IMAGE AND VIDEO RETARGETING USING ADAPTIVE SCALING FUNCTION

Jin-Hwan Kim, Jun-Seong Kim, and Chang-Su Kim

Media Communications Laboratory, School of Electrical Engineering, Korea University, Seoul, Korea
e-mail: {arite, junssi153, changsukim}@korea.ac.kr

ABSTRACT

An image and video retargeting algorithm using an adaptive scaling function is proposed in this work. We first construct an importance map which uses multiple features: gradient, saliency, and motion difference. Then, we determine an adaptive scaling function, which represents a scaling factor of each column in the source image. Finally, the target image is constructed with a weighted average filter using those scaling factors. Moreover, we extend this algorithm to video sequence. Simulation results demonstrate that the proposed algorithm provides better results than conventional retargeting methods.

1. INTRODUCTION

Recently, as users can access multimedia contents with various devices, including mobile phones, portable multimedia players, and televisions, the demands for effective resizing techniques have been increased. For example, movie contents are often manufactured with an aspect ratio of 2.35:1, but may be consumed on multimedia display devices with different aspect ratios such as 4:3 or 16:9. Retargeting methods are employed to fit the sizes of contents into those of devices. Conventional approaches include the scaling, cropping, and letter box methods. However, when an image is scaled, object shapes can be distorted if the aspect ratios of the original and retargeted images are different. The letter box method preserves the aspect ratio, but it makes it difficult to perceive objects when the target display is small. The cropping method also has a problem that the visual information of cropped regions is lost entirely. Figure 1 illustrates the scaling and cropping methods.

Several algorithms have been proposed recently to overcome these limitations and resize images and videos in a content-aware manner. Liu and Gleicher [1] proposed an image retargeting algorithm using fish-eye view warping. Their algorithm detects a region of interest (ROI) based on a saliency map and face detection results, and then warps the region outside the ROI while preserving the ROI. It is simple but causes distortions in the warped region, yielding unnatural images. Avidan and Shamir [2] proposed the seam carving algorithm, which finds a monotonic and connective path, called seam, that is the least noticeable. Seams are carved out iteratively, until the remaining image has the target size. In [3], the seam carving is extended to video retargeting using two-dimensional seam manifolds. The seam carving provides impressive results, but it also has limitations. When a target size is too small, important objects are carved out and the image becomes distorted. To avoid carving out important regions, a hybrid algorithm, which switches modes between the seam carving and the conventional scaling, is proposed in [4]. Recently, Kim *et al.* [5] proposed a retargeting algorithm based on Fourier analysis. It first divides an image into strips according to image contents. Then, it scales each strip differently to minimize the sum of distortions, which are modeled in the frequency domain.

This research was supported by the Ministry of Knowledge Economy, Korea, under the Information Technology Research Center support program supervised by the Institute for Information Technology Advancement (IITA-2009-C1090-0902-0017).

(a) (b) (c)

Figure 1: An original image in (a) is retargeted using (b) the scaling method and (c) the cropping method.

Liu and Gleicher [6] proposed an adaptive video cropping algorithm, which moves a cropping window based on image saliency, motion saliency, and face detection results. Deselaers *et al.* [7] improved the adaptive cropping by employing zooming operations as well. Wolf *et al.* [8] described the retargeting process from a source image to a target image as a system of linear equations and solved the system in the least square manner. Their algorithm also uses local saliency, face detection, and motion detection results to define the system of equations.

In this work, we propose an image and video retargeting algorithm. The proposed algorithm first computes an importance map based on gradient, saliency, and motion difference features. Then, it determines the scaling factor of each column adaptively so that more important columns are preserved, while less important columns are downsampled. The target image is constructed using a weighted average filter, which employs the adaptive scaling factors as weights. Simulation results demonstrate that the proposed algorithm resizes images and video more effectively than the conventional algorithms.

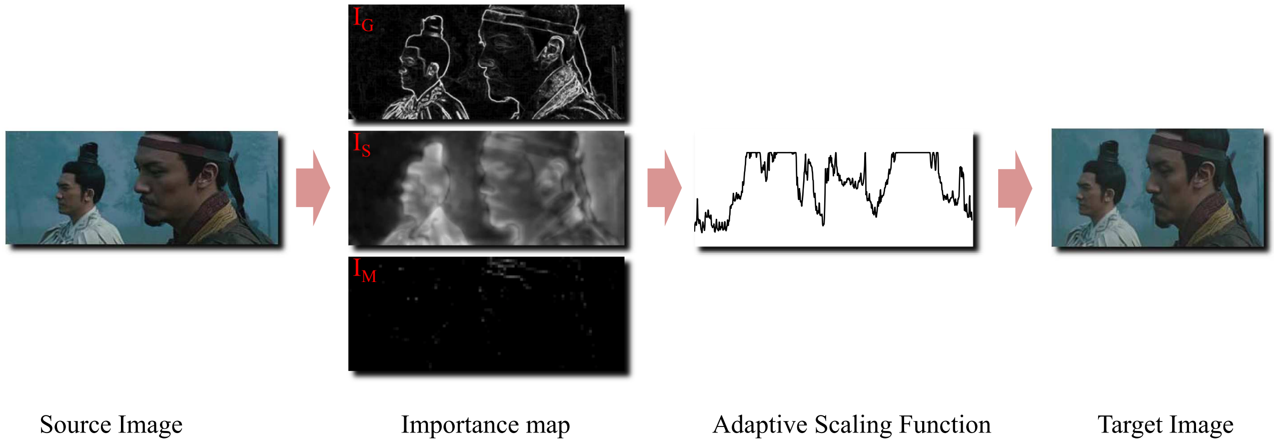


Figure 2: An overview of the proposed algorithm. The proposed algorithm computes the adaptive scaling function based on the importance map, and then constructs the target image using a weighted average filter.

The paper is organized as follows. Section 2 describes the proposed algorithm, and Section 3 provides retargeting results in comparison with the conventional algorithms. Finally, Section 4 concludes the paper and discusses future work.

2. PROPOSED ALGORITHM

Figure 2 shows an overview of the proposed algorithm. First, the proposed algorithm extracts an importance map, describing the regional importance of the source image. Second, based on the importance map, the proposed algorithm computes an adaptive scaling function. Third, the proposed algorithm constructs the target image using a weighted average filter.

In this section, for the sake of simplicity, we assume that a source image of width W_{in} is resized in the horizontal direction only to make a target image of width W_{out}, where $W_{\text{out}} < W_{\text{in}}$. However, the extension to vertical resizing is straightforward.

2.1 Importance Map

The importance map I is defined as a weighted sum of three feature maps: gradient map I_G, saliency map I_S, and motion difference map I_M.

$$I = w_G I_G + w_S I_S + w_M I_M, \tag{1}$$

where w_G, w_I, and w_M are weighting parameters. In this work, those parameters are fixed equally to 1/3. Figure 3 illustrates how these three maps compose the importance map.

2.1.1 Gradient Map

The human visual system is more sensitive to complex regions containing edges than to flat regions. Therefore, we extract a gradient map from the source image to represent the edge information. We acquire the gradient map from the gradient magnitude of each pixel, given by

$$\|\nabla F(x,y)\| = \sqrt{\left(\frac{\partial}{\partial x}F(x,y)\right)^2 + \left(\frac{\partial}{\partial y}F(x,y)\right)^2} \tag{2}$$

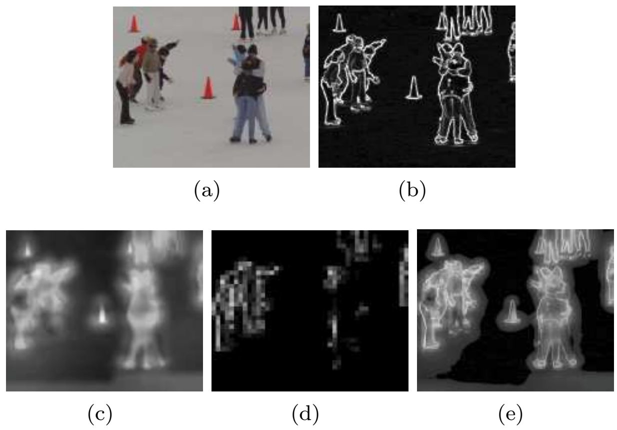


Figure 3: An example of importance map: (a) input frame, (b) gradient map, (c) saliency map, (d) motion difference map, and (e) importance map.

where the partial derivatives are approximated by the Sobel operators.

2.1.2 Saliency Map

We also use a saliency map, which has been proposed in various forms [9, 10, 11]. Ma and Zhang [9] used contrast information to extract a saliency map, Hou and Zhang [10] used the log-spectrum, and Itti *et al.* [11] used luminance, color, and orientation features. In this work, we adopt the Itti *et al.*'s algorithm, in which feature differences are computed in multiple scales with a Gaussian pyramid. Then, the differences are combined to construct the saliency map.

2.1.3 Motion Difference Map

In the case of video signals, the human visual system is also sensitive to object motions. Thus, we detect object motions and assign higher importance values to moving objects. For computational simplicity, we obtain frame differences instead of estimating the optical flow. In other words, absolute pixel differences between two adjacent frames represent motion activities in this work.

2.2 Adaptive Scaling Function

Using the importance map, we derive an adaptive scaling function $s(x)$, which represents the scaling factor of the xth column in the source image.

2.2.1 Initialization

First, we add up the importance values within each column of the importance map I by

$$I(x) = \sum_{y} I(x,y). \tag{3}$$

The column sum $I(x)$ represents the importance of the xth column. Thus, the scaling factor $s(x)$ of the xth column should be proportional to the column sum $I(x)$. Thus, it is initialized by

$$s_i(x) = \frac{I(x)}{\sum_x I(x)} W_{\text{out}}, \tag{4}$$

where W_{out} denotes the width of the target image.

2.2.2 Normalization

The initial scaling factor $s_i(x)$ may be greater than 1. However, in this work, we assume that the target image has a narrower width than the source image. Therefore, the scaling factor should be normalized to have a value between 0 and 1.

If we simply normalize all initial scaling factors by dividing them by the maximum factor, the sum of all normalized factors may not be equal to the target width. Therefore, in this work, we normalize the initial scaling factors by

$$s_n(x) = \begin{cases} s_i(x)^{\gamma} & \text{if } s_i(x) < 1 \\ 1 & \text{if } s_i(x) \geq 1 \end{cases} \tag{5}$$

where γ is a variable to be set such that

$$\sum_{x} s_n(x) = W_{\text{out}}.$$

2.2.3 Refinement

Next, we refine the normalized scaling factors to obtain the final scaling factors $s(x)$. Specifically, the factors are enhanced so that a large factor becomes even larger, whereas a small factor becomes smaller.

$$s(x) = \begin{cases} (1-\theta)\left(\frac{s_n(x)-\theta}{1-\theta}\right)^{\frac{1}{\beta^{(1/\theta)}}} + \theta & \text{if } s_n(x) > \theta \\ \theta\left[1 - \left(\frac{-s_n(x)+\theta}{\theta}\right)^{\frac{1}{\beta^{(1-\theta)}}}\right] & \text{if } s_n(x) \leq \theta \end{cases} \tag{6}$$

where the threshold θ is selected to satisfy the constraint $\sum_x s(x) = W_{\text{out}}$, and β is a controllable parameter that determines the shape of the refinement curve. In this work, β is fixed to 1.4. Figure 4 shows an example of the refinement curve, when $\theta = 0.5$ and $\beta = 1.4$. We see that the scaling factors are amplified if $s_n(x) > \theta$, and reduced otherwise.

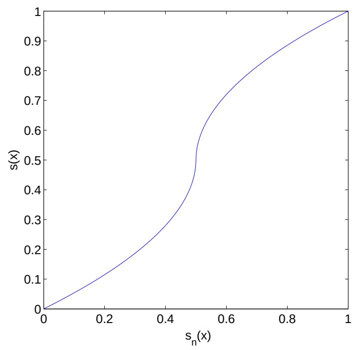


Figure 4: A refinement curve, when $\theta = 0.5$ and $\beta = 1.4$.

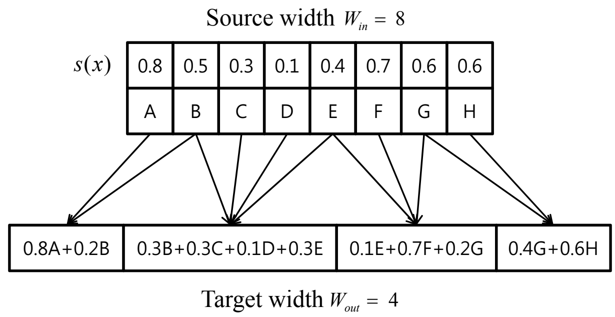


Figure 5: Target image generation using scaling factors. Each target pixel is a weighted sum of source pixels, and the weights come from the scaling factors. If a scaling factor is not consumed up in a pixel, then the remaining value is used for the next pixel also.

2.3 Target Image Generation

Given the scaling factor $s(x)$ for each column in the source image, the proposed algorithm simply fills in each target pixel with a weighted sum of source pixels, where the weights come from the scaling factors. If a scaling factor is not used up for a pixel, the remaining value is used for the next pixel also.

For example, in Figure 5, suppose that a row in the source image has 8 pixels and that we generate a target row of 4 pixels. The first pixel in the target row is filled in with the weighted sum of A and B, where the weights come from the scaling factors $s(x)$. The whole scaling factor for B is not consumed up yet, thus B is also used to generate the next pixel in the target row. In this way, all target pixels are filled in.

2.4 Video Retargeting

In video retargeting, if each frame is resized independently, the resultant target video sequence may yield severe jittering artifacts. To suppress jittering artifacts, we enforce mooth variation between the scaling functions of adjacent frames.

Let $s_k(x)$ denote the scaling function of the kth frame, which is computed independently of the other frames as described in Section 2.2. Then, we obtain a new scaling function $s'_k(x)$ of each frame sequentially

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

Figure 6: An original image in (a) is resized by (b) the scaling method, (c) the cropping method, and (d) the proposed algorithm.

using the scaling function of the previous frame by

$$s'_k(x) = \omega s_k(x) + (1-\omega)s'_{k-1}(x), \tag{7}$$

where ω is a renewal weight given by

$$\omega = \frac{\sum_x \|s_k(x) - s'_{k-1}(x)\|}{2(W_{\text{in}} - W_{\text{out}})}. \tag{8}$$

The renewal weight ω is proportional to the scaling factor difference between adjacent frames. By suppressing drastic variations of scaling functions, the proposed algorithm can provide temporally coherent video retargeting results.

3. EXPERIMENTAL RESULTS

Figure 6 compares the proposed algorithm with the standard scaling and cropping methods. We see that the proposed algorithm preserves the important region, a Japanese traditional building, more effectively than the scaling method. Moreover, the proposed algorithm contains most visual contents in the original image, including leaves and flowers, which are discarded in the cropping method.

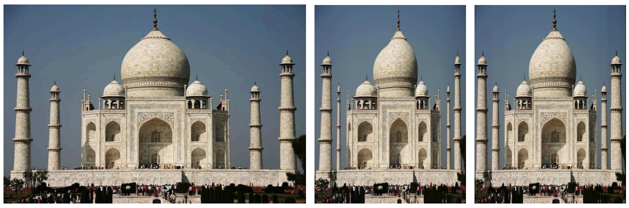
Figure 7 compares the proposed algorithm with the seam carving with forward energy [3]. Although the seam carving algorithm provides a natural rendering of the scene, it makes the main object, a black human figure, thinner and distorted. On the other hand, the proposed algorithm preserves the shape of the main object more faithfully.

Figure 8 also compares the proposed algorithm with Kim *et al.*'s algorithm[5]. On this image, the proposed algorithm provides more symmetrical and visually pleasing result. The proposed algorithm can be regarded as an extreme case of Kim *et al.*'s algorithm, when each strip consists of a single column of pixels. Therefore, the proposed algorithm can be more adaptive to image contents, but may distort object shapes. When there is an important object, Kim *et al.*'s algorithm can place it within a strip and thus can preserve its shape more reliably.

Figure 9 shows an example of changing the aspect ratio of the movie "Indiana Jones." The aspect ratio of the original movie is 2.38:1. We retarget the movie into the sizes of HDTV and SDTV, which have aspect ratios 16:9 and 4:3, respectively. We see that the proposed

(a) (b) (c)

Figure 7: Comparison of the proposed algorithm with the seam carving: (a) original image, (b) the seam carving with forward energy, and (c) the proposed algorithm.



(a) (b) (c)

Figure 8: Comparison with Kim *et al.*'s algorithm: (a) original image, (b) Kim *et al.*'s algorithm (c) the proposed algorithm.

algorithm preserves the vehicles faithfully, while scaling down less important regions. Thus, the proposed algorithm presents better results than the standard cropping and scaling methods.

Figure 10 compares the proposed algorithm with optimal cropping, which moves the cropping window to track the most salient region based on the Itti *et al.*'s saliency measure. In the bottom row, note that the cropping discards one of the characters, while the proposed algorithm preserves all three characters. The resultant video clips are available on the internet [12].

4. CONCLUSIONS AND FUTURE WORK

We proposed an algorithm for image and video retargeting, which preserves important regions while scaling down less important regions. The proposed algorithm first computes an importance map and an adaptive scaling function. Then, based on the adaptive scaling function, the target image is constructed from the source image with a weighted average filter. Experimental results demonstrated that the proposed algorithm provides better results than the conventional algorithms.

One of the future research issues is to extend the proposed algorithm so that the scaling operation can be applied in arbitrary directions, as well as horizontal and vertical directions. Also, another issue is to generalize the proposed algorithm for other applications such as image enlarging or object removal.

REFERENCES

[1] F. Liu and M. Gleicher, "Automatic image retargeting with fisheye-view warping," in *Proc. ACM UIST*, 2005, pp. 153–162.

Figure 9: Retargeting of a movie clip ("Indiana Jones" ©2008 Paramount). The original clip in the left side has an aspect ratio 2.38:1. It is resized to the aspect ratios of 16:9 and 4:3, respectively. From top to bottom, the cropping method, the scaling method, and the proposed algorithm.

(a) Input image (b) Cropping (c) Proposed algorithm

(d) Input image (e) Cropping (f) Proposed algorithm

Figure 10: Retargeting of movie clips ("Red Cliff" ©2008 CFGC and "Resident Evil" ©2008 CAPCOM).

[2] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graphics*, vol. 26, no. 3, July 2007.

[3] M. Rubinstein, S. Avidan, and A. Shamir, "Improved seam carving for video retargeting," *ACM Trans. Graphics*, vol. 27, no. 3, Aug. 2008.

[4] D. Hwang and S. Chien, "Content-aware image resizing using perceptual seam carving with human attention model ," in *Proc. ICME*, 2008, pp. 1029–1032.

[5] J.-S. Kim, J.-H. Kim, and C.-S. Kim, "Adaptive image and video retargeting technique based on Fourier analysis," in *Proc. CVPR*, June 2009

[6] F. Liu and M. Gleicher, "Video retargeting: automating pan and scan," in *Proc. ACM Multimedia*, Oct. 2005, pp. 241–250.

[7] T. Deselaers, P. Dreuw, and H. Ney, "Pan, zoom, scan – Time-coherent, tranined automatic video cropping ," in *Proc. CVPR*, 2008, pp. 1–8.

[8] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proc. ICCV*, Oct. 2007, pp. 1–6.

[9] Y. Ma and H. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. ACM Multimedia*, 2003, pp.374–381.

[10] X. Hou and L. Zhang, "Saliency detection:A spectral residual approach," in *Proc. CVPR*, 2007, pp.1–8.

[11] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11. pp. 1254–1259, Nov. 1998.

[12] Supplementary video clips, http://mcl.korea.ac.kr/Eusipco2009Result, Feb. 2009.