DSP Final Project Team 10
Remove the People: Segmentation-Based Object Removal

Members: 黃郁傑, 陳冠弘

Abstract. Removing objects in a photo is time-consuming and requires sufficient photo-editing skills. Such disadvantages may put people in a bad mood when some unwanted objects appear in their photos. In this work, we implement and improve an existing object removing method, and furthermore, come up with a novel way to generate segmentation mask based on different filters.

1. Introduction

Seam carving is a novel form of photo edition that resizes an image by defined energy function. In comparison to other traditional method like simply cropping, scaling, the property of seam carving utilizes the content of images and improves the performance of image resizing for both reduction and expansion. However, in the application of object removal, manually giving the mask is time-consuming and tedious. Such drawbacks discourage users to use this method to edit their photos. Accordingly, automatic mask creation is an important problem to be worked on. In this work, our main contribution is that we propose a novel method to automatically generate mask for object removing by using filter-based segmentation.

2. Approach

The procedure of our proposed algorithm can be described in Figure 1. The whole structure can be mainly divided into two parts── generate mask and perform seam carving. In the former part, first step is to filter out and capture the region proposal of people in the image. We use both low pass filter and median filter to find the people region proposal. Next step we can generate mask based on the region proposal from previous step. In the other part, given the image and mask, we can use seam carving alogorithm to remove people based on energy map calculated from the original image and mask.

2.1. Finding region proposal of people

2.1.1. Extracting base layer

To capture the region proposal of people, we assume that high frequency part has bad effect and it's better to be ignored. Hence, we extract the base layer of image which is shown in Figure 2. Besides, low pass filter kind of enlarges the region of people in the spatial domain, which makes it more robust to capture the whole region of people. In this work, we choose Gaussian function as the low pass filter. The function to separate base layer (L_B) from original image is formulated as,

$$L_B(i,j) = \frac{\sum_{k,l} L(k,l)w(i,j,k,l)}{\sum_{k,l} w(i,j,k,l)}, \text{where } \mathrm{w(i,j,k,l)} = \exp\left(-\frac{(i-k)^2+(j-l)^2}{2\sigma_s^2}\right),$$

2.1.2. Median distance based on image histogram

By observation, we find that for a scene photo with background and foreground (the people), the

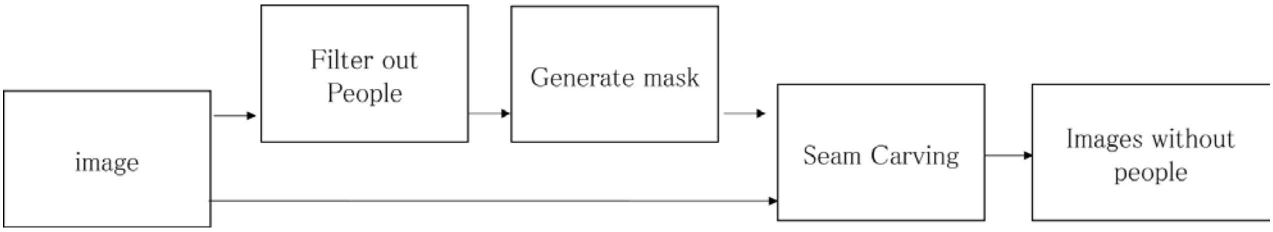


Fig. 1. Overview of our proposed algorithm

background is likely to dominate the median value of luminance and chrominance of the photo (See Figure 2). Hence, we assume that the region of people is likely to have higher difference between the median of the whole image histogram. As a result, we compute the difference between each pixel and the median of whole image in order to avoid this problem. In Figure 3., it shows the difference between each pixel and the median of whole image. The calculation of median difference (M_D) can be described as,

$$M_D = base - median(base)$$

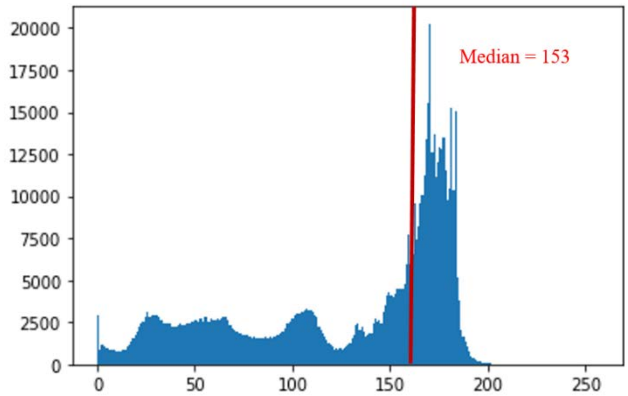


Fig. 2. Median of example base layer in Fig. 1

2.1.3. Vertical object detaction and filtering

Since we assume that the region of people is often to be a vertical object. We design a filter to check the whole image and give each pixel a vertical-ness score. In Figure 4, it shows the procedure of finding the vertical object response (V_D).

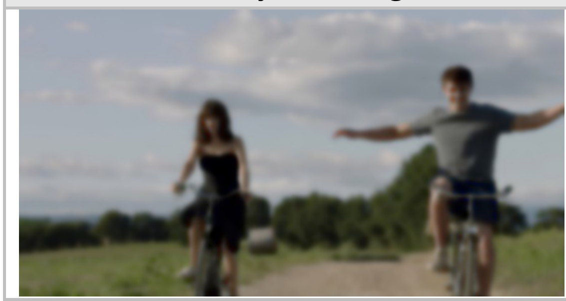
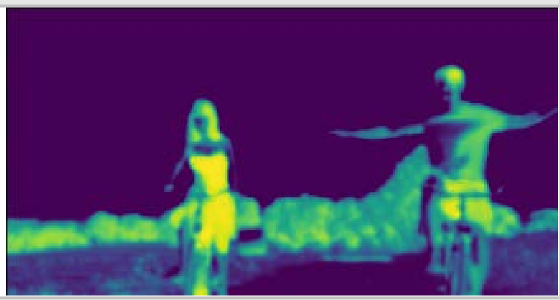
Base layer of image	Median distance of image
	

Fig. 3. Base layer and median distance of image

2.1.4. Weighting of filters

Though we design two methods to filter out the region proposal of people (using median distance and vertical object detection, respectively), the filters may be affected by some non-people object. For example, the vertical object detection filter may be affected by other vertical objects or texture of the background. To avoid this, we conclude that if a lot of pixels have high vertical-ness scores, it's not useful to use them to decide region of people. Therefore, we define a weighting mechanism to balance those two methods. That is, the final response ($F_{response}$) is fused by both vertical object detection response (V_D) and median distance response (M_D), and both with their coefficients according to their value. The formulation is as,

$$\mathrm{F}_{\mathrm{response}} = V_D * \left(1 - average(V_D)\right) + \; M_D * \left(1 - average(M_D)\right)$$

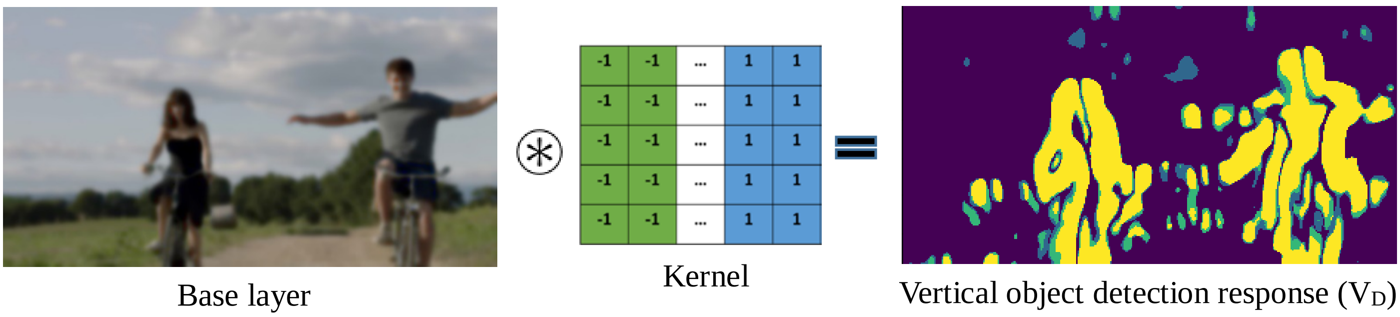


Fig. 4. Vertical object detection

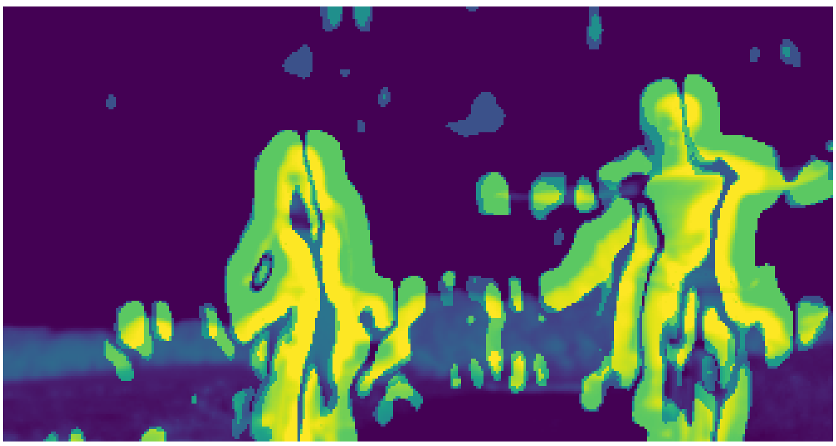


Fig. 5. Final response of whole image

2.2. Removing people by seam carving

2.2.1. Computing energy map

In this part, we are given the input image and the mask that we just obtain from the previous part.

Now we need to use the seam carving technique to iteratively decide which seam we want to remove. At the beginning, we need to use the energy function to compute the energy map of the image. In this work, we choose the Sobel filter as the energy function, which is to find the difference between adjacent pixels. If the difference is small, it's more likely to be removed. The computation of energy map can be formulated as,

$$p'_u = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * \mathbf{G} \quad \text{and} \quad p'_v = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * \mathbf{G}$$

The reason that there are two kinds of filters is because we want to emphasize both horizontal direction and vertical direction. The final score of a pixel will be the sum of the two scores.

2.2.2. Cooperating with the segmentation mask

After computing the energy map, we want to utilize the segmentation mask that we obtain from the previous part. One of the intuitive cooperation between them is that we just multiply them element-wisely. This is reasonable because the values in segmentation mask can play a role of indicating how the pixel is likely to be part of people. Suppose the region of people will have lower values in segmentation mask, after multiplication the same position of final energy map will have a scaled-down effect, which makes it more possibly to be chosen. In Figure 6., it shows the operation between energy map and segmentation mask.

Energy map Mask Final Energy Map

Fig. 6. Cooperation between energy map and mask

2.2.3. Carving the minimum seam iteratively

After calculating the energy map, we can now iteratively find the minimum seam, remove it and then keep these steps. The way we choose between all seams is that we start at each pixel of the top edge of the image, and try to traverse down until the bottom edge. During the traversal, we need minimize the total score that we visit. Once this step is done, we have scores of all seams, and we can now choose the seam with smallest score to be the victim seam. After removing it (each row will exactly be eliminated by 1 pixel), we keep doing the same thing until the number of pixels that we remove satisfies the condition.

3. Experimental Results

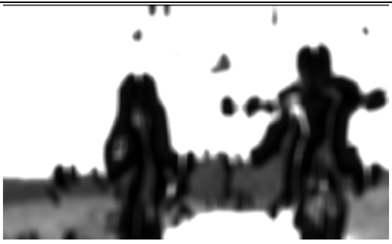
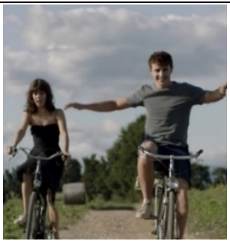
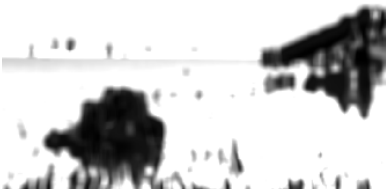
			
			
			
			
Input image	**Segmentation mask**	**Result (w/o mask)**	**Result (with mask)**

Table.1. Examples of generated images using proposed method. (Column 1 to 4 are input images, segmentation masks, results without using mask, results with using mask, respectively)

Table.2. Example of good results with segmentation mask
(left: original image; right: object-removed image)

Table.3. Example of bad results with segmentation mask
(left: original image; right: object-removed image)

4. Conclusion

In this work, we first implement an existing image resizing method based on seam carving, which utilize the content of image by constructing an energy map. When it comes to the application of object removing, it needs a mask to distinguish foreground and background which is tedious and not convenient for users. To achieve better convenience, we propose a method that generate the mask automatically based on segmentation by using filters. Our method integrates the object removing into an end-to-end procedure which first compute energy map and segmentation mask, then perform seam carving. Finally, some experimental results and analysis show the effectiveness and limitation of our method.

5. Reference

1. Shai Avidan, Ariel Shamir. Seam Carving for Content-Aware Image Resizing

2. David D. Conge, Mrityunjay Kumar, Rodney L. Miller, Jiebo Luo, Hayder Radha. Improved seam carving for image resizing