

MULTI-OPERATOR RETARGETING BASED ON PERCEPTUAL STRUCTURAL SIMILARITY

Yuming Fang¹, Weisi Lin², Zhou Wang³, Zhijun Fang¹, Long Xu⁴, and Yong Yang¹

¹School of Information Technology, Jiangxi University of Finance and Economics,
Nanchang 330032, Jiangxi, China

²School of Computer Engineering, Nanyang Technological University, Singapore 639798

³Dept. of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L3G1, Canada

⁴National Astronomical Observatories, Chinese Academy of Science, Beijing, China

ABSTRACT

We propose a new multi-operator retargeting algorithm by using three resizing operators of seam carving, scaling, and cropping iteratively. To determine which operator should be used at each iteration, we adopt structural similarity (SSIM) to evaluate the similarity between the original and retargeted images for the dynamic programming. Since the sizes of original and retargeted images are different, SIFT flow is used for dense correspondence between the original and retargeted images for similarity evaluation. Additionally, visual saliency is used to weight SSIM results based on the characteristics of the Human Visual System (HVS). Experimental results on a public image retargeting database show the promising performance of the proposed multi-operator retargeting algorithm.

Index Terms— Multi-operator Retargeting, Image Retargeting, Structural Similarity, Image Saliency

1. INTRODUCTION

With the fast spread of various mobile devices, maximizing the visual experiences of end users on small display screens has become ever important. Image/video retargeting techniques are essential to adapt the original visual content to various types display screen with different sizes or aspect ratios. A traditional method is to rescale the image by down-sampling. However, this method may result in worse viewing experience as some salient objects turn to be too small [1, 2]. Image cropping is an alternative solution preserving the Regions of Interest (ROIs) in images by deleting other background regions. The problem with this method is that it will cause context information loss [1, 2]. To overcome these problems, many content-aware image retargeting algorithms have been proposed to preserve important image content with minimal distortions during the image resizing process [1]-[15].

One early content-aware image retargeting algorithm is named seam carving proposed in [1], where a seam is defined

as eight-connected path of pixels (from top to down or left to right) including only one pixel in each row or column. Seam carving reduces width (or height) of images by removing the unimportant seams. A gradient map is used to determine the visual importance of each pixel in images. Later, Rubinstein *et al.* extended the seam carving algorithm to video retargeting by introducing the forward energy method [3]. The removed 1D seam from 2D images is extended into the 2D seam manifolds for 3D spatio-temporal volumes in video [3]. In [5], Grundmann *et al.* extended the seam carving technique through the discontinuous seam carving in both space and time for video retargeting. Instead of using geometrically smooth and continuous seams, that study adopts temporally discontinuous seams based on a designed appearance-based temporal coherence calculation method.

Other advanced image retargeting algorithms have also been proposed. Wolf *et al.* introduced a video retargeting algorithm through introducing a linear system to determine the new pixel position [6]. In that study [6], the visual importance of each image pixel is measured by the visual importance map composed of local saliency map, facial map and motion map. Jin *et al.* presented a content-aware image resizing algorithm through warping a triangular mesh over images by regarding salient line features and curved features as important regions [7]. A standard quadratic programming method is utilized to resize images [7]. Guo *et al.* advanced an image retargeting algorithm through utilizing saliency-based mesh parametrization [4].

In [8], Cho *et al.* described a patch transform method and demonstrated its application in image editing. The method divides the input image into non-overlapping patches and the new image is reconstructed based on the defined constraints [8]. A Markov network is devised to reconstruct the image with specified constraints [8]. Barnes *et al.* proposed the PatchMatch algorithm for image editing by finding approximate nearest-neighbor matches between patches [12]. That algorithm can obtain substantial performance improvement over the existing ones and thus can be used in interactive edit-

ing tools [12]. Pritch *et al.* designed an image editing algorithm based on an optimal graph labeling [11]. In that study, the defined data term and smoothness term are used to calculate the optimal shift-map [11].

Recently, Rubinstein *et al.* conducted a user study and found that applying multi-operators (such as seam carving, cropping and so on) can obtain better results than those from only a single operator in image retargeting [14]. In that study, the authors proposed a multi-operator media retargeting algorithm which combines seam carving, scaling and cropping operators to resize images. The size amount for each operation is determined by the optimal result for maximizing the similarity between the input image and the retargeted image. Bi-Directional Warping (BDW) is used to determine the number of each resizing operator [14]. In [15], Dong *et al.* introduced an image retargeting algorithm by combining seam carving and image scaling. The authors utilized a bidirectional similarity (BDS) function of image Euclidean distance, a dominant color descriptor similarity and seam energy variation to determine the best number of seam carving operation [15]. However, the methods (such as BDW, BDS, etc.) used to determine the number of each resizing operator in these multi-operator retargeting algorithms are not effective enough [16]. They can not obtain the comparable results with subjective results [16].

Based on our previous study [18], we present a new multi-operator algorithm by using seam carving, scaling and cropping iteratively based on perceptual structural similarity (SSIM). SSIM [17] is adopted to calculate the similarity between the original and retargeted images to determine the operator used at each iteration. In addition, inspired by the fact that the human visual system (HVS) always focus on the most important regions in images, we use visual saliency to weight the similarity measure. Experimental results demonstrate that the proposed multi-operator retargeting algorithm can obtain promising retargeted results.

2. PROPOSED FRAMEWORK

The proposed framework is demonstrated as Fig. 1. From this figure, we can see that the proposed framework iteratively apply a sequence of resizing operations dynamically selected from the candidate resizing operators including seam carving, scaling or cropping. Perceptual SSIM is used to evaluate the similarity between the original and retargeted images to determine the appropriate ratio among seam carving, scaling and cropping operators. Different from the study [14] which searches for the best operator in the entire multi-operator space, the proposed framework searches for the best one at each iteration. The whole image retargeting task is divided into subtasks, each associated with one iteration with only small changes in image size. In the n -th iteration, perceptual SSIM between the original and retargeted image is employed to pick the best quality image among those created by seam

carving, scaling, and cropping operators applied to the input image generated from the last iteration (in the first iteration, the original image is used as the input image). The iteration continues until the target resizing factor is obtained.

2.1. Perceptual Similarity Measure Method

The critical step of the proposed framework is to determine the optimal image resizing operator sequence for image resizing. Our previous study has demonstrated that the perceptual SSIM can be used to well evaluate the distortion of retargeted images [18]. In this study, we adopt the perceptual SSIM to calculate the similarity between the original and retargeted images to determine the optimal image resizing operator sequence.

Since the original and retargeted images have different sizes, we employ SIFT flow [19] to estimate the dense correspondence between the original and retargeted images. For each pixel p in the original image, we find its matching pixel p' in the retargeted image by using SIFT flow. Once the pixel correspondence is established, the local structures in the corresponding regions can be compared for similarity evaluation. SSIM index [17, 21] is used to predict the structural similarity between image patches. Given two image patches x and x' centered at p and p' in the original and retargeted images, respectively, the SSIM index between them is given by

$$\text{SSIM}(\mathbf{x}, \mathbf{x}') = \frac{(2\mu_{\mathbf{x}}\mu_{\mathbf{x}'} + C_1)(2\sigma_{\mathbf{x}\mathbf{x}'} + C_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\mathbf{x}'}^2 + C_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{x}'}^2 + C_2)}, \quad (1)$$

where $\mu_{\mathbf{x}}$, $\mu_{\mathbf{x}'}$, $\sigma_{\mathbf{x}}^2$, $\sigma_{\mathbf{x}'}^2$, and $\sigma_{\mathbf{x}\mathbf{x}'}$ denote the local mean of \mathbf{x} and \mathbf{x}' , the local variance of \mathbf{x} and \mathbf{x}' , and the local covariance between \mathbf{x} and \mathbf{x}' , respectively. C_1 and C_2 are small positive stability constants that account for the saturation effects of the visual system at low luminance and contrast [17]. Similarly with the study [18], we use multi-scale operation to calculate the overall SSIM value by using the SSIM measures across different scales:

$$\text{SSIM}(\mathbf{x}, \mathbf{x}') = \frac{1}{M} \sum_{j=1}^M w_j \text{SSIM}_j(\mathbf{x}, \mathbf{x}'), \quad (2)$$

where SSIM_j is the SSIM measure at the j -th scale, and w_j for $j \in \{1, 2, \dots, M\}$ are weighting parameters determining the relative importance of different scales. Here, we up-sample the SSIM maps from different scales based on bilinear interpolation to match the size of the final SSIM map. As in [18], we set $M=5$ and adopt the weighting factors w_j obtained through psychovisual experiments.

It is well known that the Human Visual System (HVS) would pay more attention on salient regions than other regions in the natural scene. In this study, we calculate the saliency map by considering both bottom-up and top-down mechanisms. The bottom-up saliency map S_b is computed by

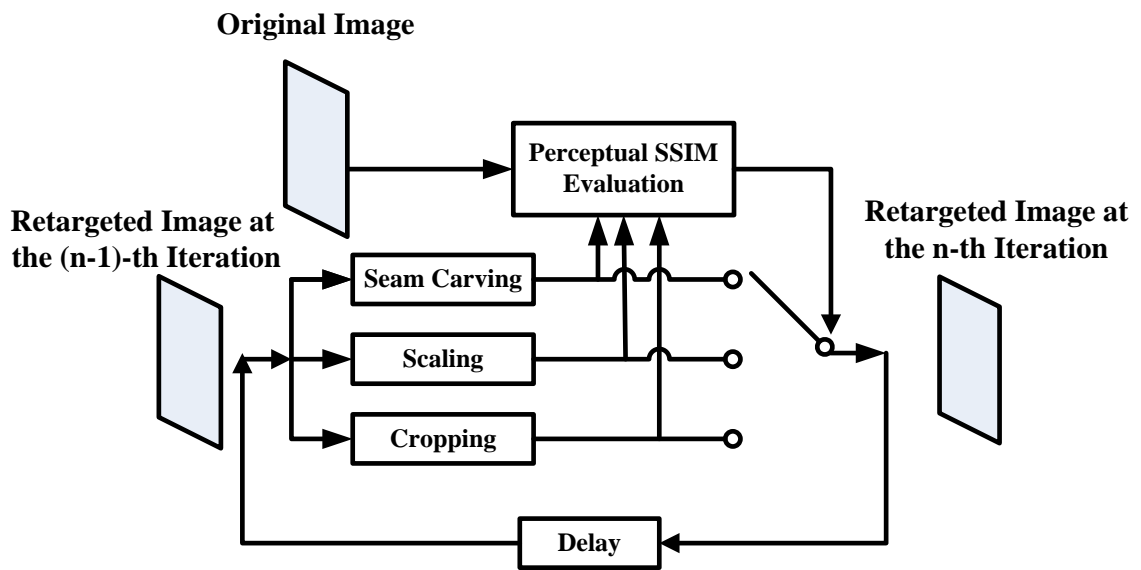


Fig. 1: The proposed framework.

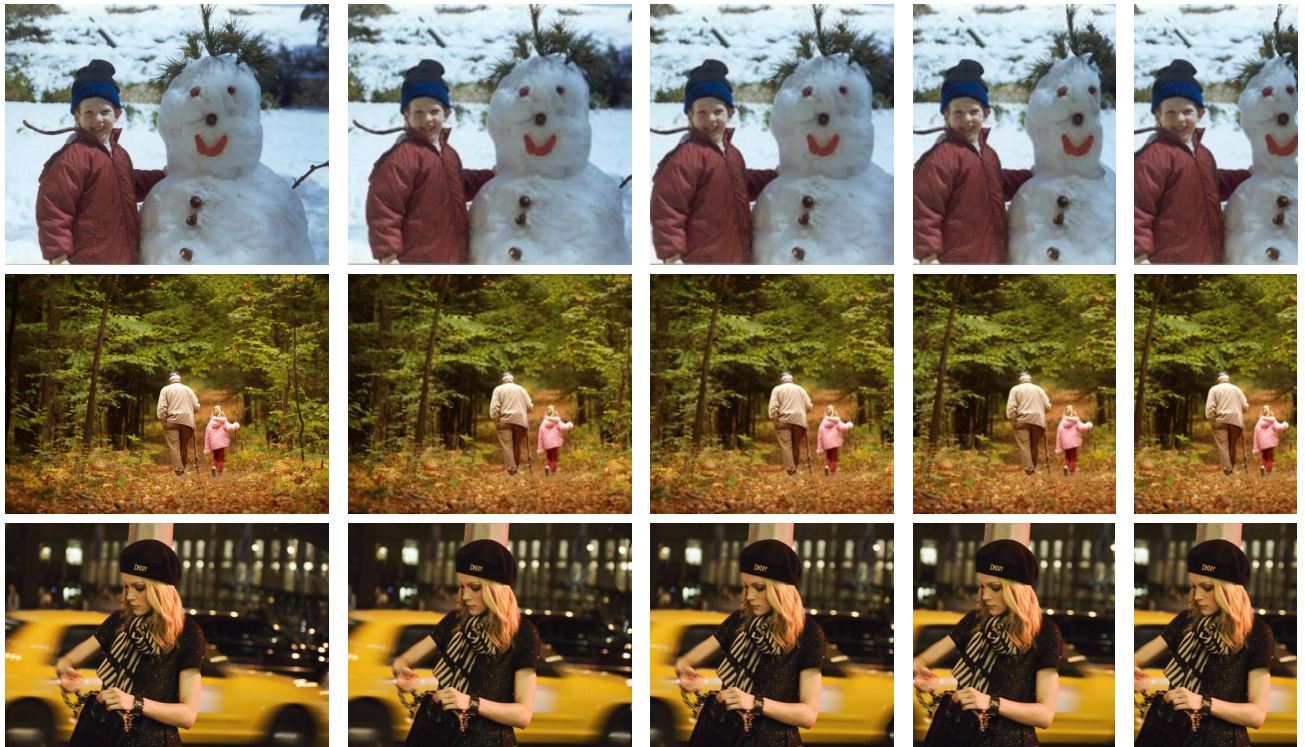


Fig. 2: Sample images created by the proposed method. Column 1: reference images. Columns 2-5: retargeted images for resizing factors of 12.5%, 25%, 37.5%, and 50%, respectively.



Fig. 3: Left: reference images; middle: retargeted images by multi-operator retargeting [14]; right: retargeted images by the proposed method.

our previous model in compressed domain [13], while the top-down saliency map S_t is estimated by a face detector [22]. Of course, other top-down factors can be also integrated into the proposed method. The linear combination method is used to combine the bottom-up and top-down saliency map to obtain the final saliency map for images as:

$$S = (S_b + S_t)/2 \quad (3)$$

The overall SSIM can provide a useful local indicator about the similarity between the original and retargeted images. Meanwhile, recent studies have demonstrated that even strong artifacts may be unnoticed in the locations outside the regions of interest [23]. Therefore, we use the saliency map to weight the overall SSIM to calculate the SSIM for image retargeting (IR-SSIM) [18] as follows.

$$\text{IR-SSIM} = \frac{\sum_i S_i \text{SSIM}_i}{\sum_i S_i} \quad (4)$$

where SSIM_i and S_i are the SSIM value and saliency value at each spatial location i .

2.2. Optimized Image Resizing

Assume we want to resize the image with width of m into the desired one with width of m' by using a set of k operators O_1, O_2, \dots, O_k and the similarity calculation from IR-SSIM for each retargeting iteration. There are $O(k^w)$ (where

$w = m - m'$) different multi-operator sequences that resize the original image into the expected result. The search space is exponential in the size change w . To speed up the search process, we resize the image by the width of C at each resizing iteration. Assume $g_{i-1} = O_1, O_2, \dots, O_{i-1}$ is the used operator set in the previous $i-1$ iterations, the operator at i th iteration O_i can be determined by the following function:

$$e = \arg \max_{1 \leq i \leq k} \text{IR-SSIM}(I, \langle g_{i-1} \cup O_i \rangle(I)) \quad (5)$$

where k represents the retargeting operator; I denotes the original image.

3. EXPERIMENTAL RESULTS

To validate the proposed method, we use a public benchmark image retargeting database [16] to conduct the experiment. The benchmark was collected by a user study from 210 participants based on database composed of 37 source reference images. Fig. 2 provides visual examples that demonstrate the iterative process of the proposed method. For images in this figure, the original images are gradually resized until the target size is reached, generating a series of retargeted images as mid-results. From this figure, we can see that the proposed method can preserve the salient regions in original images well during image resizing operation.

To demonstrate the performance of the proposed multi-operator retargeting method, we have conducted a comparison experiment by using the existing multi-operator method [14]. Fig. 3 gives some comparison samples from the proposed method and the multi-operator algorithm [14]. From the first row of this figure, we can see that the boy in the retargeted image from [14] suffer from much distortion, while it is well preserved in the retargeted result from the proposed method. Other two comparison samples also demonstrate that the retargeted results from the multi-operator [14] suffer from much distortion with the salient objects, while the proposed algorithm can obtain much better retargeted results.

4. CONCLUSION

We propose a multi-operator retargeting algorithm based on a perceptual similarity measure, which is designed based on the SIFT flow, visual saliency and SSIM index. The SIFT flow is adopted to calculate the dense correspondence between the original and retargeted images. SSIM index weighted by visual saliency is used to evaluate the similarity between the original and retargeted images to determine the operator used at each iteration. Three operators including seam carving, scaling, and cropping are used to resize the original images. Experimental results on a public database demonstrate the promising performance of the proposed method. In the future, we will further investigate the multi-operator retargeting by integrating other top-down factors such as specific objects.

5. ACKNOWLEDGE

This work was supported by the Research Foundation from Department of Education of Jiangxi Province (No.GJJ14347).

6. REFERENCES

- [1] S. Avidan, and A. Shamir. Seam carving for content-aware image resizing. *ACM Transactions on Graphics*, 26(3), 2007.
- [2] D. Vaquero, M. Turk, K. Pulli, M. Tico, and N. Gelfand. A survey of image retargeting techniques. in *SPIE Applications of Digital Image Processing XXXIII*, vol. 7798, 2010.
- [3] M. Rubinstein, A. Shamir, and S. Avidan, Improved seam carving for video retargeting. *ACM Transactions on Graphics*, 2008.
- [4] Y. Guo, F. Liu, J. Shi, Z.-H. Zhou, and M. Gleicher. Image retargeting using mesh parameterization. *IEEE Transactions on Multimedia*, 11(5), 856-867, 2009.
- [5] M. Grundmann, V. Kwatra, M. Han, and I. Essa. Discontinuous seam carving for video retargeting. *IEEE International Conference on Computer Vision and Pattern Recognition*, 2010.
- [6] L. Wolf, M. Guttman, and D. CohenOr. Non-homogeneous content driven video retargeting. *IEEE International Conference on Computer Vision*, 2007.
- [7] Y. Jin, L. Liu, and Q. Wu, Nonhomogeneous scaling optimization for realtime image resizing, *The Visual Computer*, vol. 26, no. 6-8, pp. 769-778, 2010.
- [8] T. S. Cho, M. Butman, S. Avidan, and W. T. Freeman, The patch transform and its applications to image editing, *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [9] J.-S. Kim, J.-H. Kim, and C.-S. Kim. Adaptive image and video retargeting technique based on Fourier analysis. *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [10] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee. Optimized scale-and-stretch for image resizing. *ACM Transactions on Graphics (Proc. of SIGGRAPH ASIA)*, 27(5), 2008.
- [11] Y. Pritch, E. Kav-Venaki, and S. Peleg. Shift-map image editing. in Proc. *IEEE International Conference on Computer Vision*, 2009.
- [12] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (Proc. of SIGGRAPH)*, 28(3), 2009.
- [13] Y. Fang, Z. Chen, W. Lin, and C. Lin. Saliency detection in the compressed domain for adaptive image retargeting. *IEEE Transactions on Image Processing*, 21(9): 3888-3901, 2012.
- [14] M. Rubinstein, A. Shamir, and S. Avidan. Multi-operator media retargeting. *ACM Transactions on Graphics*, 28(3): 1-11, 2009.
- [15] W. Dong, N. Zhou, J. C. Paul, and X. Zhang. Optimized image resizing using seam carving and scaling. *ACM Transactions on Graphics*, 28(5): 1-10, 2009.
- [16] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir. A comparative study of image retargeting. *ACM SIGGRAPH Asia*, 29(6): 1-9, 2010.
- [17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13 (4): 600-612, 2004.
- [18] Y. Fang, K. Zeng, Z. Wang, W. Lin, Z. Fang, and C.-W. Lin. Objective Quality Assessment for Image Retargeting Based on Structural Similarity. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, To appear in 2014.
- [19] C. Liu, J. Yuen and A. Torralba. SIFT flow: dense correspondence across different scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5), 2011.
- [20] D. G. Lowe. Object recognition from local scale-invariant features. *IEEE International Conference on Computer Vision*, 1999.
- [21] Z. Wang, and A. C. Bovik. Mean squared error: love it or leave it? - a new look at fidelity measures. *IEEE Signal Processing Magazine*, 2009.
- [22] P. A. Viola, and M. J. Jones, Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2): 137-154, 2004.
- [23] S. Castillo, T. Judd, and D. Gutierrez. Using eye-tracking to assess different image retargeting methods. *ACM SIGGRAPH Symp. Applied Perception in Graphics and Visualization*, 2011.