

# Depth Estimation by Finding Best Focused Points Using Line Fitting

Aamir Saeed Malik and Tae-Sun Choi

Gwangju Institute of Science and Technology,  
1 Oryong-Dong, Buk-Gu, Gwangju, 500712, Korea  
{aamir, tschoi}@gist.ac.kr

**Abstract.** This paper presents a method for depth estimation using image focus based on the linear regression model. For each pixel, we select two datasets based on the maximum value which is calculated using Laplacian operator. Then linear regression model is used to find lines that approximate these datasets. The best fit lines are found using least squares method. After approximating the two lines, their intersection point is calculated and weights are assigned to calculate the new value for the depth map. The proposed method is compared with four depth estimation algorithms.

**Keywords:** Depth Map, 3D shape recovery, Shape from focus, Line fitting.

## 1 Introduction

There are many methods for the calculation of depth leading to 3D shape recovery. In this paper, we limit our discussion to one of such methods, i.e., Shape From Focus (SFF). The objective of shape from focus is to find out the depth of every point of the object from the camera lens. Hence, finally we get a depth map which contains the depth of all points of the object from the camera lens where they are best focused or in other words, where they show maximum sharpness.

The basic problem of imaging systems, such as the eye or a video-camera, is that depth information is lost while projecting a 3D scene onto 2D image plane. Therefore, one fundamental problem in computer vision is the reconstruction of a geometric object from one or several observations. Shape information that is obtained from the reconstruction of a geometric object is of critical importance in many higher level vision applications like mobile robot systems. For example, an unmanned spacecraft, in order to land safely on lunar surface, needs to estimate depth details of the terrain. Various image processing techniques retrieve the lost cue and shape information from the pictorial information. Shape from focus (SFF) is one of such image processing techniques that are used to recover such information.

Various techniques and algorithms have been proposed in the literature for the implementation of SFF. They include methods using focus image surface, Lagrange polynomial, neural networks, dynamic programming etc. But almost all the techniques start with the estimation of the depth map. Hence, the techniques for the estimation of this initial depth map become quite significant.

Generally SFF scheme relies on a Focus Measure operator and an approximation technique. Focus Measure operator plays a very important role for three dimensional shape recovery because it is the first step in calculation of the depth map. So a focus measure operator needs to show robustness even in the presence of noise. Hence it should provide a very good estimate of the depth map.

## 2 Related Work

### 2.1 Focus Measure Methods

A Focus Measure operator is one that calculates the best focused point in the image. And focus measure is defined as a quantity to evaluate the sharpness of a pixel locally. Franz Stephan Helml and Stefan Scherer [1] summarized the traditional focus measures while introducing three new focus measure operators.

Laplacian, the most commonly used operator, is suitable for accurate shape recovery because of being a symmetric operator, and is obtained by adding second derivatives in the x and y directions. Modified Laplacian (ML) [2,3] is computed by adding squared 2nd derivatives. In order to handle possible variations, Shree K. Nayar and Yasuo Nakagawa suggested a variable spacing (step) between the pixels used to compute derivatives. In order to improve robustness for weak-texture images, Shree K. Nayar and Yasuo Nakagawa [2] presented focus measure at  $(x,y)$  as sum of ML values in a local window (about  $5 \times 5$ ).

Tenenbaum Focus Measure is gradient magnitude maximization method that measures the sum of squared responses of horizontal and vertical Sobel masks. Variance Focus Measure is based on the variance of gray-level which is higher than that in a blur image. Mean Method Focus Measure [1,3] depends on the ratio of mean grey value to the center grey value in the neighborhood. The ratio of one shows a constant grey-level or absence of texture. Curvature Focus Measure [1] exploits that the curvature in a sharp image is expected to be higher than that in a blur image. Point Focus Measure [1] is approximated by a polynomial of degree four.

### 2.2 Approximation Methods

A more accurate depth range image can be obtained by using some optimization and approximation method. The results of the focus measures, defined in section 2.1, are refined using such a reconstruction scheme. First we discuss the traditional SFF method. In Traditional (TR) SFF, for each image in the sequence, focus measure at each pixel is computed by Sum Modified Laplacian in a 2D neighborhood around the pixel. The results of TR SFF are improved by Subbarao and Choi [4] who proposed a new concept termed Focused Image Surface (FIS) based on planar surface approximations. The FIS of an object is defined as the surface formed by the set of points at which the object points are focused by a camera lens. Aamir Malik and Tae-Sun Choi [5] summarized various approximation techniques.

FIS can be improved by a piecewise curved surface rather than piecewise planar approximation. This was proposed by Choi and J. Yun [6]. They estimated the piecewise curved surface by interpolation using second order Lagrange polynomial. Asif and Choi [7] used Neural Networks to learn shape of FIS by optimizing the focus

measure over small 3-D windows. Bilal and Choi [8] proposed the use of Dynamic Programming (DP) to handle the complexity of FIS. DP is motivated by the Bellman's principal of optimality. A direct application of DP on a 3D data is impractical due computational complexity. Therefore, a heuristic model based on DP was proposed by Bilal and Choi.

### 3 Method

An initial estimate of depth map is obtained using the SML focus measure. For every pixel, its maximum value is obtained by considering the entire sequence of image frames. A set of successive points are selected leading upto the maximum value (with maximum value being the last point in the set). Then we use linear regression to find a line that approximates this data set. Again another set of successive points are selected following the maximum value (with maximum value being the first point in the set). Procedure is repeated to find a line that approximates this data set. After approximating the two lines, their intersection point is calculated. The new value for the depth map is calculated by assigning weights based on the distance of intersection point from the maximum value and the distance of intersection point from the closest value in the two data sets. An averaging filter is used to minimize noise effects. We also used optimum window size 5x5 [9].

Let  $I(x,y,k)$  be the acquired image sequence where  $k$  represents the total number of frames.  $I_m(x,y,k)$  is acquired by processing each of these frames using sum of modified Laplacian method [2,3]. Let  $P$  be any pixel  $(x,y)$  then  $P_{\max}$  is calculated as:

$$P_{\max}(x, y) = \max[I_m(x, y, k)], \text{ where } k = 1, \dots, n \quad (1)$$

The value of  $k$  that gives  $P_{\max}$  is referred to as  $k_{\max}$ . Now, two datasets ( $DS_1$  and  $DS_2$ ) are generated with one leading upto  $P_{\max}$  while the other following  $P_{\max}$ .

$$DS_1(t) = P(x, y, k_{\max} - t); \text{ where } t = 0, 1, \dots, j \ \& \ j < k_{\max} \quad (2)$$

$$DS_2(t) = P(x, y, k_{\max} + t); \text{ where } t = 0, 1, \dots, j \ \& \ j < k_{\max} \quad (3)$$

$$\text{when } t = 0, DS_1 = DS_2 = P_{\max}$$

In this paper, we used  $t=0,1,\dots,4$ . We found that  $j=4$  results in optimum performance for our algorithm. After generating the two datasets, we used the least square linear regression for the data sets to approximate two lines. The method of finding the line that best fits the collected data is known as linear regression while the least-squares regression line minimizes the sum of the squares of the residuals. The best fit line associated with the  $n$  points  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  has the form:

$$y = mx + b \quad (4)$$

$$\text{slope: } m = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$\text{Intercept : } b = \frac{\sum y - m \sum x}{n}$$

The best fit line is calculated using the least squares method. After calculating the best fit lines for the two datasets ( $DS_1$  and  $DS_2$ ), the intersection point of the two lines is found. Let  $P_i$  be the point of intersection of two lines and  $P_c$  be the point in the datasets  $DS_1$  and  $DS_2$  that is closest to  $P_i$ . Let  $x_1$  be the distance of  $P_i$  from  $P_c$  and  $x_2$  is the distance of  $P_i$  from  $P_{\max}$ , then:

$$x_1 = |P_i - P_c| \quad (5)$$

$$x_2 = |P_i - P_{\max}| \quad (6)$$

Total distance,  $x$ , is given as:

$$x = x_1 + x_2 \quad (7)$$

Finally, the new focused pixel,  $P$ , is calculated as:

$$P = w_1 P_c + w_2 P_{\max} \quad (8)$$

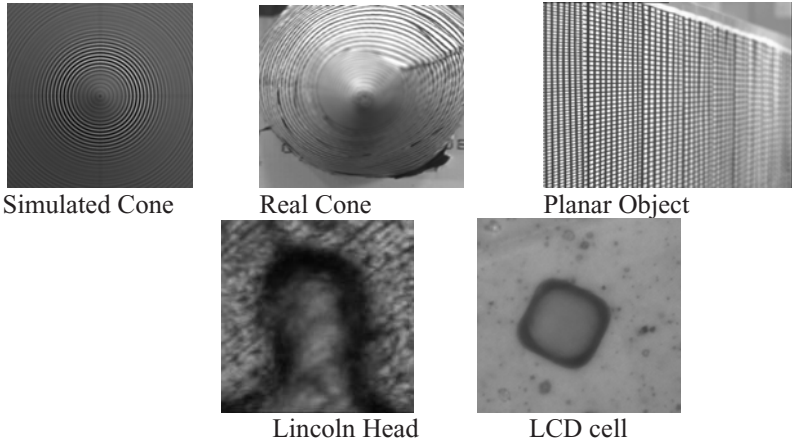
$$\text{where } w_1 = \frac{x_2}{x} \text{ and } w_2 = \frac{x_1}{x}$$

Hence more weight is given for the point closest to the intersection point as compared to  $P_{\max}$ . In case  $P_{\max}$  is closest point then the second closest point is taken as  $P_c$ .

## 4 Results and Discussion

The proposed algorithm is applied on a sequence of 97 simulated cone images, 97 real cone images, 87 real planar object images, 68 microscopic coin (Lincoln head on US penny) images and 60 microscopic TFT-LCD cell images. The resolution of the images is 360x360 pixels for cone images, 200x200 for planar object and 300x300 pixels for the microscopic images. Fig 1 shows one of the frames for all objects.

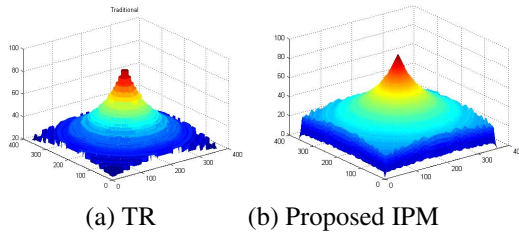
The real cone was taken from the CCD camera system. The real cone object was made of hardboard with black and white stripes drawn on the surface so that a dense texture of ring patterns is viewed in the images. The absolute displacement between two consecutive image frames was about 0.03mm. Similarly, a sequence of 97 simulated cone images were generated corresponding to 97 lens positions. The third object used was a slanted planar object. A total of 87 images were acquired corresponding to 87 lens positions for this object, which has only horizontal slope and equally spaced strips to make a poor texture.



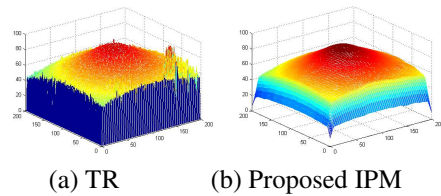
**Fig. 1.** Test Objects

Apart from the normal images acquired by CCD camera, we also acquired microscopic object images using microscopic image acquisition setup. Using the motor driver to control the step motor, object is displaced in z-direction. Using this experimental setup, we obtained images for 2 objects, namely, Lincoln head part in one cent coin and TFT-LCD cell.

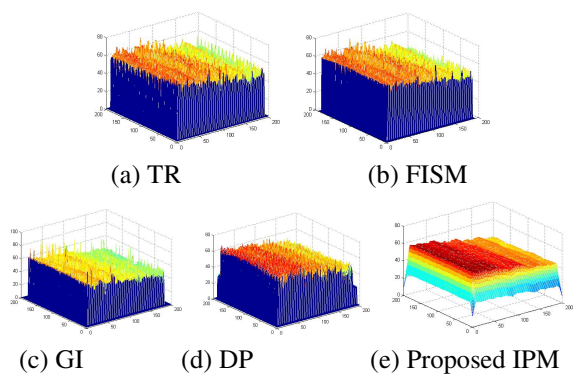
Figures 2 to 6 show the depth maps of simulated cone, real cone, slanted planar object, microscopic coin head and microscopic TFT-LCD cell respectively. The difference



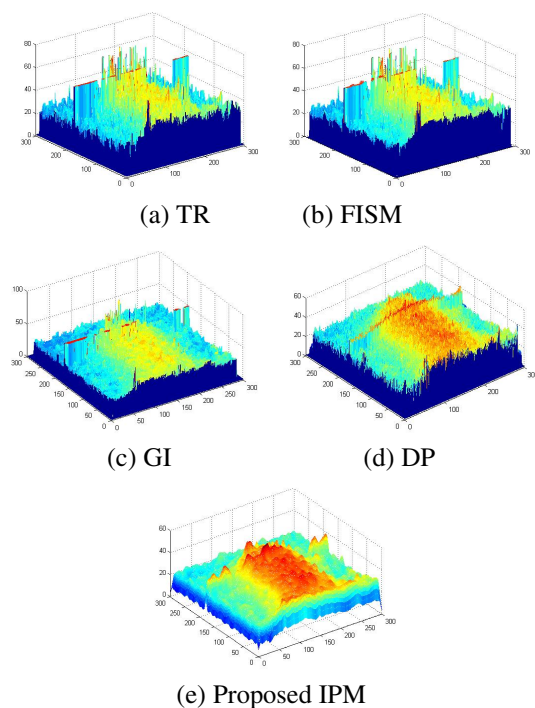
**Fig. 2.** Depth maps for the simulated cone



**Fig. 3.** Depth maps for the real cone

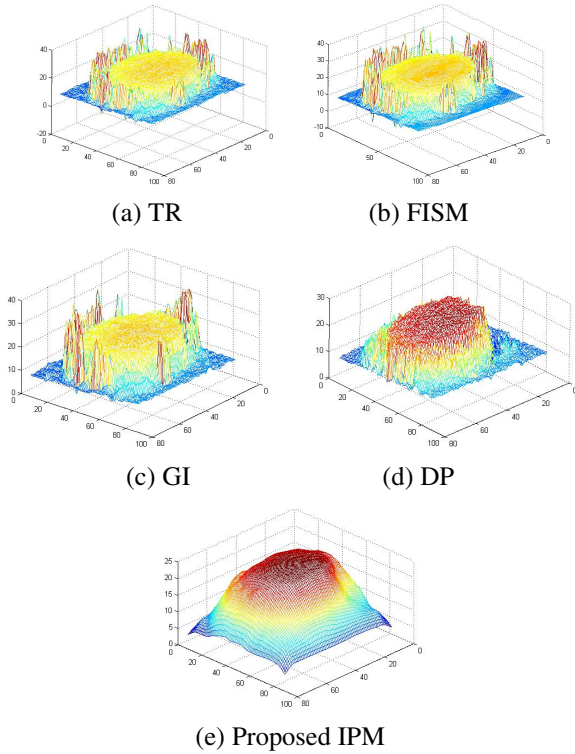


**Fig. 4.** Slanted planar object



**Fig. 5.** Microscopic coin object

in the figures of the proposed method with other methods is quite clear. We can see that the surface of depth map from the proposed Intersection Point Method (IPM) is much smoother as compared to that of the other methods, namely, the Traditional



**Fig. 6.** Microscopic TFT-LCD Cell

(TR) method, Focused Image Surface Method (FISM), Gaussian Interpolation (GI) method and Dynamic Programming (DP) method. For details of these methods, please see section 2.

## 5 Conclusions

In this paper, we presented a method for depth estimation using image focus based on the linear regression model. The intersection point is found for two lines that are approximated using least squares linear regression and the new value for the depth map is calculated by assigning weights to points based on their distance from the intersection point. As for computational efficiency, this method is computationally more efficient than FISM however less efficient than GI and DP.

## Acknowledgement

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korean government (MOST) (No. R01-2007-000-20227-0).

## References

1. Helmlí, F.S., Scherer, S.: Adaptive Shape from Focus with an Error Estimation in Light Microscopy. In: 2nd Int'l Symposium on Image and Signal Processing and Analysis, Croatia, pp. 188–193 (2001)
2. Nayar, S.K., Nakagawa, Y.: Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16(8), 824–831 (1994)
3. Malik, A.S., Choi, T.-S.: A Novel Algorithm for Estimation of Depth Map using Image Focus for 3D Shape Recovery in the Presence of Noise. *Pattern Recognition* 41(7), 2200–2225 (2008)
4. Subbarao, M., Choi, T.-S.: Accurate recovery of three dimensional shape from image focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(3), 266–274 (1995)
5. Malik, A.S., Choi, T.-S.: Application of Passive Techniques for Three Dimensional Cameras. *IEEE Transactions on Consumer Electronics* 53(2), 258–264 (2007)
6. Choi, T.S., Asif, M., Yun, J.: Three-dimensional shape recovery from focused image surface. In: *IEEE International Conference of Acoustics, Speech and Signal Processing*, vol. 6, pp. 3269–3272 (1999)
7. Asif, M., Choi, T.-S.: Shape from focus using multilayer feedforward neural network. *IEEE Transactions on Image Processing* 10(11), 1670–1675 (2001)
8. Ahmad, M.B., Choi, T.-S.: A Heuristic approach for finding best focused shape. *IEEE Transactions on Circuits and Systems for Video Technology* 15(4), 566–574 (2005)
9. Malik, A.S., Choi, T.-S.: Consideration of Illumination Effects and Optimization of Window Size for Accurate Calculation of Depth Map for 3D Shape Recovery. *Pattern Recognition* 40(1), 154–170 (2007)