International Journal of Innovative
Computing, Information and Control
Volume 9, Number 4, April 2013

ICIC International ©2013 ISSN 1349-4198
pp. 1727–1736

NOVEL STEREO IMAGE AND VIDEO RETARGETING APPROACH

Chin-Chen Chang[1,*], Huei-Yung Lin[2,3] and Chia-Hao Hsieh[2]

[1]Department of Computer Science and Information Engineering
National United University
No. 1, Lienda, Miaoli 36003, Taiwan
*Corresponding author: ccchang@nuu.edu.tw

[2]Department of Electrical Engineering
[3]Advanced Institute of Manufacturing with High-Tech Innovations
National Chung Cheng University
No. 168, Sec. 1, University Rd., Min-Hsiung Township, Chia-yi County 621, Taiwan

Received February 2012; revised July 2012

Abstract. *In this paper, a novel stereo image and video retargeting approach is proposed. The proposed approach first extracts four feature maps: disparity map, gradient map, saliency map, and motion difference map. After that, the proposed approach constructs an importance map which combines the four feature maps by the weighted sum. Finally, the proposed approach constructs the target image using the seam carving method based on the importance map. Moreover, the proposed approach is extended to the retargeting of video sequences. Considering the images with time continuity, the proposed approach constructs an accumulated map by calculating the average of the importance maps for a period of time. The experimental results show that the proposed approach performs well in terms of the resized quality.*

Keywords: Image resizing, Image retargeting, Stereo image, Video retargeting, Seam carving, Feature map

1. **Introduction.** Numerous and varied devices for displaying multimedia contents exist, from CRTs to LCDs, and from plasma to LEDs. Display device has moved from the two-dimensional plane toward 3D TV. To meet various demands, changing display content has facilitated the development of a highly dynamic range of display devices. Regarding display screen size, two commonly used display specifications (aspect ratios) are 4:3 and 16:9.

Nonstandard screen aspect ratios will be applied more extensively because of cellular phones, portable multimedia players and so on. In such cases, different image sizes are required to adapt to the display devices. Scaling and cropping are two standard methods for resizing images. Scaling resizes the image uniformly over an entire image. However, when the display screen is too small, the image loses some of its detail in adjusting to the limitations of the display screen. Cropping resizes the image by discarding boundary regions and preserving important regions. This method provides a close-up of a particular image section, but prevents users from viewing the rest of the image.

Recently, several retargeting techniques [1,8,11,12] for resizing image based on image contents have been proposed. These methods require a certain understanding of image content and do not adjust the size of the image as a whole. Retargeting preserves important regions and discards less important regions to achieve a target image size. Since the creation of stereo images for a 3D display from the 2D images is important, developing techniques for stereo image retargeting is essential.

In this paper, a novel stereo image and video retargeting approach is proposed. The proposed approach first extracts four feature maps: disparity map, gradient map, saliency map, and motion difference map. After that, the proposed approach constructs an importance map which combines the four feature maps by the weighted sum. Based on the importance map, the important regions are preserved and less important regions are discarded. Finally, the proposed approach constructs the target image using the seam carving method [1] based on the importance map. Moreover, the proposed approach is extended to the retargeting of video sequences. Considering the images with time continuity, the proposed algorithm computes an accumulated map by calculating the average of the importance maps for a period of time. The experimental results show that the proposed approach resizes stereo images and videos effectively.

The remainder of this paper is organized as follows. Section 2 reviews related works. In Section 3, the proposed approach is introduced. Section 4 describes the experimental results. Lastly, Section 5 briefly describes conclusions.

2. **Related Works.** Avidan and Shamir [1] proposed a method for adjusting image size based on image content. They analyzed the relationships of energy distribution in the image and compared methods of image resizing. The proportion of residual energy after image resizing indicates the quality of the resizing. Moreover, they proposed a simple method for image processing using seams, which are 8-connected lines that vertically or horizontally cross images. By iteratively adding or removing seams, their approach can alter the size of images. However, because the content of images is often complex, how to determine the correct subject position according to image features is a goal for future research.

Kim et al. [11] used the adaptive scaling function utilizing the importance map of the image to calculate the adaptive scaling function for image and video retargeting. Kim et al. [12] used Fourier analysis for image resizing. After constructing the gradient map, they divided the image into strips of various lengths, and then used Fourier transform to determine the spectrum of each strip. The spectrums are then used as a low-pass filter to obtain an effect similar to smoothing. The level of horizontal reduction for each strip is then determined according to the influence of the filter.

Detecting visually salient areas is a part of object detection. The traditional method for determining the most conspicuous objects in an image is to set numerous parameters and then use the training approach to determine image regions that may correspond to the correct objects [2,3,9,14]. However, the human eye is capable of quickly locating common objects [20,25]. Various approaches have proposed for simulating the functions of the human eye; for instance, Saliency ToolBox [22] and Saliency Residual (SR) [7]. The Saliency ToolBox requires a large amount of computation. By comparison, SR is the fastest algorithm. SR transforms the image into Fourier space and determines the difference between the log spectrum and averaged spectrum of the image. The area, which shows the difference, is the potential area of visual saliency.

Hwang and Chien [8] used a neural network method to determine the subject of images. They also used face recognition techniques to ensure the human faces within images. For ratios that could not be compressed using the seam carving method, they used proportional ratio methods to compress the subject of images. Rubinstein et al. [18] proposed a method of improvement for the procedure of seam carving. This method utilizes techniques of forward energy and backward energy to reduce discontinuity in images.

Wang et al. [23] proposed a method that simultaneously utilized techniques of stereo imaging and inpainting. This method can remove image objects that caused occlusion, restoring original background image and depth information. Kim and Kim [10] presented

an approach for resizing a source image to fit the dimensions of arbitrary displays. They introduced an image importance model based on the local dominance for image resizing. Their approach can be applied to a wide range of image classes. To cope with some artifacts in the retargeted images, Wang and Lai [24] proposed a compressibility assessment approach for media retargeting by combining the entropies of image gradient magnitude and orientation distributions. The resized media is obtained to preserve the image content and structure.

3. **The Proposed Approach.** Figure 1 illustrates the flowchart of the proposed approach. First, the proposed approach extracts four feature maps, namely, disparity map, gradient map, saliency map, and motion difference map from input stereo images. After that, the proposed approach integrates all the feature maps to an importance map by the weighted sum. Finally, the proposed approach constructs the target image using the seam carving method [1]. Moreover, the proposed approach is extended to the retargeting of video sequences by using an accumulated map.

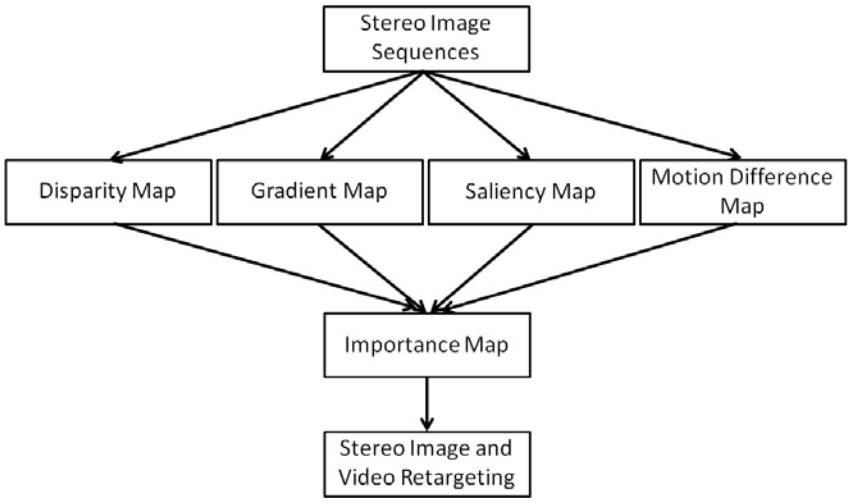

FIGURE 1. Flowchart of the proposed approach

3.1. **Importance map.** Several image features [1,3,4,7,9-12,14] have been extracted for image analysis. In the proposed approach, the importance map E_{imp} is defined as a weighted sum of the four feature maps as

$$E_{imp} = \alpha_1 E_{disparity} + \alpha_2 E_{gradient} + \alpha_3 E_{saliency} + \alpha_4 E_{diff}$$

where α_1, α_2, α_3 and α_4 are the weights for disparity map $E_{disparity}$, gradient map $E_{gradient}$, saliency map $E_{saliency}$, and motion difference map E_{diff}, respectively.

3.1.1. *Disparity map.* The proposed approach adopts the method of Konolige [13], which applies the sum of absolute difference (SAD) to find the corresponding blocks of the left-right images. The corresponding points are the centers of the corresponding blocks with the strongest SAD value. This method can be applied to any pair of stereo images [5,6] that have been rectified. The stereo matching method based on block matching is divided into three parts:

1. Preprocessing: balancing the image intensity and intensifying the texture features.
2. Using SAD window to find the corresponding points of the left-right images along the horizontal coaxial.
3. Post processing: removing the false corresponding points.

The balance of image intensity can reduce errors in SAD approximation and intensifying texture features can increase the accuracy of SAD in high texture regions. After finding the corresponding points, the proposed approach applies left-right consistency to remove all the wrong corresponding points, which results in the final disparity map as $E_{disparity}$.

However, some corresponding points could not be found by using block matching, which might be due to the inconspicuous texture around the points, occlusion, or matching errors resulting in some broken regions. The errors are divided into type I error and type II error. The cause of type I error is oversensitivity in the detector. Therefore, when the conforming probability of block matching is overset, the points that should not be corresponded occur. The cause of type II error is opposite that of type I's. Since the detector is not sensitive enough, the over loading of conforming probability brings out the removal of two corresponding blocks due to inconspicuous features.

From the abnormality on the disparity map, the errors include the following: isolated noise, broken edge, and broken object. The isolated noise results mostly from type I error. Delete the single object whose size is smaller than a certain threshold. Because the original definition of the cluster objects is based on the binary image, the original disparity map is binarized. Broken objects are caused by type II error. When the internal texture intensity of the object is low, no corresponding points are detected. The proposed approach calculates the number of every non-zero pixel in the range of its extension for distance n in east, west, south, and north direction, by centering on every zero pixel on the disparity map. When any range made by two directions has a value, the complement value of the pixel can be determined.

3.1.2. *Gradient map.* The Sobel calculation on original image I results in the gradient map. The operators of X direction and Y direction of Sobel are defined by

$$Sobel_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, \text{ and } Sobel_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}.$$

The horizontal operators, which are shown as a vertical line on the image, are used to find the horizontal gradient of the image, while the vertical operators, which are shown as a horizontal line, are used to find the vertical gradient of the image. Considering the direction of image resizing, the proposed approach offers weight individually to the Sobel horizontal and vertical directions as

$$E_{gradient} = \sqrt{w_h \cdot \left|\frac{\partial I}{\partial x}\right|^2 + (1 - w_h) \cdot \left|\frac{\partial I}{\partial y}\right|^2},$$

where w_h is the weight for the Sobel horizontal direction.

3.1.3. *Saliency map.* The computer vision [2,3,9,14] tries to imitate the possible visual perception of the human eye, from object detection, object classification to object recognition.

The most-studied feature of natural images is the invariant of extension and reduction, which is also called $1/f$ law [17,18]. After adding several images together and proceeding the fast Fourier transform, the amplitude $A(f)$ of the averaged Fourier spectrum are observed as

$$E\{A(f)\} \propto 1/f.$$

Based on the residual image in frequency space [4,7,21], Hou and Zhang [7] proposed a log spectrum representation to find the relation of polyline and visual features in log spaces. The method can rapidly detect conspicuous objects without extra references.

Given an image I, the intensity and the phase spectrum of the image can be obtained by transferring the image to Fourier space as

$$A(f) = \Re(F[I]), \text{ and } P(f) = \Im(F[I]),$$

where F denote the Fourier Transform. Let $L(f)$ be the log spectrum obtained after the reduction of the original image to one of $size_r \times size_r$ by Fourier transform as follows:

$$L(f) = \log(A(f)).$$

Therefore the spectral residual $R(f)$ can be obtained by

$$R(f) = L(f) - h_n(f) * L(f),$$

where $h_n(f)$ is an $n \times n$ matrix defined by

$$h_n(f) = \frac{1}{n^2} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix},$$

and $h_n(f) * L(f)$ is the averaged spectrum approximated by convoluting the input image.

Using inverse Fourier transform, the saliency map in spatial domain can be constructed as

$$E_{saliency} = g(x) * F^{-1}[\exp(R(f) + P(f))]^2,$$

where F^{-1} denotes the inverse Fourier transform and $g(x)$ is a Gaussian filter smoothing the saliency map.

3.1.4. *Difference map.* In the case of videos, the proposed approach detects object motions and assigns more importance to moving objects. The common way is the background subtraction in which the background could be generated by frame difference, Gaussian model [15], and Gaussian mixture model [19].

Providing a fixed camera and a background with low changes, the proposed approach applies the frame difference to extracting possible moving objects in the image. The frame difference first finds the reference image ($I_{reference}$), and then calculates the motion difference map (E_{diff}) according to each current image ($I_{current}$) as

$$E_{diff} = abs(I_{current} - I_{reference}).$$

3.2. **Image retargeting.** The proposed approach applies the method proposed by Avidan and Shamir [1] for image retargeting. Let I be an $n \times m$ image and the vertical seam is defined as

$$\mathbf{s^x} = \{s_i^x\}_{i=1}^n = \{(x(i), i)\}_{i=1}^n, \text{ s.t. } \forall i, \ |x(i) - x(i-1)| \leq 1,$$

where x is a mapping $x : [1, \dots, n] \to [1, \dots, m]$.

A vertical seam is an 8-connected line. Every row only contains a single pixel. Carving the seam interactively is considered an advantage because it can prevent horizontal displacement during the deleting process. Horizontal displacement appears if the number of deleted pixels in each row is different, resulting in changes in the shape of the object. Therefore, the pixels of the path of the vertical seam $\mathbf{s}$ (e.g., vertical seam $\{s_i\}$) is indicated as $\mathbf{I_s} = \{I(s_i)\}_{i=1}^n = \{I(x(i), i)\}_{i=1}^n$. All pixels will move leftward or upward to fill the gaps of deleted pixels.

Horizontal reduction can be equated with deleting the vertical seam; the energy map is used to select seams. Given an energy function e the energy $E(\mathbf{s}) = E(\mathbf{I_s}) = \sum_{i=1}^{n} e(\mathbf{I}(s_i))$ of a seam is determined by the energy occupied by the positions of each pixel. When

cutting a particular image horizontally, deleting the seam with the lowest energy $s^* = \min_{\mathbf{s}} E(\mathbf{s}) = \min_{\mathbf{s}} \sum_{i=1}^{n} e(\mathbf{I}(s_i))$ first is essential.

Dynamic programming can be employed to calculate s^*. The smallest accumulated energy M is calculated with every possible point on the seam (i, j) from the second to the last row of the image as

$$M(i,j) = e(i,j) + \min(M(i-1,j-1), M(i,j-1), M(i+1,j-1)).$$

Then, the backtracking method is adopted to iteratively delete the seams with relatively weak energy by gradually searching upward for the seams with a minimum energy sum from the point with the weakest energy in the last row.

3.3. **Video retargeting.** In video retargeting, if each frame is resized independently, the target video may result in jittering artifacts. Therefore, the proposed approach considers image continuity within a period of time when processing videos.

A simpler method involves using a sliding window, a W_{slide}. The width of the W_{slide} is set at the number of frames per second (fps), to improve the steadiness of the frame. Only gradual changes are made within one second. Indeed, the fps can be obtained to determine the size of the W_{slide}. Therefore, accounting for continuity of time when forming the E_{imp} into a sliding window, the accumulated map E_{acc} is defined as

$$E_{acc}(t) = \begin{cases} E_{imp}(0), & \text{if } t = 0 \\ \dfrac{\sum_{i=t-p}^{t+p} E_{imp}(i)}{2p}, & \text{if } t \geq 1 \end{cases},$$

where p is fps.

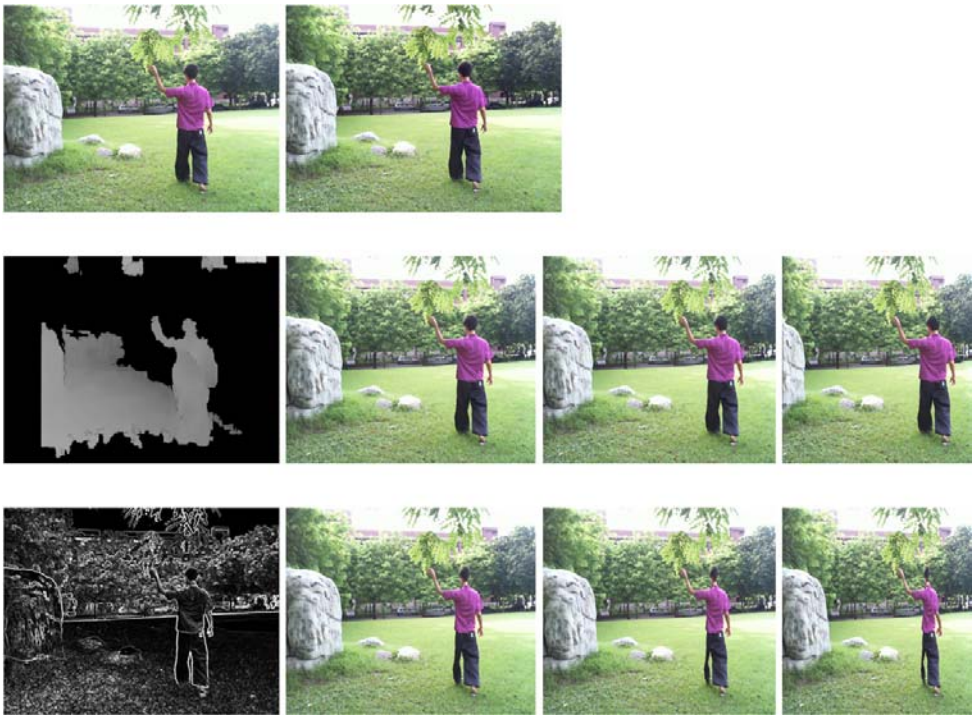

FIGURE 2. Original left-right images (the top row), the disparity map and resized results by the proposed approach (the middle row), and the standard gradient map and resized results by the seam carving approach (the bottom row)

4. **Results.** Two web cameras are connected to a digital camera binocular. The digital camera binocular is installed onto a laptop to create a dolly shot when the scene requires the camera to shift.

First, the proposed algorithm using the disparity map is compared with the seam carving incorporating the standard gradient map. Figure 2 shows the original left-right images (the top row), the disparity map and resized results by the proposed approach (the middle row), and the standard gradient map and resized results by the seam carving approach (the bottom row). Generally, objects which are closer are noticeable to human beings. The results significantly reveal the features of the disparity map, which is not determined by the textural information of the image. Therefore, the disparity map is not affected by the complex background of the image, and can protect the subject from being destroyed by the seam carving algorithm based on the standard gradient map.

Then, the proposed algorithm based on the weighted gradient map is compared with the seam carving which uses the standard gradient map. Figure 3 shows the original image (the top row), the weighted gradient map and resized results by the proposed approach (the middle row), and the standard gradient map and resized results by the seam carving approach (the bottom row). The results show that the proposed approach performs better

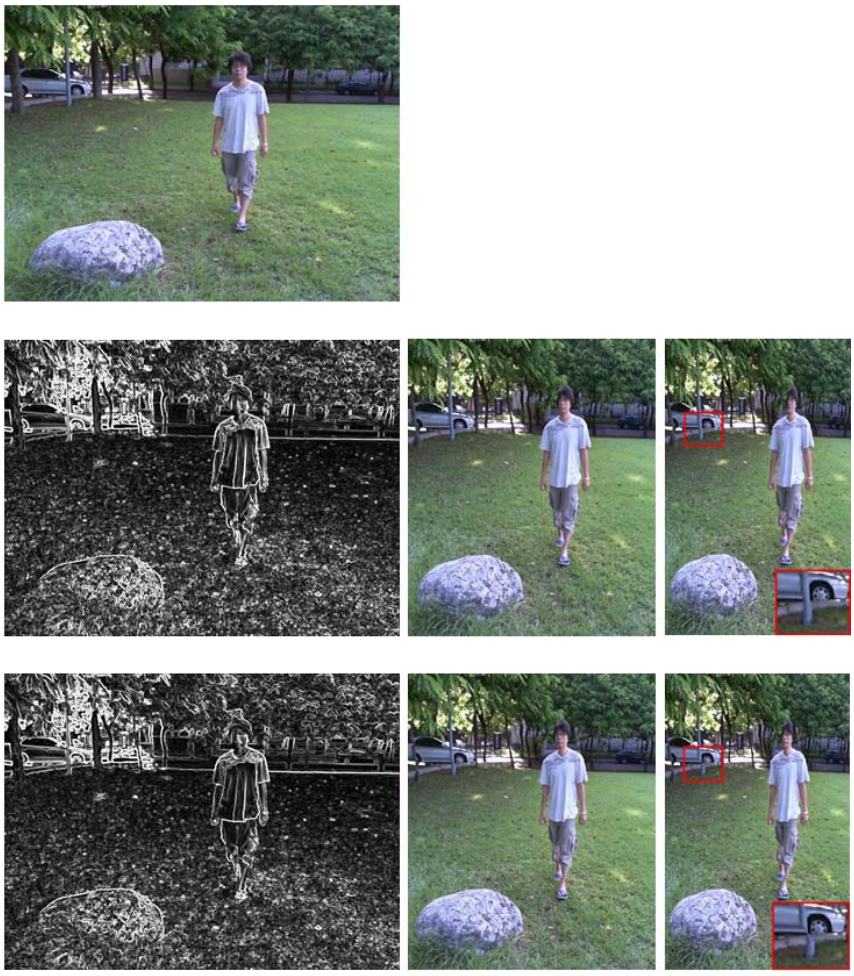

FIGURE 3. Original image (the top row), the weighted gradient map and resized results by the proposed approach (the middle row), and the standard gradient map and resized results by the seam carving approach (the bottom row)

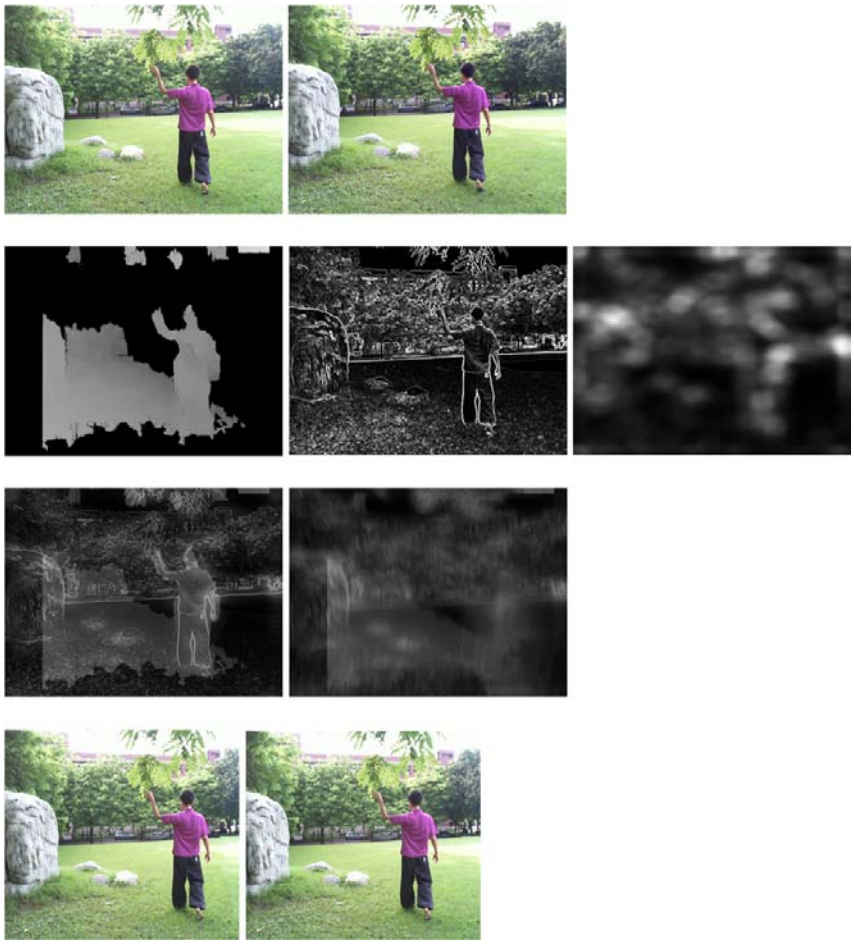

FIGURE 4. Original left-right images (the top row), the disparity map, gradient map and saliency map (the second row), the importance map and accumulated map (the third row), and resized results from left to right based on the importance map and the accumulated map, respectively (the bottom row)

than the seam carving. If the standard gradient map is used, the object would not be able to protect the vertical line in the image during the horizontal reduction.

Finally, the proposed approach based on accumulated maps is compared with the approach based on importance maps for video retargeting. Figure 4 shows original left-right images (the top row), the disparity map, gradient map and saliency map (the second row), the importance map and accumulated map (the third row), and the results from left to right using the importance map and accumulated map, respectively (the bottom row). Figure 5 shows the resized image sequences based on importance maps (the top two rows) and resized image sequences based on accumulated maps (the bottom two rows). Continuous images can be verified that the frame has become more stable after the accumulated map is adopted. When the fps is too large, the faster object is more difficult to adapt. Moreover, maintaining the steadiness of the frame when the fps is too small is difficult. Therefore, the fps can be selected according to the video content to obtain a balanced effect.

It is not easy to adequately determine the parameters for the proposed algorithm. In the experiments, the parameters are set heuristically. Any other values of parameters would not contribute to the quality of resized images.

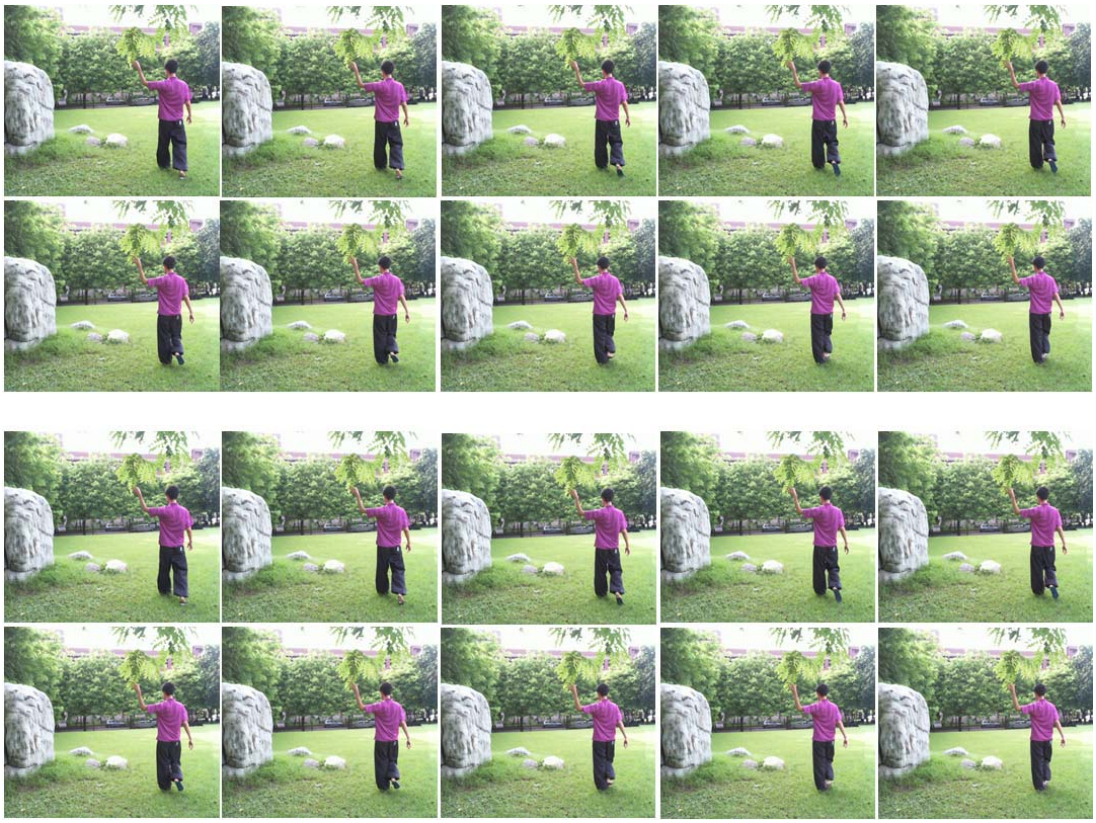

FIGURE 5. Resized image sequences based on importance maps (the top two rows) and resized image sequences based on accumulated maps (the bottom two rows)

5. **Conclusions.** This paper has proposed a novel method for stereo image and video retargeting. Several analyses are conducted, including the energy of disparity, gradient, visual saliency, and motion detection. Moreover, different types of energy are integrated as importance maps for stereo image and video retargeting. Therefore, a perfect protection of the subject is achieved. Finally, considering the images with time continuity and calculating the accumulated map for a period of time can reduce the pixel dislocation caused by selecting different seams from neighboring frames.

Acknowledgment. The authors would like to thank the National Science Council of Taiwan for financially supporting this research under Contract No. NSC 101-2221-E-239-033-.

REFERENCES

[1] S. Avidan and A. Shamir, Seam carving for content-aware image resizing, *ACM Transactions on Graphics*, 2007.

[2] R. Fergus, P. Perona and A. Zisserman, Object class recognition by unsupervised scale-invariant learning, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.2, pp.II-264-II-271, 2003.

[3] D. Gao and N. Vasconcelos, Integrated learning of saliency, complex features, and object detectors from cluttered scenes, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.2, pp.282-287, 2005.

[4] C. Guo, Q. Ma and L. Zhang, Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8, 2008.

[5] R. Hartley, Theory and practice of projective rectification, *International Journal of Computer Vision*, vol.35, pp.115-127, 1999.
[6] R. Hartley, R. Gupta and T. Chang, Stereo from uncalibrated cameras, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.761-764, 1992.
[7] X. Hou and L. Zhang, Saliency detection: A spectral residual approach, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8, 2007.
[8] D. S. Hwang and S. Y. Chien, Content-aware image resizing using perceptual seam carving with human attention model, *Proc. of IEEE International Conference on Multimedia and Expo*, pp.1029-1032, 2008.
[9] L. Itti, C. Koch and E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.20, pp.1254-1259, 1998.
[10] W. Kim and C. Kim, A novel image importance model for content-aware image resizing, *Proc. of the 18th IEEE International Conference on Image Processing*, pp.2469-2472, 2011.
[11] J. H. Kim, J. S. Kim and C. S. Kim, Image and video retargeting using adaptive scaling function, *Proc. of the 17th European Signal Processing Conference*, 2009.
[12] J. S. Kim, J. H. Kim and C. S. Kim, Adaptive image and video retargeting technique based on Fourier analysis, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.1730-1737, 2009.
[13] K. Konolige, Small vision systems: Hardware and implementation, *Proc. of Eighth International Symposium on Robotics Research*, vol.8, pp.111-116, 1997.
[14] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang and H. Shum, Learning to detect a salient object, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.
[15] P. Power and J. Schoonees, Understanding background mixture models for foreground segmentation, *Proc. of Image and Vision Computing*, 2002.
[16] M. Rubinstein, A. Shamir and S. Avidan, Improved seam carving for video retargeting, *ACM Transactions on Graphics*, pp.1-9, 2008.
[17] D. Ruderman, The statistics of natural images, *Network: Computation in Neural Systems*, vol.5, no.4, pp.517-548, 1994.
[18] A. Srivastava, A. Lee, E. Simoncelli and S. Zhu, On advances in statistical modeling of natural images, *Journal of Mathematical Imaging and Vision*, vol.18, no.1, pp.17-33, 2003.
[19] C. Stauffer and W. Grimson, Adaptive background mixture models for real-time tracking, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.2, 1999.
[20] A. Treisman and G. Gelade, A feature-integration theory of attention, *Cognitive Psychology*, vol.12, pp.97-136, 1980.
[21] A. vander Schaaf and J. VanHateren, Modelling the power spectra of natural images: Statistics and information, *Vision Research*, vol.36, pp.2759-2770, 1996.
[22] D. Walther and C. Koch, Modeling attention to salient proto-objects, *Neural Networks*, vol.19, pp.1395-1407, 2006.
[23] L. Wang, H. Jin, R. Yang and M. Gong, Stereoscopic inpainting: Joint color and depth completion from stereo images, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8, 2008.
[24] S. F. Wang and S. H. Lai, Compressibility-aware media retargeting with structure preserving, *IEEE Transactions on Image Processing*, vol.20, no.3, pp.855-865, 2011.
[25] D. Wang, A. Kristjansson and K. Nakayama, Efficient visual search without top-down or bottom-up guidance, *Perception & Psychophysics*, vol.67, p.239, 2005.