Depth Estimation of A Monocular Image from Image Blur

Zhonghai Deng, Jingyuan Zhang

1Computer Science Dept. University of Alabama Tuscaloosa

Abstract— In computer vision, estimating the depth information of object from an image/video has always been a hot topic, and finds many applications. In this paper, we make use of the image blurs caused by defocus aberration to estimate the relative distance in an object. To be more computationally efficient, we focus on the blurs along image edges. Experiments shows the effectiveness of our proposed method.

1. Introduction

In computer vision, one essential problem is to reconstruct the 3-D scene from images. This could be simplified to estimate the distance of an object from captured still image. However, due to the information loss from 3D scene into 2D color image, this task seems un-solvable.

Robert etc. [1] uses the blur to analysis its affect on the perceived distance and size. They propose a probability model to help us understand how the blur indicates the apparent scale of the image’s contents. Based on this assumption, they use a Bayesian model to understand the distance. However, their result are judged, and compared with the human viewers responds, thus cannot give a confident and convincing result.

Another way, to estimate the image depth is done by Saxena etc. [3], where they use the image structure to find the relative distance for each pixels. By using a Markov Random Field (MRF) model, the give a more accurate and visually more pleasing result.

In optics and camera industry, the traditional way to compute depth is taking two images, one is a sharp image of the scene taken with a larger depth of field, while the other is blurred with varying the focal length.

One solution require using one image as a benchmark, and estimate blurs from the other one in various image points and compute its depth map/image of the scene accordingly. However in most of the times, people hardly have the benchmark image. Even in the real life, instead of like a which is actually a monocular image due to the fact that digital camera has only one ‘eye’.

The ‘focal length’ is the distance between the lens plane and the sensor plane which is usually stated in millimeters. For cameras with zoom lenses, both the minimum and maximum focal lengths would be provided in the EXIF header file, for example 18âÅâ55 mm. Though sometimes, they would be replaced by the Depth of Field (DOF), indicating the distance range that objects could be sharp enough. Another property of the camera is the angle of view, which is the visible area of the scene captured by the lens, stated as an angle. Wide angle of views captures larger areas, small angles gets smaller areas. Also, changing the focal length would automatically changes the angle of view.

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

Previously, researchers have used the defocus to compute depth or range, but mainly for the purpose of lens focus adjustment. [1] used the gradient in blur to compute range. [2] generalized it and extended the application of the defocus information for range computing while p241-cho present a matrix based method to estimate the camera shake blur. To compute depth from image blur, all these approaches utilize special features in scenes (or require texture in a small band of spatial frequencies). Also, all of them aim to adjust the focus of the camera, i.e. they use more than one image to do the range estimation, and adjust the camera settings accordingly.

\[
e = (s, f, D)
\]

if we define

\[
q = \frac{2R}{D} \frac{s - v}{v} = s \left[ 1 - \frac{1}{s} \frac{1}{v} \right]
\]

Our method works in the opposite way. By revisiting the current problem, we propose a depth re-estimation based on an assumption that all image blur comes from the lens ill-focus. The method does not place any restrictions on the scene other than what is normal for passive depth imaging methods, i.e. we do not require that the scene must contain some texture or other visual features. For any method lack of features or texture should cause graceful degradation, however. We introduce a confidence which indicates the reliability of the computed depth estimates. Actually, we believe that the use of a confidence measure is a key point for the usefulness of the idea.

2. Blur Measurement

In today’s commercial digital camera, the most important part is the sensor, which transfer the light signals into the electric ones. Before that, camera need a lens system to focus the lights onto the sensor. To better control this focusing process, the lens system has four degrees of freedom: shutter speed (exposure time to sensor), focus (position the sensor should be), focal length (zoom), and aperture (diameter of the lens cover).
For simplicity, we assume an optical system is only a thin convex lens (see Figure 1). Here, the lights coming from the out-world scene, are focused on to the light detector (sensor). We use the $f$ for the focal length, $s$ for the distance from the sensor to the lens, $v$ for the distance from the lens to the focal plane. The scene object is placed at a distance $u$ from the lens.

From Figure 1, we can see that sometimes, when the lights are focused on position $p'$, the image plane is positioned behind, thus the lights coming from $p$ are longer focused on a single point on the image plane, forming a blurred circle, with diameter $R$. The similar case when the object plane is closer to the lens than the focal plane $p'$.

Fig. 1: Imaging Formation in a Convex Lens (Reproduced from [4])

We know that the lens can only focus at a certain point, therefore, when recording images of 3D scenes we have the situation in Figure 1, the objects in the scene appear more or less blurred on sensor. Thus many researchers take two consecutive images, and measure the change in blur from the first image to the next we may compute the depth. However, in our setting, we have only one image available, i.e., without the benchmark image, we can only use this monocular image for depth estimation.

In our experiments, instead of estimating every pixel point for each pixel, in this paper, we try to estimate the absolute image distance for each image object. And this distance within objects are interpolated for computational consideration.

2.1 Blur Model

An edge is modeled as a step function $Au(x) + B$ of unknown amplitude $A$ and pedestal offset $B$, which, for the purposes of this discussion, will be aligned with the y-axis of the image coordinate frame. The focal or penumbral blur of this edge is modeled by the Gaussian blurring kernel

$$g(x, y, \lambda_b) = \frac{1}{2\pi\lambda_b^2}e^{-\frac{(x^2+y^2)}{2\lambda_b^2}}$$

where $g(x, y)$ is the pixel value at an image, $\lambda_b$ is the distance constant factor that generate the blur function. Estimating the sensor noise statistics for an imaging system is relatively straightforward. For the system used in Fig. 2, a region of a de-focused image of a plain flat surface was first selected.

Fig. 2: Imaging Blur Type due to Imperfection Focusing

Our model predicts that the camera imaging system estimates absolute distance of the object at focus by finding the depth of focus from header file.

three types of blur: (1) blur that is completely consistent with the relative distances in a scene (consistent-blur condition),

Thus, to simplify our computation, and get a more smooth result, we assume that in the perfect condition, all edges should be sharp, i.e., edges of objects at focus should be only one pixel size. (In real world, perfect edge should be of no width theoretically.)

3. Depth Estimation via Image Blur

Theoretically speaking, for each pixel values of an image, there exist an depth value. However, due to the various types of constrains, it is impossible, and unnecessary to do the depth estimation for each pixel.

So, in this paper, we first segment the image into different objects, and estimate these value according to their focused point. And then, we estimate objects in other parts to access their depth value. We do it by using the blurs of their contouring edges.

3.1 Extract Image Focal Length

Every image file from a digital camera comes naturally with a header file. Though its format varies from brand to
brand, model to model, it generally contains some basic information, like image file name, image size, number of colors, file size, file type and so on. Another feature that must be included is the 'Focal Length', which usually uses ‘mm’ as the measurement unit.

In our experiments, we use the ExifTool to extract the this focal length. There are two other measures that we are also interested. One is the $F$ number, meaning the $\frac{f}{D}$ value. The other value is the image depth of field (DOF) value of an image, though not

In optics, DOF is the distance between the nearest and farthest objects in a scene that appear acceptably sharp in an image. Although a lens can precisely focus at only one distance at a time, the decrease in sharpness is gradual on each side of the focused distance, so that within the DOF, the unsharpness is imperceptible under normal viewing conditions.

3.2 Inference Distance from Blur

So, we first expose the image part, or objects that is on focus, which is achieved by comparing the sharpness metric inside each image part.

And then, the absolute depth value could be calculated as

$$u = \frac{fv}{v - f}$$

Where $f$ is the focal length, which can be extracted from the image EXIF header. And from figure 1 we have

$$\frac{2R}{D} = \frac{S}{v}$$

Where $D = \frac{f}{F_{num}}$, and $F_{num}$ could also be extracted from the EXIF info. Thus we have

$$v = s * \frac{2R + D}{D}$$

As we are only interested in the relative depth of the image, i.e. the absolute value of the object at focus is not needed. Therefore, in the following sections, we shrink the value $s$, so that distance of the object at focus is set to be 1, and other distances are compared with it.

From equation 3, we can see the relationship between camera optics and image blur. Also, the human visual system uses of retinal-image blur to adjust the focal length. For instance, if our target object is blurred, our pupils would automatically adjust the focal length so that the focused point would fall onto the retina.

From above figure, we can see that the more blur occurs, the larger radius $R$ projected on the sensor plane.

Also, these blurs contain other depth cues if we can determine the objects at focus, and compare them with the rest, (i.e. linear perspective, relative size, etc.). This information help to specify the relative distances among objects in the scene. Although this distance are scale ambiguous, thus we seems cannot directly calculate the absolute distances of the interested objects. We can actually, use the $f$ number to help determine this value.

4. Experimental Result and Analysis

In this part, we apply our algorithms on the images taken by commercial cameras. However, to be more comparable with current depth estimation result, we use the well-known peppers image.

4.1 Determine the Beacon Distance of Object at Focus

From the image EXIF header, we can always extract the image depth of field (DOF) value of an image, as well as the focal length. Also, there is also a $F$ number, which means the $\frac{f}{D}$ value.

So, we first expose the image part, or objects that is on focus, which is achieved by comparing the sharpness metric inside each image part.

And then, the absolute depth value could be calculated as

$$u = \frac{fv}{v - f}$$

Where $f$ is the focal length, which can be extracted from the image EXIF header. And from figure 1 we have

$$\frac{2R}{D} = \frac{S}{v}$$

Where $D = \frac{f}{F_{num}}$, and $F_{num}$ could also be extracted from the EXIF info. Thus we have

$$v = s * \frac{2R + D}{D}$$

As we are only interested in the relative depth of the image, i.e. the absolute value of the object at focus is not needed. Therefore, in the following sections, we shrink the value $s$, so that distance of the object at focus is set to be 1, and other distances are compared with it.
4.2 Determine the Beacon Distance of Object out of Focus

In our experiments, we set only one object at focus, and the rest would be depth estimated from their blur value. For computational consideration, we infer the depth for the edges, and the inner point would be interpolated from edges.

Due to the fact that traditional edge finding algorithms tends not to find a closed area for an object, we use the contour detection and hierarchical segmentation algorithm borrowed from ??, to find the boundary, and estimates how much blurs along the boundary.

4.3 Analysis of Detection

Due to the fact that we do not have the ground truth depth map, we only print the depth information from the image. But this depth estimation method produce a convincing result on its object depth estimation. However, our method does not perform well on the case that for image parts out of DOF (Depth of Field), it would produce a unreliable result. Also, no bench-mark database could be used to gain the comparison with other methods.

5. Conclusions and Future Work

In this paper, we propose a novel image blur estimation method, and estimate the depth from a single blurred image. We have also reviewed the achievements, discuss the experimental result. Even though our proposed method can give an exciting estimation result, the depth maps it produces are not always satisfying. Also, one drawback of our proposed method is that it cannot differ the blurs caused by before/behind focal plane (i.e. too near, and too far from focused point). This problem might be solved by combining other methods, like structure inference.

References


