

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,500

Open access books available

118,000

International authors and editors

130M

Downloads

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



Chapter

Depth Extraction from a Single Image and Its Application

Shih-Shuo Tung and Wen-Liang Hwang

Abstract

In this chapter, a method for the generation of depth map was presented. To generate the depth map from an image, the proposed approach involves application of a sequence of blurring and deblurring operations on a point to determine the depth of the point. The proposed method makes no assumptions with regard to the properties of the scene in resolving depth ambiguity in complex images. Since applications involving depth map manipulation can be achieved by obtaining all-in-focus images through a deblurring operation and then blurring the obtained images, we have presented methods to derive all-in-focus images from our depth maps. Furthermore, 2D to 3D conversion can also be achieved from the estimated depth map. Some demonstrations show the performance and applications of the estimated depth map in this chapter.

Keywords: depth estimation, blur estimation, depth from defocus, all in focus, refocusing, defocus magnification, 2D to 3D

1. Introduction

Derivation of depth information from 2D images is one of the most important issues in the field of image processing and computer vision. The depth information can be applied in 2D to 3D conversion, image refocusing, scene interpretation, the reconstruction of 3D scenes, and depth-based image editing. There are some techniques used to derive depth information, such as depth from focus [1], stereo vision [2], and depth from motion [3]. Nevertheless, these techniques are complicated by the need to acquire multiple images, thereby making them impractical when only one image is available or the features corresponding between the images cannot be resolved well. To this end, a number of approaches have been proposed to acquire depth information from a single image, such as the computational photography approach [4], which modifies the shape of the aperture of a traditional lens, and the Kinect approach [5], which uses a structured light to derive depth maps.

An image captured by a conventional camera contains a blurred version of a scene that is out of focus. The blurriness of a pixel is called the “circle of confusion” (COC) and is usually modeled as a 2D Gaussian function. When a single image is taken by a conventional camera with a fixed focal length, aperture size, and distance from the image plane to the lens, a pixel’s COC is related only to the depth of the corresponding scene point. In such cases, depth estimation corresponds to blur estimation. Theoretically, if the depth map of an image can be accurately estimated, applications that manipulate depths of objects can be run by first applying

deblurring followed by blurring operations. This is because the deblurring operation will move the objects closer and the blurring operation will move them farther away from the camera.

Blurring operation is more robust to depth map inaccuracy than the deblurring one, and many applications have been successfully built based on this operation. For example, defocus magnification [6] increases the out-of-focus area in an image by magnifying the existing blurriness to keep the shape of sharp regions and by modifying the depths of objects that are not in the focal plane to move the objects farther away from the plane. Deblurring operation, on the other hand, can be very sensitive to the accuracy of a depth map. A deblurring operation usually highlights the edged and textured points in an image. If their depths are overestimated, the operation can generate ringing artifacts that severely degrade the perceptual quality of an image.

The depth map estimation from an image is a fundamentally ill-posed problem. For example, in a single image, we cannot resolve the ambiguity between out-of-focus edges and the original smooth edges, we cannot determine whether blurriness of a point is in the front of focal plane or behind the focal plane, and we cannot estimate the depths of points in a smooth area. These problems cannot be resolved without the assumptions between the local image features and the scene. In this context, a widely adopted assumption is that a blurred edge is obtained via smoothing a step edge with a Gaussian kernel [7]. Although this assumption has been adopted in an autofocus system of a camera, the goal of the autofocus is to derive the depths of manually selected scene points rather than the depth map of an image. The approach that is based on the assumption on scene points has also been used in estimating a depth map. Edges in a scene are first modeled, and the depths of the blurred edge points are then derived from the degree of blurriness that has been applied on the scene to obtain the points. However, because many types of

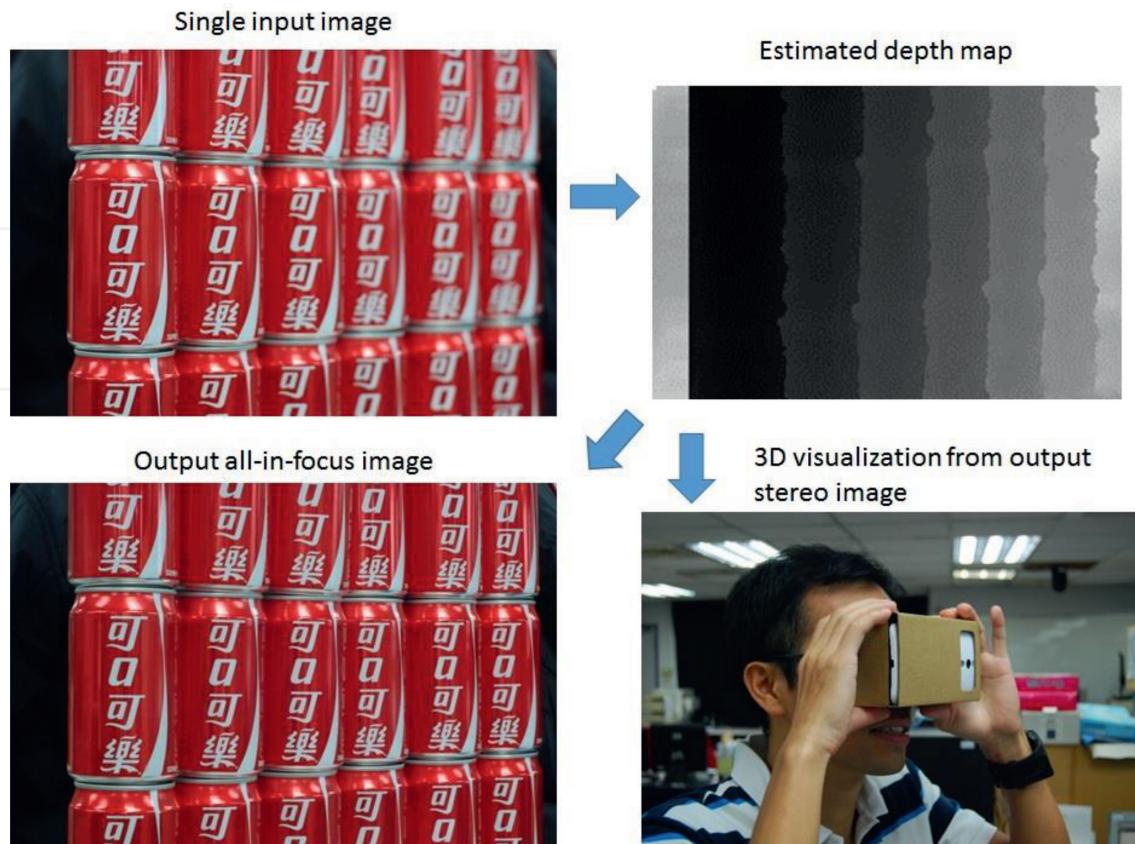


Figure 1.
Applications from the depth map of a single image.

singularities far beyond the step edges in the scene can appear in an image, the approach based on the scene modeling can be too restricted, as only a few types of singularities can be modeled, to derive precise depths of all points. As a result, the depth precision derived based on scene modeling is usually limited to images with two depth layers, foreground and background.

In this chapter, we propose a blurring-deblurring method that does not require the modeling of edge points. In the blurring process, a point is blurred by increasing its COC to the limit of a camera. However, in the deblurring process, a point is deblurred by reducing its COC to the limit in the other end. Combine the results of these two processes and derive the depths of edges. Therefore, the approach estimates the depths of edge points based on the characteristic curve of COC vs. the depth characteristic curve of a camera. We demonstrated that the proposed approaches can reliably derive depth maps of complex images and synthesize all-in-focus images. Furthermore, the depth maps can also be applied to synthesize the stereo image to 3D visualization through the mobile device. **Figure 1** shows a diagram of applications from the depth map of a single image.

The remainder of this chapter is organized as the following. The relationship between the depth of a point and out-of-focus blurriness in images obtained by the thin-lens camera model is reviewed in Section 2. The proposed blurring-deblurring approach is presented in Section 3. The depth refinement approach and image deblurring process are presented in Section 4. In Section 5, we demonstrated the depth map results and applications. Section 6 contains some conclusions.

2. Camera model and out-of-focus blurriness

The out-of-focus blurriness is defined by the COC if an object is not in the camera's focal plane. However, it is impossible to determine whether an object is behind or in front of the focal plane based on the blurriness of an object [8]. In the following, we consider a case of the condition.

In a thin-lens model,

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}, \quad (1)$$

where u is the distance between the lens and the scene point, v is the distance between the focal plane and the lens, and f is the focal length of the lens. If a light point is not in the focal plane but placed in front of the camera, the source's image will be a circular disk with diameter D_{COC} instead of a point, as shown in **Figure 2**. Let d be the distance between the lens and the image sensor. Then, the in-focus scene distance $u_{in-focus}$ can be derived as follows:

$$\frac{1}{u_{in-focus}} + \frac{1}{v} = \frac{1}{f}. \quad (2)$$

For a particular lens, the focal length f and the aperture A are constants; the F-number $N = \frac{f}{A}$ is also a constant. Given the geometric relationship $D_{COC}(u)$ shown in **Figure 2** and the lens formula, the COC's diameter of a scene point at distance u from the lens depends on whether $u > u_{in-focus}$ (the scene point is farther from the lens than the focal plane) or $u < u_{in-focus}$ (the scene point is closer to the lens than the focal plane).

In the case where $u > u_{in-focus}$, we can derive the following relationship from the similar triangles shown in **Figure 2(a)**:

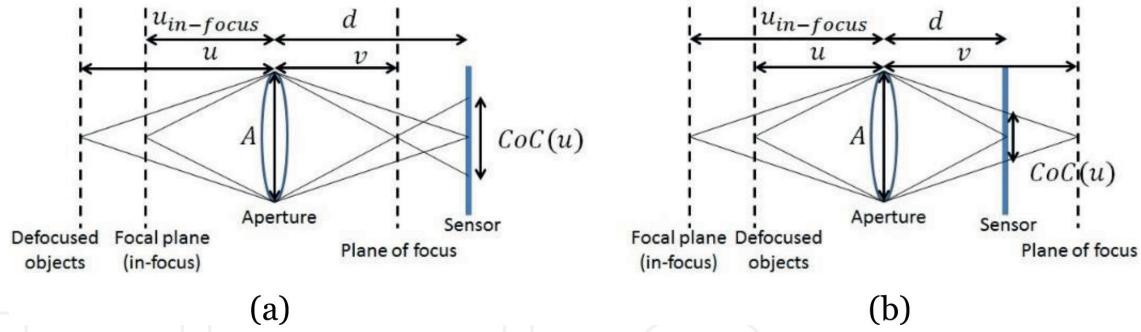


Figure 2.
The geometry of imaging: u is the distance of a scene point from the lens, $u_{in-focus}$ is the distance of the focal plane from the lens, d is the distance between the lens and the image sensor, and the diameter of the lens' aperture is A .
(a) $u > u_{in-focus}$ and (b) $u < u_{in-focus}$.

$$\frac{D_{COC}(u)}{A} = \frac{d - v}{v}. \quad (3)$$

Using Eq. (1) and $N = \frac{f}{A}$, we obtain

$$D_{COC}(u) = \left(-1 + \frac{d}{f} - \frac{d}{u} \right) \frac{f}{N}. \quad (4)$$

In the case where $u < u_{in-focus}$, we can derive the following relationship from similar triangles shown in **Figure 2(b)**:

$$\frac{D_{COC}(u)}{A} = \frac{v - d}{v}. \quad (5)$$

Using Eq. (1) and $N = \frac{f}{A}$, we obtain

$$D_{COC}(u) = \left(1 - \frac{d}{f} + \frac{d}{u} \right) \frac{f}{N}. \quad (6)$$

From Eqs. (4) and (6), we can derive D_{COC} of a scene point; however, the equations do not allow us to determine whether the scene point is in front of or behind the focal plane. To remove the ambiguity, the assumption that all the scene points are behind the focal plane is adopted.

An image is usually modeled as the convolution of the scene and a camera-relative PSF. The Pillbox function is an ideal PSF, which is a box function with support σ and constant value $\frac{1}{\sigma}$. Usually, the Gaussian function is the approximation of the Pillbox function, and the standard derivation of the Gaussian function is $\frac{\sigma}{\sqrt{2}}$. Due to the factor, the difference between their frequency domain magnitudes is small, and the latter is easier to do analysis. In this chapter, we will use the Gaussian function to characterize the PSF of a camera.

3. Blurring and deblurring approach

Using the proposed approach, scene depths will be determined from the estimated blurriness in a single image by using a combination of blurring and deblurring processes. In this section, we will explain the rationale for combining the

blurring and deblurring processes and provide the formulation of the combined approach. The depth of a scene point is defined as the distance between the camera lens and the point. In addition, the proposed method assumes that all the interested scene points are behind the focal plane (it matches to the case: $u > u_{in-focus}$) as [9].

3.1 Concept

The (D_{COC} vs. u) curve of Eq. (4), illustrated in **Figure 3(a)**, gives the relationship between the depth of a scene point and its D_{COC} value of a camera. The latter increases with the depth of the scene point. When D_{COC} reaches its limit (D_{COC}^*), the point can be assumed to be at infinity.

Let D_{COC} be the blurriness of a point. A blurring operator can be defined to add an increment of blurriness to the point to obtain a new blurriness $D_{COC} + \delta(D_{COC})$, with $\delta(D_{COC}) > 0$. This can be regarded as increasing the depth of the point by moving it along the (D_{COC} vs. u) curve toward the right end point of the figure. If the blurring operation is applied repeatedly, the blurriness can reach D_{COC}^* and the point is at infinity.

If the increment in the blurriness of a point to reach D_{COC}^* can be determined, we can convert this increment into the increment in depth by referring to the curve (D_{COC} vs. u) in **Figure 3(a)** and derive the true depth of the point. However, as shown in the curve $\frac{\partial u}{\partial D_{COC}}$ of **Figure 3(b)**, a small increment in blurriness close to D_{COC}^* yields a substantially large increment of depth. This means that the depth determination close to D_{COC}^* is relatively unstable and inaccurate.

On the other hand, the deblurring operator is defined to reduce the blurriness of the point to obtain $D_{COC} - \delta(D_{COC})$. If a deblurring operation is repeatedly applied to a point, the latter will become sharper. The deblurring process gradually reduces the depth of the point by moving it along the (D_{COC} vs. u) curve toward the left end point, corresponding to move the point to the focal plane or be in focus. If the decrement in blurriness, by making a point in-focus, can be determined, we can convert the decrement to the decrement in depth of the point to the focal plane. Then, we refer to the curve (D_{COC} vs. u) in **Figure 3(a)** to acquire the true depth of the point. However, as shown in the curve $\frac{\partial u}{\partial D_{COC}}$ of **Figure 3(b)**, a small decrement

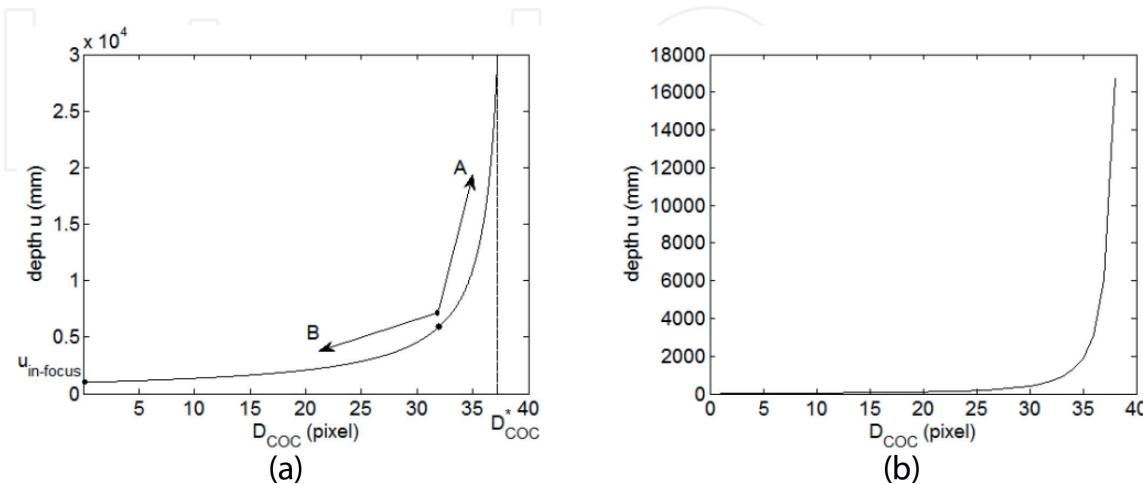


Figure 3.

The relation between the out-of-focus blurriness and the distance u in Eq. (4) in the pixel domain, where $u_{in-focus} = 1000$ mm, $f = 50$ mm, $N = 5.6$, $d = 52.6316$ mm, 0.0061 mm per pixel, and D_{COC}^* is 38.5184. (a) Plot of D_{COC} versus u , the depth. The limit of D_{COC} is denoted by D_{COC}^* . In the blurring process, the dotted point is moved along curve A. In the deblurring process, the point is moved along curve B. (b) The derivation of curve (a). When u approaches $u_{in-focus}$, the blurring process fails to estimate the depth. When u is close to the focal plane, $u_{in-focused}$, the deblurring process fails in the area.

in depth close to the focal plane can yield a substantial decrement in D_{COC} , which means that D_{COC} cannot be reliably and accurately obtained when the point moves closer to the focal plane by a deblurring operation.

Since the depth estimation at large D_{COC} and the D_{COC} estimation of a point close to focal plane are unreliable, we were motivated to propose the blurring and deblurring approach that combines the differential blurring and deblurring operations to yield a more robust depth estimation of a scene point than only using one of them.

3.2 Formulation

Let u_0 be the true depth of the point at x ; and let $F_b(x, u - u_0 + u_\infty)$ and $F_d(x, u - u_0 + u_{in-focused})$ be the blurring measurement and deblurring measurement, respectively. The blurring measurement measures whether a point is blurred to D_{COC}^* , and the deblurring measurement measures whether that point is deblurred to be in-focus. We define that $F_b(x, u - u_0 + u_\infty)$ is a proper function with a (local) minimum near u_∞ and $F_d(x, u - u_0 + u_{in-focused})$ is a proper function with a (local) minimum near $u_{in-focused}$. The following formula is used to determine the true depth u_0 of the point x :

$$\min_u \lambda F_b(x, u - u_0 + u_\infty) + F_d(x, u - u_0 + u_{in-focused}), \quad (7)$$

with the constraint that

$$\Delta D_{COC}^b(u) + \Delta D_{COC}^d(u) = D_{COC}^* - D_{COC}(u_{in-focused}), \quad (8)$$

where $D_{COC}^* - D_{COC}(u_{in-focused})$ is a camera-dependent constant and λ is the Lagrangian parameter that balances the blurriness and deblurriness measurements, and

$$\Delta D_{COC}^b(u) = D_{COC}^* - D_{COC}(u) \quad (9)$$

is the increment of blurriness to D_{COC}^* and

$$\Delta D_{COC}^d(u) = D_{COC}(u) - D_{COC}(u_{in-focused}) \quad (10)$$

denotes the decrement of the D_{COC} assuming the point at depth u to the focal plane. The constraint in Eq. (8) is necessary because it indicates that the sum of the added blurriness from the current guess u to D_{COC}^* and the reduced blurriness from u to $D_{COC}(u_{in-focused})$ is a constant, $D_{COC}^* - D_{COC}(u_{in-focused})$.

3.3 Blurring and deblurring measurements

For a simplified analysis but without any loss of generality, the following derivations were based on one-dimensional signals and neglecting the boundary conditions.

3.3.1. Blurring measurement

The objective of the blurring process is to determine the amount of blurriness required for a point to reach D_{COC}^* . When edged or textured patches are gradually placed at far distance, the details of the patches become faint, their variances

decrease, and only their mean brightness can be derived at infinity. Thus, the variance of a patch can be used as the blurriness measurement. Specifically, when a patch is blurred to reach D_{COC}^* , its variance can be assumed to be 0.

Let the true depth of the scene point x be u_0 , and let the image of the point be

$$f_0(x) = g(\sigma(u_0)^2) * s(x) \quad (11)$$

where g is the Gaussian function and $\sigma(u_0)^2$ is the variance of g at depth u_0 . We define the blurriness measurement as follows:

$$F_b(x, u - u_0 + u_\infty) = \|g(\sigma(u_\infty)^2 - \sigma(u)^2) * f_0(u) - E\{f_0(x)\}\|^2, \quad (12)$$

where $E\{f_0(x)\}$ is the mean on a neighborhood of $f_0(x)$ and $\sigma(u_\infty) = D_{COC}^*$. Assuming that $\sigma_b^2 = \sigma(u_\infty)^2 - \sigma(u)^2$, we obtain from Eqs. (11) and (12)

$$\|g(\sigma_b^2) * f_0(x) - E\{f_0(x)\}\|^2 = \|g(\sigma_b^2 + \sigma(u_0)^2) * s(x) - E\{f_0(x)\}\|^2. \quad (13)$$

The above equation is derived by using the fact that the convolution of two Gaussians of variances σ_1^2 and σ_2^2 is a Gaussian of variance $\sigma_1^2 + \sigma_2^2$. If u is equal to u_0 , then $\sigma_b^2 + \sigma(u_0)^2 = \sigma(u_\infty)^2$ and it becomes

$$\|g(\sigma(u_\infty)^2) * s(x) - E\{f_0(x)\}\|^2 \approx 0, \quad (14)$$

since $g(\sigma(u_\infty)^2) * s(x)$ can be approximated as the mean of $f_0(x)$. Thus, $F_b(x, u - u_0 + u_\infty)$ reaches a local minimum when u is equal to u_0 .

3.3.2 Deblurring measurement using blurring-deblurring operator

In contrast to blurring, deblurring is extremely unstable, and it usually assumes some prior knowledge of the scene so that the high-frequency (edge and texture) information can be recovered. Because of the prior assumption, when the estimated depth is overestimated, ringing artifact occurs in the result of the deblurring process.

We denote $g(\sigma(u)^2) * g^{-1}(\sigma(u)^2)$ as the blurring-deblurring operator, where $g^{-1}(\sigma(u)^2)$ is the reduction in blurriness of a Gaussian kernel with variance $\sigma(u)^2$. An image is first deblurred and then blurred by the same Gaussian kernel with variance $\sigma(u)^2$. A deblurring process will tend to over-enhance the high-frequency information in the image if u is an overestimated depth. As shown by the subfigures in the second row of **Figure 4(c)**, it causes severe artifacts. So, the measurement of the error from the blurring-deblurring operator of a given point is proposed in the following:

$$S(x, u) = \|f_0(x) - g(\sigma(u)^2) * g^{-1}(\sigma(u)^2) * f_0(x)\|^2. \quad (15)$$

As shown in **Figure 4(d)**, when the estimation of the blurring scale is over the true scale (the scale is 4), the result of $S(x, u)$ increases dramatically because of the artifacts in the neighborhood of the edge points.

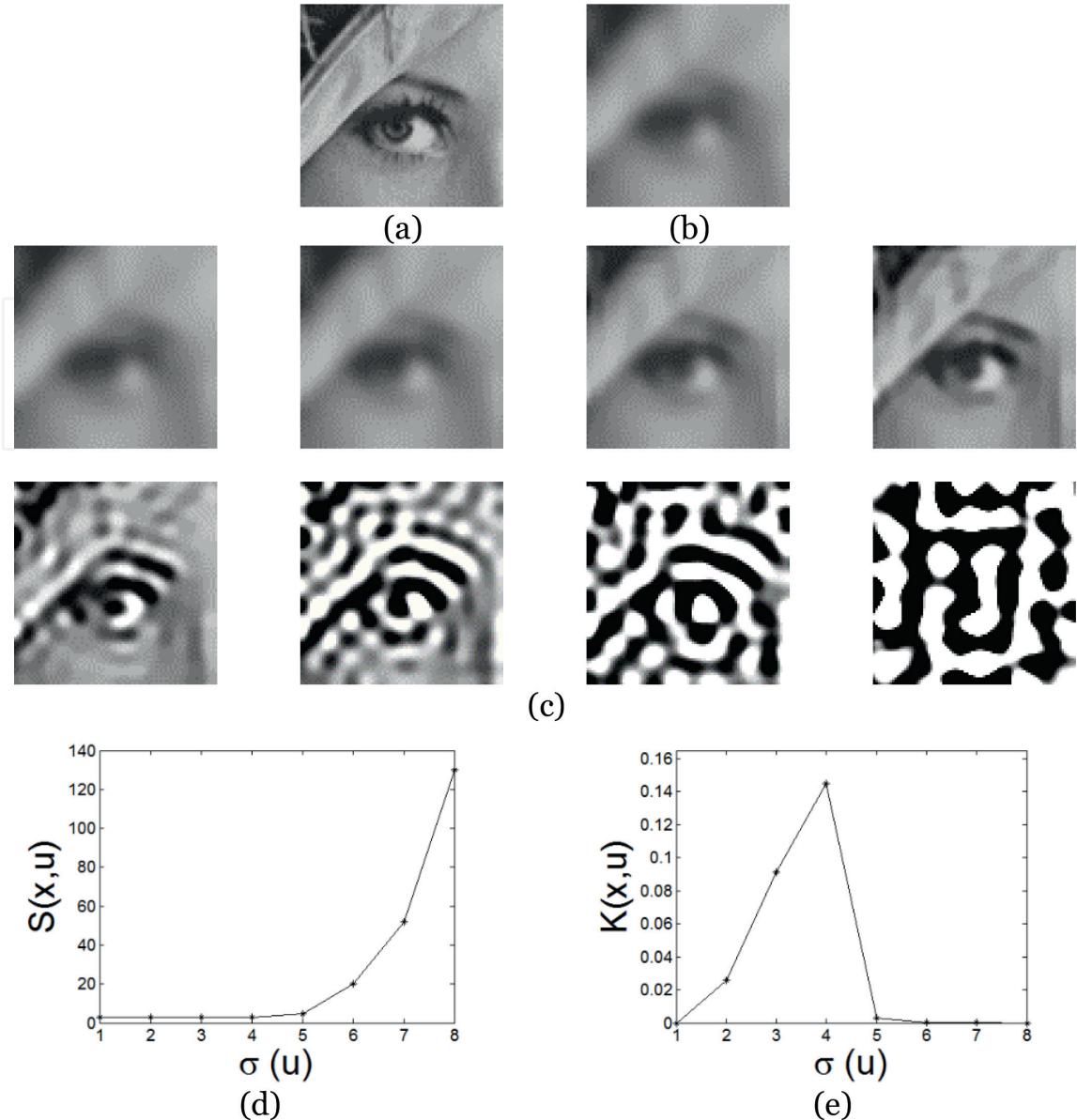


Figure 4.

(a) The patch was taken from the eye of the ‘‘Lena’’ image, (b) the blurred patch of (a) with blurring scale 4, (c) the candidate scene patches obtained by deblurring the patch in (b) with a TV-based method (section 4.3) and different blurring scales whose standard deviations ranged from 1 to 8, and (d) the curve of $S(x, u_j)$ and (e) $K(x, u_j)$, with $j = 1, \dots, 8$.

From **Figure 4(d)**, $S(x, u)$ is asymmetric with respect to over- and underestimation of u_0 , where u_0 is the true blurring scale or true depth. To capture the transition point from small to large values of $S(x, u)$, we calculated the curvature at u_j of the smooth curve as follows:

$$K(x, u_j) = \frac{S(x, u_{j+1}) - 2S(x, u_j) + S(x, u_{j-1})}{\left(1 + (S(x, u_{j+1}) - S(x, u_j))^2\right)^{1.5}}, \quad (16)$$

as shown in **Figure 4(e)**. The larger the value of the curvature, higher is the probability that it is the pivot point for the transition. Thus, we define the deblurring measurement as follows:

$$F_d(x, u - u_0 + u_{in-focused}) = -K(x, u). \quad (17)$$

The measure $F_d(x, u - u_0 + u_{in-focused})$ has a local minimum at the transition of $S(x, u)$. Thus, when u is equal to u_0 , $F_d(x, u - u_0 + u_{in-focused})$ becomes the minimum.

3.3.3 Depth estimation

The blurring and deblurring measurements, F_b and F_d , defined in Eqs. (12) and (17), respectively, can be substituted in the objective function in Eq. (7) to obtain

$$H(u) = \lambda \left\| g(\sigma(u_\infty)^2 - \sigma(u)^2) * f_0(u) - E\{f_0(x)\} \right\|^2 - K(x, u). \quad (18)$$

The blurring and deblurring approach can now be used to derive the solution for

$$\min_u H(u), \quad (19)$$

based on the constraint in Eq. (8). The complexity of the problem relies on how precise the depth is measured for each point. Although depth is an important cue, it seems that the relative depths, such as which object is in foreground and which is in background, are more important than the accurate depths. The blurring and deblurring approach is a point-wise optimization method. We used a method by deriving the best solution from the d candidate depths in a sequence, u_1, \dots, u_d , to save the computational cost. To the optimization problem, the solution is to pick a candidate depth that has the minimization. The candidate depths were chosen so that $\sigma(u_i) - \sigma(u_{i-1}) = 0.2$ for $i = 2, \dots, d$ in implementation. This procedure takes $\mathcal{O}(\gamma N^2 d)$, where γN^2 is the ratio of edged and textured pixels in an image of N^2 pixels, point-wise blurring and deblurring operations to determine the depth map of the image.

4. Depth refinement and image deblurring

The blurring and deblurring measurements are only able to determine the depths of edge and texture points. However, if blurring and deblurring with Gaussian kernel of any variance are applied to a sufficiently large patch of constant value, it will yield a patch of constant value. Therefore, the proposed approach cannot reliably determine the depths of points in smooth regions. Thus, we resort to another approach to derive the depths of smooth scene points.

On the other hand, all in focus is to generate an image that is focused everywhere or to transfer the depth map of an image to a depth map in focal plane. Therefore, an all-in-focus image can be generated by the deblurring process. Practically, since a deblurring process is very sensitive to overestimated depth, if there are overestimated depths in images, the deblurring process can hamper the all-in-focus result and render a visually unacceptable image.

This problem cannot be trivially solved by subtracting a depth from all the points because the value to subtract is not easy to determine. This value should be large enough to stabilize the deblurring process and at the same time small enough so that the depths are not underestimated too much, rendering a blurred all-in-focus image. We used two methods, viz., depth quantization and TV deblurring process, to rectify the effects caused from the depth estimation error.

4.1 Depths of smooth scene points

We followed [6] to estimate the depth at edges and texture followed by propagation of the results to other points. In our method, we use Canny edge detector [10] to decide whether a point is an edged or textured point. The depths of these points (called Canny points) were then estimated from the blurring and deblurring approach. For convenience, we called the remaining points as the smooth points.

The propagation algorithm to derive the depths of smooth points was based on the solution of the Dirichlet problem [11], which addresses the temperature distribution from the boundary to the interior of a medium. The solution of the Dirichlet problem is based on two principles: the maximum principle and the uniqueness principle. The maximum principle states that the interior temperature lies between the maximum boundary temperature and the minimum boundary temperature, and the uniqueness principle states that the solution of the problem is unique. In our approach, we regarded the temperature as the depth and defined the boundary points as the union of the smooth points at the border and the non-smooth points of an image. The depth was first assigned to the smooth points at the border of the image. Then, we used the solution of the Dirichlet problem to derive the depth of the smooth points inside the image. By this approach, the depths of the smooth points were never larger than those of the enclosing points.

The steps of the depth propagation procedure are as follows. First, we normalized the depths of the Canny points by setting the depth of the point farthest from the camera as 1. Then, we assigned depths to the smooth points on the borders of the image. Because the top border of an image is usually the background, the depth of the smooth points on that border was assigned the value 1. For a smooth point on the left-hand, right-hand, and bottom borders, we assigned the depth of the closest non-smooth point. **Figure 5** demonstrates an example how the depths are propagated to an image.

4.2 Depth quantization

For the quantization process, a depth can be approximated by a layer. The motivation of the idea was from the scalar quantization in compression. Via the quantization process, a coefficient can be approximated. In the process, the layer L is a parameter. From the results in Section 3.3.3, the histogram of the depths was calculated firstly. The representative (anchor) depth, a_1 , for layer 1 was always assigned as the minimum depth of all the depths. When the depths are partitioned into two layers, we proposed the following optimization to determine the anchor depth of layer 2, a_2 :

$$\min_{a_2} \sum_{z_i \in \text{layer 1}} (z_i - a_1)^2 + \sum_{z_i \in \text{layer 2}} (z_i - a_2)^2, \quad (20)$$

where a_2 is subject to $a_2 > z_i > a_1$, for each z_i in layer 1, and $z_i \geq a_2$, for each z_i in layer 2. By recursively subdividing a depth layer to acquire two more depth layers, the procedure can be applied to acquire L layers. The depths in the layer are updated to be the depth of the anchor if the anchor depth of a layer is determined. For instance, the depth z_i in layer j is assigned as a_j . Hence, $z_i - a_j$ for z_i in layer j is the error in approximating z_i with a_j . In the approximation, the anchor depth can be found from Eq. (20), which has the minimization of the error.

By sampling a few depths as candidate anchor depths firstly, the anchor depth a_2 in Eq. (20) can be derived. Next, we set each candidate as a_2 to calculate the average error from Eq. (20). With the help of the histogram of the depths, this process can

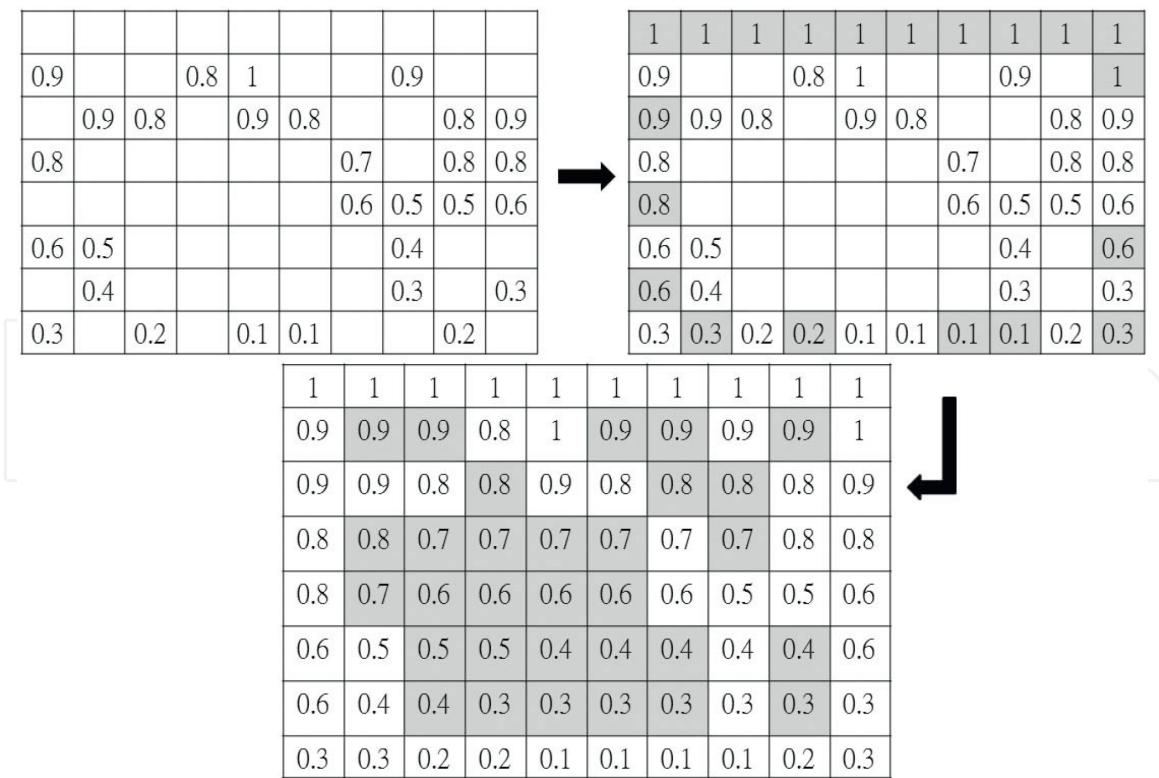


Figure 5.

An example of depth propagation to the interior smooth patches achieved by solving the Dirichlet problem. (top left) The depth map of non-smooth patches (the depth of patches farthest from the camera was set at 1), (top right) the depths assigned to the border patches, and (bottom) the depths of the interior smooth patches derived by solving the Dirichlet problem.

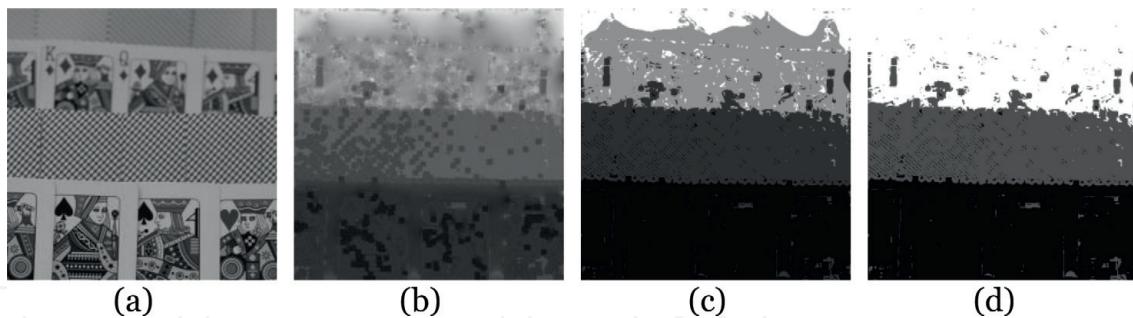


Figure 6.

(a) The image composed of four depth layers. (b) Depth map derived by multi-scale blurring and deblurring approach. The top two layers of depths can hardly be distinguished. (c) The depth map is quantized to four layers. (d) The depth map is quantized to three layers.

be efficiently achieved. So, the anchor depth is the depth in the candidate depths that yields the smallest average error. **Figure 6** shows the estimated depth map and the depth quantization results on an image, composed of four layers of depths. After quantization, some outliers in **Figure 6(b)** are removed.

4.3 Deblurring process

Patch y can be modeled as $g(\sigma) * x$ where x and y are vectors, x is the vector of the scene patch X ($x = \text{vec}(X)$), and σ is the out-of-focus blurriness of x in y . The deblurring process restores x by

$$\min_x \frac{\mu}{2} \|y - g(\sigma) * x\|^2 + \|D_1 X\|_1 + \|D_2 X\|_1, \quad (21)$$

where μ is a Lagrangian multiplier and $\|D_1X\|_1$ and $\|D_2X\|_1$ denote the discrete total variation of X in horizontal and vertical directions, respectively. The $g(\sigma) * x$ can be represented as a matrix–vector multiplication. Eq. (21) can be solved by an efficient variable splitting technique, as described in Ref. [12].

5. Experimental results

We will demonstrate the depth estimation results and some applications related to the estimated depth map including all in focus, refocusing, defocus magnification, and 2D to 3D conversion. We use synthetic images, real images, and video frames to show the performance of the method.

5.1 Synthetic image

First, we evaluated the depth estimations on synthetic images. The scene images with the ground truth depth map are from [13], but the images and depth maps are not aligned. The proposed method is allowed to estimate the blurriness from a blur image; however, the given scene image is near to an all-in-focus image. Therefore, we align the depth map and the original image by nearest neighbor scaling method and then apply Eq. (4) to generate a blur map according to the given depth map with some fixed camera parameters. The transformed Gaussian blur kernels are controlled in the range from 0 to 8. A blur image is synthesized when the original scene image is convolved with the corresponding blur map. Two sets of the original scene images, depth maps of the ground truth, synthetic blurred images, and estimated depth maps are shown in **Figure 7** (a), (b), (c), and (d), respectively. Perceptually, the estimated depth maps correspond to the ground truth depth maps with outlines.

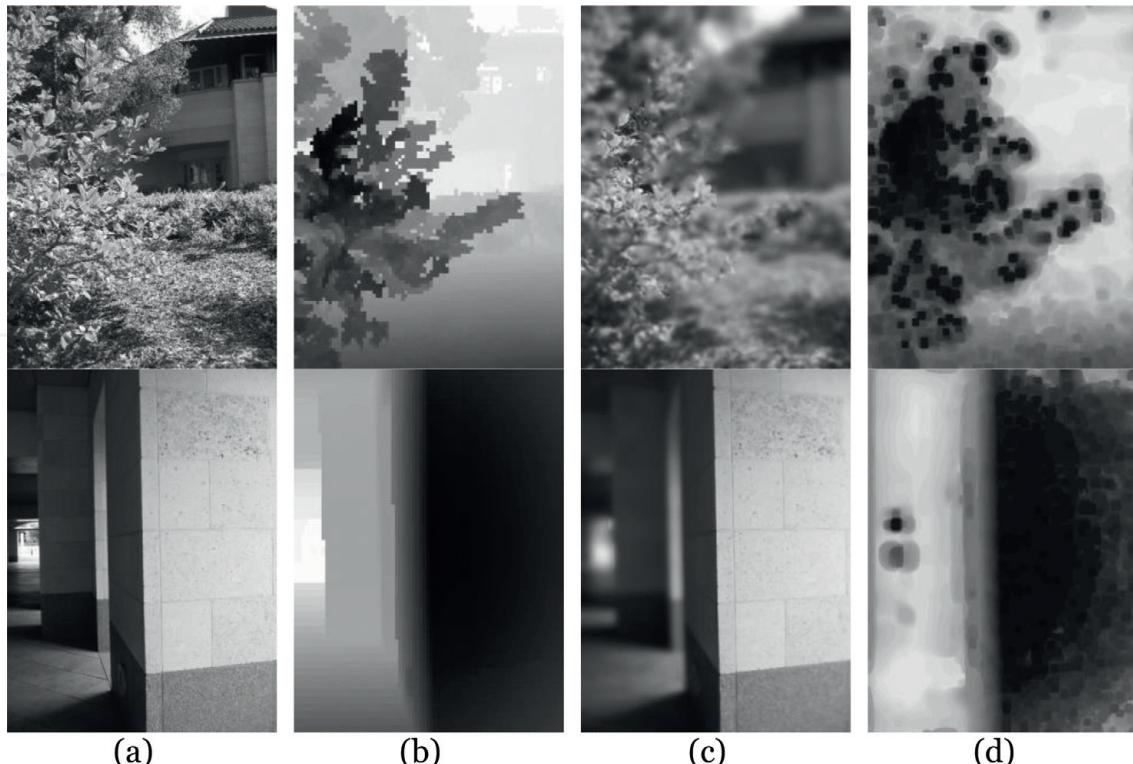


Figure 7.
Datasets for the blurriness estimation. (a) Original scene image. (b) The ground truth depth map. (c) Synthetic blurred input images. (d) Estimated depth map.

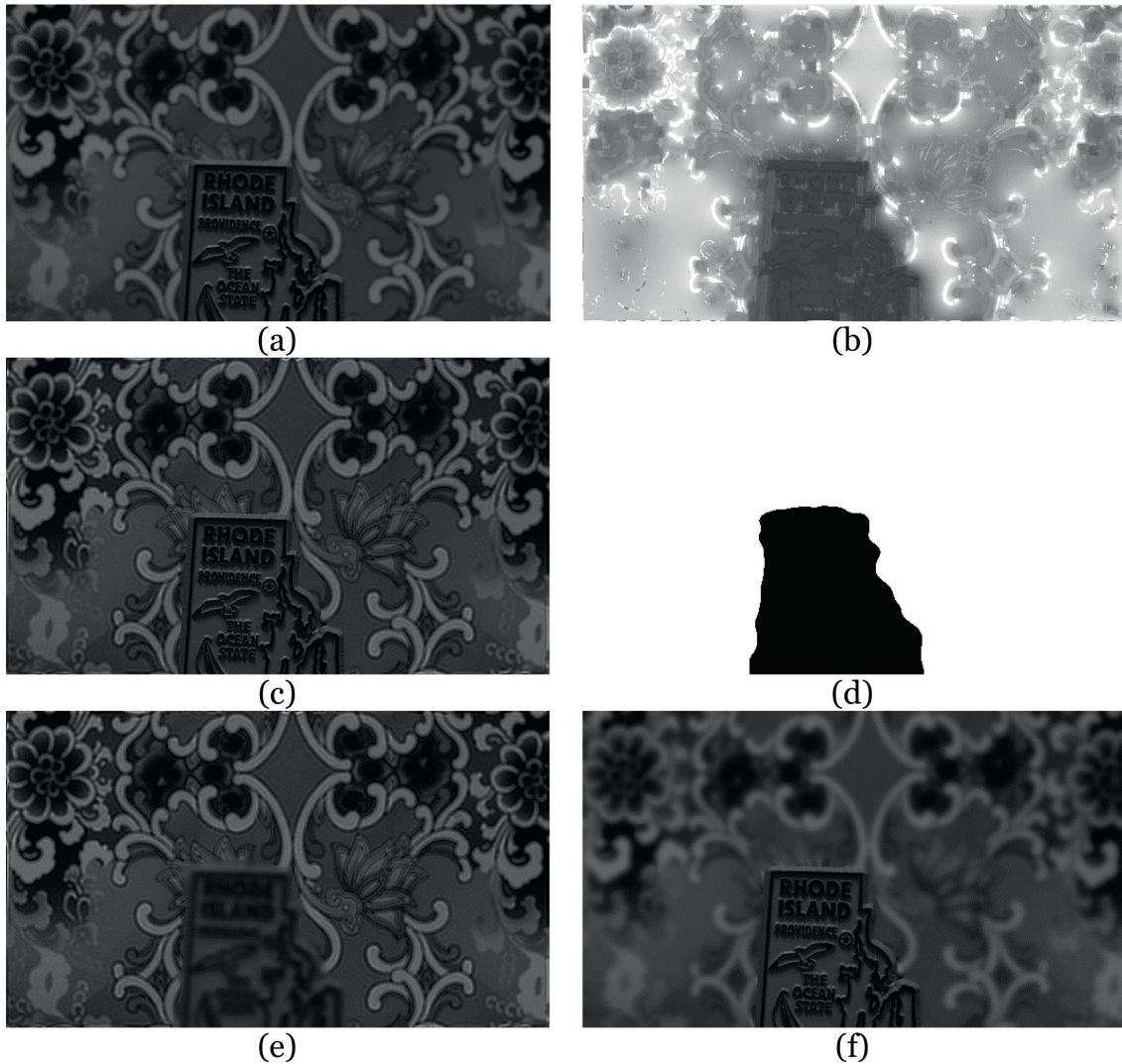


Figure 8.

(a) The original image. (b) Estimated depth map. (c) All-in-focus image, derived with the depth map in (a). (d) Quantized depth map into two layers. The darker area is the foreground, and the bright area is the background. (e) Refocus image. (f) Defocus magnification image.

5.2 Real image

Second, we set real images to the depth estimation method. **Figure 8** shows the results on a two-layer image, and the input image (a) is from [14]. The estimated depth map, synthesized all-in-focus image, quantized depth map, refocus image,

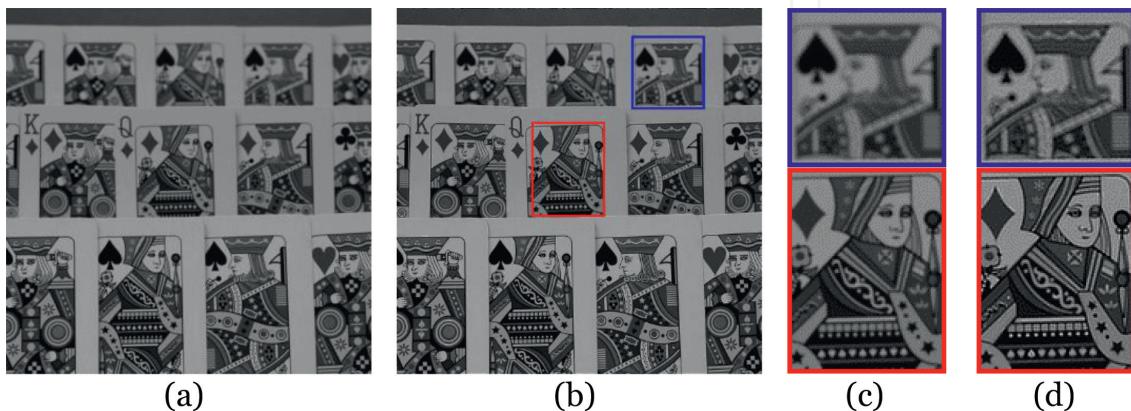


Figure 9.

(a) The poker card image. (b) Synthetic all-in-focus image from the estimated depth map. (c) Magnified regions from (a). (d) Magnified regions from (b).

and defocus magnification image are shown in **Figure 8(b)–(f)**, respectively. With the quantized depth map, the applications of refocusing and defocus magnification can be manipulated from the all-in-focus image. Therefore, if obtaining a high quality of the depth map and all-in-focus image, the performance of these applications will be great.

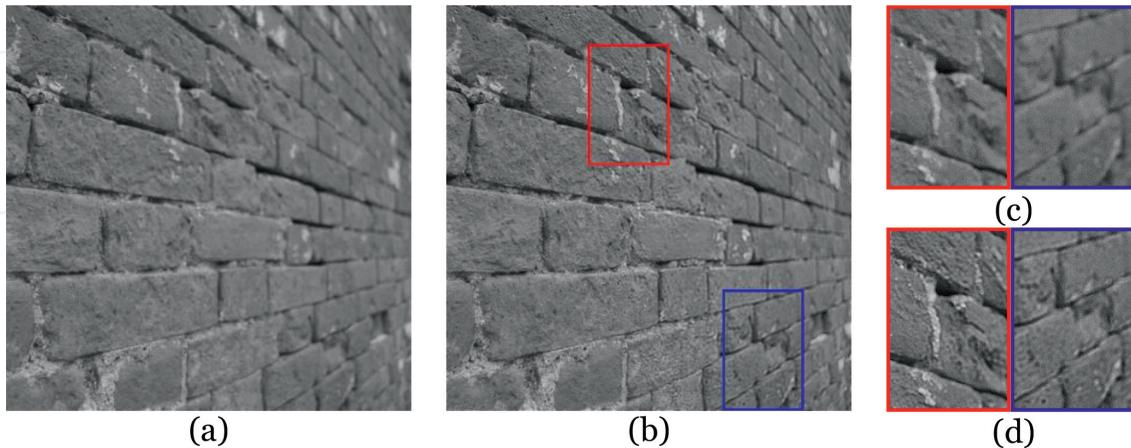


Figure 10.

(a) The slanting brick wall image. (b) Synthetic all-in-focus image from the estimated depth map. (c) Magnified regions from (a). (d) Magnified regions from (b).

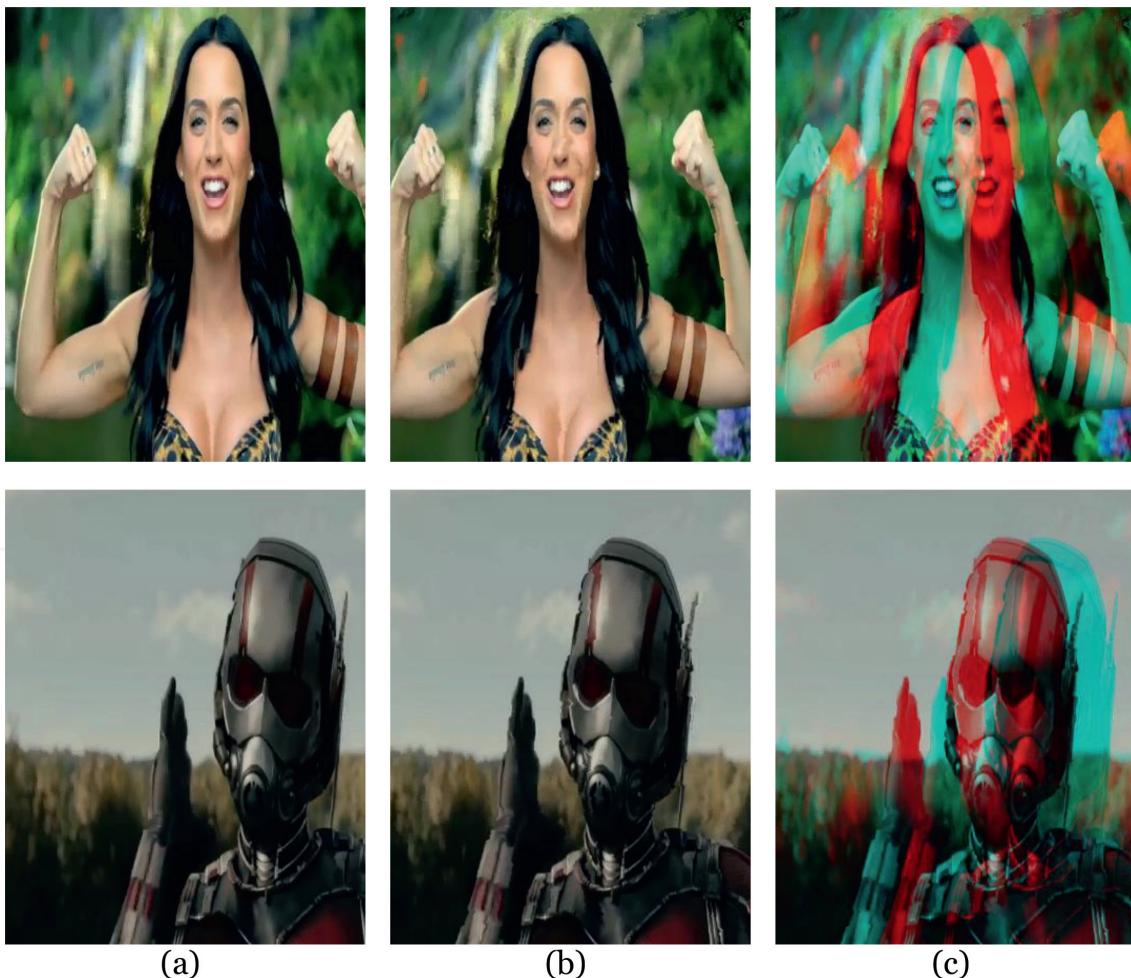


Figure 11.

2D to 3D conversion. (a) Input image (left-eye image). (b) Synthesized right-eye image. (c) Synthesized anaglyph 3D image.

Figures 9 and 10 present the performance of the all-in-focus image. **Figure 9(a)** is a three-layer poker card image, and the camera was focused on the third row of cards. The blurriness increased from the bottom part of the image to the top. So, the first and second rows of cards are out of focus. **Figure 9(b)** shows the synthesized all-in-focus image, and (c) and (d) show the corresponding magnified regions from (a) and (b), respectively. **Figure 10(a)** was captured from a ramp of brick wall. As the camera was focused on the leftmost part of the image, the blurriness of the brick wall progressively increased from the left part of the image to the right. **Figure 10(b)** shows the synthesized all-in-focus image, and (c) and (d) show the magnified regions. From these two sets of image, the results show a significant improvement when comparing to the original images.

5.3 Video frame

Third, we apply the method to the video frames. In this subsection, we show the performance of 2D to 3D conversion. The video frames in **Figure 11(a)** are used as the input left-eye images, which are from YouTube; (b) are the synthesized right-eye images, which are from left-eye images and corresponding depth maps; and (c) are the synthesized anaglyphs, which are from (a) and (b). With the anaglyph glasses, the 3D effect will be visualization. Combine the mobile device with the stereo image and VR device, such as Google Cardboard; the 3D effect also will be visualized.

6. Conclusions

In this chapter, a single-image depth estimation method was presented. The depth map was derived based on the characteristic curve of COC vs. depth of a camera. Applications to manipulate depth maps can be achieved by first deblurring an image to all in focus and then blurring the all-in-focus image. Thus, generation of an all-in-focus image from a depth map successfully is significant for the depth estimation method. Furthermore, the quality for 2D to 3D conversion also depends on the performance of depth estimation. The proposed depth estimation method makes it possible to produce high-quality all-in-focus images and 2D to 3D conversion, even from originals with a complex depth map layout.

Author details

Shih-Shuo Tung and Wen-Liang Hwang*
Institute of Information Science, Academia Sinica, Taipei, Taiwan

*Address all correspondence to: whwang@iis.sinica.edu.tw

IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Grossmann P. Depth from focus. *Pattern Recognition Letters*. 1987;5(1): 63-69. DOI: 10.1016/0167-8655(87)90026-2
- [2] Bulthoff HH, Mallot HA. Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America. A*. 1998;5(10):1749-1758. DOI: 10.1364/JOSAA.5.001749
- [3] Ullman S. The interpretation of structure from motion. *Proceedings of the Royal Society of London B: Biological Sciences*. 1979;203(1153): 405-426. DOI: 10.1098/rspb.1979.0006
- [4] Levin A, Fergus R, Durand F, Freeman WT. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*. 2007;26(3). DOI: 10.1145/1276377.1276464
- [5] Khoshelham K, Elberink SO. Accuracy and resolution of Kinect depth data for indoor mapping applications. *Sensors*. 2012;12(2):1437-1454. DOI: 10.3390/s120201437
- [6] Bae S, Durand F. Defocus magnification. *Computer Graphics Forum*. 2007;26(3):571-579. DOI: 10.1111/j.1467-8659.2007.01080.x
- [7] Zhuo S, Sim T. Defocus map estimation from a single image. *Pattern Recognition*. 2011;44(9):1852-1858. DOI: 10.1016/j.patcog.2011.03.009
- [8] Tung SS, Shao HC, Hwang WL. Extending depth of field in noisy light field photography. In: Proceedings of the IEEE International Conference on Awareness Science and Technology (iCAST '17); 8-10 November 2017; Taiwan. pp. 127-131. DOI: 10.1109/ICAwST.2017.8256430
- [9] Tung SS, Hwang WL. Multiple depth layers and all-in-focus image generations by blurring and deblurring operations. *Pattern Recognition*. 2017; 69:184-198. DOI: 10.1016/j.patcog.2017.03.035
- [10] Canny J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1986;8(6):679-698. DOI: 10.1109/TPAMI.1986.4767851
- [11] Doyle PG, Snell JL. *Random Walks and Electric Networks*. USA: The Mathematical Association of America; 1984. DOI: 10.5948/UPO9781614440222
- [12] Wang Y, Yang J, Yin W, Zhang Y. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*. 2008;1(3):248-272. DOI: 10.1137/080724265
- [13] Saxena A, Chung SH, Ng AY. 3d depth reconstruction from a single still image. *International Journal of Computer Vision*. 2008;76(1):53-69. DOI: 10.1007/s11263-007-0071-y
- [14] Zhang W, Cham WK. Single-image refocusing and defocusing. *IEEE Transactions on Image Processing*. 2012; 21(2):873-882. DOI: 10.1109/TIP.2011.2162739