Ph.D. DISSERTATION

DEPTH MAP GENERATION USING DEFOCUS BLUR ESTIMATOR AND ITS CONFIDENCE

탈초점 흐림 정도의 예측 및 그 신뢰도를 이용한 깊이 맵 작성 기법

BY

WONSEOK SONG

FEBRUARY 2016

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Ph.D. DISSERTATION

DEPTH MAP GENERATION USING DEFOCUS BLUR ESTIMATOR AND ITS CONFIDENCE

탈초점 흐림 정도의 예측 및 그 신뢰도를 이용한 깊이 맵 작성 기법

BY

WONSEOK SONG

FEBRUARY 2016

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

DEPTH MAP GENERATION USING DEFOCUS BLUR ESTIMATOR AND ITS CONFIDENCE

탈초점 흐림 정도의 예측 및 그 신뢰도를 이용한 깊이 맵 작성 기법

지도교수 김 태 정

이 논문을 공학박사 학위논문으로 제출함

2015년 12월

서울대학교 대학원

전기컴퓨터 공학부

송 원 석

송원석의 공학박사 학위 논문을 인준함

2015년 12월

위 원 장: _____

부위원장: _____

위 원: _____

위 원: _____

위 원: _____

Abstract

The depth map is an absolute or relative expression of how far from a capturing device each region of an image is, and a popular representation of the 3D (three-dimensional) structure of an image. There are many depth cues for depth map estimation using only a 2D (two-dimensional) image, such as the defocus blur, the geometric structure of a scene, the saliency of an object, and motion parallax. Among them, the defocus blur is a popular and powerful depth cue, and as such, the DFD (depth from defocus) problem is important for depth estimation. This paper aims to estimate the depth map of a 2D image using defocus blur estimation. It assumes that the focus region of an image is nearest, and therefore, the blur radius of the defocus blur increases with the distance from the capturing device so that the distance can be estimated using the amount of defocus blur. In this paper, a new solution for the DFD problem is proposed. First, the perceptual depth, which is based on human depth perception, is defined, and then the (true) confidence values of defocus blur estimation are defined using the perceptual depth. Estimation methods of confidence values were designed for the gradient- and second-derivative-based focus measures. These estimated confidence values are more correct than those of the existing methods. The proposed focus depth map estimation method is based on the segment-wise planar model, and the total cost function consists of the data term and the smoothness term. The data term is the sum of the fitting error costs of each segment at the fitting process, and the confidence values are used as fitting weights. The smoothness term means the amount of decrease of total cost function by merging two adjacent segments. It consists of the boundary cost and the similarity term. To solve the cost optimization problem of the total cost function, iterative local optimization based on the greedy algorithm is used. In experiments to evaluate the proposed method and the existing DFD methods, the

synthetic and real images are used for qualitative evaluation. Based on the results, the proposed method showed better performances than the existing approaches for depth map estimation.

keywords: Depth Map, Defocus Blur, Focus Measure, Confidence Value
student number: 2008-30227

Contents

Abstract **i**

Contents **iii**

List of Figures **v**

List of Tables **ix**

1 Introduction **1**

1.1 Focus Depth Map 1

1.1.1 Depth from Defocus Blur 2

1.1.2 Absolute Depth vs. Relative Depth 3

1.2 Focus Measure 4

1.3 Approaches of the Paper 5

2 Blur Estimation Methods Using Focus Measures **6**

2.1 Various Blur Estimation Methods 6

2.1.1 Gradient-based Methods 6

2.1.2 Laplacian-based Methods 8

2.1.3 Gaussian-filtering-based Methods 12

2.1.4 Focus Measure Based on Adaptive Derivative Filters 12

2.2 Comparison of the Blur Estimators 15

3 Confidence Values of Focus Measures **21**

3.1 True Confidence Value 21

3.1.1 Perceptual Depth by the Parallactic Angle 21

3.1.2 True Confidence Value Using the Perceptual Depth and Blur Radius 23

3.1.3 Examples of True Confidence Values 26

3.2 Confidence Value Estimation Methods for Various Focus Measures . . 27

3.2.1 Blur Estimator Based on the Gradient Focus Measure 27

3.2.2 Blur Estimator Based on the Second Derivative Focus Measure 29

4 Focus Depth Map Estimation **31**

4.1 Piecewise Planar Model 31

4.2 The Proposed Focus Depth Map Estimation Method 34

4.2.1 Cost Function 34

4.2.2 Depth Map Generation Algorithm 38

5 Experimental Results **40**

5.1 Comparison of the Confidences Value Estimation Methods of Focus Measures 40

5.2 Performances of the Proposed Depth Map Generation Method 70

5.2.1 Experiments on Synthetic Images 70

5.2.2 The Experiments on Real Images 73

5.2.3 Execution Time 81

6 Conclusion **84**

Abstract in Korean **91**

List of Figures

1.1 Defocus blur by the thin-lens model. 2

1.2 Relation between the blur radius (r) and the depth (d). 3

2.1 Estimated σ by (a) MG and (b) TEN. 9

2.2 Estimated σ by Laplacian when the step size is (a) 0.1 and (b) 0.01. . 11

2.3 Estimated σ by Gaussian filtering when σ_0 is (a) 20, (b) 33, and (c) 50. 13

2.4 Comparison results of the blur estimators. (a) Results in noiseless conditions and (b) its highlighted image on the low σ_b region. 17

2.5 Comparison results of the blur estimators on (a) $\sigma_n = 0.1$, (b) $\sigma_n = 1$, and (c) $\sigma_n = 3$. 19

2.6 MSE of the blur estimators. 20

3.1 Binocular parallax and the parallactic angle. 22

3.2 Perceptual depth d_p(d) when (a) $d_m = 1$ and (b) $d_m = 5$. 24

3.3 True confidence value of the gradient-based focus measure. 26

3.4 True confidence value of the second-derivative-based focus measure. . 27

3.5 Estimated confidence value of the gradient-based focus measure. . . . 29

3.6 Estimated confidence value of the second-derivative-based focus measure. 30

4.1 Simple explanation of the existing depth map estimation methods when each object has a constant depth. 32
4.2 Simple explanation of the existing depth map estimation methods when the depth changes gradually. 33
4.3 Simple explanation of the concept of boundary cost computation. . . . 36
4.4 Example of the processes of the proposed depth map generation algorithm. 39

5.1 Comparison of the confidence values of the gradient-based focus measure for the step function image ($\sigma_n = 0$). 43
5.2 Comparison of the confidence values of the gradient-based focus measure for the step function image ($\sigma_n = 1$). 44
5.3 Comparison of the confidence values of the gradient-based focus measure for the step function image ($\sigma_n = 2$). 45
5.4 Comparison of the confidence values of the gradient-based focus measure for the step function image ($\sigma_n = 3$). 46
5.5 Comparison of the confidence values of the second-derivative-based focus measure for the step function image ($\sigma_n = 0$). 47
5.6 Comparison of the confidence values of the second-derivative-based focus measure for the step function image ($\sigma_n = 1$). 48
5.7 Comparison of the confidence values of the second-derivative-based focus measure for the step function image ($\sigma_n = 2$). 49
5.8 Comparison of the confidence values of the second-derivative-based focus measure for the step function image ($\sigma_n = 3$). 50
5.9 Mean square error of the confidence values estimation methods for the step function images in decimal scale. 51
5.10 Mean square error of the confidence values estimation methods for the step function images in log scale. 52

5.11 Ratio of the mean square error of the estimated confidence values to the sum of squared true confidence values for the step function images in decimal scale. 53
5.12 Ratio of the mean square error of the estimated confidence values to the sum of squared true confidence values for the step function images in log scale. 54
5.13 Comparison of the confidence values of the gradient-based focus measure for the random rectangles image ($\sigma_n = 0$). 57
5.14 Comparison of the confidence values of the gradient-based focus measure for the random rectangles image ($\sigma_n = 1$). 58
5.15 Comparison of the confidence values of the gradient-based focus measure for the random rectangles image ($\sigma_n = 2$). 59
5.16 Comparison of the confidence values of the gradient-based focus measure for the random rectangles image ($\sigma_n = 3$). 60
5.17 Comparison of the confidence values of the second-derivative-based focus measure for the random rectangles image ($\sigma_n = 0$). 61
5.18 Comparison of the confidence values of the second-derivative-based focus measure for the random rectangles image ($\sigma_n = 1$). 62
5.19 Comparison of the confidence values of the second-derivative-based focus measure for the random rectangles image ($\sigma_n = 2$). 63
5.20 Comparison of the confidence values of the second-derivative-based focus measure for the random rectangles image ($\sigma_n = 3$). 64
5.21 Mean square error of the confidence values estimation methods for the random rectangles images in decimal scale 65
5.22 Mean square error of the confidence values estimation methods for the random rectangles images in log scale. 66

5.23 Ratio of mean square error of the estimated confidence values to the sum of squared true confidence values for the random rectangles images in decimal scale. 67
5.24 Ratio of mean square error of the estimated confidence values to the sum of squared true confidence values for the random rectangles images in log scale. 68
5.25 Estimated depth (standard deviation of the blur) maps for the step function image. 71
5.26 Estimated depth (standard deviation of the blur) maps for the random rectangles image. 72
5.27 Capturing configuration and captured images (a) camera and object configuration; and (b) captured images. 74
5.28 True distance maps and true perceptual depth maps of the images. . . 75
5.29 True distance maps and estimated depth maps obtained through Chen's method, Lee's method, and the proposed method (gradient-based). . . 76
5.30 True distance maps and estimated depth maps obtained through Chen's method, Lee's method, and the proposed method (second-derivative-based). 76
5.31 Mean square error of the depth map estimation methods for the real images in decimal scale 77
5.32 Mean square error of the depth map estimation methods for the real images in log scale 78
5.33 Estimated depth maps using practical images. 80
5.34 Effect of the number of initial segments on the execution time. 82
5.35 Effect of the number of final segments on the execution time. 83

List of Tables

2.1 Comparison results of the blur estimators. 18

5.1 Mean square errors of confidence values for the step function image (average of 5 trials). All the values of the table are 100 times the MSE. 55

5.2 Mean square errors of the confidence values for the random rectangles image (average of 5 trials). All the values of the table are 100 times the MSE. 69

5.3 Mean square errors of estimated depth (σ_b) maps using synthetic images. 72

5.4 Mean square errors of estimated depth maps using the real images. All the values of the table are 100 times the MSE. 79

Chapter 1

Introduction

1.1 Focus Depth Map

The depth map represents the distances of the regions of an image from a capturing device, and is used to show the 3D (three-dimensional) structure of the scene. Stereoscopic or multi-view images can be generated through the DIBR (depth-map-image-based rendering) method using 2D (two-dimensional) monoscopic images and a depth map. Other problems of computer vision can also be solved using a depth map.

The estimation of the 3D structure from a 2D image can be interpreted as a depth map estimation problem using only a 2D image. Most 2D-3D conversion methods use a depth map and DIBR methods, and as such, the performances of 2D-3D conversion methods depend on the accuracy of the depth map estimation.

There are several monocular depth cues for depth map estimation, such as the motion parallax, motion information, linear perspective, relative size, and occlusion. Among them, the focus depth cue is the representative cue and is thus widely used for depth map estimation. It assumes that the focus point of a capturing device is the nearest point of a scene; as such, the well-focused regions of an image are regarded as the near regions, and the blurred regions are regarded as the far regions. The estimation

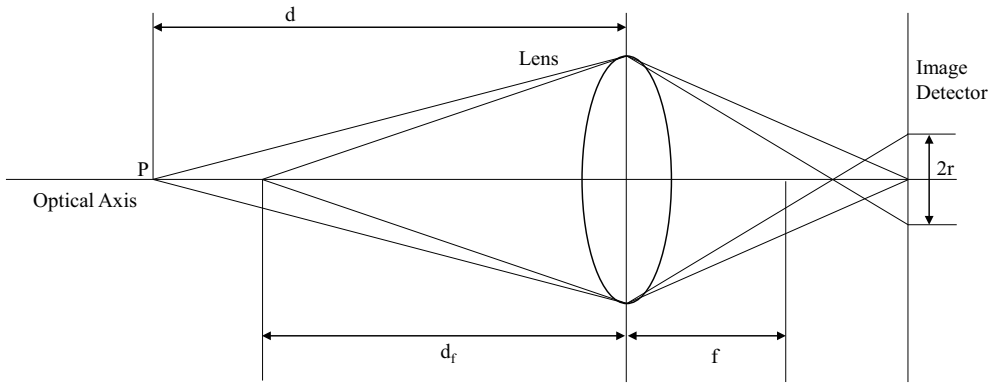


Figure 1.1: Defocus blur by the thin-lens model.

problem of the focus depth map is finding the degree of focusing of all the regions or pixels of an image.

1.1.1 Depth from Defocus Blur

The relation between the depth and the defocus blur is shown in Fig. 1.1. Suppose object P at depth d is blurred on the image detector, and another object at depth d_f is focused on the detector. Then, the radius of blur r of object P is obtained using the following formula of the thin lens-model [12]:

$$r = \frac{d - d_f}{2d} \frac{f^2}{N(d_f - f)}, \tag{1.1}$$

where f and N are the focal length and the stop number of the camera, respectively.

Fig. 1.2 shows the relation between r and d when $f = 4$, $d_f = 10$, and $N = 2$. For convenience, the units of f and d_f are omitted. As shown in the figure, r is zero on focused depth d_f, and increased when d was apart from the focused depth. Therefore, the estimation of d from r has the problem of ambiguity.

Most of the depth from defocus (DFD) methods assume that the focused point is the nearest point on a scene. Based on this assumption, this paper deals with the region where d is larger than the focused depth. Now, the problem of estimating d can

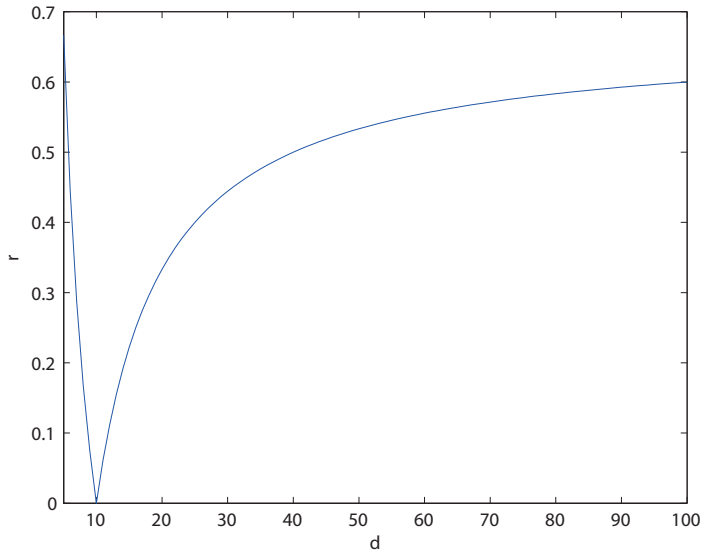


Figure 1.2: Relation between the blur radius (r) and the depth (d).

be regarded as the estimation of r. In the next chapter, the existing focus measures are compared from the viewpoint of the blur radius estimation.

1.1.2 Absolute Depth vs. Relative Depth

There are two different representations of the image depth of a depth map. The absolute depth is the physical distance between the focused point and the object point. Therefore, the absolute depth is measured using length (e.g., meters) as the unit. On the other hand, the relative depth is the relative value of the absolute depth on the scene. The relative depth is a ratio and therefore has no unit. For example, the typical representations of the relative depth are the real number between zero and one, and the 8bit integer value.

For the estimation of the absolute depth, several camera parameters, including the focal length, are needed. On the problem of depth map estimation, however, the

camera parameters are assumed to be generally unknown. On the other hand, the relative depth of a point on the scene can be estimated without the camera parameters.

The relative depth gives much information about a scene structure even though it is not a physical distance. The near region, far region, and overall 3D shape can be recognized by the relative depth map. In addition, most of the 3D image formats use the relative depth map to avoid the quantization problem yielded by the differences in the depth dynamics of the scenes. In the real 3D sequences, some scenes have too small overall absolute depths compared to the other scenes. Nevertheless, fine 3D structures can be described by the relative depth map while the whole absolute depth map is quantized to zero. In this study, therefore, "depth" means the relative depth in all the following chapters, except "absolute depth."

1.2 Focus Measure

Focus measure is the quantitative measure of focusing for a region or pixel of an image. Generally, high-frequency components are measured to get the focus measure because defocus blurring shows a low-pass property. The local variance, the magnitude of the gradient, the Laplacian, and the Gaussian filter are examples of focus measures. Many focus depth map estimation methods estimate the depth value of a pixel using focus measures. In other words, a focus measure can be converted to the depth value. A low focusing value of a region indicates a far region, and vice versa.

The simple approach estimates the depth value of a pixel by converting the focus measure of the pixel directly. There is a serious problem, however, in estimating the depth value of a pixel using only the focus measure of the pixel. The local region has to have a sharp texture or edge to obtain an accurate focus measure because defocus blurring does not show the artifacts on plane regions. The focus measure is not accurate on a textureless region; as such, it is important to calculate the level of confidence of the focus measure of each pixel. Furthermore, a method for spreading the local

focusing information to a whole image is needed because the reliable pixels of the image are sparse and inhomogeneous.

1.3 Approaches of the Paper

In this paper, the definitions of the perceptual depth and the confidence value are proposed at first. Statistical properties of focus measures are used to estimate the confidence values of the focus measures of each pixel, and then the depth values of an image segment are estimated using blur estimators with their confidence values. A segment-wise planar model is assumed, and the weighted fitting method is adopted for the estimation of an image segment. Through adaptive image segmentation, lack of focussing information due to oversegmentation and boundary loss due to undersegmentation are prevented. As a result, the focus depth map is smooth in objects with discontinuities on object boundaries.

Chapter 2

Blur Estimation Methods Using Focus Measures

In this chapter, researches on various focus measures and their comparison are presented. In the first section, focus measures and blur estimation methods are defined. Then examples of the results of blur estimations methods on simple example images are put forth.

In the second section, the performances of the blur estimators are compared. First, the previous works on the results of the comparison of the performances of the focus measures and on the finding of the optimal focus measure are introduced. Then, the blur estimators are compared by applying them on the test images.

2.1 Various Blur Estimation Methods

2.1.1 Gradient-based Methods

The gradient is the basic feature representing the degree of definition of the edges or textures of an image. In the continuous domain, a gradient vector is calculated using

the following equation:

$$\nabla i_c = \frac{\partial i_c}{\partial x}\hat{x} + \frac{\partial i_c}{\partial y}\hat{y} \tag{2.1}$$

where i_c is a continuous image. For discrete images, the discrete form of a gradient can be represented by the filters. The following filters are used to get the gradient vector:

$$\mathbf{g_x} = \frac{1}{2}\begin{bmatrix}-1 & 0 & 1\end{bmatrix}, \quad \mathbf{g_y} = \frac{1}{2}\begin{bmatrix}-1 & 0 & 1\end{bmatrix}^T \tag{2.2}$$

There are various focus measures based on the gradient, such as squared gradient (SG), magnitude of gradient (MG), and Tenenbaum focus measure (TEN) [3]. SG and MG are obtained using the following equations, and each measure may be combined with the appropriate thresholding functions:

$$\text{SG} = G_x^2 + G_y^2 \tag{2.3}$$

$$\text{MG} = \left(G_x^2 + G_y^2\right)^{\frac{1}{2}} \tag{2.4}$$

where

$$G_x = \mathbf{g_x} * i(x,y),\ G_y = \mathbf{g_y} * i(x,y) \tag{2.5}$$

TEN is a famous focus measure and is widely used in various applications. TEN is the local sum of the squared Sobel gradient magnitude on window W.

$$\text{TEN} = \sum_{x,y \in W} \left(\nabla S_x(x,y)^2 + \nabla S_y(x,y)^2\right)^{\frac{1}{2}} \tag{2.6}$$

where $\nabla S_x(x,y)$ and $\nabla S_y(x,y)$ are obtained using following Sobel gradient filters:

$$\nabla S_x(x,y) = \begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix} * i(x,y), \nabla S_y(x,y) = \begin{bmatrix}1 & 2 & 1\\ 0 & 0 & 0\\ -1 & -2 & -1\end{bmatrix} * i(x,y) \tag{2.7}$$

The relation between the gradient and blur radius r is obtained as follows. Suppose a blurred image i where unblurred image u is a horizontal step function with height

a on a continuous space, and defocus blur function b is a Gaussian with standard deviation σ.

$$i = u * b \tag{2.8}$$

$$u(x, y) = \begin{cases} 0, & \text{if } x < 0 \\ a, & \text{if } x \geq 0 \end{cases} \tag{2.9}$$

$$b(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2.10}$$

Then the gradient of blurred image i is

$$G_x(x, y) = \frac{a}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}, \quad G_y(x, y) = 0, \tag{2.11}$$

and σ is

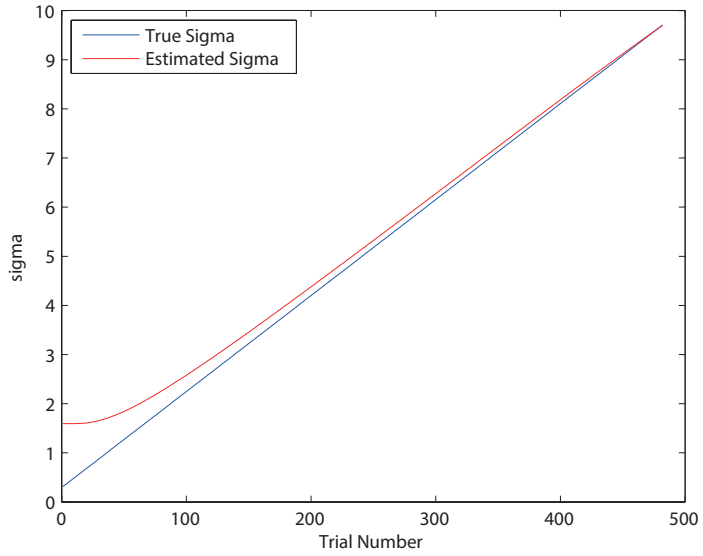
$$\sigma = \frac{a}{\sqrt{2\pi} G_x(0, 0)}. \tag{2.12}$$

In this case, r is proportional to the σ of the blur function on $\sigma = \rho r$, where ρ is a constant that can be approximated by camera calibration [8]. In this paper, therefore, σ is estimated by the gradient to get r. Note that the unknown parameter ρ does not affect the estimation of the relative depth.

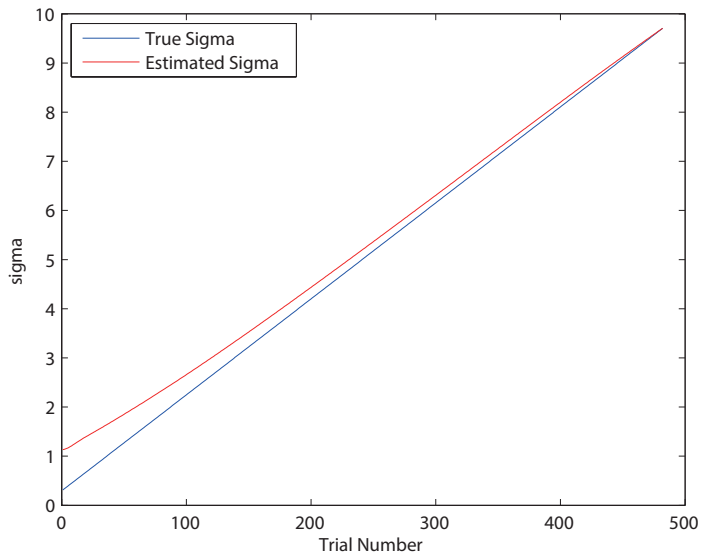
In the discrete domain, the estimated σ has an error because the blur function has to be discretized. Fig. 2.1 shows the experimental results of the estimation of σ on the discrete domain. It is shown that the error is larger when σ is small because the discretization error of the Gaussian blur function is larger. In the figure, for squared gradient-based methods like TEN, the square root of a focus measure is plotted with an appropriate normalizing constant for each method.

2.1.2 Laplacian-based Methods

The Laplace operator is a second-order differential operator in the Euclidean space. In image processing, Laplacian shows a high-pass filter and is a popular feature for edge



(a)



(b)

Figure 2.1: Estimated σ by (a) MG and (b) TEN.

detection, sharpening, etc. Laplacian is defined as the following equation:

$$\nabla^2 i_c = \frac{\partial^2 i_c}{\partial x^2} + \frac{\partial^2 i_c}{\partial y^2} \tag{2.13}$$

and the following filter is implemented as a discrete form of Laplacian:

$$\mathbf{L} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{2.14}$$

Laplacian is suitable for measuring image sharpness, and several focus measures based on Laplacian have been proposed [6]. The sum of modified Laplacian (SML) was proposed by Nayar and Nakagawa [2]. Modified Laplacian (ML) is designed to prevent the cancelation of the x and y directions, which have opposite signs.

$$\nabla^2_{ML} i_c = \left| \frac{\partial^2 i_c}{\partial x^2} \right|^2 + \left| \frac{\partial^2 i_c}{\partial y^2} \right|^2 \tag{2.15}$$

SML is the local sum of ML through the window W.

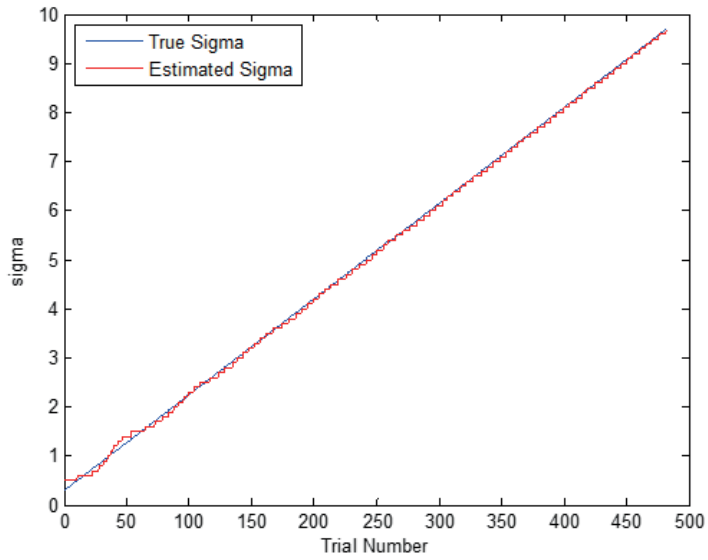
$$\text{SML} = \sum_{x,y \in W} \nabla^2_{ML} i_c(x, y) \tag{2.16}$$

For example, for Gaussian blur b and horizontal step function u, the Laplacian of blurred image i is

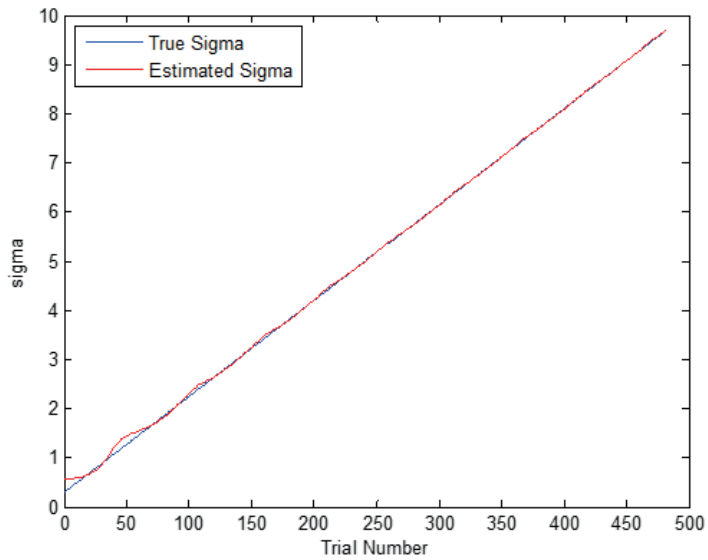
$$\nabla^2 i(x, y) = \frac{x}{\sqrt{2\pi}\sigma^3} e^{-\frac{x^2}{2\sigma^2}}, \tag{2.17}$$

and the magnitude of $\nabla^2 i(x, y)$ is maximum at $x = \sigma$. Therefore, σ can be estimated by finding x, which gives a maximum $\nabla^2 i(x, y)$. In this paper, the following equation is adopted to consider both side regions of an edge:

$$\hat{\sigma} = \frac{1}{2} \left\{ \left| \arg\max_x \nabla^2 i(x, y) \right| + \left| \arg\min_x \nabla^2 i(x, y) \right| \right\} \tag{2.18}$$



(a)



(b)

Figure 2.2: Estimated σ by Laplacian when the step size is (a) 0.1 and (b) 0.01.

2.1.3 Gaussian-filtering-based Methods

Some researchers adopt the focus measure by Gaussian filtering [4, 5]. Suppose the focused image is $\mathbf{r}$, the blurred image is $\mathbf{i}$, and the PSF (point spread function) of the blur is $\mathbf{h}$, with standard deviation σ. Then $\mathbf{i}$ can be represented by the convolution of $\mathbf{r}$ and $\mathbf{h}$.

$$\mathbf{i} = \mathbf{r} * \mathbf{h} \tag{2.19}$$

Let the filtered image of $\mathbf{i}$ be $\mathbf{p}$ by Gaussian filter $\mathbf{G}$, with standard deviation σ_0,

$$\mathbf{p} = \mathbf{i} * \mathbf{G} \tag{2.20}$$

then focus measure $\mathbf{GF}$ is defined as

$$\mathbf{GF} = \mathbf{i} - \mathbf{p} = \mathbf{r} * (\mathbf{h} - \mathbf{h} * \mathbf{G}), \tag{2.21}$$

and $\mathbf{GF}$ is inversely correlated to σ.

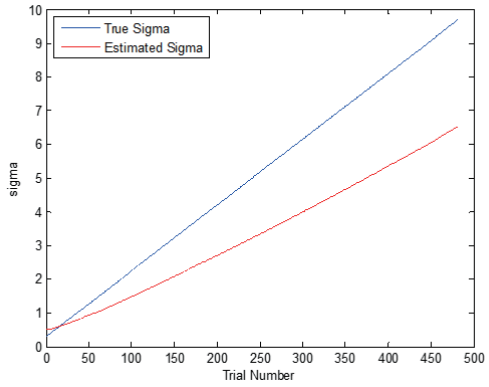
2.1.4 Focus Measure Based on Adaptive Derivative Filters

In the previous section, the focus measures based on derivative filters like gradient (first-order derivative) and Laplacian (second-order derivative) were introduced. These filters are implemented as the minimum sizes, such as 3x1 or 1x3 for the gradient and 3x3 for the Laplacian, and they act as good derivative estimators in noiseless conditions. In noisy conditions, however, the responses of small filters are easily disturbed by the noise. Therefore, some researches integrate appropriate smoothing processes to derivative filters.

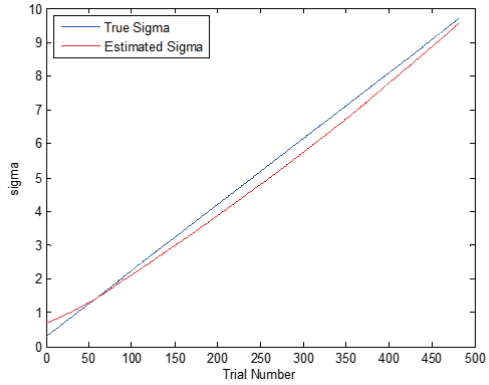
The following equations are Gaussian derivative filters for first-order derivatives (gradient) [26]:

$$g_1^x(x, y, \sigma_{f,1}) = \frac{-x}{2\pi\sigma_{f,1}^4} e^{-(x^2+y^2)/2\sigma_{f,1}^2} \tag{2.22}$$

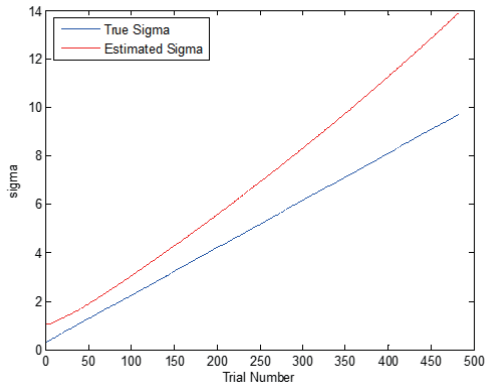
$$g_1^y(x, y, \sigma_{f,1}) = \frac{-y}{2\pi\sigma_{f,1}^4} e^{-(x^2+y^2)/2\sigma_{f,1}^2} \tag{2.23}$$



(a)



(b)



(c)

Figure 2.3: Estimated σ by Gaussian filtering when σ_0 is (a) 20, (b) 33, and (c) 50.

where $\sigma_{f,1}$ denotes the scale of the first derivative Gaussian estimator. Then, the gradient magnitude $r_1(x, y)$ and the gradient direction θ_g are

$$r_1(x, y, \sigma_{f,1}) = \sqrt{\left(r_1^x(x, y, \sigma_{f,1})\right)^2 + \left(r_1^y(x, y, \sigma_{f,1})\right)^2} \tag{2.24}$$

$$\theta_g = \arctan\left(r_1^y(x, y, \sigma_{f,1})/r_1^x(x, y, \sigma_{f,1})\right) \tag{2.25}$$

where $r_1^x(x, y, \sigma_{f,1}) = g_1^x(x, y, \sigma_{f,1}) * I(x, y)$ and $r_1^y(x, y, \sigma_{f,1}) = g_1^y(x, y, \sigma_{f,1}) * I(x, y)$.

The Gaussian second-order derivative filters are the following:

$$g_2^x(x, y, \sigma_{f,2}) = \frac{1}{2\pi\sigma_{f,2}^4}\left((x/\sigma_{f,2})^2 - 1\right)e^{-(x^2+y^2)/2\sigma_{f,2}^2} \tag{2.26}$$

$$g_2^y(x, y, \sigma_{f,2}) = \frac{1}{2\pi\sigma_{f,2}^4}\left((y/\sigma_{f,2})^2 - 1\right)e^{-(x^2+y^2)/2\sigma_{f,2}^2} \tag{2.27}$$

$$g_2^x y(x, y, \sigma_{f,2}) = \frac{xy}{2\pi\sigma_{f,2}^6}e^{-(x^2+y^2)/2\sigma_{f,2}^2} \tag{2.28}$$

where $\sigma_{f,1}$ denotes the scale of the derivative filter and the response is

$$\begin{aligned} r_2(x, y, \sigma_{f,2}) = \cos^2\theta_g g_2^x(x, y, \sigma_{f,2}) - 2\cos\theta_g \sin\theta_g g_2^x y(x, y, \sigma_{f,2}) \\ + \sin^2\theta_g g_2^x y(x, y, \sigma_{f,2}) \end{aligned} \tag{2.29}$$

When $\sigma_{f,1}$ and $\sigma_{f,2}$ are large, the derivatives become more robust to noise, but the correctness of the derivatives is lowered. Therefore, $\sigma_{f,1}$ and $\sigma_{f,2}$ have to be selected adaptively based on the effect of the noise. Elder at el. proposed the local-scale control method to decide the appropriate scales $\sigma_{f,1}$ and $\sigma_{f,2}$ [26] based on the following criteria:

$$\sigma_{f,1}(x, y) = \inf\{\sigma : r_1(x, y, \sigma) > c_1(\sigma)\} \tag{2.30}$$

$$\sigma_{f,2}(x, y) = \inf\{\sigma : r_2(x, y, \sigma) > c_2(\sigma)\} \tag{2.31}$$

where critical value $c_1(\sigma)$ and $c_2(\sigma)$ are computed using the standard deviation of the noise and the following equations:

$$c_1(\sigma) = \frac{1.1 s_n}{\sigma^2} \tag{2.32}$$

$$c_2(\sigma) = \frac{1.8 s_n}{\sigma^3} \tag{2.33}$$

In the local-scale control method, a large scale is selected to suppress the noise by getting the smoothed derivative when the response of the derivative filter is weak compared to the noise. On the other hand, if the response is dominant to the noise, a small scale is selected to get the correct derivative.

To estimate the standard deviation of the blur using the values derived through the adaptive scale filters, the blur estimator needs to be slightly modified.

$$\sigma_b = \sqrt{\tilde{\sigma_b}^2 - \sigma_f^2} \tag{2.34}$$

In the above equation, $\tilde{\sigma_b}$ is the standard deviation computed by the blur estimator introduced in the previous section, and σ_f is the scale factor of an appropriate-order derivative filter. For example, when $\tilde{\sigma_b}$ is computed using a gradient, σ_f means $\sigma_{f,1}$ and σ_f means $\sigma_{f,2}$ for the second-derivative-based blur estimator, respectively.

2.2 Comparison of the Blur Estimators

Subbarao et al. compared the performances of the focus measures to find the optimal focus measure for autofocusing [7]. They proposed the AUM (autofocus uncertainty measure) and the ARMS (autofocus root mean square) errors for the comparison and the finding of the optical focus measure. The AUM and the ARMS errors, however, depend on not only on the focus measure to evaluate but also on the test image. Therefore, the optimal focus measure also depends on the image and thus suggests the focus measure to be selected based on the AUM or ARMS error calculated on the target image. In addition, this approach is inappropriate for depth estimation from a single image because for the calculation of the AUM or ARMS error, three or more images with different lens positions are needed. Therefore, the question of what the best focus measure for DFD is has not yet been solved and needs to be investigated.

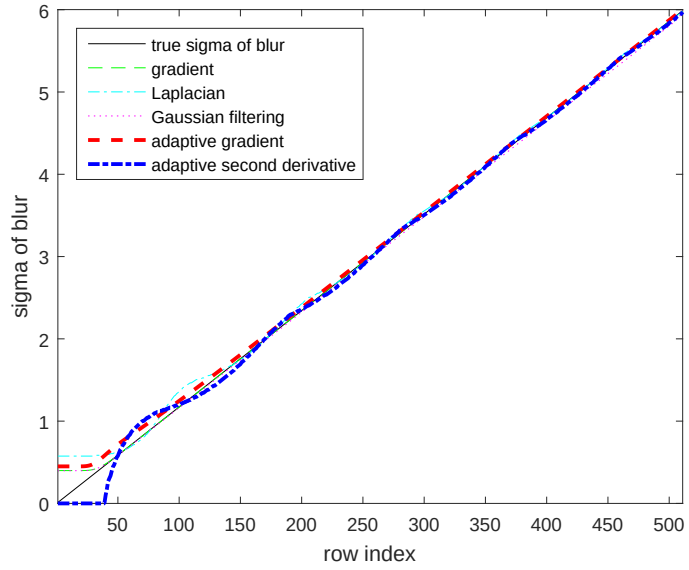
In this paper, the performances of the blur estimators introduced in the previous section are compared using the step-function-based test image. The comparison is first performed in noiseless conditions, and then the same comparison is carried out in noisy conditions using additive WGN (white gaussian noise).

For the comparison of the blur estimators, the step edge image is synthesized. The image consists of 512 rows and 128 columns of pixels. The intensities of the step function are 64 gray level for the darker region and 128 gray level for the brighter region. The blur is simulated using a 2D Gaussian blur filter, and the standard deviation of blur σ_b is varied from 0 to 10 in the vertical direction.

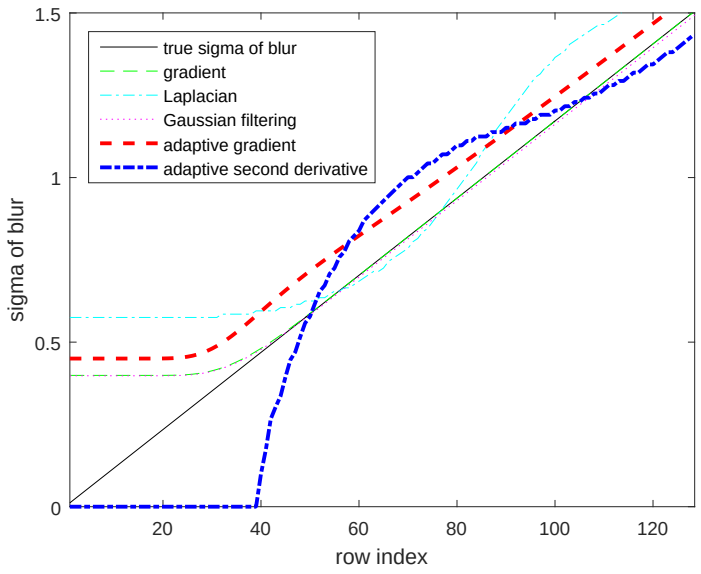
The estimators for the comparison are the gradient-based method, the Laplacian-based method, the Gaussian-filtering-based method, the gradient-based method with the local-scale control, and the second-derivative-based method with the local-scale control. The evaluation is performed using the MSE (mean squared error) between the true σ_b and the estimated σ_b on the edge pixels (64th column) of the original step function image. The standard deviations of the additive WGN are set at 0, 0.1, 1, and 3.

Fig. 2.4 shows the blur estimation results in noiseless condition ($\sigma_n = 0$). The overall performances of the estimators are similar, and no one shows outstanding results. When the σ_b is small, however, all the estimators have dicretization problems. In the Fig. 2.4(b), such discretization errors can be more easily observed in the low σ_b region.

When WGN is added to the image, the performances of the estimators are radically differentiated. It can be shown that the Laplacian-based method and the Gaussian-filtering-based method give poor results when the standard deviation of the noise is increased, as shown in Fig. 2.5. The gradient-based method and Gaussian-filtering-based method show acceptable results when σ_b is not large, but the results become worse when $s_n = 3$. The adaptive methods with the local-scale control show good



(a)



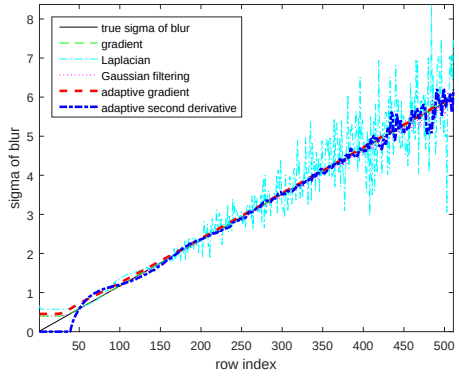
(b)

Figure 2.4: Comparison results of the blur estimators. (a) Results in noiseless conditions and (b) its highlighted image on the low σ_b region.

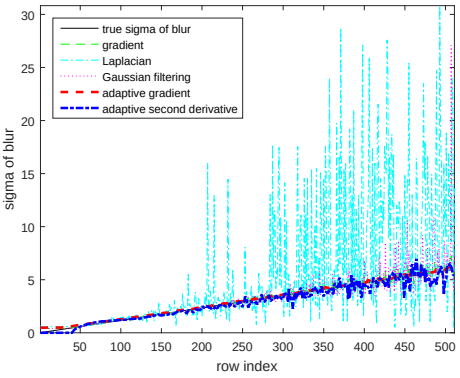
estimation results, however, compared to the non-adaptive methods, even when $s_n = 3$, due to its noise-adaptive filter. Therefore, in this paper, the adaptive filters with the local-scale control method are used to compute the derivative values and σ_b.

Table 2.1: Comparison results of the blur estimators.

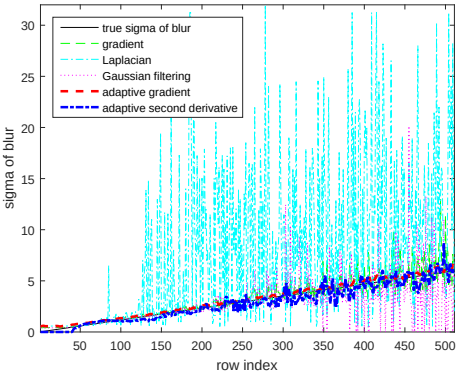
	$\sigma_n = 0$	$\sigma_n = 0.1$	$\sigma_n = 1$	$\sigma_n = 3$
gradient	0.0034	0.0042	0.0636	0.7861
Laplacian	0.0131	0.2664	29.7573	66.0224
Gaussian filtering	0.0044	0.0093	1.4041	3.7816
gradient with local scale control	0.0074	0.0079	0.0180	0.0650
second derivative with local scale control	0.0088	0.0201	0.1483	0.2538



(a)



(b)



(c)

Figure 2.5: Comparison results of the blur estimators on (a) $\sigma_n = 0.1$, (b) $\sigma_n = 1$, and (c) $\sigma_n = 3$.

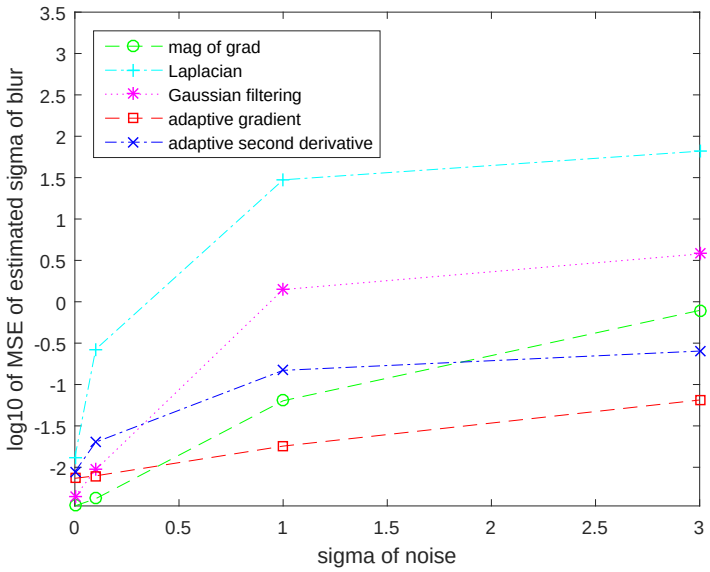


Figure 2.6: MSE of the blur estimators.

Chapter 3

Confidence Values of Focus Measures

3.1 True Confidence Value

3.1.1 Perceptual Depth by the Parallactic Angle

Human can sense the distance of an object in a scene through various depth cues. It is known that the binocular parallax is the most powerful and important cue among the different depth cues. Fig. 3.1 shows how to discriminate the distance using the binocular parallax and the parallactic angle. The near object is at distance d from both eyes, and the far object is at distance x from the near object. The corresponding visual angles of both objects are θ_1 and θ_2, and A is the interpupillary distance. A human senses the distance of an object through a visual angle, which consists of an object and both eyes.

The distance discrimination threshold is calculated as follows. To discriminate the distances of both objects, parallactic angle η of the objects has to be larger than the sensible minimum value of the parallactic angle. If x is set as the distinguishable threshold of the distance between both objects to determine that the distances of the objects are different, then eta can be approximated using the following equations

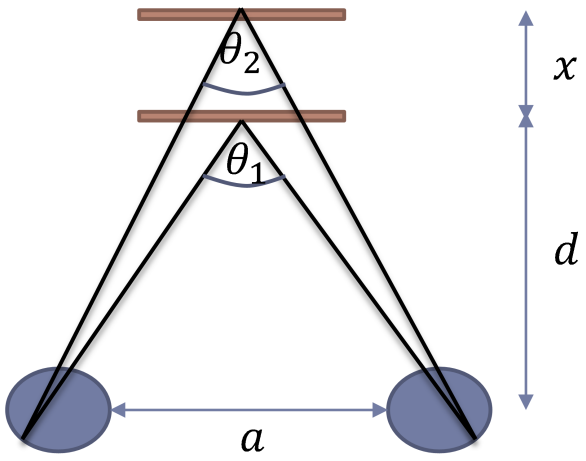


Figure 3.1: Binocular parallax and the parallactic angle.

because d is much larger than A [10]:

$$\eta = \theta_1 - \theta_2 = \frac{A}{d} - \frac{A}{d+x} = \frac{Ax}{d^2 + dx} = \frac{Ax}{d^2} \tag{3.1}$$

$$x = \frac{\eta}{A} d^2 \tag{3.2}$$

Based on equation 3.1, the noticeable threshold of distance x is proportional to d^2, and as such, the human perception of distance is less sensitive in far-view regions. For example, $x = 0.03m$ when $d = 6m$ while $x = 300m$ when $d = 600m$, where $A = 0.065m$ and $\eta = 5.42 \times 10^{-5}$. The physical distance does not well reflect the human perception of depth, and as such, the perceptual depth is proposed in this paper.

Perceptual depth is the relative depth value based on the human perception of the distance of an object. The range of perceptual depth $d_p(d)$ is $[0, 1)$, and $d_p = 0$ when the distance is minimum while d_p is converged to 1 when the distance goes to infinity. The shape of function $d_p(d)$ is derived using the equation $d_p(d+x) - d_p(d) = const$, which means that distance discrimination threshold x gives a constant difference of d_p on all the distances. Then, d_p can be approximated as follows, where d_m is the

minimum distance of a scene:

$$d_p(d) = -\frac{d_m}{d} + 1 \tag{3.3}$$

As in the above equation, $d_p(d)$ maps $[d_m, \inf)$ to $[0, 1)$. Therefore, d_m decides the shape of the function. The shape of a low d_m increases more rapidly, which means that when d_m is set to a low value, the slope of $d_p(d)$ is higher in the near region about d_m. Fig. 3.2 shows the two function examples made by different d_m values. In the earlier section, it was assumed that the focus is on the nearest object in the DFD problem. Therefore, d_m may be set to the estimated distance of the nearest object for the appropriate selection of the shape of the function.

The relation between the perceptual depth and the blur radius can be derived using equations 1.1 and 3.3, as follows:

$$r = \left\{1 - \frac{d_f}{d_m}(1 - d_p)\right\} \frac{f^2}{2N(d_f - f)} \tag{3.4}$$

$$r = pd_p + q \tag{3.5}$$

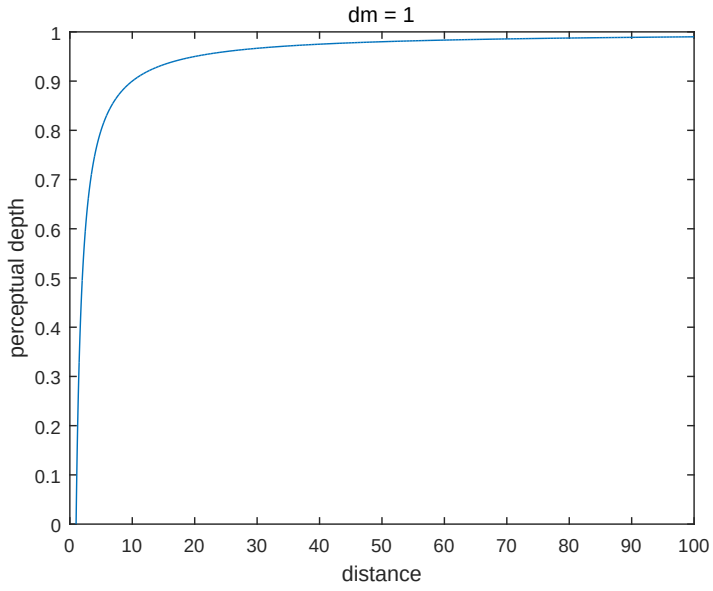
where

$$p = \frac{d_f f^2}{2d_m N(d_f - f)}, \quad q = \left(-\frac{d_m}{d_f} + 1\right) \frac{f^2}{2N(d_f - f)} \tag{3.6}$$

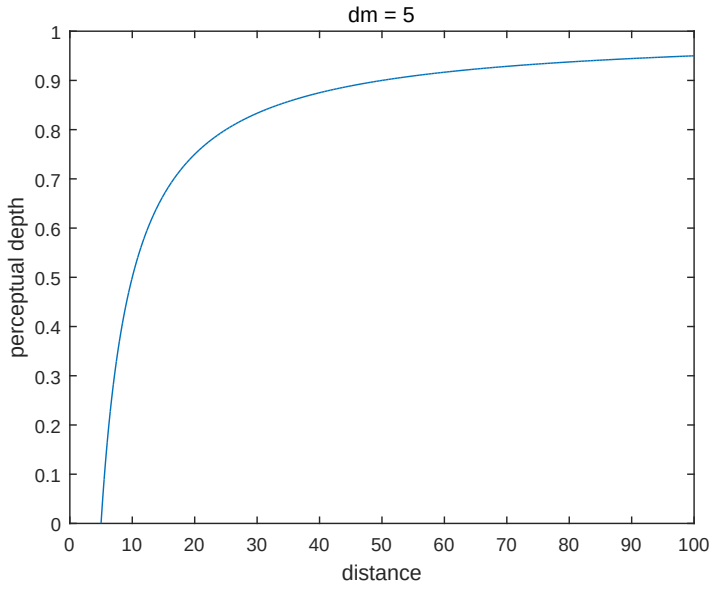
As shown in equation 3.5, d_p and r have a linear relation; therefore, the perceptual depth estimation and evaluation problem can be viewed as a problem of a blur radius or the σ of a blur if the relative depth will be concentrated on.

3.1.2 True Confidence Value Using the Perceptual Depth and Blur Radius

The confidence value of a focus measure means the correctness of a blur estimator based on a focus measure. Therefore, it depends on the type of blur estimator or focus measure used. For example, the gradient-focus-measure-based blur estimator introduced in the previous chapter has estimation results and its own confidence value.



(a)



(b)

Figure 3.2: Perceptual depth d_p(d) when (a) $d_m = 1$ and (b) $d_m = 5$.

In this chapter, the confidence value from the true values to be estimated is called a "true confidence value" to distinguish it from the "estimated confidence value" in the next chapter.

The true confidence value as an evaluation criterion for a focus-measure-based blur estimator is defined as follows in this paper. First of all, the confidence value is the quantitative measure of how reliable a blur estimation is. In this paper, this reliability is measured based on the similarity between the true perceptual depth and the estimated perceptual depth; as such, the true confidence value is defined by the true d_p and the estimated $\hat{d}_p$, as in the following equation:

$$c_{d_p} = \left| d_p - \hat{d}_p \right| \tag{3.7}$$

As the true d_p is needed for the computation of true confidence value c_{d_p}, it is assumed that the true depth information and the capturing configuration for estimating $\hat{d}_p$ from a focus measure are known. c_{d_p} is the reliability measure of a depth estimator, and as such, it is dependent on the type of blur estimator or focus measure adopted in a depth estimator.

In addition, c_{d_p} has to be calculated on valid pixels that have sufficient textures because textureless pixels cannot give any information about the blurs of an image. For example, if the magnitude of the gradient of a pixel is larger than the threshold, then the pixel can be regarded as valid.

Then, the computation of a true confidence value is derived as a function of the blur radius obtained using equation 3.5, as follows:

$$c_{d_p} = \left| d_p - \hat{d}_p \right| = \frac{1}{p} \left| r - \hat{r} \right| \tag{3.8}$$

This means that relative values of c_{d_p} of all the pixels in a given scene are equivalent to blur-radius-based confidence values c_r and standard-deviation-based confidence value c_σ, where n_r and n_σ are normalization constants.

$$c_r = \frac{1}{n_r} \left| r - \hat{r} \right|, \quad c_\sigma = \frac{1}{n_\sigma} \left| \sigma - \hat{\sigma} \right| \tag{3.9}$$

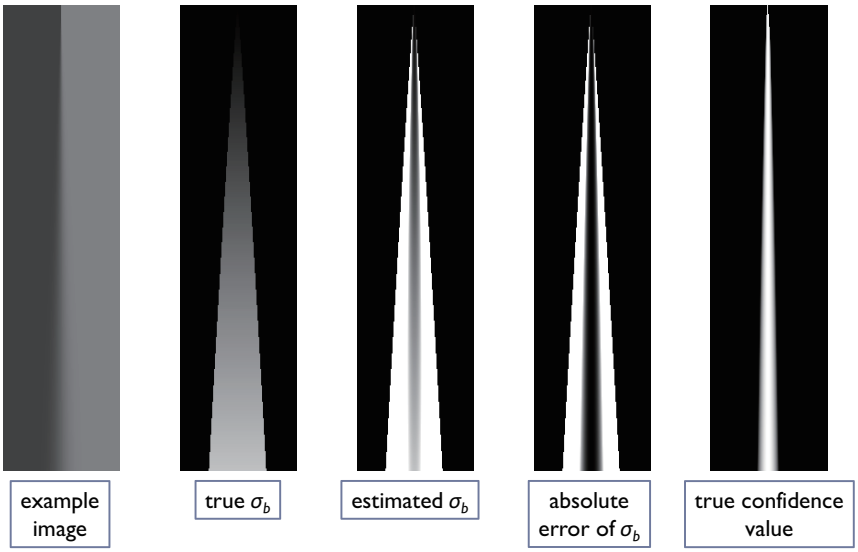


Figure 3.3: True confidence value of the gradient-based focus measure.

Furthermore, c_r and c_σ can be calculated as the relative approximation of c_{d_p} even in a case where constant p is unknown or undefined. Henceforth, true confidence value c is c_r or c_σ.

3.1.3 Examples of True Confidence Values

In this section, a simple image example is introduced, and its true confidence values are presented. Image example i is a Gaussian blurred step function obtained using equation 2.8. Fig. 3.3 shows the image example and the processes of the calculation of the true confidence values. All the images are normalized for visual convenience.

In the figure, the true σ_b is the standard deviation of the Gaussian blur kernel. As shown by the true σ_b image, every pixel of the image example is blurred by the Gaussian kernel, but its effect does not appear on the plain regions, except the near-edge regions. In this case, only the near-edge pixels are valid pixels; therefore, the true confidence values are also computed based only on the valid pixels.

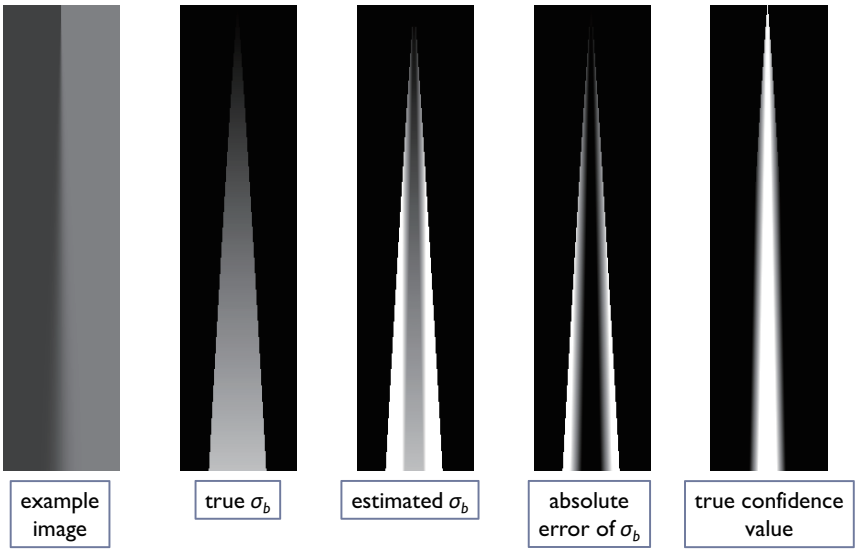


Figure 3.4: True confidence value of the second-derivative-based focus measure.

3.2 Confidence Value Estimation Methods for Various Focus Measures

In the previous chapter, the true confidence value was proposed when a blur kernel was known. A blur kernel cannot be known, however, for computing the true confidence value in a DFD problem because the problem involves the estimation of a blur kernel from an image. Therefore, the confidence value must be estimated from the image whose FDM (focus depth map) is desired to be obtained. In this section, estimation methods of a confidence value for various blur estimators that adopts various focus measures are presented.

3.2.1 Blur Estimator Based on the Gradient Focus Measure

The gradient-based focused measure was introduced in the earlier chapter. Here, the estimation method of a confidence value for the gradient-based focus measure and the

blur estimator is proposed. The estimation method of a confidence value of gradient $\hat{c}_g$ is designed to have the following properties:

$$\hat{c}_g(x,y) = \begin{cases} exp\left\{-\frac{\alpha}{\hat{a}}\left(d_e\sigma_w\right)^2\right\}\cdot\left(-k\sigma_f+1\right), & \text{if } \left(-k\sigma_f+1\right) > 0 \\ 0, & \text{if } \left(-k\sigma_f+1\right) \leq 0 \end{cases} \tag{3.10}$$

At first, the shape of the function is also Gaussian to model the Gaussian blur because the defocus blur kernel is assumed to be Gaussian. The gradient-based focus measure shows the Gaussian shape on the Gaussian blurred edge, and the difference between the focus measure of (x, y) and the nearest edge pixel is also Gaussian. d_e is the distance between (x, y) and the nearest edge pixel so that $\hat{c_g}$ is decreased when d_e is increasing, like the true confidence value. Edge pixels are detected at the local peaks of the focus measure, but various edge detectors may be employed for detecting edge pixels.

σ_w is the standard deviation of the gradients of a local window. This term reflects the degree of textureness. Its value is high at the region that has a complex texture and enough blur information, and low at the region that has no texture and no blur information. Therefore, a higher σ_w gives a higher confidence value.

σ_f is the standard deviation of the adaptive gradient filter and is related to the correctness of a gradient. If σ_f is large, the gradient value is computed from image filtered by the strong smoothing filter. Therefore, low σ_f gives a higher confidence value. The attenuation shape of σ_f is linear from the regression of the step edge.

The gradient-based blur estimator uses the height of edge a, as in equation 2.10, but a has to be estimated because it is unknown for an arbiter image. $\hat{a}$ is the estimated height of the edge. In this paper, it is estimated simply based on the difference between the maximum and minimum intensities of the local windows. Lastly, α decides the attenuation degrees of the side regions of the edges. Its value is selected empirically.

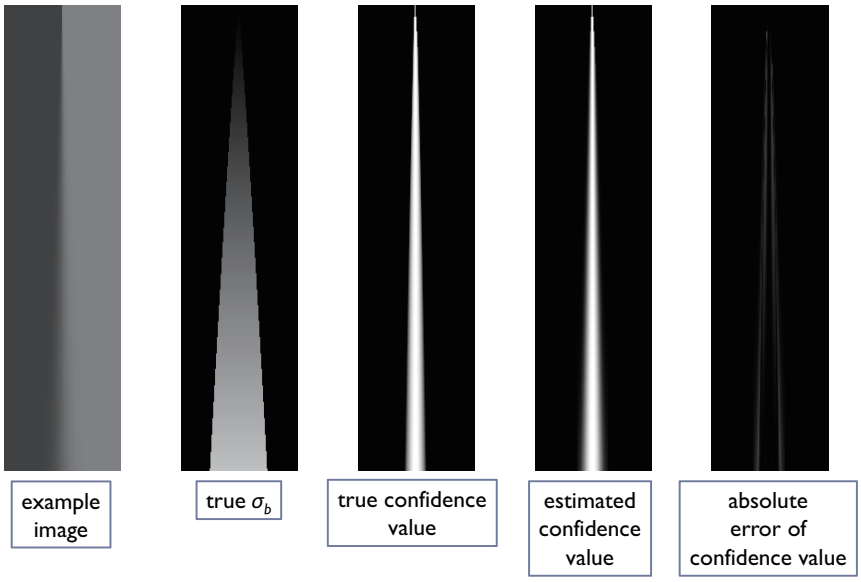


Figure 3.5: Estimated confidence value of the gradient-based focus measure.

3.2.2 Blur Estimator Based on the Second Derivative Focus Measure

The second-derivative-based focus measures was introduced in the earlier chapter. Here, the estimation method of a confidence value $\hat{c_L}$ for the second-derivative-based focus measure and the blur estimator is proposed as the following equations:

$$\hat{\sigma}_s(x,y) = \frac{1}{2}\left\{\left|\arg\max_{x_t} \nabla^2 i(x_t, y_y)\right| + \left|\arg\min_{x_t} \nabla^2 i(x_t, y_t)\right|\right\} \tag{3.11}$$

$$\hat{c}_s(x,y) = \begin{cases} exp\left(-\beta \frac{\left|\nabla^2 i(\mathbf{x})\right|\sigma_s}{|\nabla i(\mathbf{x})||\nabla^2 i(\mathbf{p})||\nabla^2 i(\mathbf{q})|}\right), & \text{if } p_x q_x \leq 0 \\ \left\{1 - \frac{\min(|p_x|,|q_x|)}{\sigma_{max}}\right\} \cdot exp\left(-\beta \frac{\left|\nabla^2 i(\mathbf{x})\right|\sigma_s}{|\nabla i(\mathbf{x})||\nabla^2 i(\mathbf{p})||\nabla^2 i(\mathbf{q})|}\right), & \text{if } p_x q_x > 0 \end{cases} \tag{3.12}$$

where σ_s is the local standard deviation of the second derivatives and $\mathbf{p}$ and $\mathbf{q}$ are the positions where the second derivative values are the peak on both sides of (x, y) along gradient direction pixels (x_t, y_t).

$\hat{c}_s$ is high when the peak values of the side regions are high and the second deriva-

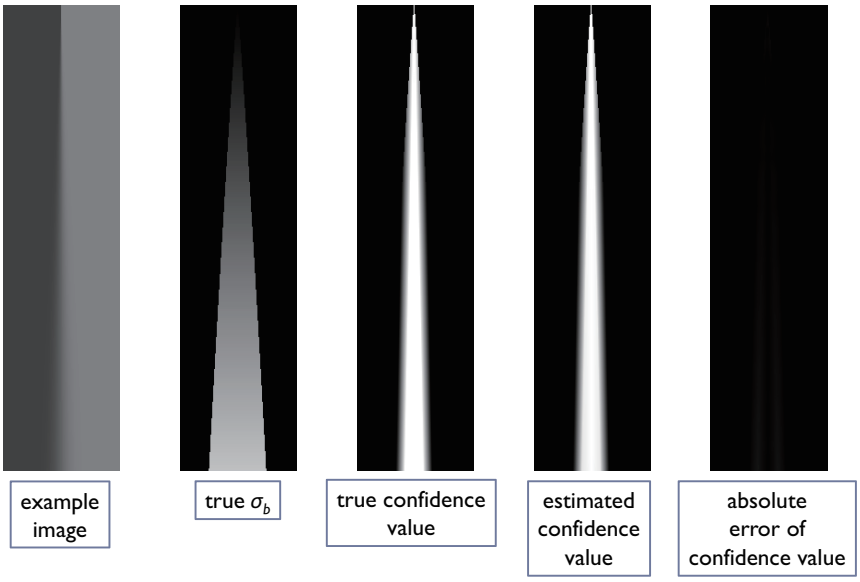


Figure 3.6: Estimated confidence value of the second-derivative-based focus measure.

tive of (x, y) is low because the second derivative is zero at the center of the unit step edge. As the error of the estimated standard deviation using second-derivative-based blur estimator is increased if (x, y) is outside the regions between the two peaks of the second derivative that have opposite signs around the unit step edge, $\hat{c}_s$ is designed to decrease when $p_x q_x > 0$. The attenuation constant β is selected empirically.

Chapter 4

Focus Depth Map Estimation

4.1 Piecewise Planar Model

The information about the defocus blur of an image is concentrated on the edges or complex textures, as mentioned in the last chapter. The confidence values of focus measures are sufficiently high only on these regions, and are too low for estimating the blur on the remaining regions, which account for a much greater portion of the whole area than the edge regions. As such, to estimate the entire depth map of an image, the blur information of the edge regions has to be spread to the textureless or no-information regions.

The existing methods for spreading information can be categorized into two types of algorithms: The filtering-based method [4] and the "segmentation and averaging" method [9][5]. These types of algorithms both have strong points and weak points, as shown in Fig. 4.1 and 4.2. Each (b) of the figures shows sparsely distributed focus measures with errors.

The filtering-based methods are appropriate for representing a smooth depth map because a filter has to be large enough to spread the defocus information. Many details of a scene are lost, however, and such methods cannot represent the depth disconti-

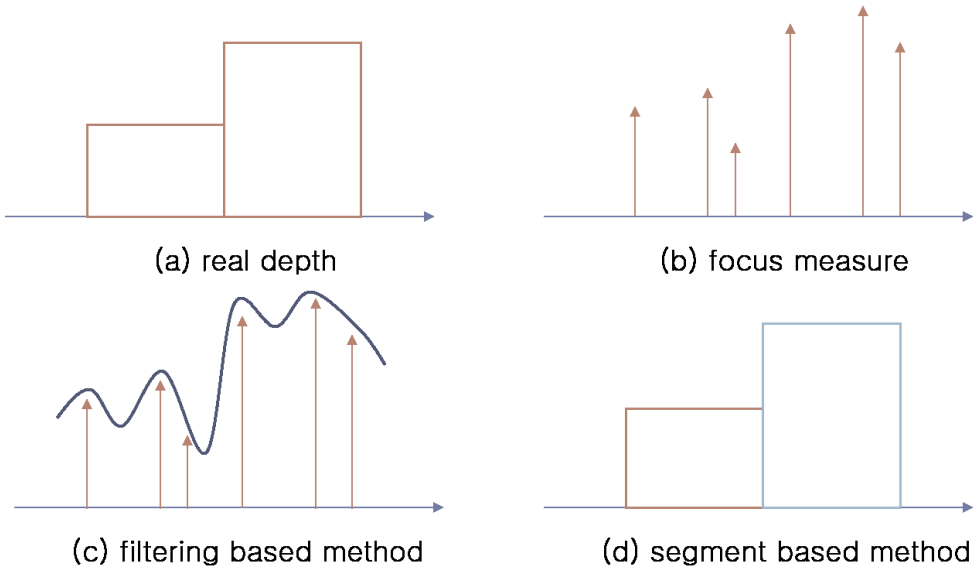


Figure 4.1: Simple explanation of the existing depth map estimation methods when each object has a constant depth.

nuities, as shown in Fig. 4.1(c), due to the large smoothing filter. On the other hand, the segmentation-based methods are outstanding for conserving the scene details and depth discontinuities, as shown in Fig. 4.1(d), but they cannot represent the gradual changes in the depth, as shown in Fig. 4.2(d). In addition, very small segments give erroneous depth segments due to lack of information.

In the recent stereo matching method, the segment-wise (piecewise) planar model is developed to estimate a disparity map [11]. It is assumed that the disparity of each segment is a plane; as such, a gradual change in the disparities in a segment can be represented and makes the segment robust against the outliers compared to the filtering methods.

The segment-wise-planar-model-based method tries to find the optimal disparity map using total cost function C explained by equation 4.1. Total cost C consists of the

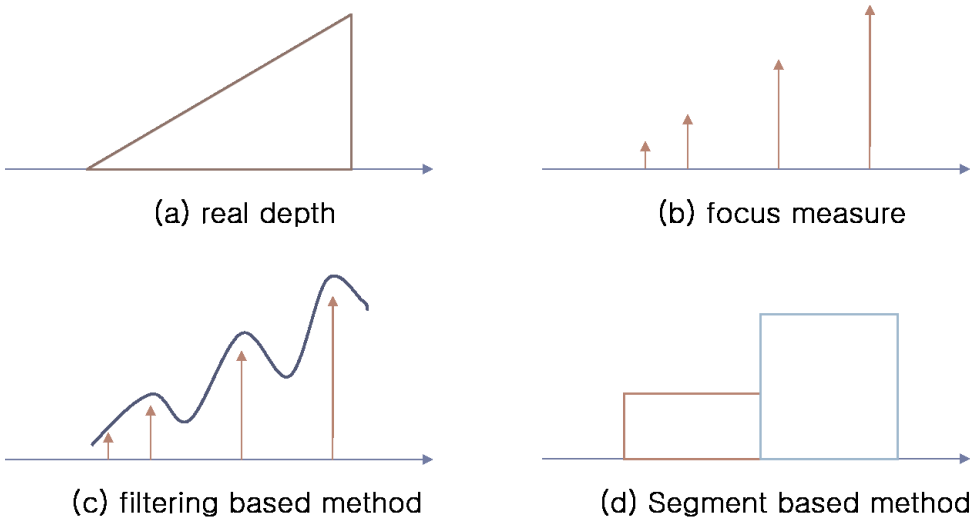


Figure 4.2: Simple explanation of the existing depth map estimation methods when the depth changes gradually.

data term C_d and smoothness term C_s.

$$C = C_d + \lambda \cdot C_s \tag{4.1}$$

where λ is a constant value and the data term C_d is

$$C_d = \sum_i e_i \tag{4.2}$$

where e_i is the fitting error of segment i.

The smoothness term is the value obtained from multiplying border length $l(i,j)$ and color similarity $s(i,j)$ for neighboring segments i and j.

$$C_s = \sum_{i,j} l(i,j) \cdot s(i,j) \tag{4.3}$$

$l(i,j)$ is the length of the border between segments i and j. This term means that the separation of the segments with a complex border entails a higher cost than that of

the segments with a simple border. Color similarity $s(i,j)$ is defined by the following equation:

$$s(i,j) = \left(1 - \frac{\min\left(|m(i) - m(j)|, 255\right)}{255}\right) \cdot 0.5 + 0.5 \tag{4.4}$$

where $m(i)$ and $m(j)$ are the mean values of the pixels in segment i and j.

4.2 The Proposed Focus Depth Map Estimation Method

4.2.1 Cost Function

The proposed cost function for a depth map is based on the segment-wise planar model explained in the last section. The following equations are the total cost function and the data term:

$$C = C_d + \lambda \cdot C_s \tag{4.5}$$

where λ is a constant value. The data term is

$$C_d = \sum_i e_i \tag{4.6}$$

where e_i is the fitting error of a segment i.

At first, the data term C_d represents the goodness of fit of all the segments. To fit the estimated depth to the plane, in this paper, the weighted linear least square regression method is used. The error function of fitting segment i to a linear plane is as follows:

$$\tilde{e}_i(\mathbf{w}) = \sum_{(x,y)\in i} c(x,y) \cdot \left\{\hat{d}(x,y) - (w_1 x + w_2 y + w_3)\right\}^2 \tag{4.7}$$

where $\mathbf{w} = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix}$ is the fitting coefficients vector. As each pixel has a different texture and edge, the confidence value $c(x,y)$ is used as the weight of fitting. Then, the fitting error cost of segment i is

$$e_i = \tilde{e}_i(\mathbf{w_m}) \tag{4.8}$$

where $\mathbf{w_m} = \mathrm{argmin}_{\mathbf{w}} \tilde{e}_i(\mathbf{w})$. Finally, the data term C_d can be computed by adding the fitting error costs of all the segments.

The proposed method has two different modes and data terms in fitting a segment: (1) the linear plane mode, which is introduced in the segment-wise-planar-model; and (2) the constant depth mode. The purpose of the constant depth mode is to prevent overfitting. Unlike the disparity information of stereo matching, defocus information is very sparse. Therefore, the slope of a segment may yield a large error on the sparse region. In this paper, the constant depth mode is applied to a segment that has sparse depth information, using the confidence value density d_c, which is defined as follows:

$$d_c(x, y) = \sum_{i,j} c(x + i, y + j) \cdot w(i, j) \tag{4.9}$$

where c is the confidence value and w is the Gaussian window. Then, the normalized variance of confidence density $V_c = Var(d_c)/\sum d_c$ is computed. A high V_c means that the area of the sparse information region of a segment is large, and as such, the constant depth mode is applied to the segment. On the other hand, the linear plane mode is applied to the segment that has a small V_c.

$$\mathbf{w} = \begin{cases} \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} & \text{for linear plane mode} \\ \begin{bmatrix} 0 & 0 & w_3 \end{bmatrix} & \text{for contant depth mode} \end{cases} \tag{4.10}$$

The smoothness cost C_s is modified by replacing boundary cost $b(i, j)$ with border length $l(i, j)$.

$$C_s = \sum_{i,j} b(i, j) \cdot s(i, j) \tag{4.11}$$

Fig. 4.3 shows the concepts of boundary cost computation. Fig. 4.3(a) and (b) have different boundaries on the same images. In this case, Fig. 4.3(b) has a high boundary cost than Fig. 4.3(a) due to the complex shape of the boundary of Fig. 4.3(b). In the next case, Fig. 4.3(c) and (d) have the same complex boundaries, but Fig. 4.3(c) looks like two different segments only. To reflect this concept, the boundary cost

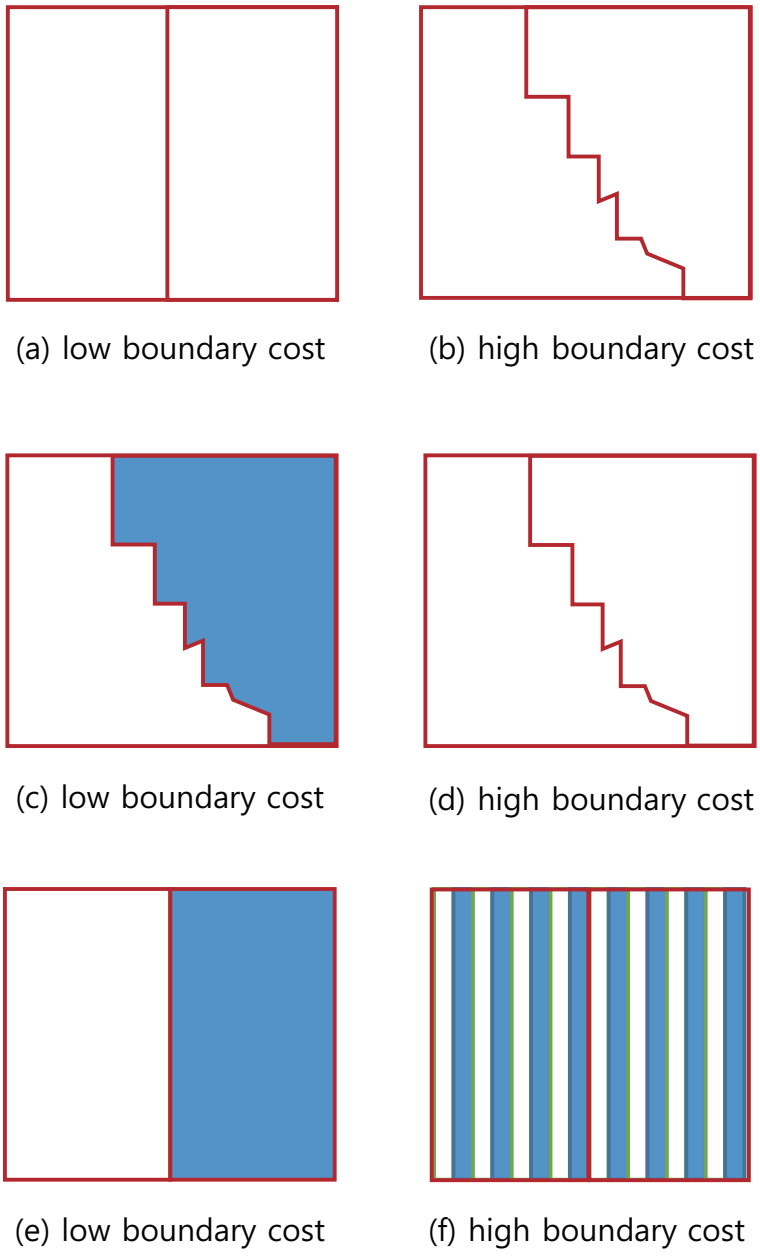


Figure 4.3: Simple explanation of the concept of boundary cost computation.

considers the gradients of the boundary pixels. Fig. 4.3(e) and (f) have the same simple boundaries and the same gradients of the boundary pixels, but Fig. 4.3(e) looks like two different segments while Fig. 4.3(f) looks like one segment because the interior textures of the figures are quite different. The proposed boundary cost uses a non-boundary gradient mean to solve this problem. Briefly, the boundary cost is high when the boundary is complex, when the gradients of the boundary pixels are low, and when the textures in both segments are complex.

The complete form of the boundary cost is as follows:

$$b(i,j) = \sum_{k \in B} \frac{255 - |grad(k) - \rho \cdot (m_g(i) + m_g(j))|}{255} \tag{4.12}$$

and the non-boundary gradient mean m_g is

$$m_g(i) = \frac{1}{N_i} \sum_{k \notin B} grad(k) \tag{4.13}$$

where N_i is the number of non-boundary pixels of segment i. Boundary cost $b(i,j)$ accumulates the values that are inversely related to the gradients of the boundary pixels instead of counting the boundary pixels. This term prevents merging through the object boundaries by a small $b(i,j)$. On the other hand, a boundary on a smooth plane gives a large $b(i,j)$; as such, the segments around the boundary are induced to merge. In addition, the subtraction of m_g from accumulates values prevents a low boundary cost on the false boundary of the complex texture regions, and therefore, the segments in the complex texture can also be merged.

The color similarity $s(i,j)$ is also modified as follows:

$$s(i,j) = \exp(-\eta \|mean(i) - mean(j)\|_1) \tag{4.14}$$

Through the modified term, the color similarity of the adjacent segments has a greater influence at the segment merging stage, which will be explained in the next section.

4.2.2 Depth Map Generation Algorithm

The proposed depth map generation algorithm uses greedy algorithm based iterative local cost optimization [33]. Although the greedy algorithm cannot guarantees to get the global optimum solution, it is very fast and effective to reach the local optimum solution. At the setup process, first, an image is divided into N_s segments through entropy-based oversegmentation [15]. The $C_d(i)$ values of segments i and $\triangle C(i,j)$ for neighboring segment pairs i and j are computed. $\triangle C(i,j)$ refers to the cost that can be reduced when i and j are merged.

$$\triangle C(i,j) = C_d(i) + C_d(j) + C_s(i,j) - C_s(i \cup j) \tag{4.15}$$

In the iteration process, $(i_m, j_m) = \arg\max_{(i,j)} \triangle C(i,j)$ is calculated, and if $\triangle C(i_m, j_m) > 0$, then segment i_m and j_m are merged into i_m. After merging, $C_d(i_m)$, $C_s(i_m, j)$, and $\triangle C(i_m, j)$ are updated for neighboring segment j of i_m. If no merging occurs, the iteration stops.

Fig. 4.4 shows an example of the processes involved in the proposed focus depth map estimation algorithm. At the initial step, the image is divided into 10 segments. In this step, each segment has its own depth plane, but the planes of the segments are not consistent due to the lack of information on each segment. Through iteration, similar segments are merged to decrease the cost function, and the consistency of the map is gradually increased. The figure shows that the consistency is improved after five iterations. Finally, all the segments are merged into one segment, and the map becomes natural.

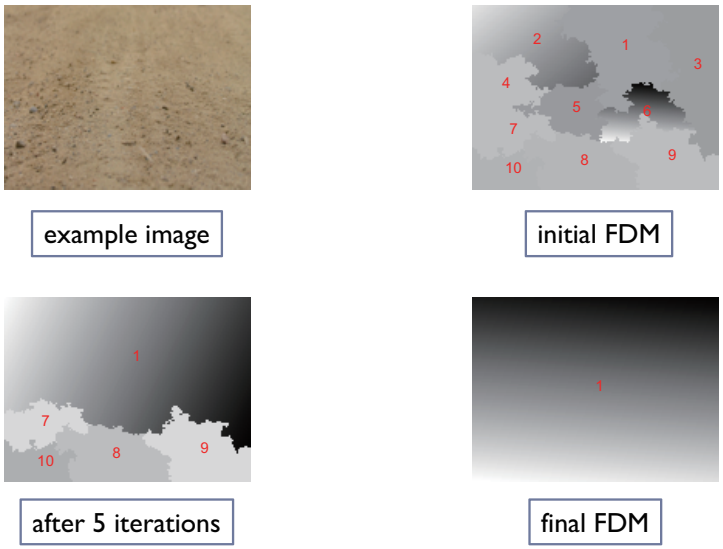


Figure 4.4: Example of the processes of the proposed depth map generation algorithm.

Chapter 5

Experimental Results

The experiments that were performed in this research consisted of two main categories: (1) comparison of the confidence value estimation methods of focus measures; and (2) comparison of the depth map estimation methods. To compare the existing methods and the proposed method, some images were synthesized artificially, and some images were captured with a camera. All the images had enough focus information to estimate a depth map.

5.1 Comparison of the Confidences Value Estimation Methods of Focus Measures

To verify the confidence value estimation methods, artificial images were synthesized using a step function and random rectangles. The step function image was explained in the earlier chapter, and there was a step function on the horizontal center of all the lines of the image with Gaussian blur. The standard deviation of the blur σ_b was vertically increased linearly from 0 to 6. The step function image is shown in Fig. 5.1. In the image, WGN is added, and its standard deviation σ_n is varied to 0, 1, 2, and 3.

The random rectangles image consisted of a plane background and of 20 randomly located rectangles. The widths, heights, locations, directions, and intensities were all random. As overlapping of the rectangles was allowed, the number of visible rectangles could be lower than 20, as shown in Fig. 5.13. The blur of the image was the same as that of the step function image.

For the synthetic images, quantitative evaluation could be performed because the true confidence value of the focus measures and the true σ_b were known. Therefore, the evaluation criteria are the MSE (mean square error) for the confidence value and the MSE of σ_b for the depth (standard deviation) map.

Fig. 5.1 - 5.8 compare the confidence values of the gradient- and second-derivative-based focus measures for the step function image. The true confidence values seem to form a thick line on the vertical center, and their widths are increasing with the blur. The error images are represented by the red and green colors, which mean opposite error signs. In the figures, the ratio of the amounts of the green and red color looks like similarl as such, every empirical coefficient or normalizer is sufficiently tuned for all the methods.

Fig. 5.9 - 5.12 show the MSE of the proposed and of Jing's method on various σ_n. Fig. 5.9 represents the MSE in decimal scale, and Fig. 5.10 is the same data in log scale. In the log scale graph, the difference of the performances on the small error region can be easily shown. Fig. 5.11 shows the ratio of the MSE of the estimated confidence values to the sum of squared true confidence value. It reflects the decrease of the confidence values when σ_n is high because the MSE can be decreased despite the increase of the noise due to the very low confidence value. Fig. 5.11 is its log scale version.

It was shown that the proposed method estimates the confidence values well so that the absolute errors between the estimated confidence values and the true confidence values are quite small, especially when σ_n is high. The errors are amplified for

visualization in the figures, the accurate values of the errors can be seen in Table. 5.1. Compared to Jing's method [9], the proposed method estimates the confidence values of the side lobes better. The application of Jing's method to the second-derivative-based blur estimator seems slightly inappropriate because such method is based on gradients. The reason for applying the method, however, is that it seems that there is no confidence value that fits the second-derivative-based blur estimator.

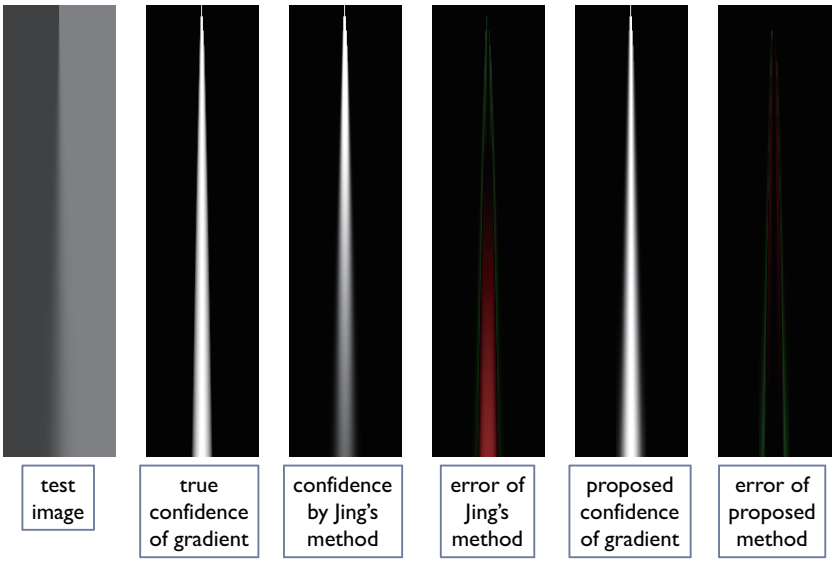


Figure 5.1: Comparison of the confidence values of the gradient-based focus measure for the step function image ($\sigma_n = 0$).

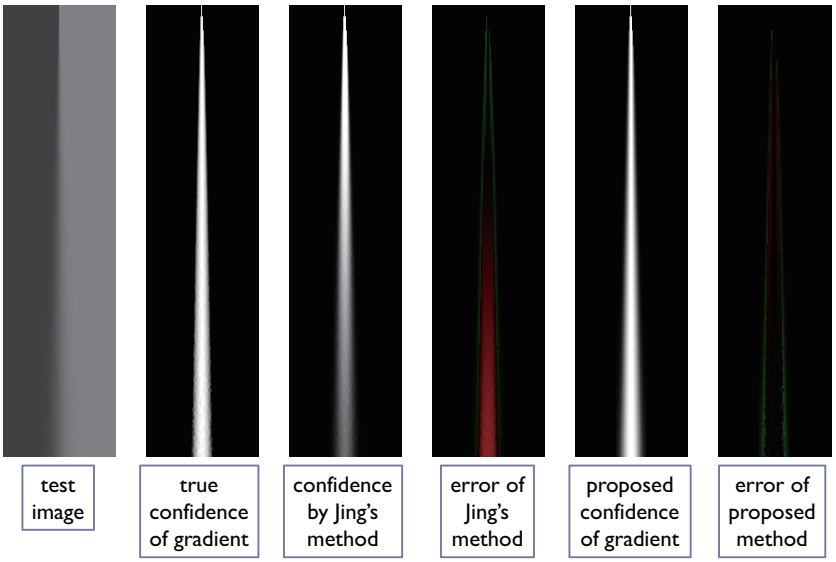


Figure 5.2: Comparison of the confidence values of the gradient-based focus measure for the step function image ($\sigma_n = 1$).

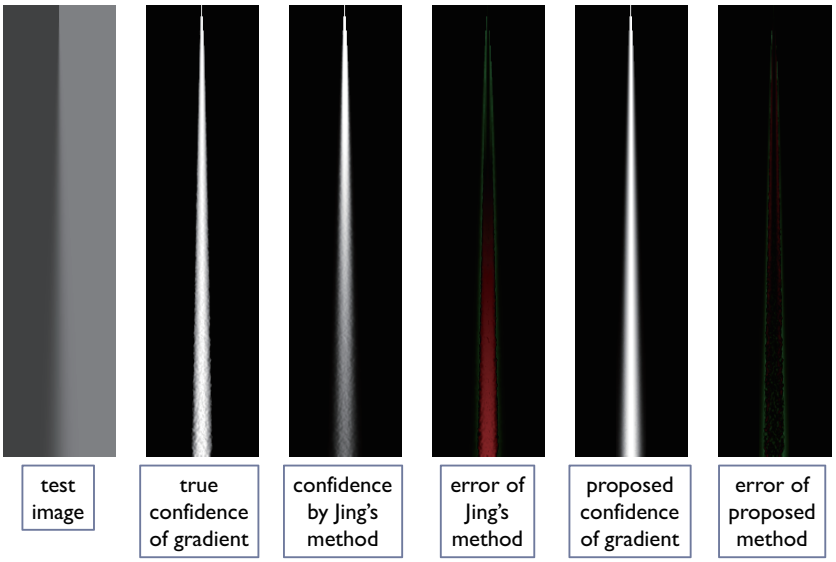


Figure 5.3: Comparison of the confidence values of the gradient-based focus measure for the step function image ($\sigma_n = 2$).

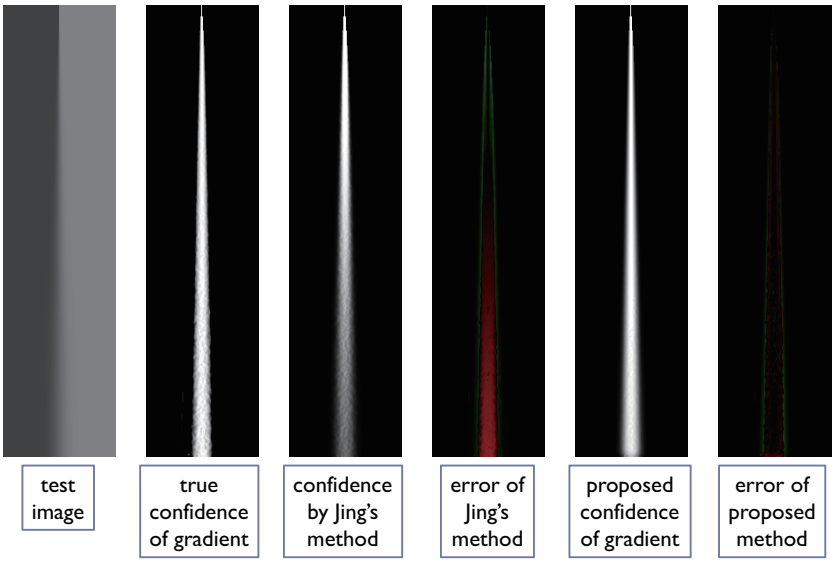


Figure 5.4: Comparison of the confidence values of the gradient-based focus measure for the step function image ($\sigma_n = 3$).

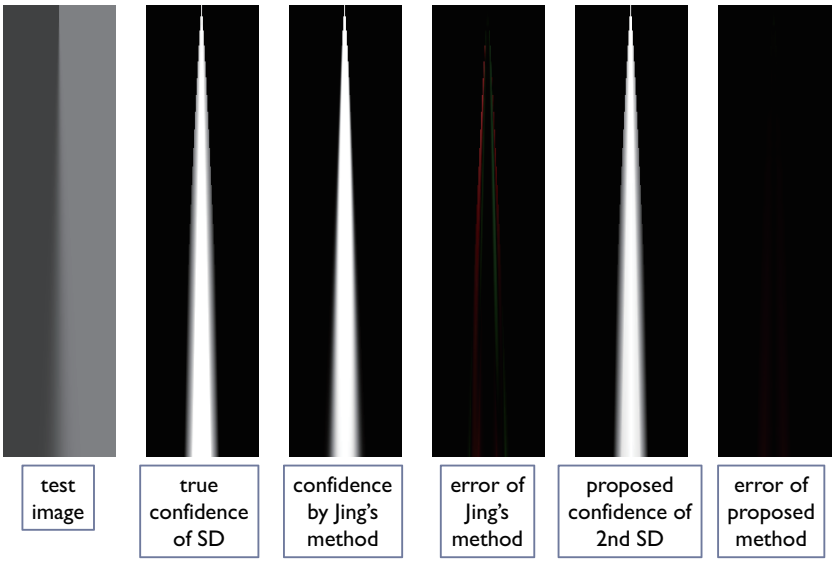


Figure 5.5: Comparison of the confidence values of the second-derivative-based focus measure for the step function image ($\sigma_n = 0$).

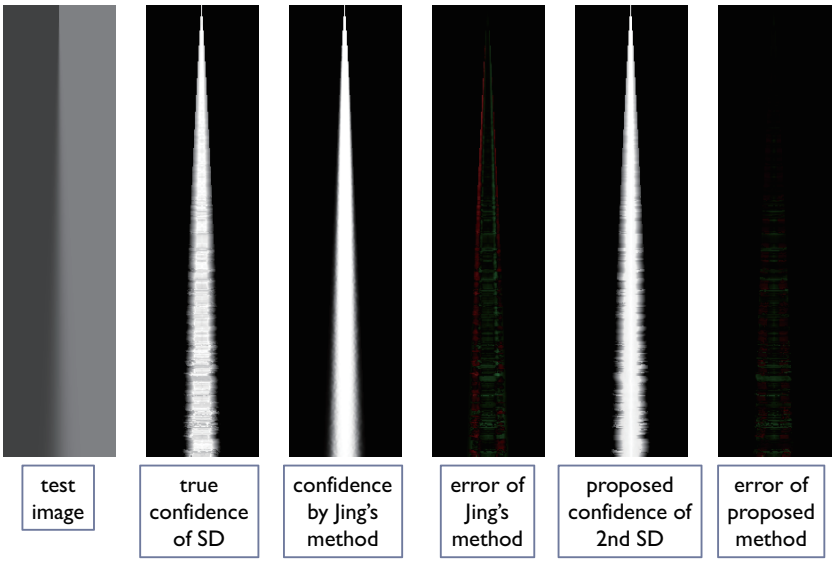


Figure 5.6: Comparison of the confidence values of the second-derivative-based focus measure for the step function image ($\sigma_n = 1$).

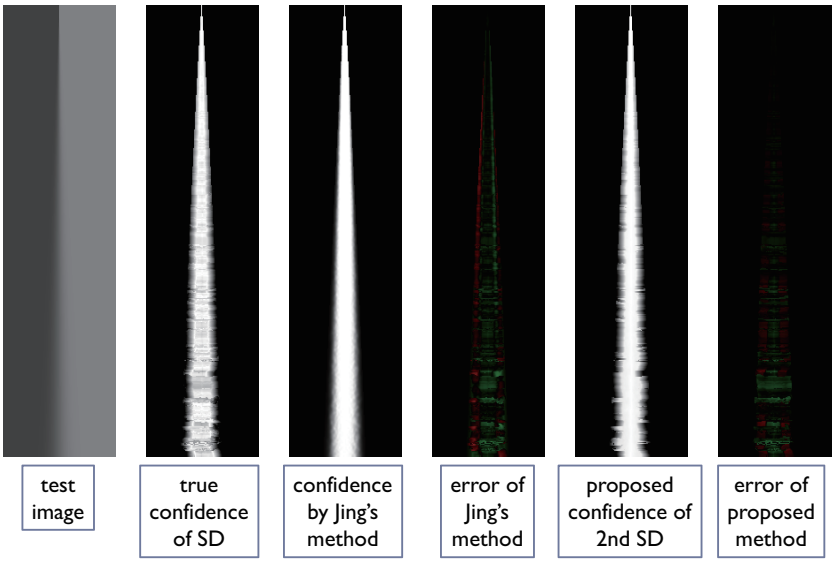


Figure 5.7: Comparison of the confidence values of the second-derivative-based focus measure for the step function image ($\sigma_n = 2$).

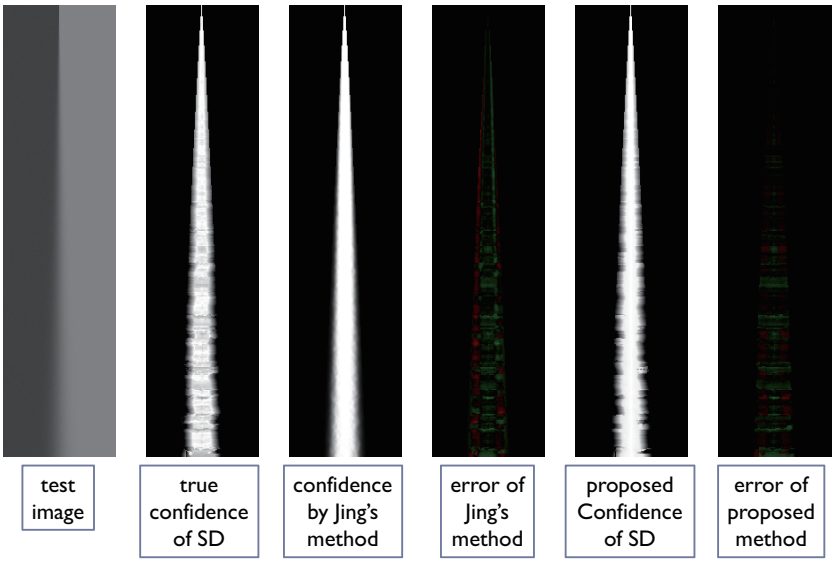


Figure 5.8: Comparison of the confidence values of the second-derivative-based focus measure for the step function image ($\sigma_n = 3$).

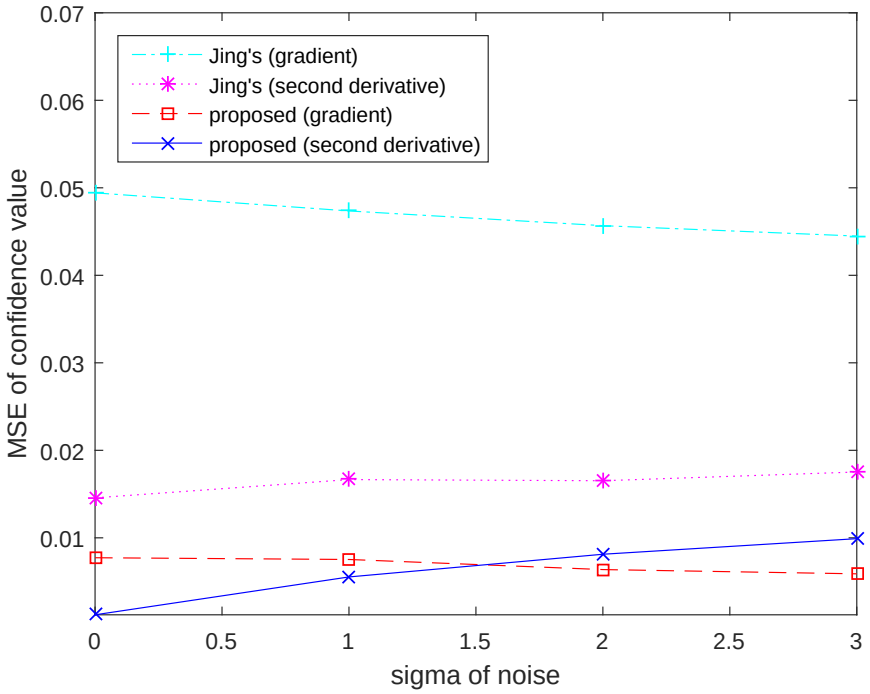


Figure 5.9: Mean square error of the confidence values estimation methods for the step function images in decimal scale.

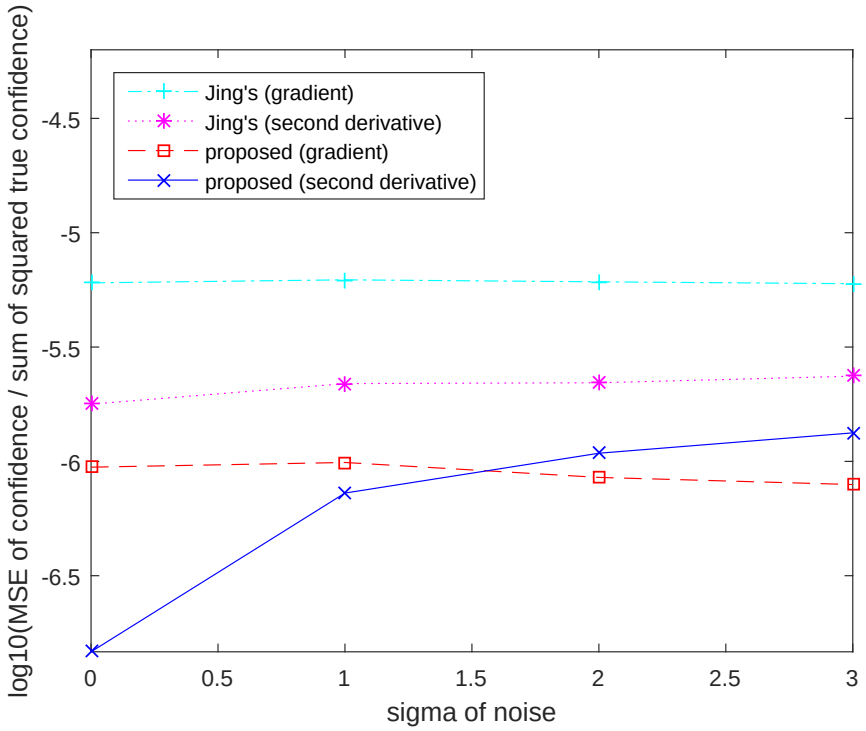


Figure 5.10: Mean square error of the confidence values estimation methods for the step function images in log scale.

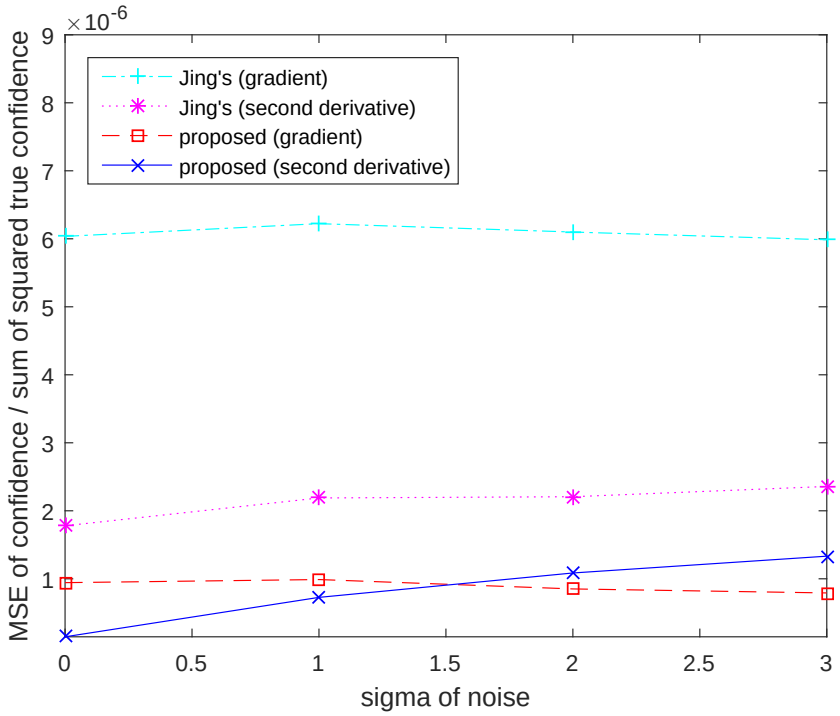


Figure 5.11: Ratio of the mean square error of the estimated confidence values to the sum of squared true confidence values for the step function images in decimal scale.

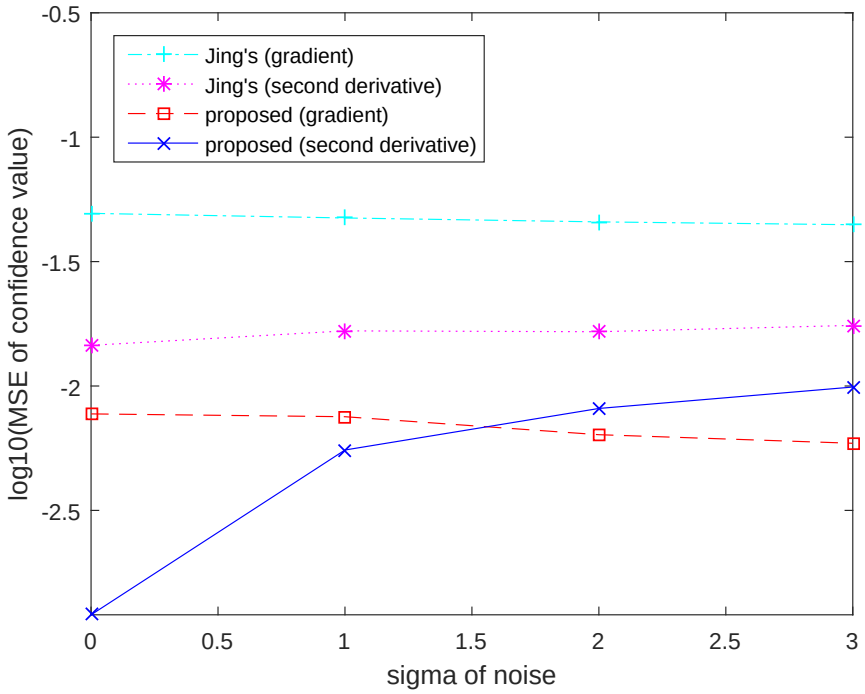


Figure 5.12: Ratio of the mean square error of the estimated confidence values to the sum of squared true confidence values for the step function images in log scale.

Table 5.1: Mean square errors of confidence values for the step function image (average of 5 trials). All the values of the table are 100 times the MSE.

	Jing's method (gradient)	proposed method (gradient)	Jing's method (second derivative)	proposed method (second derivative)
$\sigma_n = 0$	0.4942	0.0772	0.1457	0.0120
$\sigma_n = 1$	0.4735	0.0752	0.1665	0.0552
$\sigma_n = 2$	0.4567	0.0637	0.1652	0.0812
$\sigma_n = 3$	0.4448	0.0588	0.1752	0.0990

Fig. 5.13 - 5.20 compare the confidence values of the gradient- and second-derivative-based focus measures for the random rectangles image. The figures show the results of one example, and Table. 5.2 shows the average MSE of the five trials through randomly generated images. In the figure, the true confidence values are high on the boundaries of the rectangles. Like the step function image, the proposed method was shown to provide better results for the random rectangles images.

Fig. 5.9 - 5.12 show the MSE of the proposed and Jing's method on various σ_n for the random rectangles images. The explanation of each graphs is the same as those of the graphs for the step function images.

With regard to the correctness of the confidence value estimation of the gradient- and second-derivative-based focus measures, the estimator for the second-derivative-based focus measure showed better results in the simple regions, like the step image or an isolated rectangle in the random rectangles image. The estimator for the second-derivative-based focus measure, however, showed weakness at the complex region, like the overlapped rectangles. With regard to the overall correctness, the estimator for the second-derivative-based focus measure showed better results.

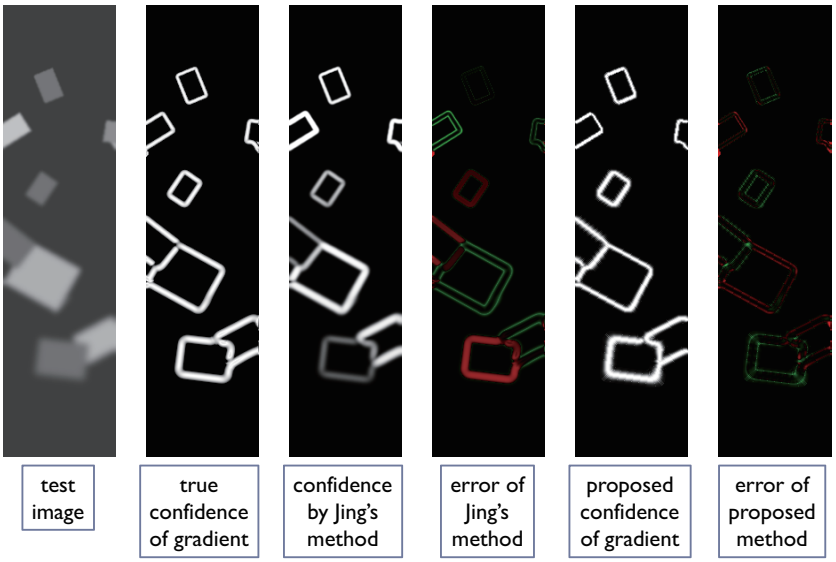


Figure 5.13: Comparison of the confidence values of the gradient-based focus measure for the random rectangles image ($\sigma_n = 0$).

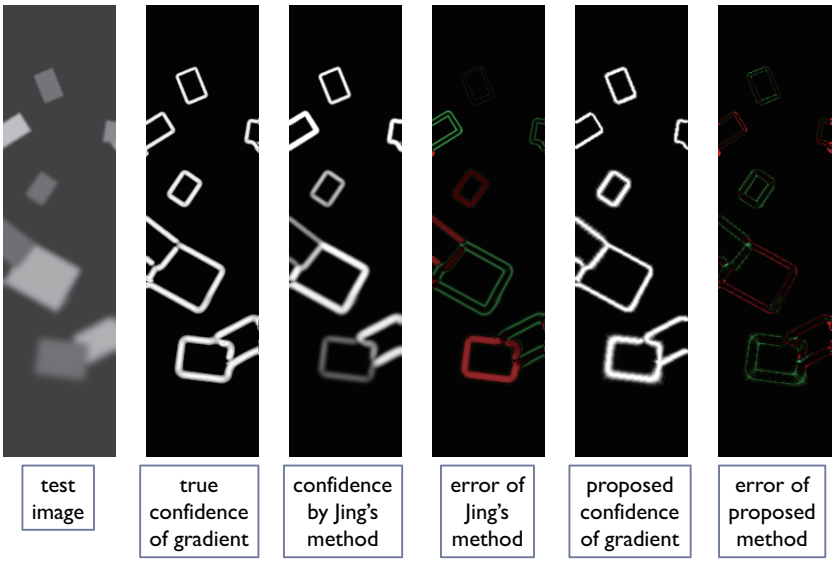


Figure 5.14: Comparison of the confidence values of the gradient-based focus measure for the random rectangles image ($\sigma_n = 1$).

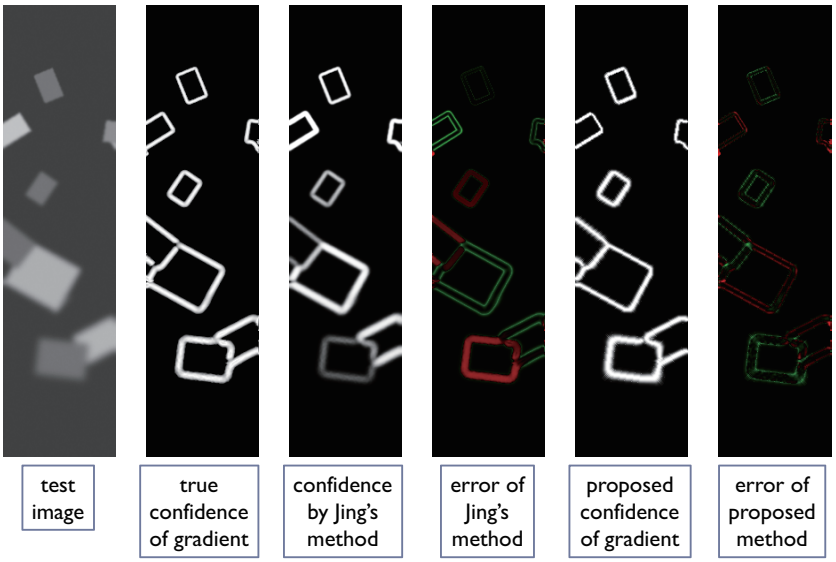


Figure 5.15: Comparison of the confidence values of the gradient-based focus measure for the random rectangles image ($\sigma_n = 2$).

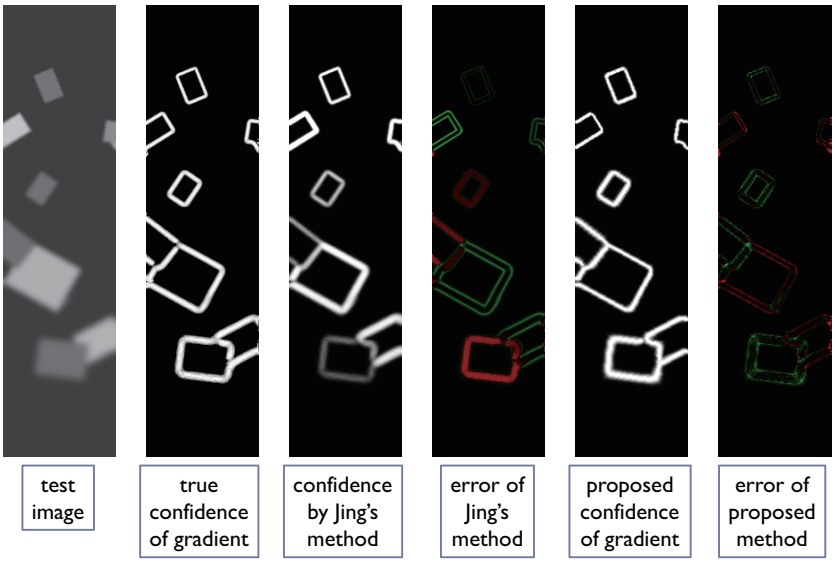


Figure 5.16: Comparison of the confidence values of the gradient-based focus measure for the random rectangles image ($\sigma_n = 3$).

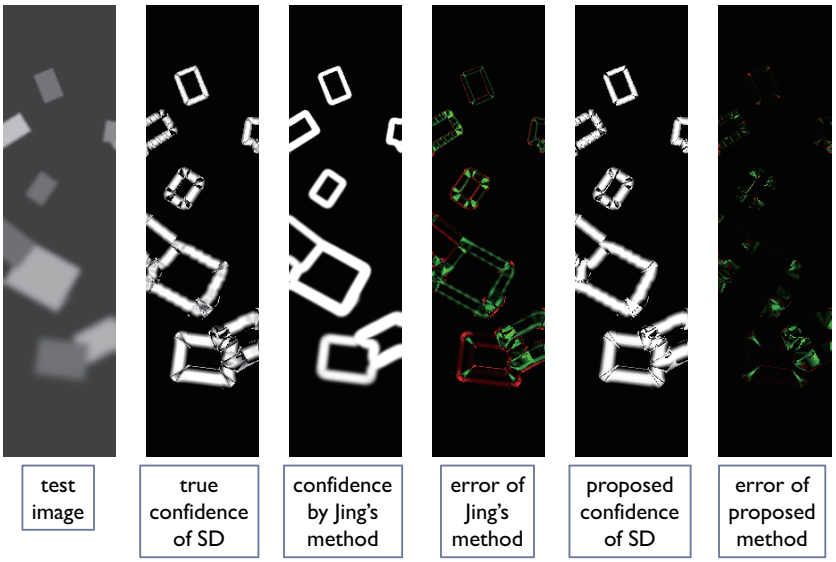


Figure 5.17: Comparison of the confidence values of the second-derivative-based focus measure for the random rectangles image ($\sigma_n = 0$).

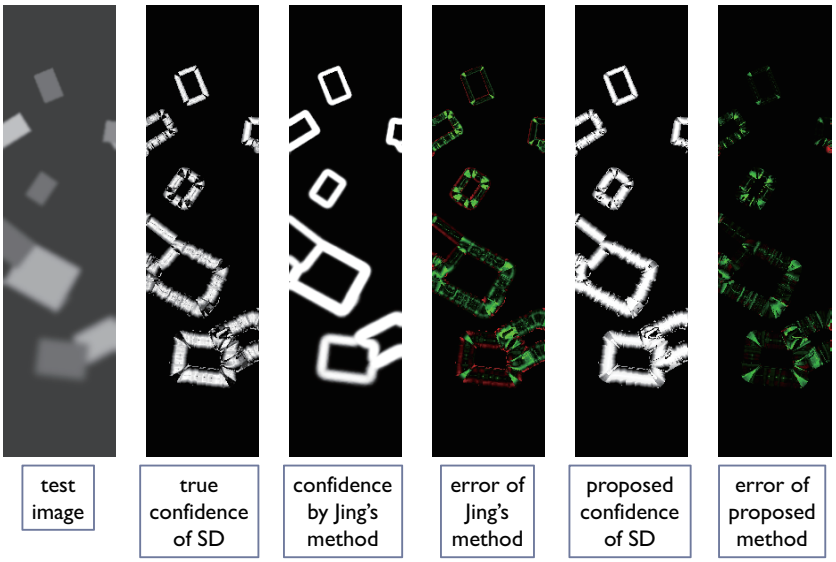


Figure 5.18: Comparison of the confidence values of the second-derivative-based focus measure for the random rectangles image ($\sigma_n = 1$).

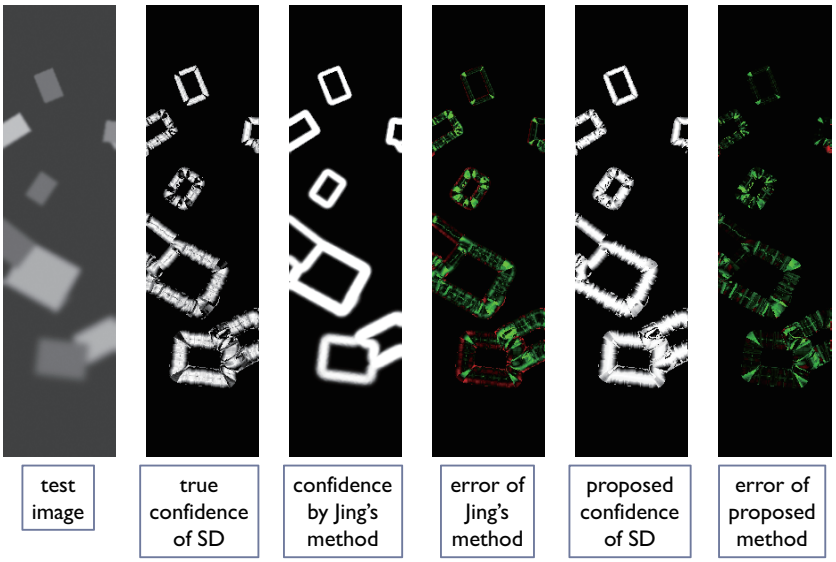


Figure 5.19: Comparison of the confidence values of the second-derivative-based focus measure for the random rectangles image ($\sigma_n = 2$).

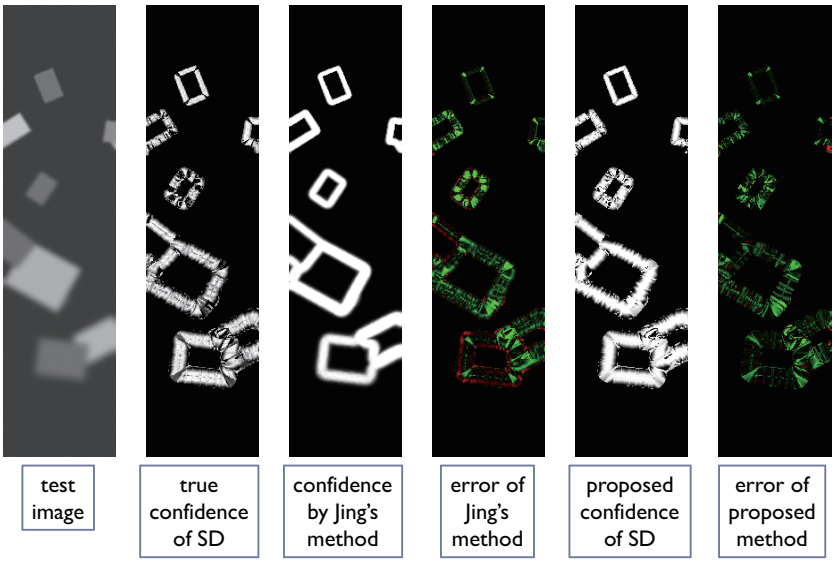


Figure 5.20: Comparison of the confidence values of the second-derivative-based focus measure for the random rectangles image ($\sigma_n = 3$).

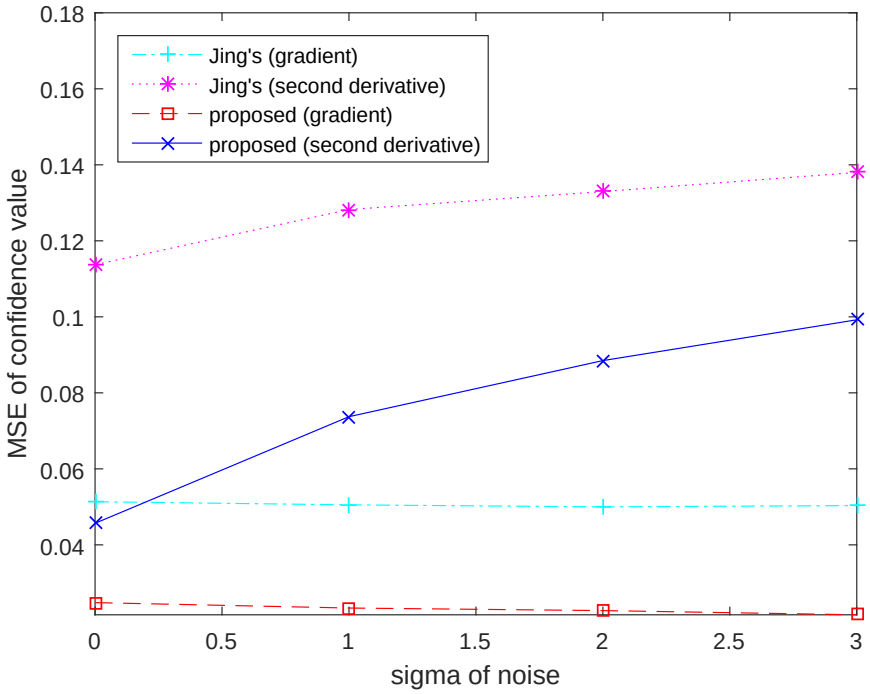


Figure 5.21: Mean square error of the confidence values estimation methods for the random rectangles images in decimal scale

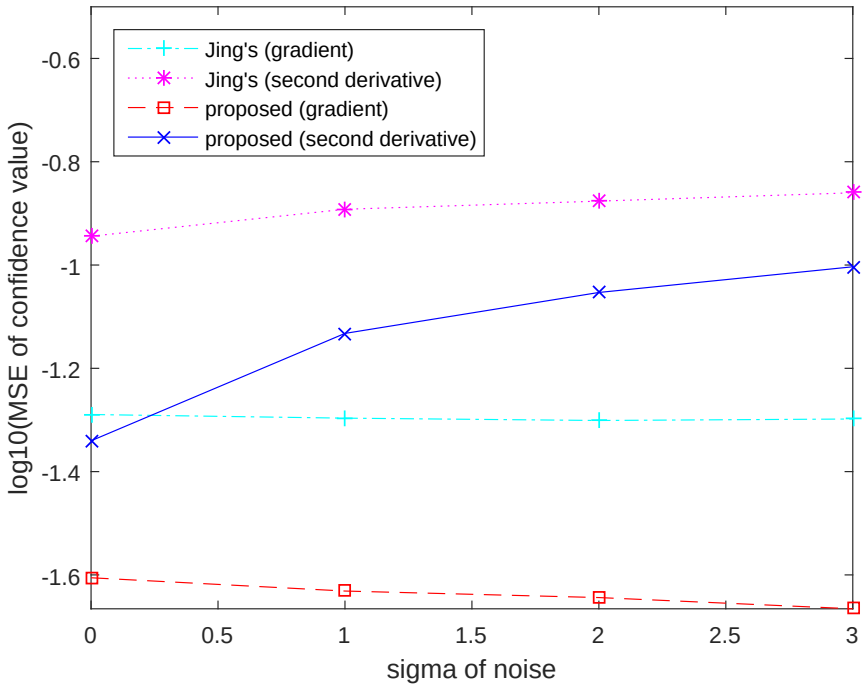


Figure 5.22: Mean square error of the confidence values estimation methods for the random rectangles images in log scale.

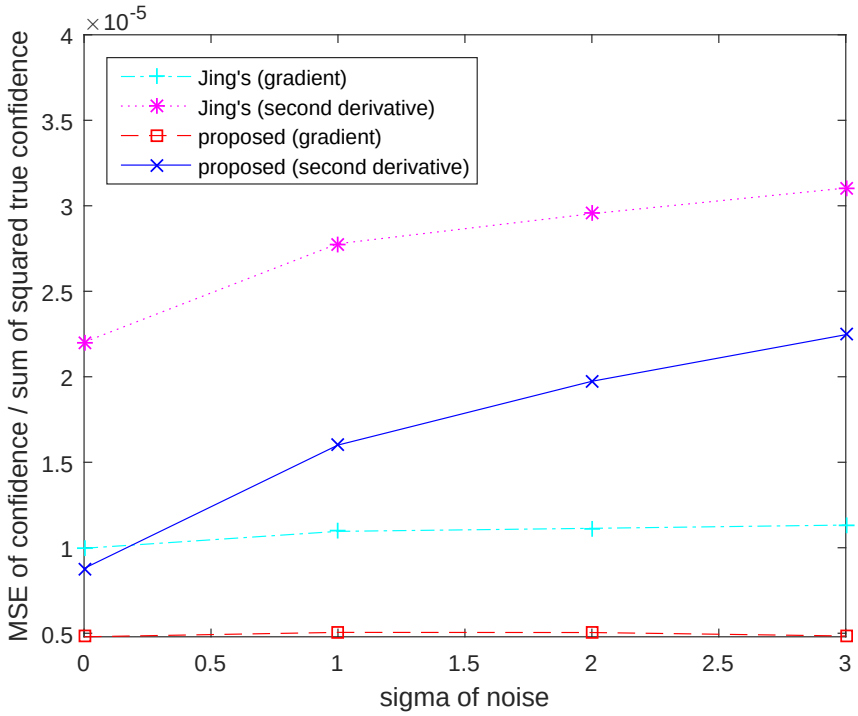


Figure 5.23: Ratio of mean square error of the estimated confidence values to the sum of squared true confidence values for the random rectangles images in decimal scale.

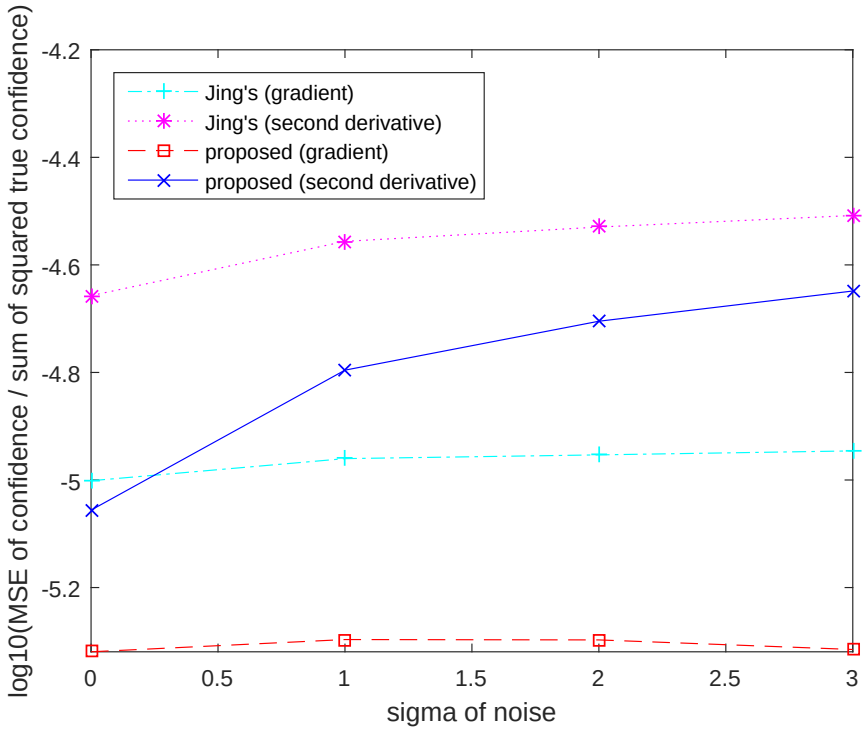


Figure 5.24: Ratio of mean square error of the estimated confidence values to the sum of squared true confidence values for the random rectangles images in log scale.

Table 5.2: Mean square errors of the confidence values for the random rectangles image (average of 5 trials). All the values of the table are 100 times the MSE.

	Jing's method (gradient)	proposed method (gradient)	Jing's method (second derivative)	proposed method (second derivative)
$\sigma_n = 0$	5.135	2.480	11.371	4.567
$\sigma_n = 1$	5.053	2.337	12.823	7.372
$\sigma_n = 2$	5.000	2.271	13.296	8.849
$\sigma_n = 3$	5.033	2.160	13.800	9.922

5.2 Performances of the Proposed Depth Map Generation Method

Depth map estimation experiments were performed on synthetic and real images. The focus depth maps of the same images with the past section were estimated using the proposed and existing methods to compare their performances. Chen's method [4] is based on the Gaussian filter, and Lee's method [5] is based on oversegmentation and averaging. Although Lee deals not only with the focus depth map but also with the pseudo depth map and depth fusion, in this paper, only the focus depth map generation parts of Lee's method are compared.

With regard to the term "focus depth map," in principle, σ_b and the depth are not the same. In the previous chapter, however, it was explained that perceptual depth d_p, blur radius r, and σ_b have a linear relationship. Therefore, with undefined or unknown parameters, the σ_b map can be regarded as the scaled perceptual depth map. In accordance with this interpretation, the focus depth map means the σ_b map in the experiments described in this section.

5.2.1 Experiments on Synthetic Images

In Fig. 5.25, the true σ_b map is a uniformly increasing map in the vertical direction, but there is no blur information on the textureless regions. Therefore, Chen's method cannot estimate the σ_b of the plane regions because it adopts the smoothing Gaussian filter. Lee's method also cannot obtain an appropriate map because the edge and plane regions are separated by the oversegmentation process. On the other hand, the proposed method estimates the planar map by iteratively merging all the segments into one segment. These tendencies are shown for the random rectangles image, as in Fig. 5.26. The errors of the proposed method are fewer than those of the existing methods. This is because the proposed method shows strong points for the gradual changes of σ_b.

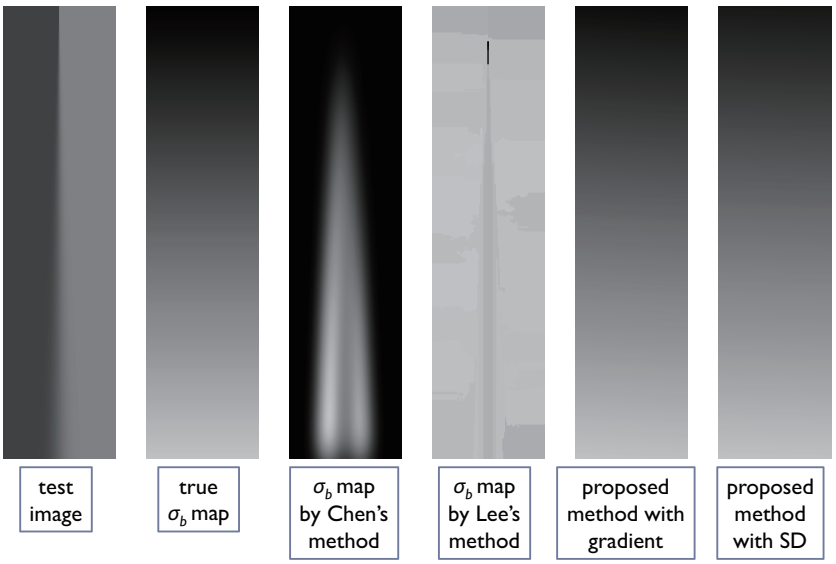


Figure 5.25: Estimated depth (standard deviation of the blur) maps for the step function image.

In addition, the iterative segment merging process gathers local information to build a robust σ_b plane. The overall performance of the gradient-based method is better than that of the second-derivative-based method. It seems that the reason for this is the larger error of the second-derivative-based blur estimator when σ_b is low.

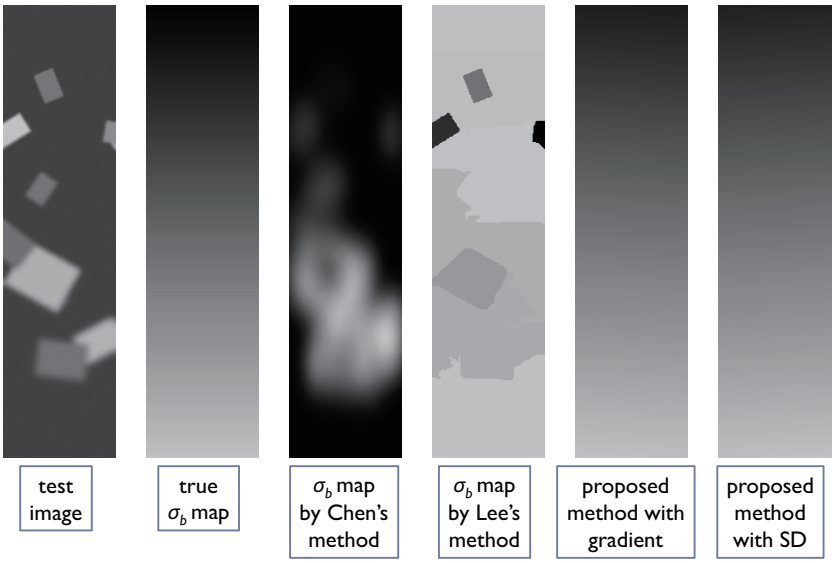


Figure 5.26: Estimated depth (standard deviation of the blur) maps for the random rectangles image.

Table 5.3: Mean square errors of estimated depth (σ_b) maps using synthetic images.

	Chen's method	Lee's method	proposed method (gradient)	proposed method (second derivative)
step function	17.8115	29.2947	0.1426	0.4222
random rectangles (average of 5 trials)	4.4321	6.5277	0.8327	0.9180

5.2.2 The Experiments on Real Images

Unlike with synthetic images, generally, there are no known true distance or depth maps of existing images. Therefore, images are captured under a controlled configuration to know the true distance and depth for quantitative evaluation using real images. In this research, five images are captured under known configurations described in Fig. 5.27. In this section, the focus depth maps are estimated using the proposed method from the real images and are compared to the results of Chen's method and Lee's method.

Fig. 5.28 shows the true distance maps and the true perceptual depth maps of the captured images. For visual convenience, the distance and depth of the gray images are inverted so that the brighter region means the nearer region. Every true distance is physically measured by meter, and the true distance images of the figure are normalized for visual purposes. The true perceptual depth maps are computed using equation 3.3 from the true distance maps. In addition, the measurement of the distance considers the capturing angle of the camera and the curve of the object; as such, a slightly curved surface of the floor and the object can be shown in the 3D visualized versions of the perceptual depth mapss.

Images 1-3 include a floor whose depth changes gradually, as with synthetic images. Chen's and Lee's methods show a coarse tendency of depth changes but also include unnatural details due to the smoothing filter and oversegmentation. The proposed method, on the other hand, shows the natural depth maps based on the gradual depth changes. As appropriate to the shapes and depths of objects like a plate and a book, Chen's and Lee's methods do not represent the smooth depths of objects. The proposed method, however, shows the smooth planes of the depths of objects, although some errors of the slopes are visible.

Images 4-5 consist of a focused foreground object and a blurred background. Chen's method cannot emphasize the depth difference between the foreground and

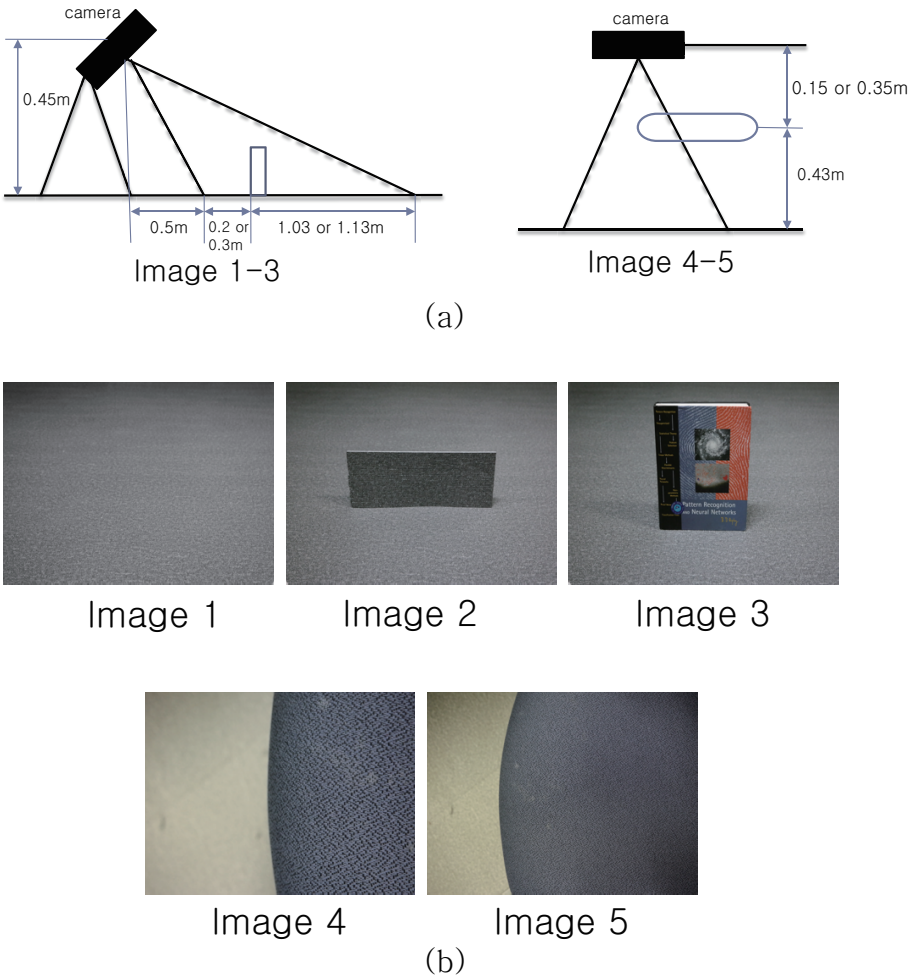


Figure 5.27: Capturing configuration and captured images (a) camera and object configuration; and (b) captured images.

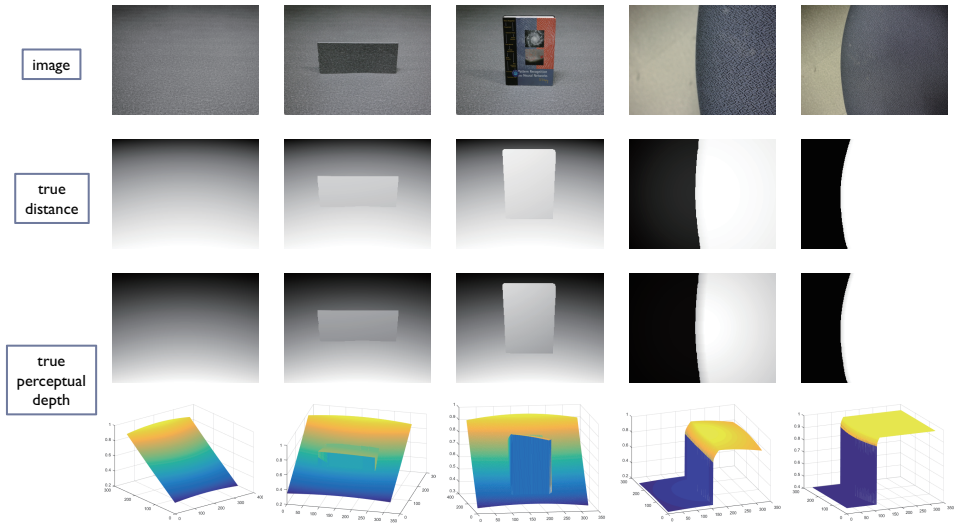


Figure 5.28: True distance maps and true perceptual depth maps of the images.

the background due to the very large gradient of the object boundaries. Lee's method shows moderate qualities but still contains unnatural depth discontinuities on the borders of the segments. On the other hand, the proposed method shows consistent and natural depth maps on both the foreground objects and the background regions.

Fig. 5.31 - 5.32, and Table. 5.4 show quantitative comparisons of the depth map estimation results. For every image, the results of the proposed method are better than those of the other methods. Especially, the proposed method outperforms the other methods at image 3, which includes the gradual depth changes and the flat region simultaneously.

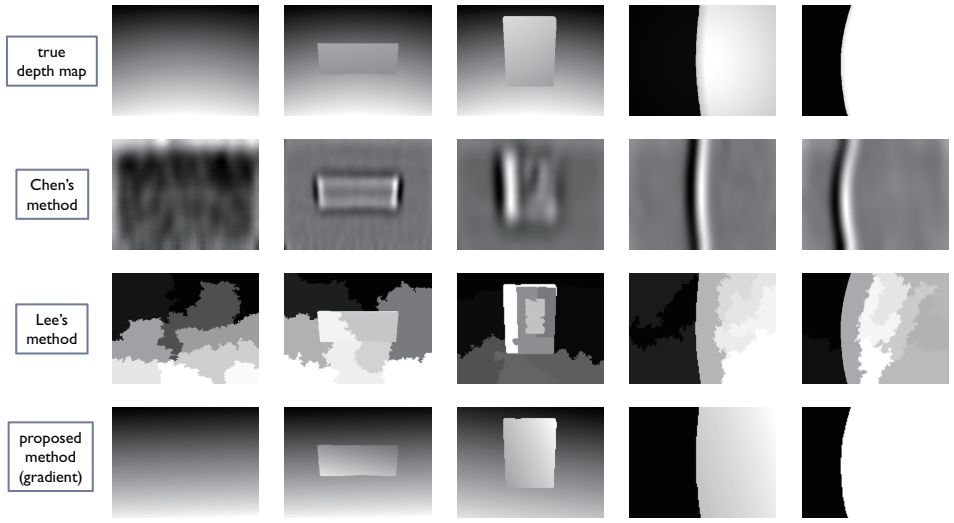


Figure 5.29: True distance maps and estimated depth maps obtained through Chen's method, Lee's method, and the proposed method (gradient-based).

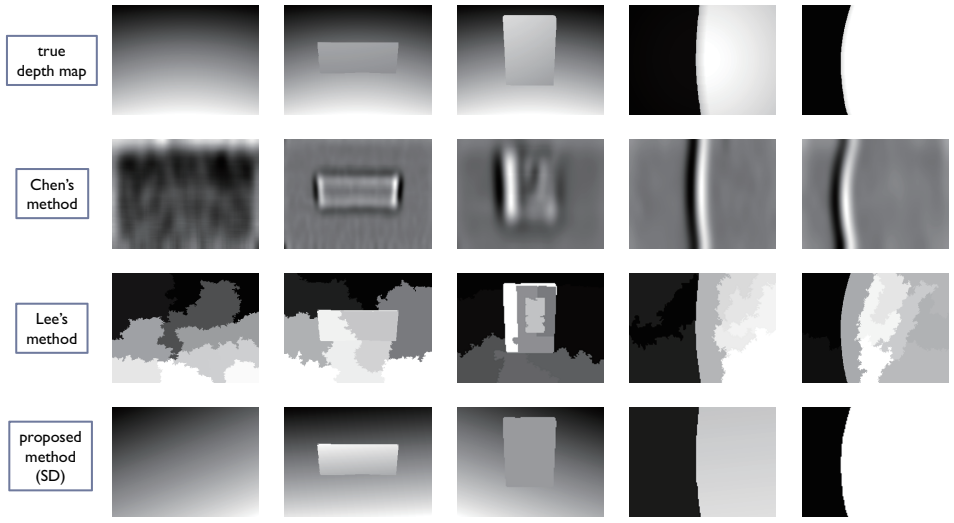


Figure 5.30: True distance maps and estimated depth maps obtained through Chen's method, Lee's method, and the proposed method (second-derivative-based).

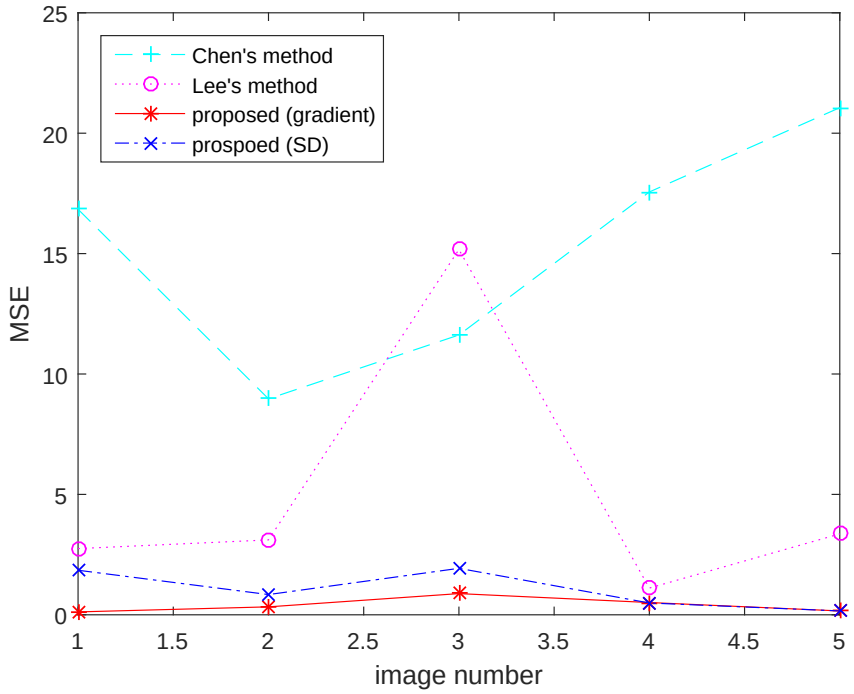


Figure 5.31: Mean square error of the depth map estimation methods for the real images in decimal scale

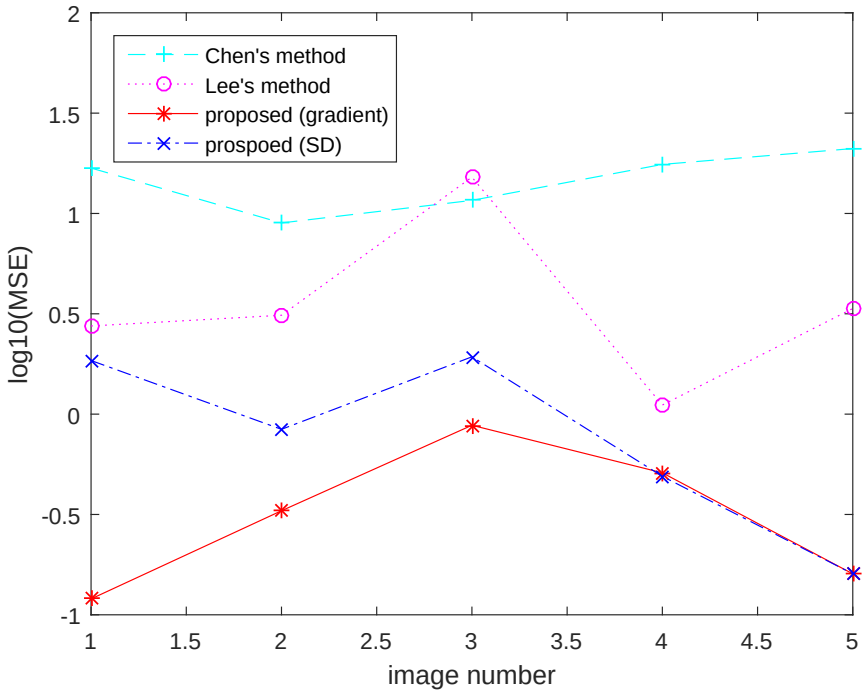


Figure 5.32: Mean square error of the depth map estimation methods for the real images in log scale

Table 5.4: Mean square errors of estimated depth maps using the real images. All the values of the table are 100 times the MSE.

	Chen's method	Lee's method	proposed method (gradient)	proposed method (second derivative)
image 1	16.9	2.75	0.12	1.85
image 2	8.97	3.11	0.33	0.84
image 3	11.6	15.2	0.88	1.93
image 4	17.6	1.11	0.51	0.49
image 5	21.1	3.36	0.16	0.16
average	15.2	5.10	0.40	1.05

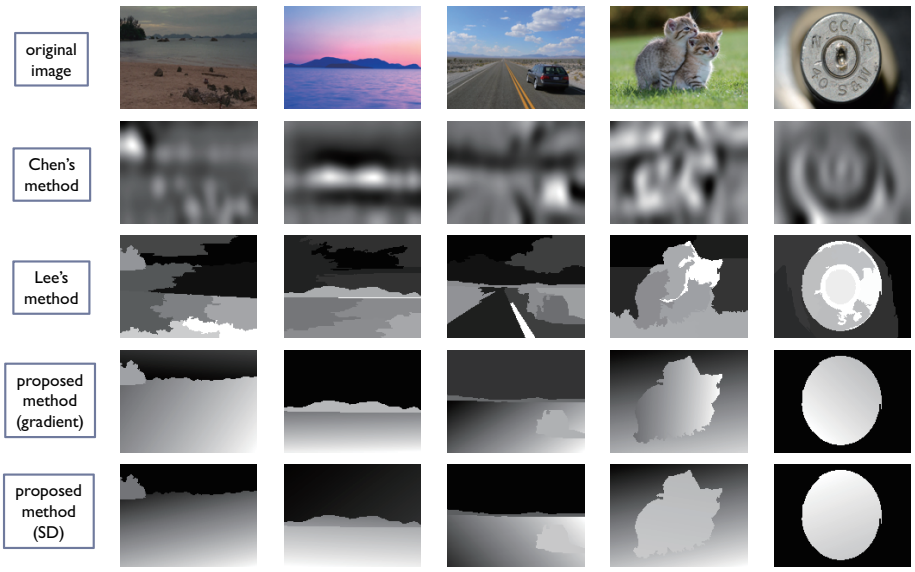


Figure 5.33: Estimated depth maps using practical images.

Fig. 5.33 shows the results of real images with unknown true depths. These images represent landscape scenes, outdoor scenes, and arbitrary objects; as such, their true distances or depths are unknown. To show the results using practical images, however, experiments using images gathered from the Internet are performed. Note that all the images have sufficient defocus information and meet the condition that the nearest region is focused. As shown in Fig. 5.33, the overall properties of the resulting images are similar to those in the previous experiments using images with known true depths. The results of the use of the proposed method show smooth and natural depth maps.

5.2.3 Execution Time

In the previous chapter, the proposed method showed better results than the existing methods, but the proposed method is in fact quite complex compared to the other methods. In other words, although the existing methods do not give correct depth maps, their execution times are very short. The following figures show the execution time of the proposed method in various cases. The size of the test image is QVGA (240 rows, 320 columns), and the testing system is core i7 3.2Ghz with MATLAB.

Fig. 5.34 shows the execution times when the number of initial segments for the oversegmentation process is changed with the assumption that a whole image is merged into one segment. The second-derivative-based method needs more time than the gradient-based method does due to the blur estimation step. The execution time increases when the number of initial segments is high, but the amounts of time increase by the number of initial segments are similar for both. For comparison, Chen's method records $0.694s$, and Lee's method records $0.946s$, and the proposed method is about 30-60 times slower when the number of initial segments is around 15.

Fig. 5.35 shows the execution times when the number of final segments for the oversegmentation process is changed. The number of final segments is dependent on an image; that is, an image that has many objects or segments needs a shorter time, and vice versa. The second-derivative-based method also needs more time than the gradient-based method does. The execution time decreases when the number of final segments is high because the number of iterations is low at the merging step. In the experiments herein, the number of initial segments is fixed to 15.

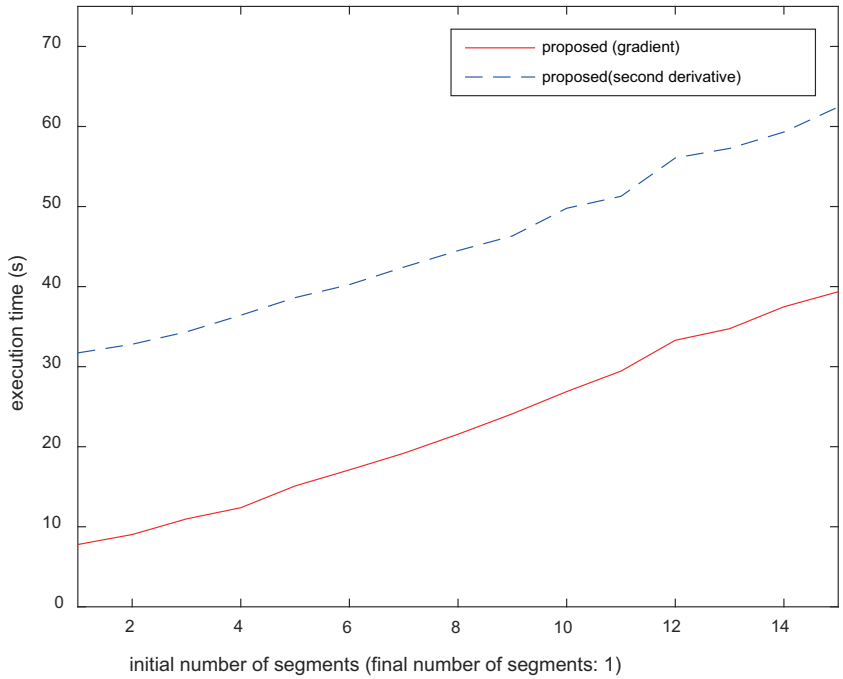


Figure 5.34: Effect of the number of initial segments on the execution time.

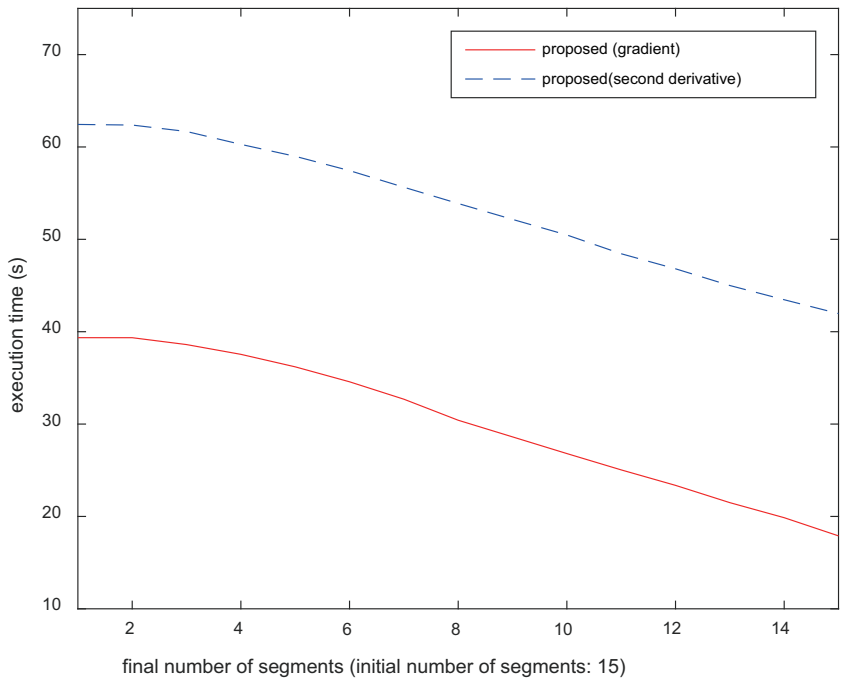


Figure 5.35: Effect of the number of final segments on the execution time.

Chapter 6

Conclusion

In this paper, a new solution for the DFD (depth from defocus) problem is proposed. The typical DFD method assumes that the focus is on the nearest region of a scene to avoid ambiguity in the relation between the depth and the defocus blur, and the proposed method follows such assumption. First, the perceptual depth, which is based on human depth perception, is defined, and then the (true) confidence values of a focus measure are defined using the perceptual depth. The estimation methods of confidence values are designed for the gradient- and second-derivative-based focus measures. These estimated confidence values are more correct than those of the existing method. The proposed focus depth map estimation method is based on the segment-wise planar model. In the fitting process of each segment, confidence values are used as fitting weights of the weighted linear least square fitting method. To solve the cost optimization problem of focus depth map estimation, iterative local optimization based on the greedy algorithm is used.

To verify the performance of the proposed method, comparative experiments are performed using the existing methods and the proposed method. The test set, which is used in the experiments, consists of artificial and real images whose true distances and

depths are known. The proposed method shows better performances than the existing approaches for defocus blur estimation and depth map estimation.

Some problems of DFD remain, however, or need to be alleviated. First, the results of the estimation of the confidence value of the gradient-based method cannot be considerably improved when the edge detection results have poor quality. To achieve better results in poor environments, the edge detector has to be more robust to noise and blur. Next, sometimes the density of the confidence values does not give correct information about the shape of a segment. For example, in the case of the perspective view of a wide and less textured floor, the algorithm decided that it is a flat segment due to the high variance of the confidence value densities. Lastly, oversegmentation does not give the exact object boundaries for some images. The segmentation results of some images show merged objects and unnatural boundaries, making the resulting depth maps also look unnatural locally.

Furthermore, some images do not meet the essential assumption that the focus is on the nearest region of a scene, and some images have motion blur due to the moving capturing devices that were used in obtaining them. In these cases, the correct depth map cannot be generated from the amounts of blur. To solve these problems, depth fusion using another depth cue is widely used. For example, motion estimation can give approximated information about motion blur; as such, more accurate defocus blur estimation is possible from the blur of an image. As no single depth cue is perfect for all images, the adaptive depth fusion by image property has to be improved in the future.

Bibliography

[1] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 493-500, 2007.

[2] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 16, no .8, pp. 824-831, 1994.

[3] J. M. Tenenbaum, *Accommodation in computer vision*, Stanford University, 1970.

[4] Y. C. Chen, Y. C. Wu, C. H. Liu, W. C. Sun, and Y. C. Chen, "Depth map generation based on depth from focus,". *IEEE International Conference on Electronic Devices, Systems and Applications*, pp. 59-63, 2010.

[5] J. Lee, S. Yoo, C. Kim, and B. Vasudev, "Estimating scene-oriented pseudo depth with pictorial depth cues," *IEEE Transactions on Broadcasting*, vol.59, no.2, pp. 238-250, 2013.

[6] M. T. Mahmood, A. Majid, and T. Choi, "Optimal depth estimation by combining focus measures using genetic programming," *Information Sciences*, vol. 181, no. 7, pp. 1249-1263, 2011.

[7] M. Subbarao, and J-K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 864-870, 1998.

[8] A. P. Pentland, "A new sense for depth of field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 4, pp. 523-531, 1987.

[9] B. Jing and D. S. Yeung, "Recovering depth from images using adaptive depth from focus," *2012 IEEE International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 3, pp. 1205-1211, 2012.

[10] M. J. Hirsch and F. W. Weymouth, "DISTANCE DISCRIMINATION: II. Effect on Threshold of Lateral Separation of the Test Objects," *Archives of Ophthalmology*, vol. 39, no. 2, pp. 224-231, 1948.

[11] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," *IEEE 18th International Conference on Pattern Recognition*, vol. 3, 2006.

[12] S. Zhuo and T. Sim, "Defocus map estimation from a single image," *Pattern Recognition*, vol. 44, no. 9, pp. 1852-1858, 2011.

[13] D. Hoiem, A. A. Efros, and M. Hebert, "Recovering surface layout from an image," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 151-172, 2007.

[14] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167-181, 2004.

[15] M. Liu, et al, "Entropy rate superpixel segmentation," *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[16] H. J. Howard, "A test for the judgment of distance," *Transactions of the American Ophthalmological Society*, vol. 17, pp. 195-235, 1919.

[17] J. I. Pascal, "Parallactic angle in binocular depth perception," *Archives of Ophthalmology*, vol. 28, no. 2, pp. 258-262, 1942.

[18] C. J. Erkelens, and R. van Ee, "A computational model of depth perception based on headcentric disparity," *Vision Research*, vol. 38, no. 19, pp. 2999-3018, 1998.

[19] L. L. Sloan, and A. Altman, "Factors involved in several tests of binocular depth perception," *AMA archives of ophthalmology*, vol. 52, no. 4, pp. 524-544, 1954.

[20] A. S. Malik and T. Choi, *Consideration of various Noise Types and Illumination Effects for 3D shape recovery*, INTECH Open Access Publisher, 2008.

[21] J. Yang and D. Schonfeld, "Virtual focus and depth estimation from defocused video sequences," *IEEE Transactions on Image Processing*, vol. 19, no. 3, pp. 668-679, 2010.

[22] Y. Tai and M. S. Brown, "Single image defocus map estimation using local contrast prior," *2009 16th IEEE International Conference on Image Processing (ICIP)*, 2009.

[23] S. Zhuo and T. Sim, "Recovering depth from a single defocused image," *International Journal of Pattern Recognition*, 2011.

[24] H. Bouma, J. Dijk, and A. WM van Eekeren, "Precise local blur estimation based on the first-order derivative," *SPIE Defense, Security, and Sensing. International Society for Optics and Photonics*, 2012.

[25] V. P. Namboodiri, and S. Chaudhuri, "On defocus, diffusion and depth estimation," *Pattern Recognition Letters*, vol. 28, no. 3, pp. 311-319, 2007.

[26] J. H. Elder and S. W. Zucker, "Local scale control for edge detection and blur estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, pp. 699-716, 1998.

[27] W. Zhang and F. Bergholm, "Multi-scale blur estimation and edge type classification for scene analysis," *International Journal of Computer Vision*, vol. 24, no. 3, pp. 219-250, 1997.

[28] M. T. Mahmood and T. Choi, "Focus measure based on the energy of high-frequency components in the S transform," *Optics letters*, vol. 35, no. 8, pp. 1272-1274, 2010.

[29] C. Tang, C. Hou, and Z. Song, "Defocus map estimation from a single image via spectrum contrast," *Optics letters* vol. 38, no. 10, pp. 1706-1708, 2013.

[30] S. Shim and T. Choi, "A novel iterative shape from focus algorithm based on combinatorial optimization," *Pattern Recognition*, vol. 43, no. 10, pp. 3338-3347, 2010.

[31] A. S. Malik and T. Choi, "A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise," *Pattern Recognition*, vol. 41, no. 7, pp. 2200-2225, 2008.

[32] M. T. Mahmood and T. Choi, "3D shape recovery from image focus using kernel regression in eigenspace," *Image and Vision Computing*, vol. 28, no. 4, pp. 634-643, 2010.

[33] S. Birchfield and C. Tomasi, "Multiway cut for stereo and motion with slanted surfaces," *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, pp. 489-495, 1999.

[34] P. F. Felzenszwalb and D. P. Huttenlocher, "Image segmentation using local variation," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 98-104, 1998.

[35] J. Andrade-Cetto and Albedo Sanfeliu, "Integration of perceptual grouping and depth," *15th International Conference on Pattern Recognition*, vol. 1, pp. 295-298, 2000.

[36] A. Fahad and T. Morris, "A faster graph-based segmentation algorithm with statistical region merge," *Advances in Visual Computing. Springer Berlin Heidelberg*, pp. 286-293, 2006.

[37] J. C. Pichel, D. E. Singh, and F. F. Rivera, "Image segmentation based on merging of sub-optimal segmentations," *Pattern recognition letters*, vol. 27, no. 10, pp. 1105-1116, 2006.

[38] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov, "Fast approximate energy minimization with label costs," *International journal of computer vision*, vol. 96, no. 1, pp. 1-27, 2012.

[39] B. Peng, L. Zhang, and D. Zhang, "Automatic image segmentation by dynamic region merging," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3592-3605, 2011.

[40] Z. Ren and G. Shakhnarovich, "Image segmentation by cascaded region agglomeration," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2011-2018, 2013.

[41] T. T. Htar and S. L. Aung, "Enhancement of Region Merging Algorithm for Image Segmentation," *International Conference on Advances in Engineering and Technology (ICAET)*, 2014.

국문 초록

깊이 맵이란 영상 내에서 촬영 장치로부터 가깝고 먼 정도를 수치적으로 나타낸 것으로서 영상의 3차원 구조를 나타내기 위해 널리 쓰이는 표현 방식이다. 2차원 영상으로부터 깊이 맵을 예측하기 위해서는 탈초점 흐림, 장면의 기하학적 구조, 객체의 주목도 및 움직임 등 다양한 종류의 깊이 정보가 활용된다. 그 중에서도 탈초점 흐림은 널리 이용되는 강력한 정보로서 탈초점 흐림으로부터 깊이를 예측하는 문제는 깊이를 예측하는 데 있어서 매우 중요한 역할을 한다. 본 연구는 2차원 영상만을 이용하여 깊이 맵을 예측하는 것을 목표로 하며 이 때, 촬영 장치로부터 영상 내 각 영역의 거리를 알아내기 위해 탈초점 거리 예측을 이용한다. 먼저 영상을 촬영할 때 영상 내 가장 가까운 곳에 초점이 맞춰져 있다고 가정하면 촬영 장치로부터 멀어짐에 따라 탈초점 흐림의 정도가 증가하게 된다. 탈초점 거리 기반 깊이 맵 예측 방법은 이를 이용하여 탈초점 흐림의 정도를 측정함으로써 거리를 예측하는 방식이다. 본 연구에서는 탈초점 거리로부터 깊이 맵을 구하는 새로운 방법을 제안한다. 먼저 인간의 깊이 지각 방식을 고려한 지각 깊이를 정의하고 이를 이용하여 탈초점 거리 예측의 (실제) 신뢰도를 정의하였다. 다음으로 그래디언트 및 2차 미분 값에 기반한 탈초점 거리 예측 결과에 대하여 신뢰도를 예측하는 방법을 설계하였다. 이렇게 예측한 신뢰도 값은 기존의 신뢰도 예측 방법으로 예측한 것에 비하여 더 정확하였다. 제안하는 깊이 맵 작성 방법은 조각 단위 평면 모델에 기반하였으며, 비용 함수는 데이터 항과 평활도 항으로 구성되었다. 깊이 맵의 전체 비용 함수를 최적화하는 과정에서는 반복적 지역 최적화 방식을 사용하였다. 제안하는 방법을 검증하기 위한 실험에는 인공 영상 및 실제 영상들을 사용하여 제안하는 방법과 기존의 탈초점 거리 기반 깊이 맵 예측 방법들을 비교하였다. 그 결과, 제안하는 방법은 기존의 방법들보다 더 나은 결과를 보여주었다.

주요어: 깊이 맵, 탈초점 흐림, 초점 값, 신뢰도
학번: 2008-30227