradient by uniform scaling (top-row) and importance filtering (bottom). Both shift-maps monotonically increase. The shift-map gradient remains constant for uniform scaling, while varying greatly for importance filtering. Here red color refers to large shifts and blue means small.

Fig.4(c) shows an example of the shift-map gradient fields using uniform scaling and importance filtering respectively. As we expected, uniform scaling leads to constant gradients while importance filtering results in greatly varying gradients across the image. Comparing the original image in Fig.3(L) and the final target image in Fig.4(a), we can see that the region of human body has small (blue) and nearly constant gradient so as to retain the part of image nearly rigid, while the background region has big (green and red) and varying gradient to allow more severe deformation. This is exactly our purpose. Below we present how our importance filtering algorithm constructs this gradient field based on the importance map.

3.3. Gradient Mapping Function

We propose to estimate the shift gradient based on the corresponding pixel importance using a non-linear mapping function. Let the importance be normalized such that $S(x, y) \in [0, 1]$. The mapping function is desired to result in bigger gradient when the importance is closer to 0 and smaller value for importance closer to 1. The mapping is non-linear so that the value drops faster when the importance gets closer to 1. An intuitive choice for such a function is the zero-mean Gaussian,

$$G(x, y) = \ell \cdot e^{-\left(\frac{S(x, y)}{\sigma}\right)^2} \quad (2)$$

where σ is the variance and ℓ is the normalization term. A typical value of σ we use is 0.5.

Based on Eqn.(1), the integral of shift gradient in a row equals the total pixel shift from the original width X to the target width X' , i.e. $|X' - X| = \sum_{x=0}^{X-1} G(x, y)$. Thus ℓ can be computed as,

$$\ell = \frac{|X' - X|}{\sum_{x=0}^{X-1} e^{-\left(\frac{S(x, y)}{\sigma}\right)^2}} = C \cdot |\alpha - 1| \quad (3)$$

where

$$C = \frac{X}{\sum_{x=0}^{X-1} e^{-\left(\frac{S(x, y)}{\sigma}\right)^2}}, \quad \alpha = X'/X \quad (4)$$

α is the scaling factor. Apparently using the mapping function in Eqn.(2) the shift gradient will be uniformly scaled by $|\alpha - 1|$. The left column in Fig.5 shows an example of shift gradient varying with different α (blue line), the top for a pixel with high importance (1.5σ) and the bottom for a low-importance one (0.5σ). We can see that, when $\alpha = 1$, i.e. copying, both gradients are zeros. When $\alpha > 1$, i.e. enlarging, the gradient for the important pixel increases linearly with α getting bigger. As a result the distortion in prominent areas becomes more noticeable, as shown at the right top of Fig.5 ($\alpha = 1.5$). On the other hand, when $\alpha < 1$, i.e. shrinking, the gradient for the less-important pixel increases linearly with α getting smaller. Since the shrinking basically squeezes or even removes less-important pixels, the up-scaled gradient leads to more severe cut on less-important areas in the prominent contents, as shown at the right top of Fig.5 ($\alpha = 0.4$).

To avoid such distortion, the mapping function should be designed in such a way that, when $\alpha > 1$, the gradient for an important pixel starts to drop quickly with α getting bigger; when $\alpha < 1$, the gradient for a less-important pixel grows more slowly than that by Eqn.(2) with α getting smaller. This way both the undesired deformation and cut in prominent areas can be reduced. Accordingly we propose three different designs for the mapping function,

$$G(x, y, \alpha) = C_1 \cdot |\alpha - 1| e^{-\alpha^{2\alpha^2} \cdot (\frac{S(x,y)}{\sigma})^2} \quad (5)$$

$$G(x, y, \alpha) = C_2 \cdot |\alpha - 1| e^{-\frac{\alpha \cdot S(x,y)}{\sigma}} \quad (6)$$

$$G(x, y, \alpha) = C_3 \cdot |\alpha - 1| e^{-|\alpha - 1| \cdot (\frac{S(x,y)}{\sigma})^2} \quad (7)$$

where the normalization terms can be computed similarly.

As shown in Fig.5, these mapping functions satisfy our need but serve for different cases. When $\alpha > 1$, the gradients for the important pixel drop to below Eqn.(2) before α reaches 2. Among them Eqn.(5) drops the fastest and (7) the slowest. The gradient by Eqn.(2) is always bigger than both (5) and (6), but smaller than (7) when $\alpha < 2$. As a result, Eqn.(5) retains the prominent contents the best, (6) works better than (2), and (7) allows the most stretching on prominent objects when $\alpha < 2$, as shown in Fig.5.

When $\alpha < 1$, the gradients for less-important pixels by all the three functions are always lower than that by Eqn.(2). Among them Eqn.(5) is the highest. Eqn.(6) is higher than (7) when $\alpha > 0.6$, but lower with smaller α . As a result, all three functions achieve less cut in the prominent areas than that by Eqn.(2). Eqn.(7) leads to the least cut when $\alpha > 0.6$ and (6) saves prominent contents the most when further down-scaling, as shown in Fig.5.

These mapping functions can be easily further combined into one function with more stable performance. But in our system we provide all of them and leave it for users to choose the best one, since they can serve for different purposes, e.g. retaining important image regions or reducing

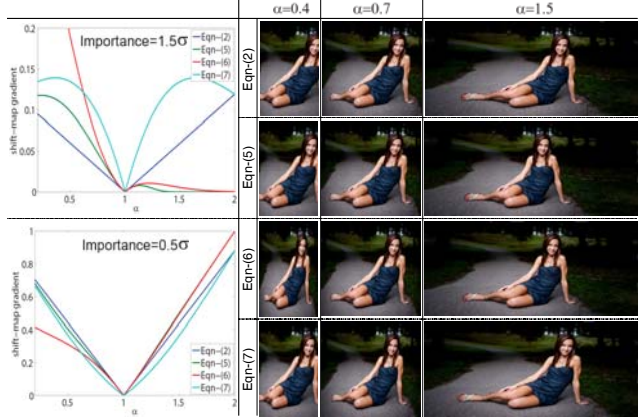


Figure 5. Left shows the response of the four gradient mapping functions with respect to the scaling factor α , top for an important pixel and bottom for a less-important one. Right shows the result images using respective functions at three resizing scales.

the cut on them. In the rest of the paper, we use Eqn.(5) as the mapping function to illustrate our method, since it works the best to preserve the prominent contents as well as minimize the structural distortion. An example of the gradient field by Eqn.(5) is shown in Fig.4(c), where $\alpha = 0.5$.

3.4. Gradient Integration by Importance-Weighted Filtering

Once the shift gradients G are constructed, we can integrate them to estimate the shift-map M and render the target image. A straightforward solution to the integral is to optimize the objective as follows,

$$M_{opt} = \min_M \quad |\nabla_x M - G| \quad (8)$$

$$s.t. \quad M(0, y) = 0, \quad M(X, y) = |X' - X|$$

Such an optimization process is often computationally expensive. Furthermore, since the pixel shift is only one-dimensional, integration along individual rows may still be inconsistent with each other. This inconsistency will cause undesired visual distortion in the target image. A naive solution to this may be direct gradient integration followed by shift-map smoothing using box filters. However smoothing the shift-map directly has disadvantages. First, it often leads to artifacts such as blurriness, holes, and pixel swaps, especially cross the object borders. Second and more importantly, smoothing by box filters can hardly rectify the shift inconsistency across the image. The inconsistency is accumulated over the columns during integration and can be big everywhere in the later part of the image. The box filters smooth the pixel shifts in small local neighborhood and thus can unlikely restore the global consistency in the target image, as shown in Fig.6 (L).

To retain consistency, we propose an efficient algorithm that incorporates importance-weighted filtering into the integration process. Specifically, at each step we integrate one pixel in each row at the present column. The shift integral



Figure 6. (L) Enlarged image using the shift-map computed by direct integration followed by box filter smoothing; (R) Enlarged image using the shift-map by our method.

at this pixel equals its gradient plus an importance-weighted average of the pixel shifts in a large neighborhood within the previous column. The formulation is,

$$M(x, y) = \frac{\sum_{j=y-r}^{y+r} w(x, j)[M(x-1, j) + G(x, j)]}{\sum_{j=y-r}^{y+r} w(x, j)} \quad (9)$$

It defines a 1D column filter of size $(2r + 1)$. The typical choice for r is a quarter of the image height. The high efficiency of 1D filter allows such a large kernel size, and averaging over a large neighborhood enables the integrated shift-map to be smooth and consistent in both dimensions. The weight w is designed in such a way that the averaging filter does not affect the important pixels as much as the unimportant ones. In another word, in the filtering process the shift from the important pixels should contribute more so that their shape will not be distorted by the nearby unimportant pixels. Hence we define w based on the pixel importance as follows,

$$w(x, y) = e^{S(x, y)} \quad (10)$$

Fig.4(b) shows the shift-map integrated from the gradient field in 4(c) using our method, where $\alpha = 0.5$. As expected, the pixel shifts are smooth and consistent. The target image can then be easily warped based on the pixel shifts. As shown in Fig.6(R), the enlarged image preserves the prominent contents, e.g. human, as well as retain the overall structural consistency, e.g. the relationship among objects.

4. Results and Discussion

We have tested our importance filtering algorithm on various images in a PC with Duo CPU 2.53GHz. Without code optimization, it takes less than 80ms for retargeting an image of 1024×768 , without the need of down-sampling. Our system is flexible to provide both fully automatic solution and interactive way for users to select areas to preserve. Though all the results shown in this paper are achieved by the automatic solution. Fig.7 shows an example of the entire pipeline of our importance filtering algorithm, where the important contents are well preserved without distorting the overall image structure.

We compare our method with the state-of-art methods on the *RetargetMe* dataset [14]. Fig.8 shows five examples. Due to space limit, we only show three methods, improved seam carving [15], shift map [13], and scale-and-

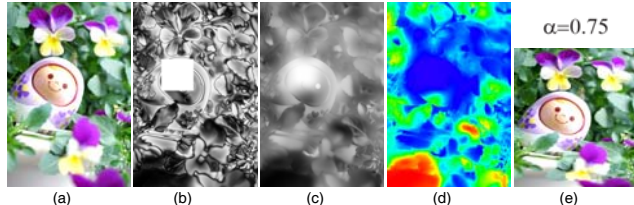


Figure 7. (a) Original, (b) Saliency, (c) Importance map, (d) Shift gradient field by Eqn.(5), (e) Retargeted image by our method.

stretch [22] here. Two more comparisons including stream video [11] and multi-operator [16] are shown in Fig.1. As expected, our method achieves the best overall balance between retaining the prominent contents and minimizing the distortion on image structure. The other methods generally work well but lead to noticeable artifacts occasionally.

For improved seam carving [15], abrupt distortions occur on the human bodies in the first row and the house shapes in the third and fifth rows of Fig.8. Stream video [11] stretches the images unnaturally and also causes undesired distortion, e.g. the leg of the man on the right and the white flags on the pillars in Figure 1. Scale-and-stretch [22] avoids abrupt distortions and achieves smooth image structure, but the contents such as humans in the first two rows, house shadow in the third row, and street sidewalk in the fifth row in Fig.8 are stretched or squeezed unnaturally. Shift map [13] produces smooth and natural images after resizing, but it results in severe cuts on important contents almost in all the examples. It can also alter the image contents significantly, e.g. the roof and chairs in Fig.1 and the shoulder of the girl in the second row of Fig.8. In many cases the resized images by these methods, especially scale-and-stretch [22] and shift map [13], may by themselves look quite realistic. But when placed together with the original image, significant changes on image structure or prominent contents can be observed. Our method tends to minimize such changes.

Among the existing works, the multi-operator method yields the most similar results with ours, as shown in Fig.1. The slight difference is that our method favors more on preserving the important contents and allows more deformation on unimportant background areas. Please refer to the supplementary materials for more results. We can also share our program for evaluation and comparison.

As shown in Fig.2, one of the critical steps in our method is image saliency estimation. It is a difficult problem by itself and none of the saliency measurements can guarantee a perfect estimation. In case the visual attention-based saliency measurement [12] fails to locate a prominent area, our method may result in undesired distortion in that area. For example, in the first row of Fig.8, one leg of the second man from the left is given a very low saliency. Our method thus shrinks it more than the other leg and leads to an unnatural local area in the resized image. A more reliable saliency measurement can further improve the performance

of our method, though it is not our key concern in this paper.

Our method directly warps the image pixels based on the integrated shift-map to render the final target image. In case the unimportant areas are squeezed a lot, the direct mapping may result in artifacts like discontinuity in the resized image. For example, in the second row of Fig.8, the right top of the image is squeezed significantly and appears divided by a couple of vertical lines. One solution may be to warp the color gradient of the resized image and then integrate the gradient to construct the target image, using gradient-domain compositing methods such as Poisson blending. This way a smooth color image can be achieved. A similar idea is presented in [20].

We straightforwardly extend our method for content-aware video retargeting, by basically frame-by-frame application of our image resizing method. The only difference is that a motion feature, motion energy image (MEI) [2], is added to the saliency cues. For each frame, the MEI is computed using a neighborhood of 20 frames and directly added to its image saliency to construct the combined saliency. The rest of the process is exactly identical to that for image retargeting. Since the original image frame is used as guidance in the filtering processes, our method is able to naturally maintain the temporal coherence in the retargeted video without the need of special care. For a video of 640×256 , Our method achieves 15 *fps* with pre-computed saliency or 6 *fps* including saliency calculation using the above-mentioned PC. Due to space limit, please refer to the supplementary materials for our results. Our method does not consider the global camera motion and thus works on videos by nearly fixed cameras. Similar to [21], the camera motion can be compensated by frame registration.

5. Conclusion and Future Work

We have presented the importance filtering algorithm for content-aware image retargeting. It directly uses the original image as the constraint to filter and estimate pixel importance so that it is consistent with the image structure. This is the key to minimize the visual distortion and yet preserve the prominent image contents. The constraint is applied on the gradient of pixel shift, instead of directly on pixel shift. This further avoids undesired distortion such as pixel swap that occurs to many earlier methods. The importance filtering operations are highly efficient and ready for real-time applications. We also show that easy extension to video retargeting is promising.

One potential improvement to the importance filtering algorithm is to extend the one-dimensional shift gradients to 2D. Even though the pixels all shift along the same dimension, the shift-map on the 2D image has a 2D gradient field. We are developing methods to estimate such 2D shift gradients and then optimize their integration to construct the shift-map by methods such as alternative 1D filtering or

Poisson blending. We believe this will further improve the 2D smoothness and consistency of the resized image.

Acknowledgement

This work was conducted in Epson Research and Development, Inc. Yu was partially supported by the National Science Foundation under grants IIS-CAREER-0845268 and IIS-RI-1016395.

References

- [1] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *Transaction on Graphics*, 2007. 89, 90, 91
- [2] A. Bobick and J. Davis. An appearance-based representation of action. In *ICPR*, 1996. 95
- [3] B. Chen and P. Sen. Video Carving. In *Eurographics*, 2008. 91
- [4] R. Gal, O. Sorkine, and D. Cohen-Or. Feature-aware texturing. In *EGSR*, 2006. 89, 91
- [5] M. Grundmann, V. Kwatra, M. Han, and I. Essa. Discontinuous seam-carving for video retargeting. In *IEEE CVPR*, 2010. 91
- [6] M. D. Gupta and J. Xiao. Bi-affinity filter: A bilateral type filter for color images. In *ECCV 2010 Workshop on Color and Reflectance in Computer Vision*, 2010. 90, 92
- [7] K. He, J. Sun, and X. Tang. Guided image filtering. In *ECCV*, pages 1–8, 2010. 90, 92
- [8] Y. Hu and D. Rajan. Hybrid shift map for video retargeting. In *IEEE CVPR*. 89, 91
- [9] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *TPAMI*, 1998. 91
- [10] Z. Karmi, D. Freedman, and C. Gotsman. Energy-based image deformation. In *Symp. on Geom. Proc.*, 2009. 89, 91
- [11] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross. A system for retargeting of streaming video. *TOG*, 2009. 90, 94
- [12] S. Montabone and A. Soto. Human detection using a mobile platform and novel features derived from a visual saliency mechanism. *Image Vision Comput.*, 2010. 89, 90, 91, 92, 94
- [13] Y. Pritch, E. Kav-Venaki, and S. Peleg. Shift-map image editing. In *ICCV*, pages 151–158, 2009. 89, 90, 91, 94, 96
- [14] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir. A comparative study of image retargeting. *ACM SIGGRAPH ASIA*, 2010. 89, 90, 91, 94
- [15] M. Rubinstein, A. Shamir, and S. Avidan. Improved seam carving for video retargeting. *TOG.*, 2008. 89, 90, 91, 94, 96
- [16] M. Rubinstein, A. Shamir, and S. Avidan. Multi-operator media retargeting. *TOG.*, 2009. 89, 90, 91, 94
- [17] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen. Gaze-based interaction for semi-automatic photo cropping. In *SIGCHI*, 2006. 90
- [18] B. Suh, H. Ling, B. B. Bederson, and D. W. Jacobs. Automatic thumbnail cropping and its effectiveness. *ACM UIST*, 2003. 90
- [19] P. Viola and M. J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 2004. 91
- [20] S.-F. Wang and S.-H. Lai. Fast structure-preserving image retargeting. *ICASSP*, 2009. 91, 95
- [21] Y.-S. Wang, H. Fu, O. Sorkine, T.-Y. Lee, and H.-P. Seidel. Motion-aware temporal coherence for video resizing. *ACM SIGGRAPH ASIA*, 2009. 95
- [22] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee. Optimized scale-and-stretch for image resizing. *ACM Transaction on Graphics*, 2008. 89, 90, 91, 94, 96
- [23] L. Wolf, M. Guttman, and D. Cohen-Or. Non-homogeneous content-driven video-retargeting. In *ICCV*, 2007. 89, 90, 91
- [24] G. Zhang, M. Cheng, S. Hu, and R. R. Martin. A shape-preserving approach to image resizing. *PG*, 2009. 90

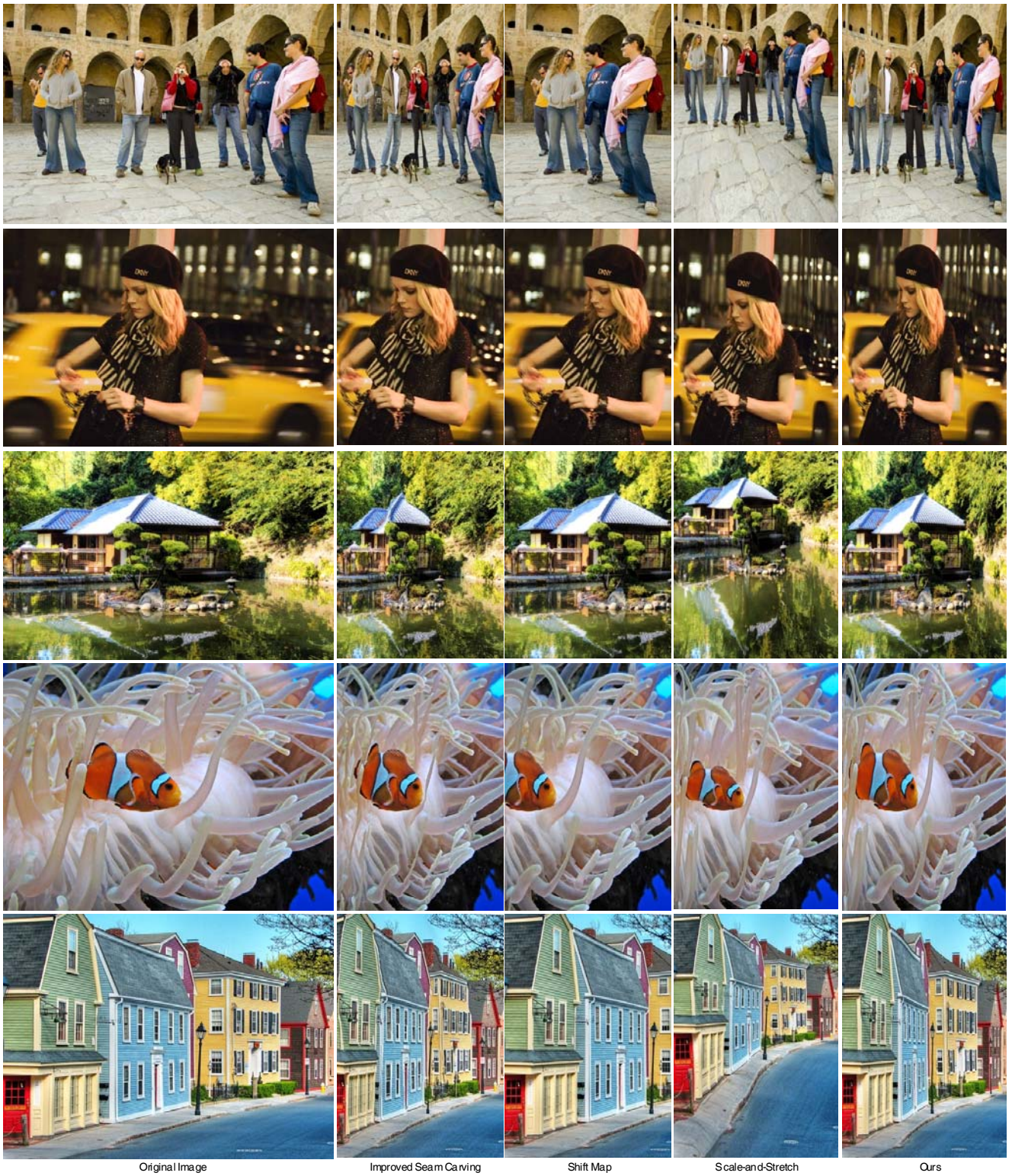


Figure 8. Comparison with the existing methods: improved seam carving [15], shift map [13], and scale-and-stretch [22]. All images are retargeted to half width ($\alpha = 0.5$). Note the distorted human bodies and buildings by [15], the unnatural stretch or squeeze on human bodies and houses by [22], and the cut and misplacement on fish, buildings, and humans by [13]. Our method tends to minimize such distortions while preserving prominent contents.