Proceedings of the Fifth IIEEJ International Workshop
on Image Electronics and Visual Computing 2017
Da Nang, Vietnam, February 28 - March 3, 2017

SIMPLE DEPTH ESTIMATION METHOD USING SINGLE IMAGE WITH TILTED LENS FOR AUTOMOTIVE APPLICATIONS

Hiroshi Ikeoka†
†Fukuyama University

Takayuki Hamamoto††
††Tokyo University of Science

ABSTRACT

We have been investigating a depth estimation method for use in automotive applications. Many conventional methods with not image sensor detect only one depth value corresponding to one sensor. On the other hand, the methods with image sensor can obtain many depth values which are useful to traffic accident prevention and automatic operation of the vehicle. However, stereo methods are too sensitive to slight variations of baseline length due to vibration and temperature, and monocular methods by focusing cannot provide a balance between wide-area estimation and real-time estimation. Therefore, for monitoring the circumstances around the vehicle, we proposed to use a monocular method that adopts tilted lens optic. In this method, the plane of sharp focus (POF) lies, and the near limit of the depth of field (DOF) is inclined to horizontal. Hence, texture of the ground is focused and objects on the ground are blur. Herewith, by the difference of the sharpness values, it can be easy to detect the position of objects on the ground. In this paper, we report our simple depth estimation method for automotive application.

1. INTRODUCTION

In the recent automotive industry, it has come to be realized that the importance of active safety technology for preventing traffic accidents and automatic operation of driverless vehicle. In these technologies, depth estimation technique is central applicstions. In fact, collision avoidance devices that utilize various sensors have already started being used.

Various sensors have been used to monitor the positon of the objects around a vehicle. Ultrasonic sensor which is cheaper has been the most used [1]. However, it obtains only one depth value corresponding to one sensor. Hence, it is needed for human to judge circumstances surrounding finally. Also, millimeter wave sensor which has high accuracy more than supersonic sensor obtains only one depth value. Moreover, it can detect only rigid bodies such as cars, and cannot detect soft bodies like human [2][3]. In recent studies, automatic operation by the laser-ladder is often introduced. Actually, it has sufficiently high accuracy for depth estimation. However, it is expensive extremely. Therefore, it does not become popular sensor soon.

While, various methods with image sensor are also used to monitor circumstances around a vehicle. Imaging

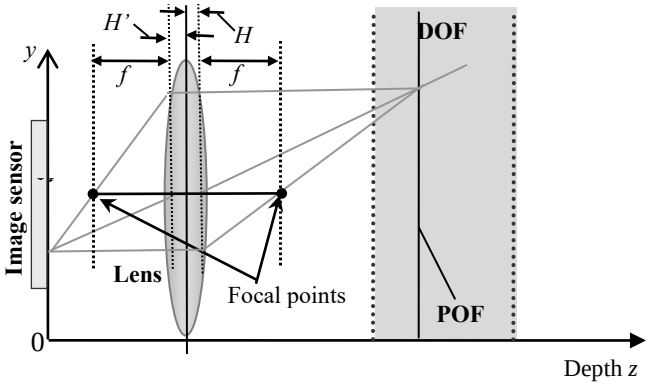


Fig.1 POF and DOF when using a conventional optics.

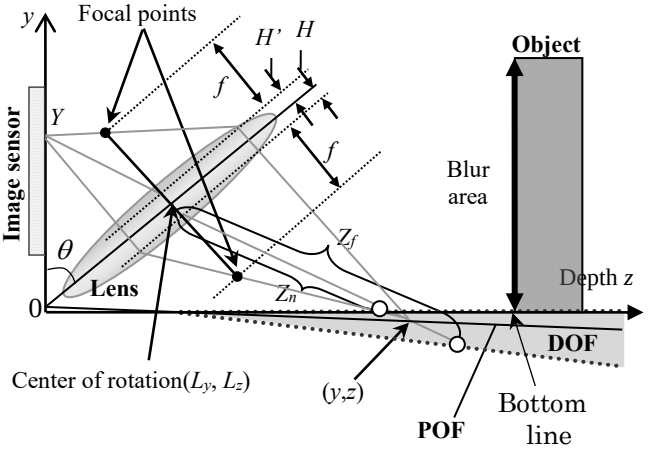


Fig.2 POF and DOF when using a tilted optics.

methods are exemplified by the stereo vision method [4][5][6]. When using this method, it is important to maintain the relative relationship between optic axes. Especially, the relationship in automotive applications is easy to be broken by disturbances such as vibration from tires, thermal expansion. Hence, the stereo method requires a tough base mount and its drawback is that its device size is larger in spite of small size camera.

Conversely, monocular vision methods, such as the Depth-from-focus and the Depth-from-defocus methods, require multi focus images that are obtained in extremely short time by a mechanical device [7][8][9][10][11][12][13]. Therefore, these methods have a disadvantage of not obtaining satisfactory amount of

incident light. In addition, a change in the depth of field that depends on the estimated depth reduces estimation accuracy.

As stated above, it is not easy to monitor circumstances around the vehicle by conventional sensing methods. Hence, existing technology does not satisfactorily answer the requirements of automotive applications. In this paper, we propose a depth estimation method which utilizes a monocular camera with tilted lens. By using this method, it is possible to achieve depth estimation that satisfies needs such as use in automotive environments.

2. IMAGING BY TITLTED LENS

When normal optical units are used as shown in Fig.1, the plane of the image sensor and the principal plane of the lens are parallel, and the near DOF limit and far DOF limits are also parallel to these planes. However, a situation in which a tilted lens is used is different: both of the DOF limits are shown by the dashed lines in Fig.2.

Moreover, we set the near DOF limit is inclined to horizontal on the ground in our proposed method. Hence, it can obtain an image in which the ground area is sharp and object areas are blur, as shown in Fig.3 (hereafter referred to as the ground-in-focus image).

2.1 POF with a tilted optics

Let us now consider the situation schematically depicted in Fig.2: an image sensor that is placed on the left side of a horizontal axis and an optical unit (lens) that has a tilted angle, θ, with respect to the image sensor. Similarly, a POF is tilted, which is different from normal optics [14][15]. The point (L_y, L_z) is the center of rotation for the lens, H is the distance between the first principal point and the center of rotation, and H' is the distance between the second principal point and the center of rotation. There are two light lines that reach the POF via each focal point, and one light line that reaches the POF via the two principal points. Hence, we can obtain a point of intersection of these light lines on the POF; the z and y coordinates are expressed with Y on the image sensor. Furthermore, by eliminating the Y from each coordinate, we can obtain the following POF expression:

$$y = \frac{1}{\sin\theta}\left(\cos\theta - \frac{f}{L_z}\right)z - \frac{L_z\cos\theta - L_y\sin\theta}{\sin\theta}. \quad (1)$$

At present, in order to simplify, we assumed that:

$$H = H' = 0\,. \quad (2)$$

Moreover, we have already considered the case in which (2) is not satisfied [14][15]. However, this case is not described in the present article. When (2) is satisfied, an image sensor plane, a lens principal plane, and a POF all intersect at the following point on the y-axis:

$$\left(-\frac{L_z\cos\theta - L_y\sin\theta}{\sin\theta}, 0\right). \quad (3)$$

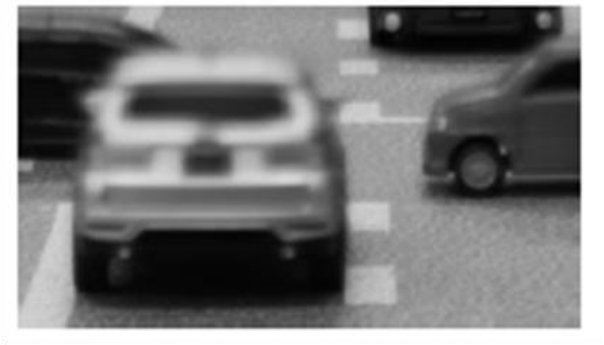


Fig.3 Ground-in-focus image.

This condition shows that the POF can be tilted with respect to the vertical axis by the angle θ of the lens; our method uses nearly horizontal POFs.

It is possible to set a tilted POF by using not only a tilted lens, but also a tilted image sensor. However, in the latter case, the image will be distorted. Hence, if the final output of an application is a depth map image, the former setting is better.

2.2 DOF with a tilted optics

When normal optical units are used, the plane of the image sensor and the principal plane of the lens are parallel, and the near DOF limit and far DOF limits are also parallel to these planes. However, a situation in which a tilted lens is used is different: both of the DOF limits shown by the dashed lines in Fig.2 are as well. If φ is the angle of the POF with respect to the ground, the gradient of the POF is:

$$\tan\varphi = \frac{1}{\sin\theta}\left(\cos\theta - \frac{f}{L_z}\right). \quad (4)$$

Strictly speaking, because the shape of the circle of confusion is oval and its size changes as a function of the distance between the lens and the POF, the depth of focus becomes variable. However, because the variation is slight, we approximate the depth of focus with a constant value ϵ. Hence, the gradients of both the near and far DOF limits are derived as follows:

$$\begin{aligned}\tan\varphi_- &= \tan\varphi - \frac{\epsilon}{f}\left(\frac{1}{\tan\theta} + \tan\varphi\right),\\ \tan\varphi_+ &= \tan\varphi + \frac{\epsilon}{f}\left(\frac{1}{\tan\theta} + \tan\varphi\right).\end{aligned} \quad (5)$$

Given that the DOF limits have gradients as in (5), the DOF limits are:

$$\begin{aligned}y &= \tan\varphi_- \cdot z - \frac{f}{\tan\theta} - L_z\cdot\tan\varphi_- + L_y\,,\\ y &= \tan\varphi_+ \cdot z - \frac{f}{\tan\theta} - L_z\cdot\tan\varphi_+ + L_y\,.\end{aligned} \quad (6)$$

As stated above, the DOF limits are approximated by linear expression. Hence, it means that it can obtain a ground-in-focus image by appropriate lens setting.

3. DEPTH ESTIMATION USING TILTED LENS

First, we obtain a sharpness map from a ground-in-focus image by using the Local Contrast Prior (LCP) [16] as the sharpness value as follows:

$$LC(x,y) = \frac{\max|Laplacian(x,y)|}{\max l(x,y) - \min l(x,y)}. \quad (7)$$

Here, l is the pixel's intensity and (x, y) are the pixel coordinates. Moreover, *Laplacian*(x, y) is different from general Laplacian. Because of easy implementation on the hardware chip, it is defined using only positive numbers as follows:

$$\begin{aligned} &Laplacian(x,y) \\ &= |l(x-1,y) + l(x+1,y) - 2l(x,y)| \\ &\quad +|l(x,y-1) + l(x,y+1) - 2l(x,y)|. \end{aligned} \quad (8)$$

Then, all pixels are classified in ground areas and object areas by a threshold value.

Second, it finds the most bottom pixel y-coordinate in object areas at each column. These pixels are located at a border between objects and the ground. Additionally, the relationships between the most bottom pixel y-coordinate of objects and depth values are inverse proportion which is provided by the camera calibration. Hence, it can obtain depth values from the pixel y-coordinates at each column.

4. EXPERIMENT

In this chapter, we explain our experiment details for performance of our method.

4.1 Our experiment environment

To test our proposed method, we performed experiments by using the camera system as shown in Fig.4. This camera system has the function which can change the position relationship between an image sensor and a lens. Herewith, we obtained the grand-in-focus image through tilted lens optics.

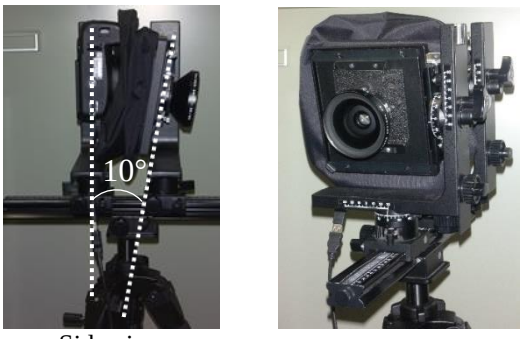
Our camera was set up using the following conditions. The position of camera was set at 440 mm in height, lens tilted angle was about 10 degrees, and interval between an image sensor to a principal point of lens is about 90 mm. In addition, we used a lens for which the focal length was $f = 90$ mm and aperture ratio (F-number) was 5.6.

Therefore, we set up that θ = 10° and $L_z = 90$ mm in (1) respectively. Moreover, the theoretical value the POF angle φ was -5°.

4.2 Experimental results

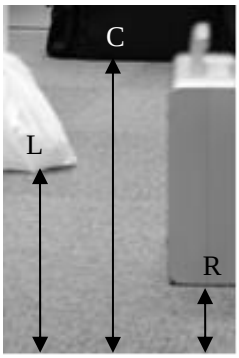
In what follows, we described the results of using our proposed method with the above experiment environment.

The input image which was taken by the 8-bit monochrome image is shown in Fig.5. We set the three



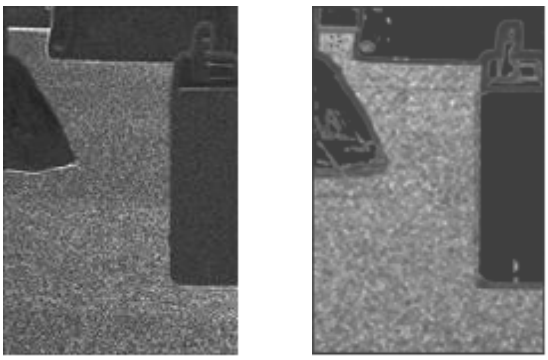
Side view Front view

Fig.4 Experimental equipment.



L: 201 mm
C: 332 mm
R: 143 mm

Fig.5 Input image.



Laplacian LCP

Fig.6 Edge detected images.

target objects on the floor which placed at 143 mm, 201 mm and 332 mm. Furthermore, images obtained by using the Laplacian and the LCP are shown in Fig.6. Our LCP range is 5 × 3 pixels. In addition to, we adopted 32 as the threshold of binarization for deciding bottom lines which are borders between the ground and each object area. We also utilized the closing and the opening process 16 times respectively for noise reduction. The sizes of these images are 864 × 1296 pixels.

As a result, we can got depth values for each object as shown in Table.1. It is get from the inversely proportional relationship between depth and y-pixel coordinate which is obtained previously like the follows:

$$depth = \frac{1}{1.26 \times 10^{-6} \times y + 1.8118 \times 10^{-3}}. \quad (9)$$

In addition to, the result of all depth values are rounded off to the first decimal place.

By comparing the ground truth of depth obtained with laser measuring, the estimation errors for each depth are within 2 mm. Hence we confirmed that our proposed method can yield about correct depth. Therefore, the performance of our method is superior to the performance of ultrasonic sensor which is the major conventional surround monitoring sensors: its depth resolution is about 10 mm. Moreover, it can obtain depth values of each object simultaneously from one sensing using only single image. Hence, it is suitable for automotive applications such as monitoring circumstances around the vehicle in real-time.

Our proposal method obtained excellent results at the experiment. However, there are several conditions: the ground is sufficient flat and has sufficient texture. Additionally, it assumed that the positon of the bottom line is the object position: the upper parts of object do not protrude than a bottom line. Nevertheless, the outdoor ground has sufficient texture generally, such as asphalt or gravel of load and park area. Also, the assumption that the bottom line is equals to the position of object is correct in most cases. In the same way, many other methods are effective only for specific height or specific surface of object. Hence, in practical usage, it would be best to use our method together other methods.

5. CONCLUSION

In this paper, we proposed a depth estimation method using single image with tilted lens. Hence, our method is suitable for monitoring circumstances around the vehicle because it satisfies the requirements of these applications, such as real-time processing and robust use in automotive environments.

In the future, we plan to confirm the performance of our method for estimation accuracy of real automotive use. Furthermore, because our proposed method is very simple, we would like to implement our method on the hardware like the FPGA.

Table.1 Experiment results.

Object label	Ground truth depth [mm]	Proposed method depth [mm] (y-coordinate)	Depth error [mm]
L	201	203 (y = 638)	2
C	332	330 (y = 214)	2
R	143	143 (y = 1051)	0

6. ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number JP15K00365.

7. REFERENCES

[1] S.Ahn, J.Choi, N.L.Doh, W.K.Chung: “A practical approach for ekf-slam in an indoor environment: fusing ultrasonic sensors and stereo camera,” Auton.Robots, vol.24, no.3, pp.315-335 (2008).
[2] M.E.Russell, A.Crain, A.Curran, R.A.Campbell, C.A.Drubin, W.F.Miccioli: “Millimeter-Wave Radar Sensor for Automotive Intelligent Cruise Control,” IEEE Transactions on Microwave Theory and Techniques, VOL.45, NO.12, pp.2444-2453 (1997).
[3] S.Tokoro, K.Kuroda, A.Kawakubo, K.Fujita, H.Fujinami: “Electronically Scanned Millimeter-wave Radar for Pre-Crash Safety and Adaptive Cruise Control System,” IEEE Intelligent Vehicles Symposium, pp.304-309 (2003).
[4] U.Franke, A.Joos: “Real-time stereo Vision for Urban Traffic Scene Understanding,” IEEE Intelligent Vehicles Symposium 2000, pp.273-278 (2000).
[5] M.Bertozzi, A.Broggi, A.Fascioli, S.Nichele: “Stereo Vision-based Vehicle Detection,” IEEE Intelligent Vehicles Symposium, IV, pp.39-44 (2000).
[6] T.B.Moeslund, C.B.Madsen, M.M.Trivedi: “Improving stereo camera depth measurements and benefiting from intermediate results,” IEEE Intelligent Vehicles Symposium, pp.935-940 (2010).
[7] E. Krotkov: “Focusing,” International Journal of Computer Vision, 1, 3, pp.223-237 (1987).
[8] S.K. Nayar, Y.Nakagawa: “Shape from Focus,” Journal IEEE Transactions on Pattern Analysis and Machine Intelligence, 16, 8, pp.824-831 (1994).
[9] M.Watanabe, S.K. Nayar: “Rational Filters for Passive Depth from Defocus,” International Journal of Computer Vision, Volume 27, Issue 3, pp.203-225 (1998).
[10] V. Aslantas, D.T. Pham: “Depth from automatic defocusing,” OPTICS EXPRESS, Vol.15, No. 3, pp.1011-1023 (2007).
[11] T.Kaneko, T.Ohmi, N.Ohya, N.Kawahara, T.Hattori: “A New Compact and Quick-Response Dynamic Focusing Lens,” TRANSDUCERS’97, vol.1, pp.63-66 (1997).
[12] H.Ikeoka, H.Kashiyama, T.Hamamoto, K.Kodama: “Depth Estimation by Smart Imager Sensor Using Multiple Focus Images,” The Institute of Image Information and Television Engineers, 62, 3, pp.384-391 (2008).
[13] H.Oku, M.Ishikawa: “High-Speed Liquid Lens for Computer Vision,” IEEE International Conference on Robotics and Automation Anchorage Convention District, pp.2643-2648 (2010).
[14] H.Ikeoka, M.Ohata, T.Hamamoto: “Real-Time Depth Estimation with Wide Detectable Range Using Horizontal Planes of Sharp Focus Proceedings,” Advanced Concepts for Intelligent Vision Systems (ACIVS), pp.669-680 (2011).
[15] H.Ikeoka, T.Murata, M.Okuwaki, T.Hamamoto: “DEPTH ESTIMATION FOR AUTOMOTIVE WITH TILTED OPTICS IMAGING,” IEEE Int. conf. on Image Processing (ICIP), CD-ROM, 5pages (2014).
[16] Yu-Wing Tai, Michael S. Brown: “SINGLE IMAGE DEFOCUS MAP ESTIMATION USING LOCAL CONTRAST PRIOR,” IEEE Int. conf. on Image Processing (ICIP), pp.1797-1800 (2009).