Combining Depth Information for Image Retargeting

Huei-Yung Lin[1], Chin-Chen Chang[2] and Jhih-Yong Huang[1]
[1]*Department of Electrical Engineering, National Chung Cheng University, Chiayi 621, Taiwan*
[2]*Department of Computer Science and Information Engineering, National United University, Miaoli 360, Taiwan*

Keywords: Image Resizing, Image Retargeting, Feature Map, Seam Carving, Ranging Camera.

Abstract: This paper presents a novel image retargeting approach for ranging cameras. The proposed approach first extracts three feature maps: depth map, saliency map, and gradient map. Then, the depth map and the saliency map are used to separate the main contents and the background and thus compute a map of saliency objects. 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. Unlike previous approaches, the proposed approach preserves the salient object well and maintains the gradient and visual effects in the background. Moreover, it protects the salient object from being destroyed by the seam carving algorithm. The experimental results show that the proposed approach performs well in terms of the resized quality.

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. These display specifications are applied to displays as large as billboards and as small as mobile phone screens. Display devices, however, have only one screen aspect ratio. This aspect ratio causes upper and lower black bands to appear when multimedia contents are displayed on screens.

Apart from the two screen aspect ratios described above, 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 (Avidan and Shamir, 2007; Hwang and Chien, 2008; Kim et al., 2009a; Kim et al., 2009b; Wang et al., 2008b; Lin et al., 2012) for resizing image based on image contents has 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.

In this paper, a novel image retargeting approach for ranging cameras is proposed. The proposed approach first extracts three feature maps: depth map, saliency map, and gradient map. Then, the depth map and the saliency map are used to compute a map of saliency objects. 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 (Avidan and Shamir, 2007) based on the importance map. The experimental results show that the proposed approach resizes image 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.

Lin H., Chang C. and Huang J..
Combining Depth Information for Image Retargeting.
DOI: 10.5220/0004232601720179
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2013), pages 172-179
ISBN: 978-989-8565-47-1

2 RELATED WORKS

Avidan and Shamir (Avidan and Shamir, 2007) 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 indicated 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. (Kim et al., 2009a) used the adaptive scaling function, utilizing the importance map of the image to calculate the adaptive scaling function, which indicated the reduction level for each row of the original image. Kim et al. (Kim et al., 2009b) 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 were then used as a low-pass filter to obtain an effect similar to smoothing. The level of horizontal reduction for each strip was 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 (Fergus et al., 2003; Itti et al., 1998; Gao and Vasconcelos, 2005; Liu et al., 2007). However, the human eye is capable of quickly locating common objects. Various approaches have proposed for simulating the functions of the human eye; for instance, Saliency ToolBox (Walther and Koch, 2006) and Saliency Residual (SR) (Hou and Zhang, 2007). 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 (Hwang and Chien, 2008) 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. (Rubinstein et al., 2008) proposed a method of improvement for the procedure of seam carving. This method utilized techniques of forward energy and backward energy to reduce discontinuity in images.

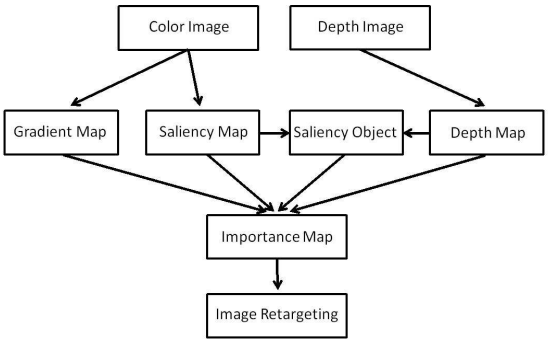

Figure 1: The flowchart of the proposed approach.

Wang et al. (Wang et al., 2008a) proposed a method that simultaneously utilized techniques of stereo imaging and inpainting. This method had the capacity to remove image objects that caused occlusion, restoring original background image and depth information. They also presented a warping approach for resizing images and preserving visually features. The deformation of the image is based on an importance map that is computed using a combination of gradient and salience features.

Achanta et al. (Achanta et al., 2009) proposed an approach for detecting salient regions by using only color and luminance features. Their approach is simple to implement and computationally efficient. It can clearly identify the main silhouettes. Also, this approach outputs saliency maps with well-defined boundaries of salient objects. Goferman et al. (Goferman et al., 2010) proposed an approach which aims at detecting the salient regions that represent the scene. The goal is to either identify fixation points or detect the dominant object. They presented a detection algorithm which is based on four principles observed in the psychological literature. In image retargeting, using their saliency prevents distortions in the important regions.

3 THE PROPOSED APPROACH

The flowchart of the proposed approach is shown in Figure 1. First, the proposed approach extracts three feature maps, namely, a depth map, a saliency map, and a gradient map from an input color image and a depth image. Then, the depth map and the saliency map are used to compute a map of saliency objects. 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 (Avidan and Shamir, 2007).

3.1 Important Map

The importance map E_{imp} is defined as

$$E_{imp} = \begin{cases} 1 & \text{if } E_o = 1 \\ \alpha_1 E_d + \alpha_2 E_s + \alpha_3 E_g & \text{if } E_o = 0 \end{cases}$$

where α_1, α_2, and α_3 are the weights for the depth map E_d, the saliency map E_s, and the gradient map E_g, respectively; E_o is the saliency object map.

3.1.1 Depth Map

The Kinect camera is used to extract depth information from an input color image. The camera uses a 3D scanner system called Light Coding using near-infrared light to illuminate the objects and determine the depth of the image. Figure 2(a) shows a color image and the corresponding depth image captured by the Kinect.

From Figure 2(a), pixel positions of the color image and the corresponding pixel positions of the depth image are not consistent. This problem can be adjusted by an official Kinect SDK, as shown in Figure 2(b). Hence, the pixel positions of the color image and the corresponding pixel positions of the depth image are consistent. However, the range covered by the depth image becomes smaller. Therefore, the original depth image of size 640×480 is cropped into a new depth image of size 585×430 by removing the surrounding area of the original depth image without the depth information and leaving the area with the usable depth, as shown in Figure 2(c). Also, as shown in Figure 2(c), black blocks in the cropped depth image are determined and the depth values of these blocks are set as 0. They cannot be measured by the Kinect due to strong lighting, reflected light, outdoor scenes, occlusions, and so on. Therefore, the depths of these regions are negligible since these areas in the whole depth image are very small.

3.1.2 Saliency Map

Visual saliency is an important factor for human visual system. Therefore, the proposed approach extracts a saliency map from the input color image. The computer vision (Avidan and Shamir, 2007; Rubinstein et al., 2008) tries to imitate the possible visual perception of the human eye, from object detection, object classification to object recognition. Recently, Achanta et al. (Achanta et al., 2009) proposed an

Figure 2: (a) Original color image and corresponding depth image captured by the Kinect; (b) Original color image and adjusted depth image by an official Kinect SDK; (c) Cropped color image and cropped depth image.

Figure 3: Original image (left), saliency map by (Achanta et al., 2009) (middle), and saliency map by (Goferman et al., 2010) (right).

approach for detecting salient regions by using only color and luminance features. Goferman et al. (Goferman et al., 2010) proposed an approach which aims at detecting the salient regions that represent the scene. Figure 3 shows the comparison of the two previous approaches (Achanta et al., 2009; Goferman et al., 2010). When a scene is complex, the approach of Achanta et al. cannot identify salient areas effectively and the method of Goferman et al. can obtain better results. Therefore, in the proposed approach, the technique of Goferman et al. is applied to complex environments for extracting a saliency map. The main concepts of the approach of Goferman et al. are described as the following.

For each pixel i, let p_i be a single patch of scale r centered at pixel i. Also, let $d_{color}(p_i, p_j)$ be the distance between patches p_i and p_j in CIE Lab color space, normalized to the range $[0,1]$. If $d_{color}(p_i, p_j)$

is high for each pixel j, pixel i is salient. In the experiment, r is set as 7. Moreover, let $d_{position}(p_i, p_j)$ be the distance between the positions of patches p_i and p_j, which is normalized by the larger image dimension. A dissimilarity measure between a pair of patches is defined as

$$d(p_i, p_j) = \frac{d_{color(p_i,p_j)}}{1 + c \cdot d_{position}(p_i, p_j)},$$

where c is a parameter, In the experiment, c is set as 3. If $d(p_i, p_j)$ is high for each j, pixel i is salient.

In practice, for each patch p_i, there is no need to evaluate its dissimilarity to all other image patches. It only needs to consider the K most similar patches $\{q_k\}_{k=1}^{K}$ in the image. If $d(p_i, p_j)$ is high for each $k \in [1, K]$, pixel i is salient. For a patch p_i of scale r, candidate neighbors are defined as the patches in the image whose scales are $R_q = \{r, \frac{1}{2}r, \frac{1}{4}r\}$.

The saliency value of pixel i at scale r is defined as

$$s_i^r = 1 - \exp\{-\frac{1}{K}\sum_{k=1}^{K} d(p_i^r, q_k^{r_k})\},$$

where $r_k \in R_q$ and K is set as 64 in the experiment. Furthermore, each pixel is represented by the set of multi-scale patches centered at it. Thus, for pixel i, let $R = \{r_1, r_2, \ldots, r_M\}$ be the set of patch sizes. The saliency of pixel i is defined as the mean of its saliency at different scales

$$\bar{S_i} = \frac{1}{M}\sum_{r \in R} S_i^r.$$

If the saliency value of a pixel exceeds a certain threshold, the pixel is attended. In the experiment, the threshold is set as 0.8. Then, each pixel outside the attended areas is weighted according to its distance to the closest attended pixel. Let $d_{foci}(i)$ be the positional distance between pixel i and the closest focus of attention pixel, normalized to the range $[0, 1]$. The saliency of a pixel i is redefined as

$$\hat{S}_i = \bar{S_i}(1 - d_{foci}(i)).$$

3.1.3 Gradient Map

The human visual system is sensitive to edge information in an image. Therefore, the proposed approach extracts a gradient map from the input color image to represent edge information.

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

$$Sobel_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

and

$$Sobel_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

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. The gradient map is defined as

$$E_{gradient} = \sqrt{(Sobel_x * I)^2 + (Sobel_y * I)^2}.$$

Using the Sobel operator can easily detect gradients of an image. However, the detected gradients do not fit the gradients perceived by the human eye. Therefore, a bilateral filter (Kim et al., 2009b) is used to reduce borders that are not visually obvious and keep borders that vary largely. The bilateral filter is a nonlinear filter and smoothes noises effectively and keeps important edges. A Gaussian smoothing is applied to an image in both spatial domain and intensity domain at the same time. The definition of the Gaussian smoothing is as follows:

$$J_s = \frac{1}{k(s)}\sum_{p \in \Omega} f(p - s) \cdot g(I_p - I_s) \cdot I_p,$$

where J_s is the result after processing pixel s by the bilateral filter. I_p and I_s are intensities of pixels p and s, respectively. Ω is the whole image. f and g are Gaussian smoothing functions for the spatial and intensity domains, respectively. $k(s)$ is a function for normalization and its definition is given by

$$k(s) = \sum_{p \in Omega} f(p - s) \cdot g(I_p - I_s).$$

Therefore, in the proposed approach, the input color image is filtered by the bilateral filter. Then, the resulting image is filtered by Sobel filter to compute the final gradient map. The proposed approach can effectively remove gradients with small changes and reserve gradients with large variations in an image. The gradients are close to human visual perception. In the experiments, the spatial domain parameters are set as 10 and the intensity domain parameters are set as 100.

Gradient information can keep the consistency of a line in the image. However, when the gradient in the image has a certain percentage of length, the use of seam carving can pass through the gradients. The gradients will be broken or distorted. Therefore, it is necessary to improve gradients for a certain length of gradients. The improved approach first uses Canny edge detection to detect edges in an image and then

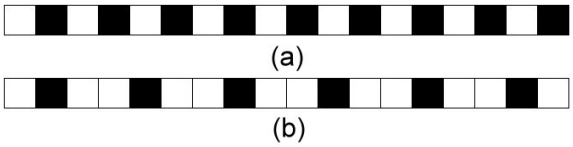

Figure 4: (a) Periodic weights of (1,0); (b) Periodic weights of (1,0,1).

Figure 5: (a) Original gradient map; (b) Improved gradient map.

uses Hough transform to find a certain length of a line. After finding straight lines by Hough transform, weights are assigned to the straight lines. When the input image is reduced to less than half of the original image, the periodic weights $(1,0)$ are used for the weighting. When an input image is reduced to more than half of the original image, the periodic weights $(1,0,1)$ are used for the weighting. See Figure 4 as an illustration.

After improving gradients, seam carving can cut gradients uniformly. Removing gradients with weight 0 can retain the gradients with weight 1. It can maintain the existing continuity and is less likely to remove the same area resulting in clear discontinuities. Figure 5 shows the original gradient map and the improved gradient map.

3.1.4 Salient Object

In an image, the human visual eye may have one or more attentions that have the greatest saliencies. Therefore, the most salient objects in the image are identified for retargeting. The salient objects are defined as the visually indistinguishable components. Since each pixel of a salient object is not necessarily a high value, image segmentation is used to find main partitions to obtain salient objects.

The depth image is segmented into depth regions. The depths are classified based on depth similarity of the scene. The image pyramids are used to split depth regions. The image pyramids down-sample the image into different scales. If pixels of *i*-th layer and farther pixels of the adjacent layer have similar colors, the father pixels and the pixels of *i*-th layer are merged into a connected component. In the same layer, if the adjacent components are too similar, they are merged into a larger connected component. After processing layer

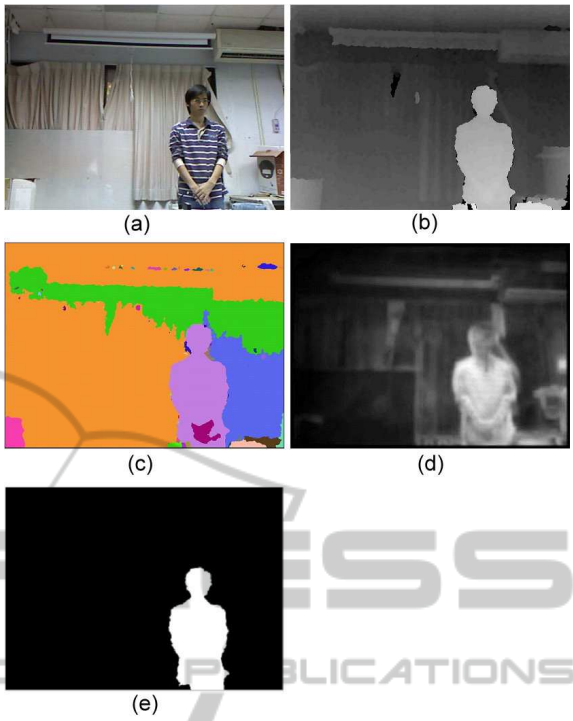

Figure 6: (a) Original image; (b) Depth map; (c) Depth regions; (d) Saliency map; (e) Saliency object.

by layer, the depth image (Figure 6(b)) of an input image (Figure 6(a)) is segmented into depth regions (Figure 6(c)).

The advantage of using the image pyramids to determine depth regions is that thresholds can be easily used for adjustment. Each component representing pixels in this region has similar depths. The depth regions segmented by the image pyramids and the saliency map (Figure 6(d)) are combined to obtain salient objects. If the salient value of a region is above a certain threshold, this region is defined as an indistinguishable object, as shown in Figure 6(e). In the experiment, if the salient object is too small, it is ignored.

3.2 Image Retargeting

The proposed approach applied the method of Avidan and Shamir (Avidan and Shamir, 2007) for image retargeting. Let I be an $n \times m$ image and the vertical seam is defined as

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

where x is a mapping $x : [1, \cdots, n] \rightarrow [1, \cdots, m]$.

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 route of the vertical seam is indicated as $I_s = \{I(S_i)\}_{i=1}^{n} = \{I(x(i), i)\}_{i=1}^{n}$. 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 $E(s) = E(I_s) = \sum_{i=1}^{n} e(I(s_i))$ 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 $s^* = \min_s E(s) = \min_s \sum_{i=1}^{n} e(I(s_i))$ 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 was 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.

4 RESULTS

Several experiments were conducted to evaluate the effectiveness of the proposed approach. The proposed algorithm was running on a laptop with a 2.40 GHz Core2 Quad CPU and 3.0 GB of memory. The camera used in this experiment was a Microsoft Kinect. If E_{object} is $0, a_1, a_2$, and a_3 are set as 0.1, 0.5, 0.4, respectively. If the Kinect cannot detect depths, a_1 and a_2 are both set as 0.5. The size of the original image is 585×430. Without loss of generality, a source image is resized in the horizontal direction only to make a target image. The extension for resizing in the vertical direction is straightforward. Therefore, the sizes of the resized images are 500×430, 400×430, 300×430, and 200×430. Moreover, the proposed approach was compared to the two previous approaches (Avidan and Shamir, 2007; Wang et al., 2008b).

The first image is an indoor environment. There is no salient object and the depths are similar. The importance map is mainly dominated by saliency map and gradient map. Figure 7 shows the original image, the depth map, the saliency map, the gradient map, the salient object and the importance map, respectively, from left to right and top to bottom. Figure 8 shows the resized results of Avidan and Shamir (the top row), of Wang et al. (the middle row), and of the proposed approach (the bottom row). For the resized images of 500×430 and 400×430, the results of the proposed approach are similar to those of the two previous approaches. However, for the resized images of 300×430 and 200×430, the results show that the proposed approach performs better than the two previous approaches. There is a serious distortion in the results of Avidan and Shamir. The size of the whiteboard in the results of Wang et al. is over reduced.

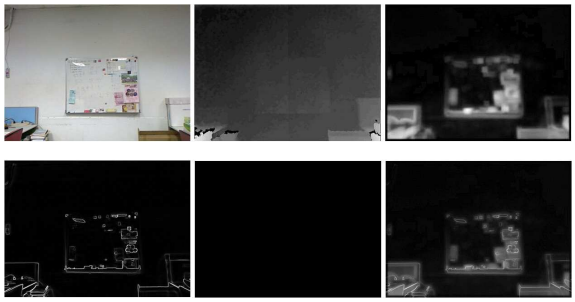

Figure 7: Original image, depth map, saliency map, gradient map, saliency object, and importance map from left to right and top to bottom, respectively.

Figure 8: Resized images by (Avidan and Shamir, 2007) (top); resized images by (Wang et al., 2008b) (middle); resized images by the proposed approach (bottom).

The second image is an indoor environment. There is a salient object. Figure 9 shows the original image, the depth map, the saliency map, the gradient map, the salient object and the importance map, respectively, from left to right and top to bottom. Figure 10 shows the resized results of Avidan and Shamir (the top row), of Wang et al. (the middle row), and of the proposed approach (the bottom row). For the resized images of 500×430 and 400×430, the results

Figure 9: Original image, depth map, saliency map, gradient map, saliency object, and importance map from left to right and top to bottom, respectively.

Figure 10: Resized images by (Avidan and Shamir, 2007) (top); resized images by (Wang et al., 2008b) (middle); resized images by the proposed approach (bottom).

of the proposed approach are similar to those of the two previous approaches. However, for the resized images of 300×430 and 200×430, the results of the proposed approach are better than those of the previous approaches. The gradients are preserved well such that gradient density is too high in the results of Avidan and Shamir. There is distortion in the face of the person. For the results of Wang et al., the difference between the salient objects and non-salient background is too large. Conversely, the proposed approach preserves the salient object well and maintains the gradient and visual effects in the background.

The final image is an outdoor environment. The depth map is not complete. Since there are strong lighting and reflected light, the Kinect cannot detect the depths well and the detected salient object is not complete. Figure 11 shows the original image, the depth map, the saliency map, the gradient map, the salient object and the importance map, respectively, from left to right and top to bottom. Figure 12 shows

Figure 11: Original image, depth map, saliency map, gradient map, saliency object, and importance map from left to right and top to bottom, respectively.

the resized results of Avidan and Shamir (the top row), of Wang et al. (the middle row), and of the proposed approach (the bottom row). For the resized images of 500×430 and 400×430, the results of the proposed approach are similar to those of the two previous approaches. However, for the resized images of 300×430 and 200×430, the proposed approach performs better than the previous approaches. The gradients are preserved well such that gradient density is too high in the results of Avidan and Shamir. There is distortion in the body of the person. The visual effects in the background are not consistent. For the results of Wang et al., the difference between the salient objects and non-salient background is too large. The legs of the salient object and the floor have similar colors such that the energy is not enough and there is distortion in the salient object. For the proposed approach, although the salient object is not complete, it can still maintain the integrity of the salient object. However, since the environment is more complex, it is difficult to achieve good visual effects for the resized results of 200×430.

From the above results, the approach of Avidan and Shamir puts more emphasis on gradient information. For making large adjustments to an image, the gradients can still be preserved well. However, gradient density is too high and the visual effects are not consistent. For the approach of Wang et al., it has good continuity for image resizing. However, for making large adjustments to an image, the salient object is too small and non-salient areas are too large. In the proposed approach, making large adjustments to an image, it can preserve the salient object well. Also, it can keep the surrounding area of the salient object on the background and remove the gradients of background far away the salient object to avoid over-concentration of the gradients. It can protect the salient object from being destroyed by the seam carving algorithm.

Figure 12: Resized images by (Avidan and Shamir, 2007) (top); resized images by (Wang et al., 2008b) (middle); resized images by the proposed approach (bottom).

5 CONCLUSIONS

This paper has proposed a novel image retargeting method for ranging cameras. Several analyses were conducted, including the energy of depth, gradient, and visual saliency. Then, the depth map and the saliency map are used to determine a map of saliency objects. Moreover, different types of energy were integrated as importance maps for image retargeting. Unlike previous approaches, the proposed approach preserves the salient object well and maintains the gradient and visual effects in the background. Moreover, it protects the salient object from being destroyed by the seam carving algorithm. Therefore, a perfect protection of the subject was achieved.

REFERENCES

Achanta, R., Hemami, S., Estrada, F., and Susstrunk, S. (2009). Frequency-tuned salient region detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1597 –1604.

Avidan, S. and Shamir, A. (2007). Seam carving for content-aware image resizing. *ACM Trans. Graph.*, 26(3):10.

Fergus, R., Perona, P., and Zisserman, A. (2003). Object class recognition by unsupervised scale-invariant learning. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–264 – II–271 vol.2.

Gao, D. and Vasconcelos, N. (2005). Integrated learning of saliency, complex features, and object detectors from cluttered scenes. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 282 – 287 vol. 2.

Goferman, S., Zelnik-Manor, L., and Tal, A. (2010). Context-aware saliency detection. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2376 –2383.

Hou, X. and Zhang, L. (2007). Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1 –8.

Hwang, D.-S. and Chien, S.-Y. (2008). Content-aware image resizing using perceptual seam carving with human attention model. In *Multimedia and Expo, 2008 IEEE International Conference on*, pages 1029–1032.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(11):1254 –1259.

Kim, J., Kim, J., and Kim, C. (2009a). Image and video retargeting using adaptive scaling function. In *Proceedings of 17th European Signal Processing Conference*.

Kim, J.-S., Kim, J.-H., and Kim, C.-S. (2009b). Adaptive image and video retargeting technique based on fourier analysis. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1730 –1737.

Lin, H.-Y., Chang, C.-C., and Hsieh, C.-H. (2012). Cooperative resizing technique for stereo image pairs. In *Proceedings of 2012 International Conference on Software and Computer Applications*, pages 96–100.

Liu, T., Sun, J., Zheng, N.-N., Tang, X., and Shum, H.-Y. (2007). Learning to detect a salient object. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1 –8.

Rubinstein, M., Shamir, A., and Avidan, S. (2008). Improved seam carving for video retargeting. *ACM Trans. Graph.*, 27(3):1–9.

Walther, D. and Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 19(9):1395 – 1407. Brain and Attention, Brain and Attention.

Wang, L., Jin, H., Yang, R., and Gong, M. (2008a). Stereoscopic inpainting: Joint color and depth completion from stereo images. pages 1–8.

Wang, Y.-S., Tai, C.-L., Sorkine, O., and Lee, T.-Y. (2008b). Optimized scale-and-stretch for image resizing. In *SIGGRAPH Asia '08: ACM SIGGRAPH Asia 2008 papers*, pages 1–8, New York, NY, USA. ACM.