

Adaptive Image and Video Retargeting Technique Based on Fourier Analysis

Jun-Seong Kim, Jin-Hwan Kim, and Chang-Su Kim
School of Electrical Engineering, Korea University, Seoul, Korea
{junssi153, arite, changsukim}@korea.ac.kr

Abstract

An adaptive image and video retargeting algorithm based on Fourier analysis is proposed in this work. We first divide an input image into several strips using the gradient information so that each strip consists of textures of similar complexities. Then, we scale each strip adaptively according to its importance measure. More specifically, the distortions, generated by the scaling procedure, are formulated in the frequency domain using the Fourier transform. Then, the objective is to determine the sizes of scaled strips to minimize the sum of distortions, subject to the constraint that the sum of their sizes should equal the size of the target output image. We solve this constrained optimization problem using the Lagrangian multiplier technique. Moreover, we extend the approach to the retargeting of video sequences. Simulation results demonstrate that the proposed algorithm provides reliable retargeting performance efficiently.

1. Introduction

Image and video contents are nowadays consumed on various multimedia devices with different display sizes and aspect ratios. For example, high definition television contents and cinema films are often watched on cellular phones or portable multimedia players with small screens. In such cases, image contents should be resized appropriately before the rendering.

Scaling and cropping are two standard techniques for resizing images. Scaling reduces the sampling rate uniformly over a whole image. It does not discard any regions during the resizing, but it causes anisotropic stretching when the aspect ratio is changed or shrinks important objects too much when the target screen is very small. On the other hand, cropping discards boundary regions while preserving important objects. But the information in the carved regions is lost entirely. Figure 1 illustrates how the two techniques reduce the horizontal resolution of an input image. We see that scaling causes unnatural stretching of the tower, while cropping carves out the cathedral that is an integral part of the photograph.

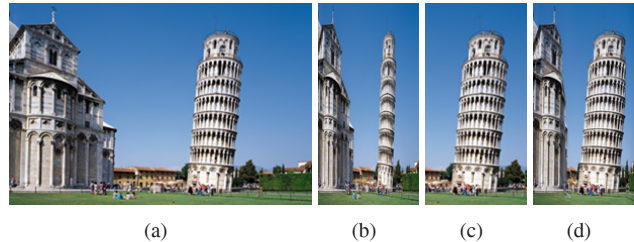


Figure 1. Image retargeting. An input image in (a) is resized by (b) the scaling technique, (c) the cropping technique, and (d) the proposed algorithm, respectively.

Recently, it has drawn much attention to develop content-aware image and video resizing techniques, also called retargeting techniques, which combine the merits of both scaling and cropping. Retargeting attempts to preserve important regions, while scaling down less important regions, to achieve a target image size. In this work, we propose a divide-and-conquer approach to the retargeting of images and video sequences. The proposed algorithm first divides an input image into several strips and then scales each strip adaptively. To protect visual contents as faithfully as possible, we analyze the scaling distortions based on the Fourier transform and formulate the resizing task as a constrained optimization problem, which is solved using the Lagrangian multiplier technique. Extensive simulation results show that the proposed algorithm resizes images in a content-aware manner more reliably, while demanding much lower computational complexity, than the conventional algorithms [2, 10].

The paper is organized as follows. Section 2 surveys previous retargeting algorithms. Section 3 describes the proposed algorithm. Section 4 extends the proposed algorithm to video retargeting. Section 5 presents experimental results. Section 6 concludes the paper and discusses future research issues.

2. Previous work

Recently, various retargeting algorithms have been proposed. Suh *et al.* [15] proposed a cropping algorithm for creating thumbnail images, which chooses cropped regions

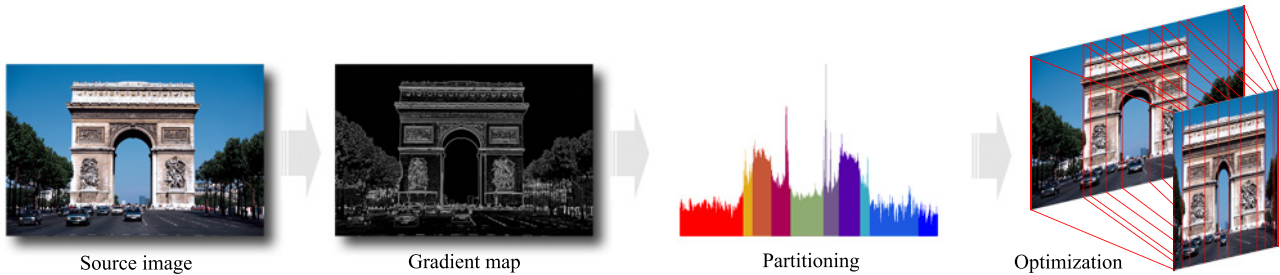


Figure 2. An overview of the proposed algorithm. The proposed algorithm partitions an input image based on the gradient map, and solves an optimization problem to decide the scaling factor for each strip.

based on the saliency map [6] and the face detection [18]. Similarly, Chen *et al.* [3] presented a system for adapting images to mobile devices. They considered a text attention model as well as saliency and face attention models to locate perceptually important regions.

Liu and Gleicher [7] proposed an image retargeting algorithm, which determines a region of interest (ROI) and then applies a fisheye-view warping to achieve a target image size. Their algorithm is simple, but the warping may cause distortions that look unnatural. Setlur *et al.* [11, 12] proposed a segmentation-based retargeting method. It segments an image into several regions, identifies important ROIs, cuts the ROIs from the image, and fills the holes with an inpainting scheme. Then, after scaling the background image, it pastes the ROIs back to the image. The quality of the resized image, however, depends on the accuracy of the segmentation, which is difficult to be done automatically.

Avidan and Shamir [2] proposed the seam carving algorithm, which removes a connected path of pixels, called a seam, from an input image repeatedly to achieve a target size. A dynamic programming method is used to find the least noticeable seam at each repetition. The seam carving provides impressive results, but it also has limitations. When an image is shrunk too much, it starts to carve out important objects, yielding unnatural artifacts. Also, when complex objects are scattered over the whole image, it is difficult to find unnoticeable seams. In [10], Rubinstein *et al.* proposed an improved energy function for the seam carving to obtain better retargeting results. They also generalized the seam carving to video retargeting by employing two-dimensional seam manifolds. In [5], Hwang and Chien proposed to use a hybrid of the seam carving and the traditional scaling. When the energy of a carved seam becomes greater than a threshold, their algorithm switches to the traditional scaling.

Pan-and-scan is a procedure to crop off the sides of widescreen films to fit them into television screens of 4:3 aspect ratio, which is often done manually. Liu and Gleicher [8] proposed a video retargeting algorithm, which automates the pan-and-scan procedure. Their algorithm ob-

tains the importance map for each frame and determines the cropping window for each frame to generate the effects of virtual pans and cuts. Tao *et al.* [16] extended the automated pan-and-scan procedure by allowing the rotation of the cropping window. Also, Deselaers *et al.* [4] introduced the zooming operation in addition to the pan-and-scan operation and proposed a dynamic programming method to provide temporally coherent retargeting results.

Wolf *et al.* [17] proposed a video retargeting algorithm, which models the mapping from an input image to a target image as a system of linear equations. Given an input image, their algorithm first computes the importance of each pixel. Then, based on the importance map, it forms the system of equations and obtains the least squares solution. Recently, Simakov *et al.* [14] presented an image and video summarization algorithm based on the bidirectional similarity measure and demonstrated that their algorithm can be employed in various applications such as retargeting, image montage, and automated cropping.

3. Proposed algorithm

Figure 2 illustrates how the proposed algorithm resizes a source image. After computing the gradient map, the proposed algorithm partitions the source image into multiple strips. Then, it models the scaling distortion of each strip and solves a constrained optimization problem to resize each strip and obtain the target image.

For the sake of simplicity, in this section, we assume that the source image is resized in the horizontal direction only. Specifically, we assume that a source image of size $W_s \times H_s$ is reduced to a target image of size $W_t \times H_t$, where $W_t < W_s$ and $H_t = H_s$. However, the proposed algorithm can be generalized straightforwardly to vertical resizing also.

3.1. Partitioning

We attempt to partition a source image I to several strips according to the complexity of regions. To measure the complexity of pixel $I(x, y)$, we employ its gradient mag-

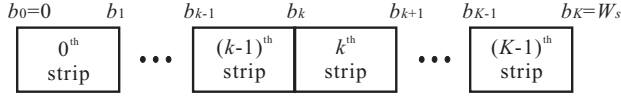


Figure 3. Partitioning of an image into K strips.

nitide

$$\|\nabla I(x, y)\| = \sqrt{\left(\frac{\partial}{\partial x} I(x, y)\right)^2 + \left(\frac{\partial}{\partial y} I(x, y)\right)^2} \quad (1)$$

where the partial derivatives are approximated by the Sobel operators. Then, we compute the complexity of the x th column by summing up the complexities of pixels in the column, *i.e.*,

$$c(x) = \sum_y \|\nabla I(x, y)\|. \quad (2)$$

The goal of our retargeting algorithm is to preserve the shapes of important objects, which are surrounded by strong edges. Thus, when we measure the column complexity in (2), we sort the gradient magnitudes of pixels and sum up only the top 10% of magnitudes, which correspond to edge pixels, instead of all magnitudes in the column.

Using the column complexities, we partition the image into K strips as shown in Figure 3. Let b_k denote the coordinate of the leftmost column in the k th strip, where $0 \leq k \leq K - 1$. Note that b_0 is fixed to 0 and b_K is fixed to W_s . Then, the partitioning problem is to determine inner boundaries b_k for $1 \leq k \leq K - 1$. Initially, the image is divided into strips of the same size, *i.e.*, $b_k = W_s \frac{k}{K}$.

Since the proposed algorithm downsamples each strip uniformly, it is desirable for a strip to consist of columns of similar complexities. Moreover, if the partitioning is successful, adjacent strips should have different complexities. These two criteria are used to decide the partitioning. Suppose that, in Figure 3, the left boundary b_{k-1} and the right boundary b_{k+1} are fixed. Then, we update the k th boundary b_k by

$$b_k = \arg \min_{b_{k-1} + \alpha \leq b < b_{k+1} - \alpha} \left(\sum_{x=b_{k-1}}^{b-1} |c(x) - s_{k-1}| + \sum_{x=b}^{b_{k+1}-1} |c(x) - s_k| + \beta \frac{1}{|s_{k-1} - s_k|} \right) \quad (3)$$

where s_{k-1} and s_k denote the averages of the column complexities in the $(k-1)$ th strip and the k th strip, given by

$$s_{k-1} = \frac{\sum_{x=b_{k-1}}^{b-1} c(x)}{b - b_{k-1}} \quad \text{and} \quad s_k = \frac{\sum_{x=b}^{b_{k+1}-1} c(x)}{b_{k+1} - b}. \quad (4)$$

In (3), the first two terms represent the variations of complexities within the $(k-1)$ th and the k th strips, respectively,

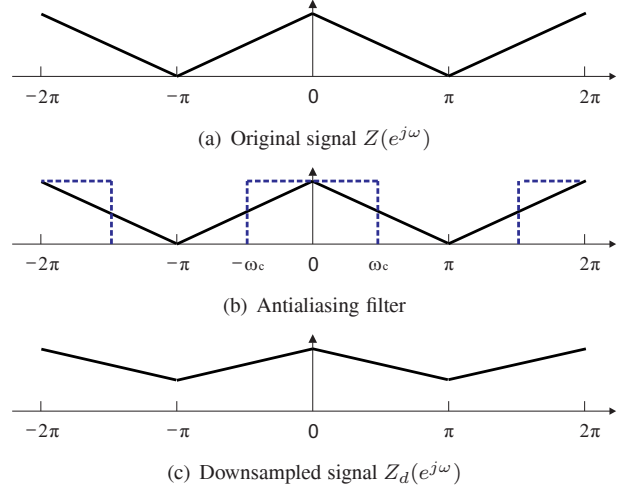


Figure 4. The frequency-domain illustration of downsampling [9]. Before reducing the sampling rate, a signal $z[n]$, whose Fourier transform is $Z(e^{j\omega})$, is lowpass-filtered to avoid aliasing artifacts.

and the last term is the inverse of the complexity discrepancy between the two strips. Therefore, the boundary b_k is selected such that the complexity variation within each strip is low, while the variation across the strips is high. Also, in (3), α is a positive constant that prevents two strip boundaries from being too close to each other. In other words, α specifies the minimum width of a strip. β is a weighting coefficient. In this work, α is fixed to 5, β is fixed to 0.05, and the number of strips K is set to 10.

Next, together with b_{k+2} , the updated b_k is used to update the next boundary b_{k+1} in a similar way. This process is iteratively applied until all boundaries converge. Figure 2 shows an example of the partitioning result. We see that the image is divided into strips according to the complexities.

3.2. Formulating scaling distortions

After the partitioning, each strip is downsampled uniformly. Let $z[n]$ denote a row signal of length l_k in the k th strip. Suppose that we reduce the length by removing r_k pixels to obtain a downsampled signal $z_d[n]$. In other words, the sampling rate is reduced by a factor of $1 - \frac{r_k}{l_k}$. In such a case, as shown in Figure 4 (b), the signal should be lowpass-filtered with a cutoff frequency $\omega_c = (1 - \frac{r_k}{l_k})\pi$ to avoid aliasing artifacts [9]. Thus, the downsampling procedure suppresses high frequency components in the original signal $z[n]$, incurring detail losses inevitably. The energy of the lost high frequency components, which we call the scaling distortion of the k th strip, can be computed in the frequency domain by

$$d_k = 2 \int_{(1 - \frac{r_k}{l_k})\pi}^{\pi} |Z(e^{j\omega})|^2 d\omega. \quad (5)$$

Next, we model the shape of the Fourier transform of an

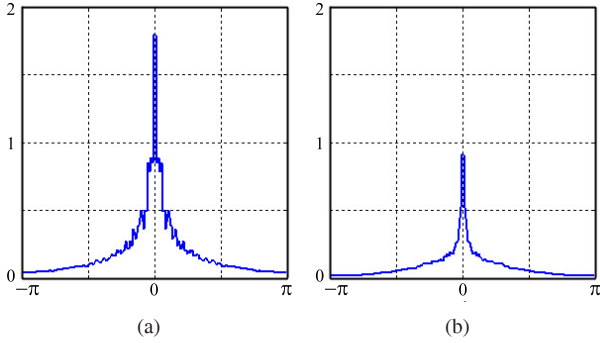


Figure 5. The average magnitudes of the Fourier transforms of (a) the “Leaning Tower of Pisa” image and (b) the “Triumphal Arch” image.

image signal. Since a typical image signal is a lowpass signal, $|Z(e^{j\omega})|$ decreases quickly as ω increases. We observe that, except for the dominant DC component $|Z(e^{j0})|$, the magnitude of the Fourier transform can be effectively approximated by an exponential function

$$|Z(e^{j\omega})| = e^{-\frac{\omega}{s_k}} \quad (6)$$

where s_k controls the decreasing rate of the exponential function. Figure 5 plots the average magnitudes of the Fourier transforms of the rows in the “Leaning Tower of Pisa” image in Figure 1 and the “Triumphal Arch” image in Figure 2. Tests on various other images also confirmed that the magnitudes exponentially decrease as ω increases.

In (6), a smaller s_k makes the exponential function decrease more quickly, which corresponds to the case of a smoother image signal. Therefore, s_k should be proportional to the complexity of the k th strip. In this work, each strip complexity s_k is first computed by averaging the column complexities within the strip as in (4), and then normalized by the average of all strip complexities. Then, by inserting (6) into (5), we obtain the scaling distortion of the k th strip, which is a function of r_k given by

$$d_k(r_k) = 2 \int_{(1-\frac{r_k}{l_k})\pi}^{\pi} e^{-\frac{2\omega}{s_k}} d\omega = s_k e^{-\frac{2\pi}{s_k} (e^{\frac{2\pi r_k}{s_k l_k}} - 1)}. \quad (7)$$

3.3. Adaptive scaling

Given the source image width W_s and the target image width W_t , $R = W_s - W_t$ columns should be removed from the source image. Therefore, the numbers r_k ($0 \leq k \leq K - 1$) of reduced columns from strips should satisfy the constraint

$$\sum_{k=0}^{K-1} r_k = R. \quad (8)$$

Table 1. The size l_k , the complexity s_k , and the number of reduced columns r_k for each strip of the “Triumphal Arch” image in Figure 2.

k	0	1	2	3	4	5	6	7	8	9
l_k	222	31	66	68	115	54	70	36	172	66
s_k	0.68	1.40	1.79	1.37	0.92	1.36	1.76	1.27	0.65	0.50
r_k	168	2	1	17	66	11	3	7	127	48

Subject to this constraint, the objective is to minimize the sum of the distortions of strips

$$\sum_{k=0}^{K-1} d_k(r_k). \quad (9)$$

This is a constrained optimization problem, which can be solved by minimizing the Lagrangian cost function

$$J = \sum_k d_k(r_k) + \lambda \sum_k r_k \quad (10)$$

$$= \sum_k \left(s_k e^{-\frac{2\pi}{s_k} (e^{\frac{2\pi r_k}{s_k l_k}} - 1)} + \lambda r_k \right) \quad (11)$$

where λ is a Lagrangian multiplier. By setting the partial derivative $\frac{\partial J}{\partial r_k}$ to 0, we obtain

$$r_k = l_k \left(1 + \frac{s_k}{2\pi} \log \mu l_k \right) \quad (12)$$

where $\mu = -\frac{\lambda}{2\pi}$. Since $0 \leq r_k \leq l_k$, r_k in (12) is clipped to

$$r_k = \max \left\{ 0, \min \left\{ l_k, l_k \left(1 + \frac{s_k}{2\pi} \log \mu l_k \right) \right\} \right\}. \quad (13)$$

Next, we find μ so that r_k 's satisfy the equality in (8). Since each r_k in (13) is a monotonic increasing function of μ , the desired μ can be computed efficiently with the bisection search method [13].

Table 1 lists the size l_k , the complexity s_k , and the number of reduced columns r_k for each strip of the “Triumphal Arch” image in Figure 2. Note that the number of reduced columns is determined adaptively according to the size and the complexity.

After deciding r_k , the proposed algorithm reduces the sampling rate of the k th strip uniformly by a factor of $1 - \frac{r_k}{l_k}$. Figure 6 shows an example of retargeting result. When the scaling factors are $\frac{5}{6}$ and $\frac{2}{3}$, the shape of the flower is protected faithfully. When the scaling factor is further reduced to $\frac{1}{2}$, the flower is slightly squeezed horizontally but the re-sized image still looks natural without severe artifacts.

4. Extension to video retargeting

We extend the proposed algorithm to video retargeting. The simplest extension is to apply the image retargeting algorithm to each frame in a video sequence independently.



Figure 6. An input image on the left side is horizontally resized with scaling factors $\frac{5}{6}$, $\frac{2}{3}$, and $\frac{1}{2}$, respectively.

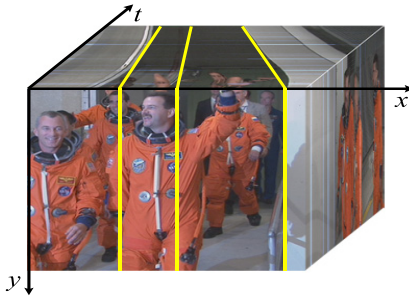


Figure 7. Partitioning of a video sequence for horizontal resizing.

Although this approach can resize each frame effectively, it cannot maintain temporal coherence, yielding annoying jitter artifacts.

An alternative approach is to apply the same partitioning and the same scaling to all frames in a video sequence. This can be achieved by treating the video sequence as a three-dimensional volume of pixels. Then, the volume is cut into parallelepipeds along the time axis, and each parallelepiped is scaled down spatially. For the partitioning and the scaling, the complexity of a parallelepiped can be defined as the sum of gradient magnitudes of the pixels inside the parallelepiped. However, similar to the static seam approach in [10], this extension may yield artifacts when the sequence contains fast object motions. Especially, an object may look unnatural, when it moves across two adjacent partitions that are downsampled with different scaling factors.

Therefore, we propose a partitioning scheme for video sequences, which takes into account object motions. Figure 7 illustrates the partitioning scheme, when a video sequence is horizontally resized. First, the whole volume is divided into two sub-volumes with a cutting plane, which is perpendicular to the $x-t$ plane. As in (3), the position and the slope of the cutting plane are determined to minimize the variation of complexities within each sub-volume and maximize the complexity discrepancy between the two sub-volumes. Each sub-volume is then recursively divided until the sequence is divided into a pre-specified number of sub-volumes.

After the partitioning, we reduce a number of planes,

which are perpendicular to the $x-t$ plane, from each sub-volume. The number of reduced planes from each sub-volume is computed by employing the Lagrangian optimization technique in Section 3.3. Notice that, within a sub-volume, the number of reduced columns is fixed over all frames and thus the scaling factors change monotonically. The proposed algorithm hence can provide temporally coherent retargeting results, without causing jitter artifacts.

5. Simulation result

Figures 1 and 8 compare the proposed algorithm with the standard scaling and cropping techniques. The proposed algorithm protects the shapes of important objects more faithfully than scaling. Furthermore, unlike cropping, the proposed algorithm does not carve out border regions entirely.

Figure 9 shows vertical resizing results, where the image height is reduced from 900 to 750, 600, 450, and 300, respectively. The proposed algorithm preserves the sizes of the flower and the butterfly, as long as they can fit into the target image. The proposed algorithm starts to scale them down, when there is no room for size reduction in the background. Figure 10 shows two examples, in which both the horizontal and the vertical sizes are halved.

Figure 11 compares the proposed algorithm with the seam carving [2], which incorporates the forward energy criterion [10] for performance improvement. The seam carving provides visually pleasing retargeting results up to a certain target width. However, if the width is further reduced, it starts to remove seams that cross foreground objects, causing severe distortions in the objects. Since seams have irregular shapes, the distortions are unpredictable. For example, in Figure 11 (a), the horse's legs are too distorted to convey the visual contents in the original image properly. On the other hand, when the target width is too narrow, the proposed algorithm downsamples the strips containing objects as well as those for the background. But, the downsampling within each strip is uniform, and the distortions are less annoying. In other words, as the target image size decreases, the proposed algorithm provides more graceful degradations than the seam carving.

Moreover, it is worthy to point out that the proposed

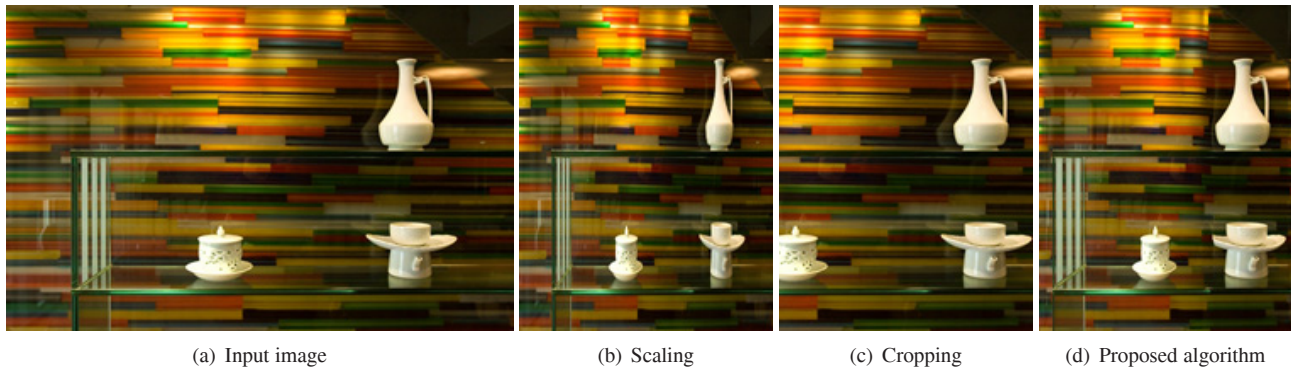


Figure 8. Comparison of the proposed algorithm with the standard scaling and cropping techniques.

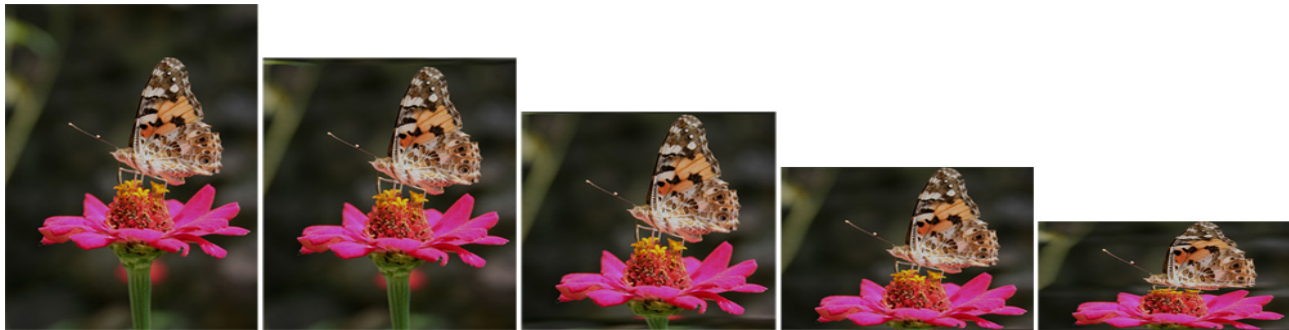


Figure 9. An input image on the left side is resized in the vertical direction with different scale factors.

algorithm requires much less computations than the seam carving. The seam carving requires a significant amount of computations to find optimal seams using dynamic programming. On the other hand, in the proposed algorithm, the main steps are the gradient map computation, the partitioning, the scaling factor decision, and the uniform scaling of each strip. The complexities for the partitioning and the scaling factor decision are negligible as compared with those for the gradient map computation and the uniform scaling. Thus, the computational complexity of the proposed algorithm is much lower than that of the seam carving. In fact, it is comparable with that of the standard scaling technique.

Figure 12 (a) shows a frame from the “Wall-E” movie clip and its resized results. The proposed algorithm preserves the robot more faithfully than the standard scaling technique. Figure 12 (b) shows similar results on the “Kung Fu Panda” clip. We have supplied retargeting results of various video clips as additional material [1]. The results exhibit excellent temporal coherence without jitter artifacts.

6. Conclusions and future work

We proposed an algorithm for image and video retargeting, which is computationally efficient but provides reliable

performance. The proposed algorithm consists of the partitioning and the scaling steps. In the partitioning, an image is divided into strips, so that the complexity variation within each strip is low. After the partitioning, a constrained optimization problem is solved to decide the scaling factor of each strip. Simulation results demonstrated that the proposed algorithm provides high image quality even when the target image size is very small.

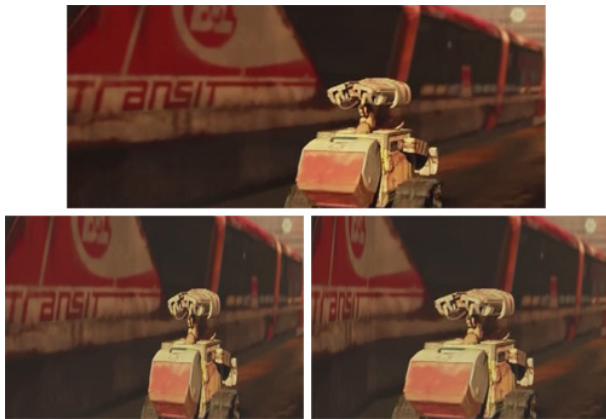
Future research issues include the development of more general partitioning schemes than the current rectangular division, which can adapt the shapes of partitions to scene contents to achieve better quality retargeting. Another issue is the sampling rate reduction in the temporal domain as well as in the spatial domain. Also, although we focused on the sampling rate reduction only, we will develop a sampling rate expansion scheme, which can interpolate images and videos in a content-aware manner.

Acknowledgements

This work was supported partly by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2008-331- D00420) and partly by the Ministry of Knowledge Economy, Korea, under the Information Technology Research Center support program supervised by the Institute of Information



(a) (b)
Figure 10. Resizing images in both horizontal and vertical directions.



(a)



(b)

Figure 12. Video clips from the “Wall-E” and “Kung Fu Panda” movies are resized. In each subfigure, the top is an original frame, the left is the result of the standard scaling technique, and the right is the result of the proposed algorithm.

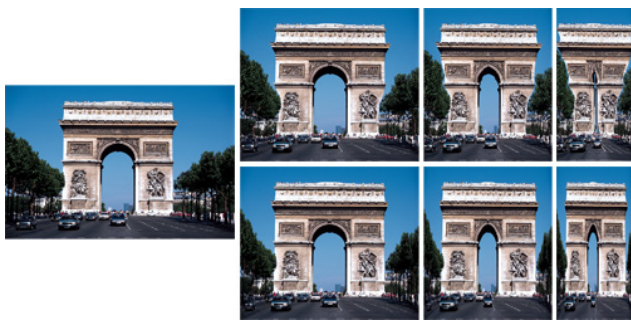
Technology Advancement (grant number IITA-2009-C1090-0902-0017).

References

- [1] Authors. Video retargeting results, 2008. Supplied as additional material.
- [2] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM Trans. Graphics*, 26(3), July 2007.
- [3] L. Q. Chen, X. Xie, X. Fan, W. Y. Ma, H. J. Zhang, and H. Q. Zhou. A visual attention model for adapting images on small displays. *ACM Multimedia Systems J.*, 9(4):353–364, 2003.
- [4] T. Deselaers, P. Dreuw, and H. Ney. Pan, zoom, scan – time-coherent, trained automatic video cropping. In *CVPR*, June 2008.
- [5] D. S. Hwang and S. Y. Chien. Content-aware image resizing using perceptual seam carving with human attention model. In *IEEE Int’l Conf. Multimedia & Expo*, June 2008.
- [6] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, Nov. 1998.
- [7] F. Liu and M. Gleicher. Automatic image retargeting with fisheye-view warping. In *ACM Symp. User Interface Software and Technology*, pages 153–162, 2005.
- [8] F. Liu and M. Gleicher. Video retargeting: Automating pan and scan. In *ACM Int’l Conf. Multimedia*, pages 241–250, 2006.
- [9] A. V. Oppenheim and R. W. Schaffer. *Discrete-Time Signal Processing*. Prentice Hall, 1989.
- [10] M. Rubinstein, A. Shamir, and S. Avidan. Improved seam carving for video retargeting. *ACM Trans. Graphics*, 27(3), Aug. 2008.
- [11] V. Setlur, T. Lechner, M. Nienhaus, and B. Gooch. Retargeting images and video for preserving information saliency. *IEEE Computer Graphics and Applications*, 27(5):80–88, Sep.–Oct. 2007.
- [12] V. Setlur, S. Takagi, R. Raskar, M. Gleicher, and B. Gooch. Automatic image retargeting. In *ACM Symp. User Interface Software and Technology*, pages 59–68, 2005.
- [13] Y. Shoham and A. Gersho. Efficient bit allocatin for an arbitrary set of quantizers. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 36(9), Sept. 1988.
- [14] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani. Summarizing visual data using bidirectional similarity. In *CVPR*, June 2008.
- [15] B. Suh, H. Ling, B. B. Bederson, and D. W. Jacobs. Automatic thumbnail cropping and its effectiveness. In *ACM Symp. User Interface Software and Technology*, pages 95–104, 2003.



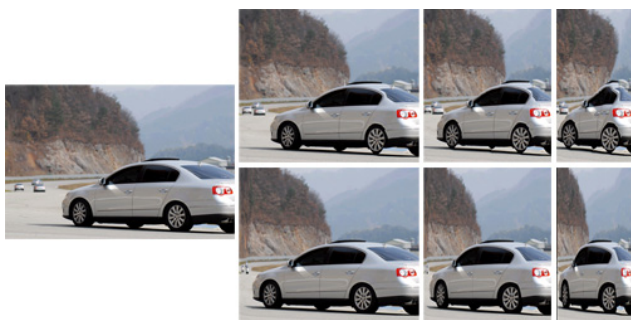
(a)



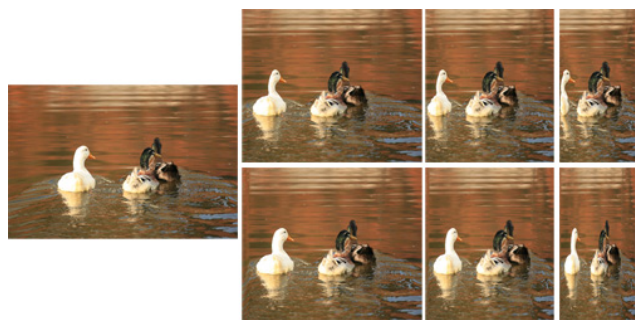
(b)



(c)



(d)



(e)

Figure 11. Comparison of the proposed algorithm with the seam carving. In each sub-figure, the top row is the result of the seam carving, and the bottom row is the result of the proposed algorithm.

[16] C. Tao, J. Jia, and H. Sun. Active window oriented dynamic video retargeting. In *ICCV Workshop on Dynamical Vision*, Oct. 2007.

[17] L. Wolf, M. Guttman, and D. Cohen-Or. Non-homogeneous content-driven video-retargeting. In *ICCV*, Oct. 2007.

[18] M. H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(1):34–58, Jan. 2002.