Interactive Background Blurring

Chih-Yu Yan Ming-Chun Tien Ja-Ling Wu
Communications and Multimedia Laboratory, National Taiwan University
{shadow, trimy, wjl}@cmlab.csie.ntu.edu.tw

ABSTRACT
Photographers usually take a shallow focus image to highlight the main subject in the picture and blur the distractions in the background region. Softening these distractions can not only produce special vision effect but also keep the privacy for people in the background. In this work, we develop a brand new defocusing algorithm to simulate the blur effect in the background caused by a wide aperture camera. We also design an easy-to-use interactive interface for user to segment foreground/background objects, adjust the depth of field for images, and assign camera settings. Techniques including lazy snapping, alpha matting, depth information generation and defocusing are integrated in the system. The experimental results show the effectiveness of the proposed methodology.

Categories and Subject Descriptors
I.3.3 [Computer Graphics]: Picture/Image Generation–Viewing algorithms

General Terms

Keywords
Image refocusing, depth of field, shallow focus, defocusing.

1. INTRODUCTION
Nowadays, we can easily take a lot of photographs by the convenient auto-focus shutter button on camera. There are usually a lot of distractions in the background of the photo we take, such as strange passengers, trash can on road, or other stuff which are not expected to appear in the picture. Some works removed these distractions by inpainting techniques [6,7], but in most cases, inpainting works poor when the removed object is large. Instead of removing these distortions, we can adjust the focus range in the picture, and make the unrelated or unimportant information softened, to highlight the subject of the photo. In the past, photographers adjusted the depth of field in the picture to soften the noisy background. They made the depth of field shallower to blur the background, and named this technique as shallow focus. Shallow focus highlights the main subject while softening distractions in the background with a shallow depth of field by using a wide aperture or long length lens. However, common digital camera with limitations in sensor and machine size/weight usually gets a deep focus which results in the failure of achieving background blurring by adjusting the depth of field.

Here we develop a system, for photo refocusing by constructing the depth information of the objects and the background in an image with simple user interaction. The depth information is then used for reconstructing a shallow focus scene to emphasize the subject while soften background distractions.

2. RELATIVE WORK
In [1], the authors gave a new way for image refocusing. They projected light dots on an indoor scene by a projector, and computed the diffusion level of light dots to estimate the depth information. And then they used the depth, color, and texture information to segment image into some parts for image refocusing. The main limitation of this work comes from the condition that the scene must be indoor with a uniform illumination; moreover, projecting light dots is inconvenient and cannot work under bright sunshine. After all, the technique is not suitable for photos taken by common travelers.

There are other ways for doing image refocusing. For instance, one can add a pattern within camera aperture of camera lens to create a coded aperture [2]. After decoding the coded aperture pattern, a high resolution depth information of the image can be obtained. If the noise ripple of the image can be removed, the decoded pattern reveals the accurate depth. Or one can estimate the optical transmission in hazy scenes [3]. During recovering a hazy-free result, depth information is also obtained. Those works still need additional hardware equipments or obeying special scene limitations. Hence, some works utilized two or more pictures to extract extra information, and then reconstruct the depth information using the motion estimation or perspective difference between a pair of images.

As depicted in the above related works, the most important feature of image refocusing is the depth information of every pixel, and therefore, researches on image refocusing have made a lot of efforts to get the required depth information. However, they still have some constraints such as demanding active projecting light, pattern on lens, special condition in the scene or multiple view images. It is very difficult and problematic to find an exactly accurate depth map of an existing image without any other hints.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MM’09, October 19-24, 2009, Beijing, China.
Copyright 2009 ACM 978-1-60558-608-3/09/10...$10.00.
3. THE PROPOSED METHOD

The proposed system consists of three modules including object segmentation module, depth map generation module, and image defocusing module. First, users take strokes to mark background and foreground objects, and then the well-known Lazy Snapping technique [5] is employed to roughly segment the image. The proposed system also combines other soft matting methods to enhance the segmentation results. Without user’s strokes, we let face-detection result to be a default setting. After giving a vanish-point and drawing a rear-wall by the user, the system will construct the depth information of the background scene. The foreground objects are then placed into the background scene with the given depth like a sequence of pop-up cards (see Figure 2).

Finally, the system simulates the lights passing through camera aperture to camera sensor, diffusing to a blur shape with its blur diameter which is determined by the object depth, camera focal length, and lens f-number. Furthermore, we can change the depth of field and the focusing part of an image to get a new view and extend the proposed system to other image browsing applications.

3.1 Object Segmentation

Lazy Snapping [5] is a hard-cut matting tool which achieves foreground/background segmentation by computing foreground/background distance probability of each pixel with few user hints. The user gives foreground and background strokes, and then a binary image map is returned in which foreground pixels are marked as 1 and background pixels are marked as 0. But, at the boundaries between objects and the background, there are often some pixels lie in an indistinct region which need applying alpha matting techniques to avoid artifacts. Those pixels are not exactly foreground or background.

We first use mean-shift clustering to cluster every pixel to a certain cluster. The features used for mean-shift clustering are position and color information of the pixel. The clustered result is then utilized in Lazy Snapping with user strokes. Considering computational complexity, we use a down-sampled image for conducting the mean-shift clustering.

If there is no user hint for segmentation, the system will use the default setting which assumes the subjects of the picture are people, hence the face-detection technique is applied to mark the positions of human faces. The system then extends each face to a subject region with a body bounding box specified by a predefined ratio of face area to body area. Foreground strokes are automatically added to the center of each face and inside the body bounding-box, while a background stroke is added outside the box.

After getting the hard-cut map by lazy snapping for each important object, we mark the boundary of an object as “uncertain”. Combining the “defined foreground” and the “defined background” parts, a tri-map is obtained. And then we use this tri-map as the initial setting data for doing alpha matting [4]. Finally, we get a new segmentation map with alpha matting information for every object.

3.2 Depth Map Generation

We designed an interface for users to easily paint a rear wall of the background and set a vanish point for the input image. The rear wall is a plane vertical to the camera, and the rear wall or the vanish point might be outside of the image. By computing the related distance between the rear wall and the vanish point, we can fold the image to a 3-D box, as shown in Figure 3.

3.3 Image defocus

Most refocusing researches use Gaussian blur filter or pill-box filter to defocus the region required to be blurred. In this work, we achieve out-of-focus blur by summing the weighted strength of all nearby light spots.

Equ.(1) is the well-known Thin Lens Equation, where \( u \) is the object distance, \( v \) is the image distance, and \( f \) denotes the focal length of the lens.

\[
\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \quad (1)
\]

If the diameter of the aperture is \( d \), lights from an object with depth \( u_b \) diffuse to a blur spot with diameter \( b \) on the image plane with distance \( v \), and its image distance is \( v_b \). As illustrated in Figure 5 we can find the relationships among \( b \), \( v \), and \( v_b \) by using the triangle similarity theorem.

We can replace the aperture diameter \( d \) with the lens f-number \( N \) and lens focal length \( f \), (i.e. \( d = f/N \)), and substitute the object distance for the image distance. Therefore, \( b \) can be written as:

\[
b = \frac{f}{N} \left( \frac{u_b - u}{u} \right) \quad (2)
\]
Figure 4. Geometry of imaging when (a) object is in front of the focus and (b) object is behind the focus.

Let $m$ denotes the image magnification which is related to the object distance and the image distance (i.e. $m = v/u$). Rearranging eqn. 1 we have $m = v/u = f/(uf)$. Finally, the diameter of the blur spot $b$ can be obtained as:

$$b = \frac{fm(u_b - u)}{u_b}$$

Moreover, the shape of the light spot depends on the aperture type. For example, a camera with f-1.4 aperture/1.4-50 mm lens has a heptagonal light spot shape. With the obtained blur spot diameter $b$ and a pre-selected aperture type, we can simulate the blur phenomenon of light ray passing through the camera aperture and projecting onto the sensor. In realistic cases, digital camera sensors respectively convert lights into red, green, and blue digital signals by applying corresponding color filters; therefore, we use RGB color space to conduct our simulation.

Assume the original image intensity $I(x,y)$ implies that the energy of light ray projecting on the pixel $(x,y)$ without being diffused to a blur spot. As shown in Figure 5, when the light ray is diffused to a blur spot of shape $s$ with diameter $b$, which covers a set of pixels denoted as $Region(x,y,s,b)$, each pixel $p(i,j)$ in $Region(x,y,s,b)$ will receive an amount of energy $E(i,j)$. Since the sum of all diffused energy inside $Region(x,y,s,b)$ should be the same as $I(x,y)$, $E(i,j)$ is inversely proportional to the spot area.

As illustrated in Figure 6, considering all diffused energy contributing to the pixel $p(i,j)$, the intensity of $p(i,j)$ after defocusing can be calculated by

$$I_{\text{defocus}}(i,j) = \frac{\sum E_n(i,j)}{\sum \text{weight}_n(i,j)} + \sum \alpha_n(i,j)$$

where

$$E_n(i,j) = I(x_n,y_n)/\text{area of Region}(x_n,y_n,s,b)$$

and

$$\alpha_n(i,j) = \begin{cases} \text{weight}_n(i,j), & I(x_n,y_n) < Th_{\text{bokeh}} \\ 0, & I(x_n,y_n) \geq Th_{\text{bokeh}} \end{cases}$$

Figure 5. A blur spot.

The diffusion condition in the dark part of an image is not as obvious as that of the bright counterpart. Hence we can also add some additional bonus in the bright part to tune the weight function.

In a realistic scene, the energy value is continuous and sometimes larger than the highest value that a digital camera sensor can represent, say $C_{\text{high}}$. “Bokeh” effect occurs when the energy overtakes $C_{\text{high}}$, and we simulate this effect by setting a threshold $Th_{\text{bokeh}}$. If the intensity value exceeds $Th_{\text{bokeh}}$, an additional weight $\alpha$ is added to emphasize the bokeh phenomenon. Furthermore, when we put focus on farther depth in realistic scenes, “shrink” effect (as shown in Figure 7) occurs on the boundaries of near foreground objects due to the near-object blurring. In this case, we apply texture synthesis for partial inpainting, as proposed in [1], to extend the lost information behind the near object.

Figure 6. Diffused energy from all around pixels.

Figure 7. When the front object is out of focus, its opaque part will “shrink”. (a) the case for focus on farther depth, (b) the case for focus on the flower, and (c) the resultant shapes of flowers in case (a) and (b).

Figure 8. Close-up results of (a) Gaussian blur, (b) our defocusing method, (c) our method considering bokeh, (d) original image taken by a digital camera (DC), and (e) the same scene taken by DSLR.

4. EXPERIMENTAL RESULT

In Figure 8, we observed that the result of our method, with considering bokeh, is more close to the picture taken by digital single-lens reflex camera (DSLR) than that of the conventional approach on the basis of Gaussian blur filters.

Figure 9 (a) shows a picture with deep focus. The vines in the background construct a rich-textured scene. According to the user hints (strokes, a rear wall and a vanish point), we construct a depth map and defocus the out-of-focus region. In this case, the aperture shape is a circle disk, the f-number is 1.4, and the focal length is 100mm. The farthest distance in the image is 50 m and the person is positioned at 1.8m from the camera. Figure 9 (c) displays the shallow focus result.
If we give the background a fixed depth information without considering its depth variation, the blur level of the background will be a constant. The results of defocusing with and without depth variation are shown in Figure 10(d) and Figure 10(e), respectively. The girl in Figure 10(e) looks like standing in front of a wall and floating on the scene. On the contrary, the girl in Figure 10(d) blends to the scene better than the other one.

In Figure 11, the flower shrubs in front of the girl can be treated as an object with depth rather than a wall. The depth variation of a foreground object can be adjusted by giving a normal vector, as shown in Figure 11 (b). In the shallow focus result (Figure 11 (c)), the flower shrubs will not in a plane. In this example, the focus object is at a depth of 16.5 m; therefore, the farther background part will be blurred in a small level and the nearby part of the flower will demonstrate a highly blur effect.

In the case of Figure 12, the focus is put on the farthest depth in the image. The opaque part of the object in front of the focus will shrink and the texture behind the object cannot be found in the original image. Figure 12 (b) is the defocusing result without considering the background texture information. The shape of the foreground object keeps unchanged despite details inside the object are blurred. In our method, we first ignore the foreground object and use texture synthesis to inpaint the region in the background occupied by the object. We defocus the inpainted image and the nearby object respectively. The weight function of the nearby object in eq.(5) is then normalized by its maximum, and utilized as an alpha map for interpolating the blurred nearby object and the inpainted background, and the result is shown in Figure 12 (c).

5. CONCLUSION

We developed an interface for user to input some hints for an image. According to these hints, the system can adjust the depth of field and produce a shallow focus image. Without the need of additional hardware or limitations in illumination, just some simple user hints the proposed system provides a large flexibility in the applicable scenes. In our future work, we will add objects from other pictures to the target image, and apply this tool to other photo browser to increase the amusement for the user.

6. REFERENCES


