Image retargeting approach for ranging cameras

H.-Y. Lin[1], C.-C. Chang*[2] and C.-Y. Huang[1]

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 gradients 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.

Keywords: Image resizing, Image retargeting, Feature map, Seam carving, Ranging camera

Introduction

Numerous and varied devices for displaying multimedia contents exist, from cathode-ray tubes to liquid-crystal displays, and from plasma to light-emitting diode. 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, non-standard 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[1–6] for resizing images based on image contents have been proposed. These methods require a certain understanding of image content. Most of these approaches try to determine the importance at each pixel to effectively preserve important regions and discard less important regions to achieve a target image size. To obtain a target image, the importance map needs to be determined. Several models for the importance map are driven by the low-level features such as gradient, visual saliency and so on. Although such approaches can obtain better resized images, they still suffer from the quality problems for complex images. Therefore, further improvements are often required to extract more reliable features.

In this paper, we propose a novel image retargeting approach for ranging cameras. To protect visual contents, we first extract a depth map, a saliency map and a gradient map from an input colour image and a depth image. In particular, we use the depth map and the saliency map to compute a map of saliency objects. After that, we construct an importance map based on the four feature maps. The idea for mixing the features is similar to the hyper-spectral content aware resizing technique presented by Scott *et al.*[5] According to the importance map, the important regions are preserved and less important regions are discarded. Finally, the target image is constructed using the seam carving method[1] based on the importance map. The results demonstrate that the proposed approach resizes images more effectively than the previous approaches.

The remainder of this paper is organised 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.

[1]Department of Electrical Engineering and Advanced Institute of Manufacturing with High-tech Innovations, National Chung Cheng University, Minhsiung, Chiayi 621, Taiwan
[2]Department of Computer Science and Information Engineering, National United University, Miaoli 360, Taiwan

*Corresponding author: Chin-Chen Chang, Department of Computer Science and Information Engineering National United University, 1, Lienda, Kungching Li, Miaoli 360, Taiwan; email: ccchang@nuu.edu.tw

Received 3 Mar 2012; accepted 19 July 2012
DOI 10.1179/1743131X12Y.0000000038

Related works

Avidan and Shamir[1] proposed a method for adjusting image size based on image content. They analysed 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 eight-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.

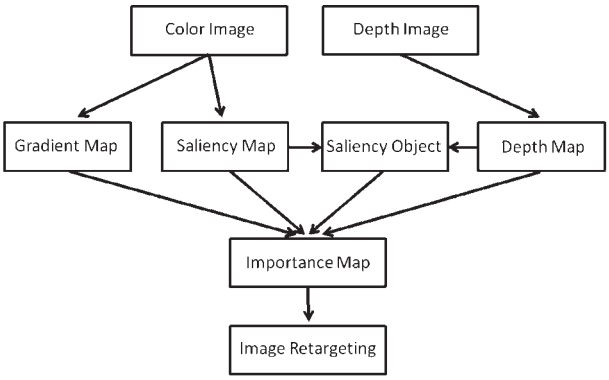
Kim *et al.*[3] used the adaptive scaling function, utilising the importance map of the image to calculate the adaptive scaling function, which indicated the reduction level for each row of the original image. Kim *et al.*[4] used Fourier analysis for image resizing. After constructing the gradient map, 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.[7–10] However, the human eye is capable of quickly locating common objects.[11,12] Various approaches have proposed for simulating the functions of the human eye, for instance, Saliency ToolBox[13] and Saliency Residual (SR).[14] 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[2] used a neural network method to determine the subject of images. They also used face detection 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.*[15] proposed a method of improvement for the procedure of seam carving. This method utilised techniques of forward energy and backward energy to reduce discontinuity in images.

Wang *et al.*[16] proposed a method that simultaneously utilised 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. Wang *et al.*[6] 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.*[17] proposed an approach for detecting salient regions by using only colour 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. These boundaries are preserved by retaining substantially more frequency content from the original image. Goferman *et al.*[18] 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.



1 Flowchart of the proposed approach

The proposed approach

The flowchart of the proposed approach is shown in Fig. 1. First, the proposed approach extracts three feature maps, namely, a depth map, a saliency map and a gradient map from an input colour 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.

Importance map

In our approach, we propose improved techniques for extracting features and integrate them to get the importance map. That is, the importance map is derived by the fusion of the four different maps. In our approach, the concept of mixing the energy from various sources is similar to the work of Scott *et al.*[5] However, our approach more focuses on the information extracted from range and intensity images.

The importance map E_{imp} is defined as

$$E_{\text{imp}} = \begin{cases} 1 & \text{if } E_{\text{object}} = 1 \\ \alpha_1 E_{\text{depth}} + \alpha_2 E_{\text{saliency}} + \alpha_3 E_{\text{gradient}} & \text{if } E_{\text{object}} = 0 \end{cases}$$

where α_1, α_2 and α_3 are the weights for depth map E_{depth}, saliency map E_{saliency} and gradient map E_{gradient}, respectively; E_{object} is the saliency object map.

Depth map

The Kinect[19] camera is used to extract depth information from an input colour 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 2a shows a colour image and the corresponding depth image captured by the Kinect.

2 ***a*** **Original colour image and corresponding depth image;** ***b*** **original colour image and adjusted depth image;** ***c*** **cropped colour image and cropped depth image**

From Fig. 2a, pixel positions of the colour 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 Fig. 2b. Hence, the pixel positions of the colour 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 Fig. 2c. Also, as shown in Fig. 2c, 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.

Saliency map

Visual saliency is an important factor for human visual system. Therefore, the proposed approach extracts a saliency map from the input colour image. The technique of Goferman *et al.*[18] is applied for extracting a saliency map since it can identify salient areas efficiently for complex environments. 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 centred at pixel i. Also, let $d_{color}(p_i, p_j)$ be the distance between patches p_i and p_j in CIE Lab colour space, normalised 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 normalised by the larger image dimension. A dissimilarity measure between a pair of patches is defined as

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

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 $d(p_i, q_k)$ is high for each $k \in [1, K]$, pixel i is salient. For a patch p_i of scale r, candidate neighbours are defined as the patches in the image whose scales are $R_q = \{r, 1/2r, 1/4r\}$.

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

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

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 centred 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, normalised to the range [0, 1]. The saliency of a pixel i is redefined as

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

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 colour 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

$$\text{Sobel}_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \text{ and}$$

$$\text{Sobel}_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$$

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. The gradient map is defined as

$$E_{gradient} = \left[(\text{Sobel}_x I)^2 + (\text{Sobel}_y I)^2\right]^{1/2}$$

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[20] is used to reduce borders that are not visually obvious and keep borders that vary largely. The bilateral filter is a non-linear 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:

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

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 normalisation and its definition is given by

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

Therefore, in the proposed approach, the input colour 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. Since the seam carving is done by iteratively deleting the seam with the lowest energy, the deleted seams may be too concentrated in the same area. Thus, the gradients will be broken or distorted. Therefore, it is necessary to improve gradients for a certain length of gradients. The proposed approach first detects edges in an image and then uses Hough transform to find a certain length of a line. Let a line be represented by

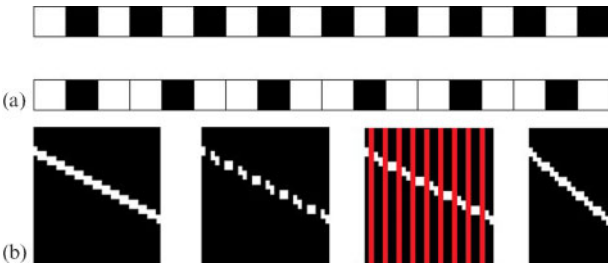
$$y = mx + c$$

where m is the slope of the line and c is the intercept. Also, the straight line can be represented as a point (m,c) in the parameter space. Each different line through the point (x,y) corresponds to one of the points on the line in the (m,c) space. For computational reasons, a line can also be represented in polar space by

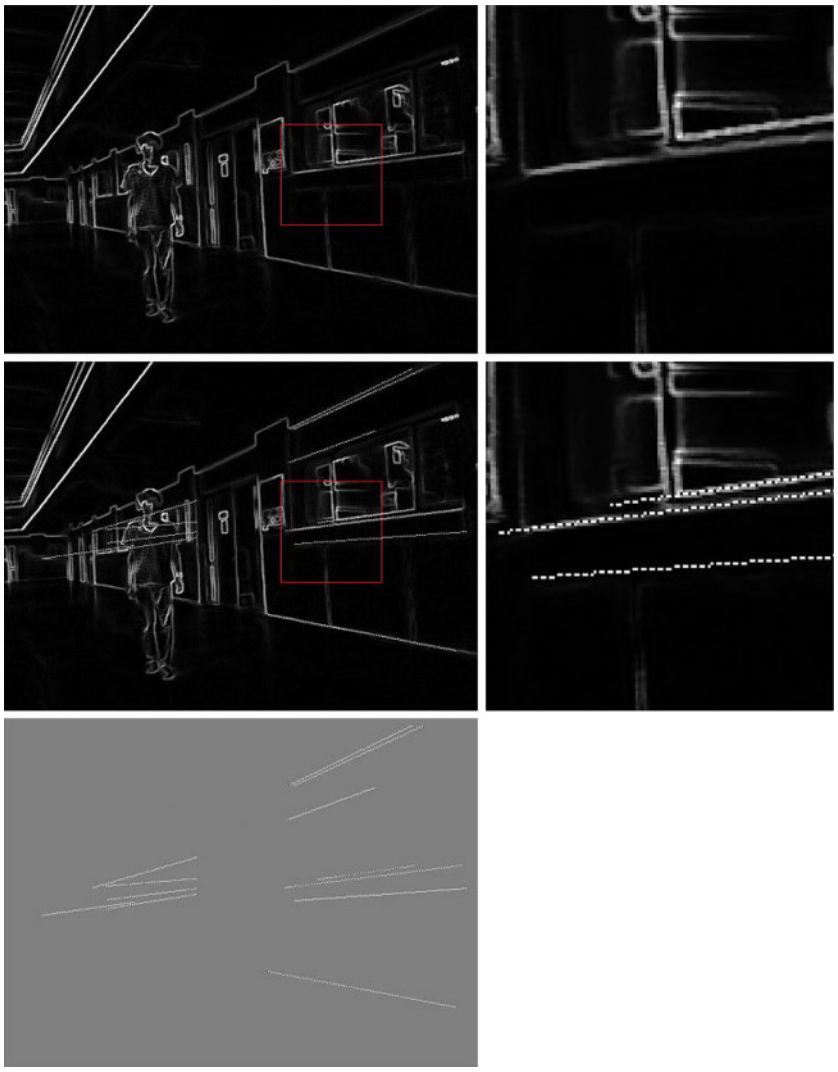
$$x\cos\theta + y\sin\theta = \rho$$

where parameter ρ represents the distance between the line and the origin; parameter θ is the angle of ρ with respect to the x-axis.

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 since more zero weights can be assigned to the straight lines. 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 since less zero weights can be assigned to the straight lines. Figure 3 shows the periodic weights of (1,0) (top) and the periodic weights of (1,0,1) (bottom) (Fig. 1a); and an original gradient map, the improved gradient map using the periodic weights (1,0,1), the image processed by the seam carving and the resized image from left to right (Fig. 1b). Figure 4 shows the original gradient map with a zoom-in view (top), the improved gradient map



3 *a* Periodic weights of (1,0) (top) and periodic weights of (1,0,1) (bottom); *b* original gradient map, improved gradient map, image processed by the seam carving and resized image (from left to right)



4 Original gradient map with a zoom-in view (top), improved gradient map with a zoom-in view (middle), and difference image of the original gradient map and the improved gradient map (bottom)

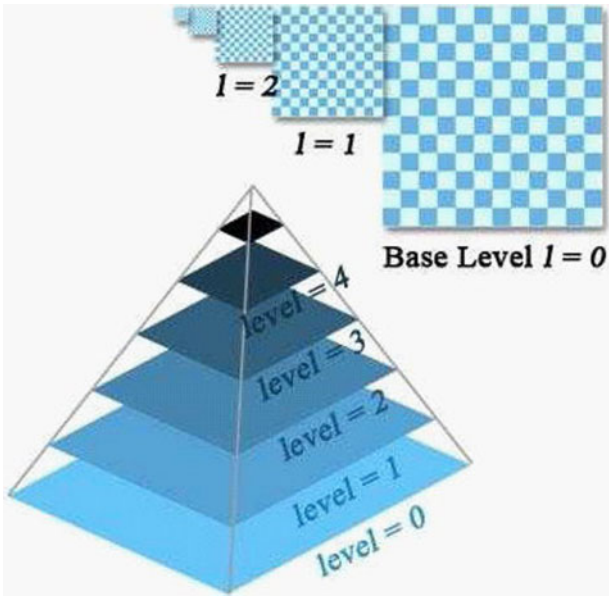
with a zoom-in view (middle), and the difference image of the original gradient map and the improved gradient map (bottom).

After improving gradients, the 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.

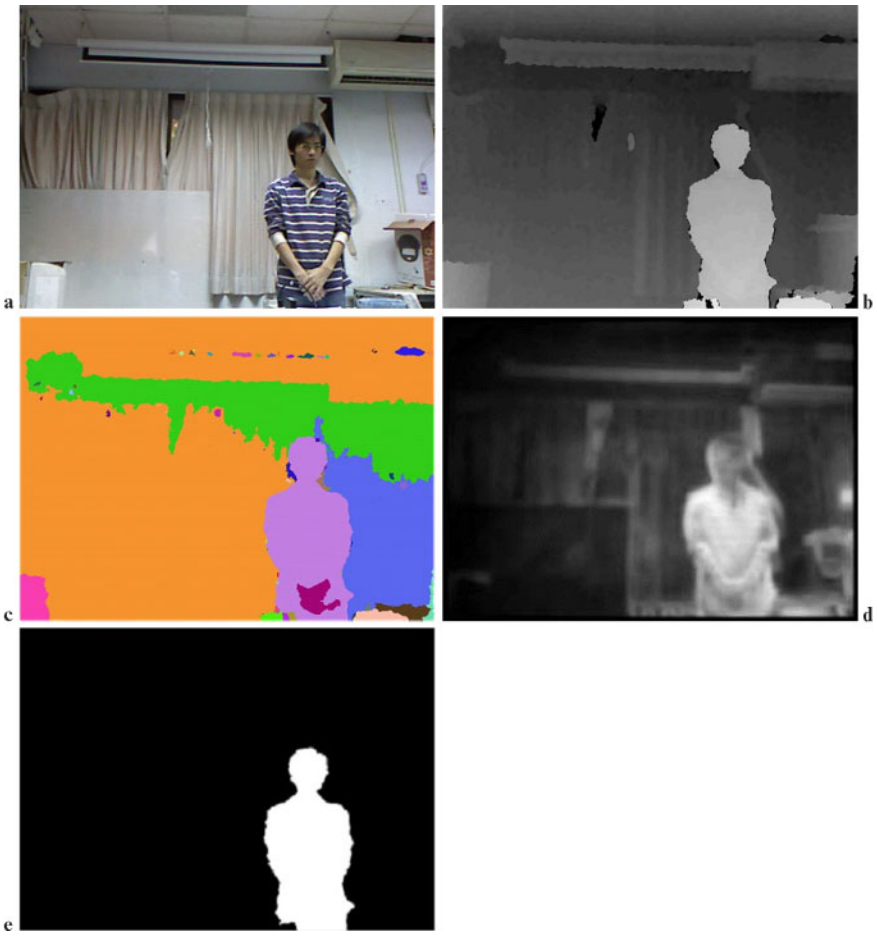
Salient object

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[21] are used to split depth regions. The image pyramids down-sample the image into different scales, as shown in Fig. 5. If pixels of the ith layer and father pixels of the adjacent layer have similar colours, the father pixels and the pixels of the ith 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 by layer, the depth image



5 Image pyramids



6 ***a*** **Original image;** ***b*** **depth map;** ***c*** **depth regions;** ***d*** **saliency map;** ***e*** **saliency object**

(Fig. 6b) of an input image (Fig. 6a) is segmented into depth regions (Fig. 6c).

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 (Fig. 6d) 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 Fig. 6e. In the experiment, if the salient object is too small, it is ignored.

Image retargeting

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

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

where x is a mapping x: [1, 2, …, n]→[1, 2, …, m].

A vertical seam is an eight-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.

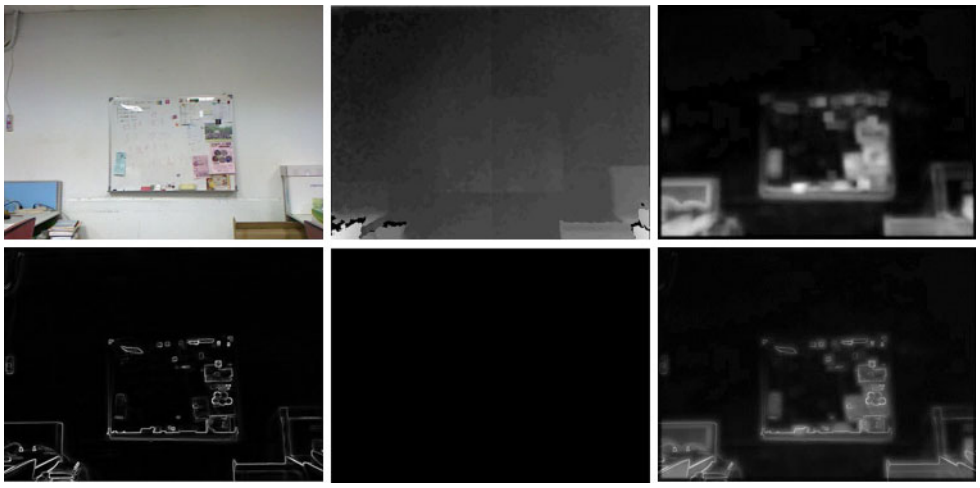
Dynamic programming can be employed to calculate s^*. The smallest accumulated energy M is calculated with every possible point on the seam (i,j) from the second to the last row of the image as

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

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.

Results

Several experiments are conducted to evaluate the effectiveness of the proposed approach. The proposed algorithm is running on a laptop with a 2·40 GHz Core2 Quad CPU and 3·0 GB of memory. The camera used in this experiment is a Microsoft Kinect. If E_{object} is 0, α_1, α_2 and α_3 are set as 0·1, 0·5 and 0·4, respectively. If the Kinect cannot detect depths, α_1 and α_2 are both set as 0·5. The size of the

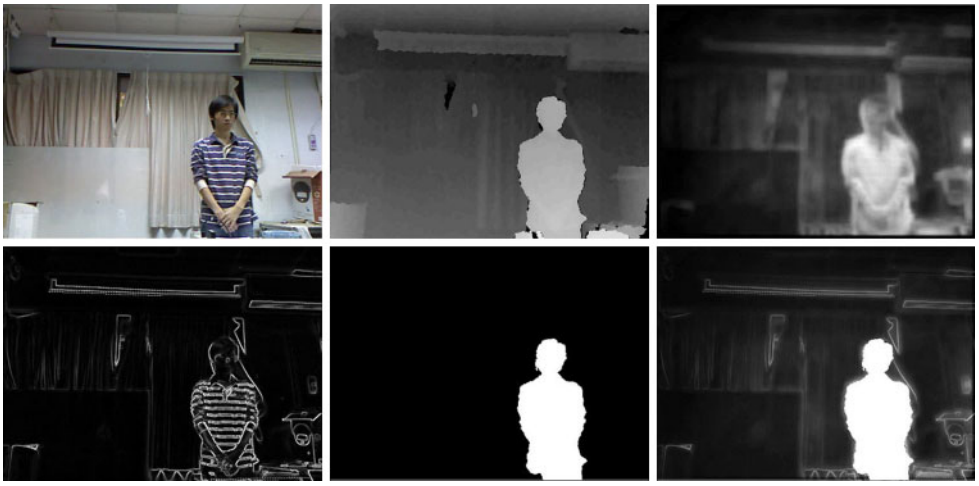


7 Original image, depth map, saliency map, gradient map, saliency object and importance map (from left to right and top to bottom)

original image is 585×430. Without loss of generality, a source image is resized in the horizontal direction only to make a target image. Resizing images in the vertical direction can be done in the similar way. Therefore, the sizes of the resized images are 500×430, 400×430, 300×430 and 200×430. Moreover, the proposed approach is compared to the two previous approaches.[1,6]

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

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

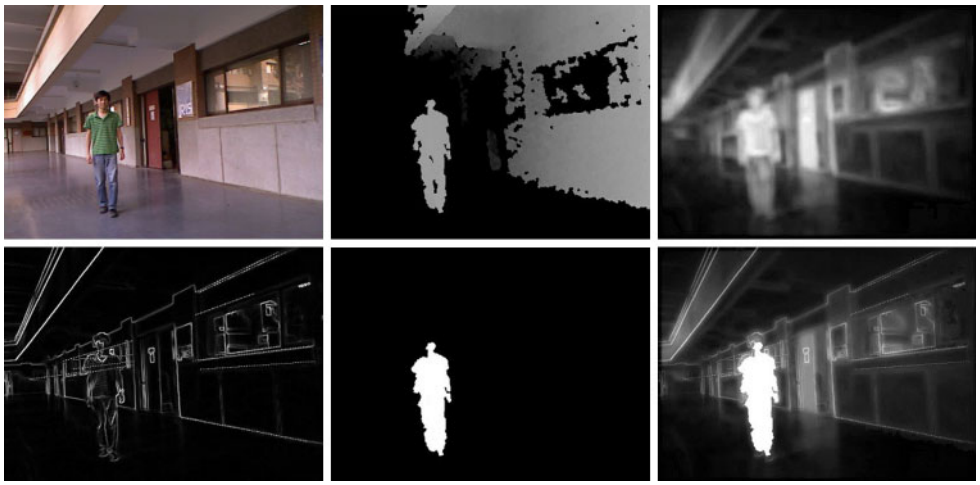


9 Original image, depth map, saliency map, gradient map, saliency object and importance map (from left to right and top to bottom)

results of Avidan and Shamir. The size of the whiteboard in the results of Wang *et al.* is over reduced.

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 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.

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



11 Original image, depth map, saliency map, gradient map, saliency object and importance map (from left to right and top to bottom)

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 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

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

colours 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, for 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.

Conclusion

This paper has proposed a novel image retargeting method for ranging cameras. Several analyses are 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 are integrated as an importance map for image retargeting. Unlike previous approaches, the proposed approach preserves the salient object well and maintains the gradients 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 is achieved.

In the future research, further studies are needed to develop an improved technique for video retargeting using the seam carving.

References

1. Avidan, S. and Shamir, S. Seam carving for content-aware image resizing. *ACM Trans. Graph.*, 2007, **26**, Article no. 10.
2. Hwang, D. S. and Chien, S. Y. Content-aware image resizing using perceptual seam carving with human attention model, Proc. 2008 IEEE Int. Conf. on *Multimedia and expo: ICME 2008*, Hannover, Germany, January 2008, pp. 1029–1032.
3. Kim, J. H., Kim, J. S. and Kim, C. S. Image and video retargeting using adaptive scaling function, Proc. 17th European Signal Processing Conf.: *EUSIPCO 2009*, Glasgow, UK, August 2009, University of Strathclyde, pp. 819–823.
4. Kim, J. S., Kim, J. H. and Kim, C. S. Adaptive image and video retargeting technique based on Fourier analysis, Proc. 2009 IEEE Computer Society Conf. on *Computer vision and pattern recognition: CVPR 2009*, 2009, IEEE Computer Society, pp. 1730–1737.
5. Scott, J., Tutwiler, R. and Pusateri, M. Hyper-spectral content aware resizing, Proc. 37th IEEE Applied Imagery Pattern Recognition Workshop: *AIPR 2008*, October 2008, IEEE Computer Society, pp. 1–7.
6. Wang, Y. S., Tai, C. L., Sorkine, O. and Lee, T. Y. Optimized scale-and-stretch for image resizing. *ACM Trans. Graph.* (*SIGGRAPH Asia 2008*), 2008, **27**, Article no. 118.
7. Fergus, R., Perona, P. and Zisserman, A. Object class recognition by unsupervised scale-invariant learning, Proc. IEEE Computer Society Conf. on *Computer vision and pattern recognition: CVPR 2003*, Madison, WI, USA, June 2003, IEEE Computer Society, Vol. 2, pp. II-264–II-271.
8. Gao, D. and Vasconcelos, N. Integrated learning of saliency, complex features, and object detectors from cluttered scenes, Proc. IEEE Computer Society Conf. on *Computer vision and pattern recognition: CVPR 2005*, San Diego, CA, USA, June 2005, IEEE Computer Society, Vol. 2, pp. 282–287.
9. Itti, L., Koch, C. and Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1998, **20**, 1254–1259.
10. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X. and Shum, H. Learning to detect a salient object. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **33**, 353–363.
11. Treisman, A. and Gelade, G. A feature-integration theory of attention. *Cogn. Psychol.*, 1980, **12**, 97–136.
12. Wang, D., Kristjansson, A. and Nakayama, K. Efficient visual search without top-down or bottom-up guidance. *Percept. Psychophys.*, 2005, **67**, 239.
13. Walther, D. and Koch, C. Modeling attention to salient proto-objects. *Neural Networks*, 2006, **19**, 1395–1407.
14. Hou, X. and Zhang, L. Saliency detection: a spectral residual approach, Proceedings of IEEE Computer Society Conf. on *Computer vision and pattern recognition: CVPR 2007*, Minneapolis, MN, USA, June 2007, IEEE Computer Society, pp. 1–8.
15. Rubinstein, M., Shamir, A. and Avidan, S. Improved seam carving for video retargeting. *ACM Trans. Graph.*, 2008, **27**, 1–9.
16. Wang, L., Jin, H., Yang, R. and Gong, M. Stereoscopic inpainting: joint color and depth completion from stereo images, Proc. IEEE Computer Society Conf. on *Computer vision and pattern recognition: CVPR 2008*, Anchorage, AK, USA, June 2008, IEEE Computer Society, pp. 1–8.
17. Achanta, R., Hemami, S., Estrada, F. and Susstrunk, S. Frequency-tuned salient region detection, Proc. IEEE Computer Society Conf. on *Computer vision and pattern recognition: CVPR 2009*, Miami, FL, USA, June 2009, IEEE Computer Society, pp. 2376–2383.
18. Goferman, S., Zelnik-Manor, L. and Tal, A. Context-aware saliency detection, Proc. IEEE Computer Society Conf. on *Computer vision and pattern recognition: CVPR 2010*, San Francisco, CA, USA, June 2010, IEEE Computer Society, pp. 1597–1604.
19. Kinect sdk(openni). Available at: <http://www.openni.org/>, accessed 31/07/2011.
20. Tomasi, C. and Manduchi, R. Bilateral filtering for gray and color images, Proc. IEEE Int. Conf. on *Computer vision: ICCV '98*, Bombay, India, January 1998, IEEE Computer Society, pp. 839–846.
21. Adelson, E. H., Anderson, C. H., Bergen, J. R., Burt, P. J. and Ogden, J. M. Pyramid methods in image processing. *RCA Eng.*, 1984, **29**, 33–41.