2012 IEEE International Conference on Multimedia and Expo Workshops

DEPTH AND GEOMETRY FROM A SINGLE 2D IMAGE USING TRIANGULATION

Yasir Salih and Aamir S. Malik, *Senior Member IEEE*
Centre for Intelligent Signal and Imaging Research
Universiti Teknologi PETRONAS, Tronoh, Perak, Malaysia
yasirutp@gmail.com and aamir_saeed@petronas.com.my

ABSTRACT

We present a novel method for computing depth of field and geometry from a single 2D image. This technique, unlike the existing ones measures the absolute depth of field and distances in the scene from single image only using the concept of triangulation. This algorithm requires minimum inputs such as camera height, camera pitch angle and camera field of view for computing the depth of field and 3D coordinates of any given point in the image. In addition, this method can be used to compute the actual size of an object in the scene (width and height) as well as the distance between different objects in the image. The proposed methodology has the potential to be implemented in high impact applications such as distance measurement from mobile phones, robot navigation and aerial surveillances.

***Index Terms*—** depth estimation, distance measurement, aerial surveillance, trigonometry.

1. INTRODUCTION

Depth of field is lost when projecting a 3D scene on a 2D imaging plane. Depth estimation is the process of retrieving the depth information from images using image contents. Depth estimation is performed by utilizing depth cues extracted from image(s) such as stereo parallax, motion parallax as well as monocular cues [1]. Stereo parallax is the spatial disparity of image points seen from different parallel cameras. Points with small depth have larger disparity than image points with large depth of field [2]. Motion parallax is the temporal disparity of image points observed from consecutive frames. Similar to stereo, near image points have larger parallax than far image points [3].

Monocular cues are depth cues extracted from a single image; this includes, focus, texture, shading as well as geometrical relationships. Focus is a measure of how accurate an object is placed from the camera; objects at the focal plane appears clear and sharp in contrast to objects outside the focal plane which appears blurred and defocused. In shape from focus multiple image are taken at a varying focus in order to reconstruct the depth of the scene [4]. Depth is also reconstructed from texture cues in image such as texture gradient and texture energy. For example distance between texels decays according to the depth of field [5]. Shading information can be used to reconstruct the reflectance map of the surface which relates the surface normal to the incoming light direction; thus depth can be computed from normal vectors of the image using various minimization methods [6]. In the recent year, multiple monocular visual cues are combined together in order to achieve better depth maps.

In this paper we introduce depth and geometry computation technique that is suitable to real time applications and can be embedded to existing image acquisition technologies which are mostly a single 2D camera. Our proposed method uses trigonometry relationships to compute the 3D absolute coordinates of any image point and also measure absolute distances in the scene.

The remaining parts of this paper are organized as follows; Section 2 discussed similar prior arts on depth and geometry computation from single image. Section 3 discussed our proposed methodology and who it is implemented as a measurement tool. Section 4 described the experiments performed to test and verifies the proposed method using different images. Finally, Section 5 concludes the paper by listing the main finding achieved by the proposed method.

2. SIMILAR WORKS

Several works have been presented before for depth computation using geometrical information in the scene. Criminisi et al. [7] used vanishing lines computed from the scene for measuring distance in the image, computing width and length as well as determining location of the camera. Rother et al. [8] computed vanishing points from consecutive frames of casual people motion. The vanishing lines are used to compute the horizon line and as a result the size of the walking person can be computed. Reibeiro and Hancock [9] presented a method for pose estimation using two vanishing points computed from textured planes. Borinova et al. [10] used vanishing lines for reconstructing the surface of outdoor scenes. They assumed the image is formed of ground and vertical object. Then they locate the ground vertical boundaries for completely defining the 3D structure. Peng et al. [11] presented 3D metric from a single uncelebrated image using orthogonal vanishing points. They utilized length ratios of lines width that have orthogonal

978-0-7695-4729-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICMEW.2012.95

direction to measure distances. Pribyl and Zemcik [12] used the size of known objects in the image (traffic signs) to calibrate the scene and measure distance and areas using the geometry of these known objects. Wang et al. [13] developed a measurement model from a single image using two orthogonal vanishing lines.

3. PROPOSED METHODOLOGY

We present a novel method for depth and geometry computation using the concept of triangulation. The proposed method works for cameras that are looking downward with a known pitch angle and height. This case is typical for surveillance cameras in order to cover a wider area and avoid occlusion from background objects. Figure 1 below shows a model for camera setup where the green trapezium region represents the area viewed by the camera. The camera is installed at a known height (h) from the ground and it has a pitch angle (θ) with the pole. The field of view of the camera is (FOV_H) in the horizontal direction and (FOV_V) in the vertical direction. The above parameters constrains the image geometry and they control how large or small the viewing area is. Depth and geometry computation in Figure 1 is performed using triangulation; thus the pitch angle is governed by Equation 1.

$$\theta < 90 - \frac{FOV_V}{2} \tag{1}$$

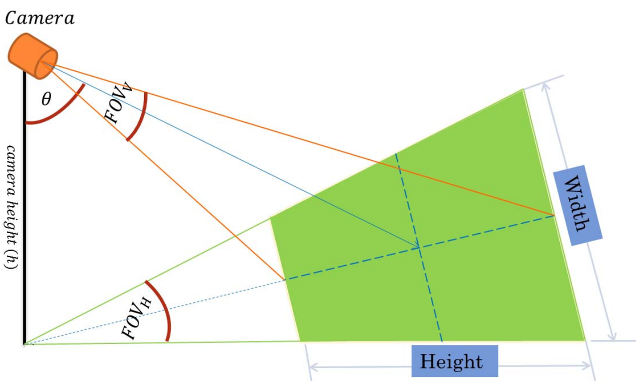

Fig. 1. Typical camera setup for surveillance cameras.

3.1 Depth from Triangulation

The geometrical structure of Figure 1 serves as a base for computing geometry from a single view. In Figure 2, let us consider an object located in the image view at point $p(i, j)$. The location of this object can be completely described by the camera height as well as the vertical angle (ψ) and rotation angle (ϕ). These two angles are computed using Equations (2) and (3) respectively where W and H are the image width and height respectively.

$$\psi = \theta + \left(\frac{H}{2} - j\right) \times \left(\frac{FOV_V}{H}\right) \tag{2}$$

$$\phi = \left(i - \frac{W}{2}\right) \times \left(\frac{FOV_H}{W}\right) \tag{3}$$

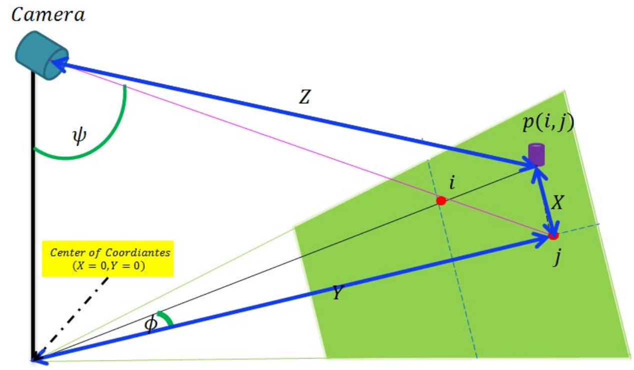

Fig. 2. Trigonometry model of an object located at point $\boldsymbol{p(i,j)}$ and its 3D coordinates.

$(\frac{FOV_V}{H})$ and $\left(\frac{FOV_H}{W}\right)$ are the angular step in vertical and horizontal direction respective. The angular step is due to one pixel displacement in the vertical or the horizontal directions. Given the angles (ψ) and (ϕ) and by assuming the center of coordinates is beneath the camera directly, the absolute ground location (X, Y) and the absolute depth of field (Z) for the image point $p(i, j)$ can be computed using Equations (4), (5) and (6) respectively.

$$Y = h \times \tan(\psi) \tag{4}$$

$$X = Y \times \tan(\phi) \tag{5}$$

$$Z = \sqrt{h^2 + Y^2 + X^2} \tag{6}$$

3.2 Height and Width Computation

In order to compute the height of an object in the scene the top most point and the bottom most point of the object is identified similar to what shown in Figure 3 and the height is computed using Equation (7).

$$X = h \times \left(\frac{Y_2 - Y_1}{Y_2}\right) \tag{7}$$

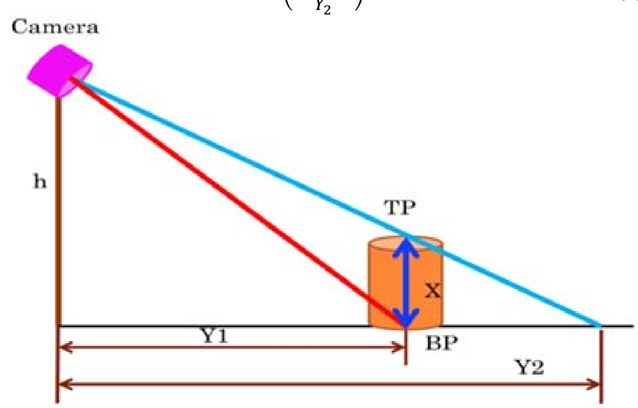

Fig. 3. Computing the height using triangulation.

Similarly, the width of any object can be computed by firstly selecting two points at its opposite ends and then computing the 3D coordinates for each of these points. After that, then width is determined by computing the distance between 3D coordinates of the selected points.

4. RESULTS AND DISCUSSIONS

The proposed geometry computation algorithm has been tested a verified using multiple images captured by different types of cameras. Before capturing any image, camera information such as height, pitch angle and field of view are recorded.

4.1 Depth Estimation

In this experiment, the proposed method has been used to measure depth of field at a varying distance from the camera. The aim is to study the effect of distance over the estimated depth of field. Figure 4 shows a 2D image with multiple 2D points selected for depth computation. In this experiment the image was captured using Canon EOS 1000D camera which has a high resolution of (3888x2597) which was installed at a height of 10.48m and with a vertical angle of 67.0^{0}. The field of view covered by this camera is 64.3^{0} horizontally and 45.3^{0} vertically.

Fig. 4. Image with multiple 2D points selected for depth measurement using the proposed method.

The selected points in Figure 4 are at a ground truth distances from 17m to 150m from the camera. Ground truth measurements were obtained using laser rangefinder. In Figure 5, the measurement error is plotted against the distance. In addition, the error has been compared against a one pixel error (we can call it quantization error). This is the error due to one pixel displacement at that depth of field for an image obtained with the above mentioned camera specifications. The graph shows that the error is increasing with the distance. This is very natural because with larger distance the size of pixel element increases. For example at 100m the error obtained is 2.1m while a one pixel error is only 0.25m; this means this is due to almost 10 pixels displacement. The maximum error obtained was 3.3m at a distance of 150m. In general, the depth of field measurement obtained using the proposed method has good accuracy.

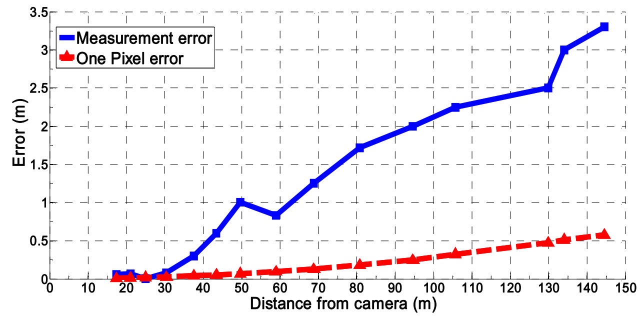

Fig. 5. Error of depth measurement at varying distance from camera compared to one pixel displacement error.

4.2 Height Measurement

The proposed methodology has been implemented for measuring the height of an object in the scene by selecting two points at its top and bottom ends. The images were captured using Dlink DCS 2120 camera which has a resolution of (320x240). This is a typical resolution for surveillance cameras in order to guarantee the frame rate. The camera was installed at a height of 2.83m and with vertical angle 60.0^{0}. The camera has a field of view of 49.6^{0} horizontally and 37.2^{0} vertically. Four images were collected to measure the height of a person at a varying distance from the camera. In Figure 6, the true height of the person is 1.76m. In the first image the measured height is 1.76m which is exactly the actual height. In the second and third images the same accurate measurement is obtained. In the fourth image the measured height is 1.77m which is only 1cm higher than the actual measurement. In this experiment, although the camera has low resolution, but accurate height measurement has been retained because the object is close to the camera.

Fig. 6. Measuring the height of a person from an image.

4.3 Width Measurement

This method has been tested for measuring the width of an object by selecting two points at the opposite ends and computing the 3D coordinates of each one of these points. After that, the width is computing the measuring the distance between these two points. In Figure 7 four images have been shown for a box and its width has been measured from multiple views. The images were captured from Logitech webcam which has a resolution of (1280x720). The camera was installed at a height of 1.88m and with vertical angle 60.0^0. This camera has a field of view of 60.0^0 horizontally and 45.0^0 vertically. The actual width of the box is 0.57m. In Figure 7 the measured width is very close to the actual one. The largest error was recorded at the first image where the measured width is 0.55m which is 2cm less than the actual one.

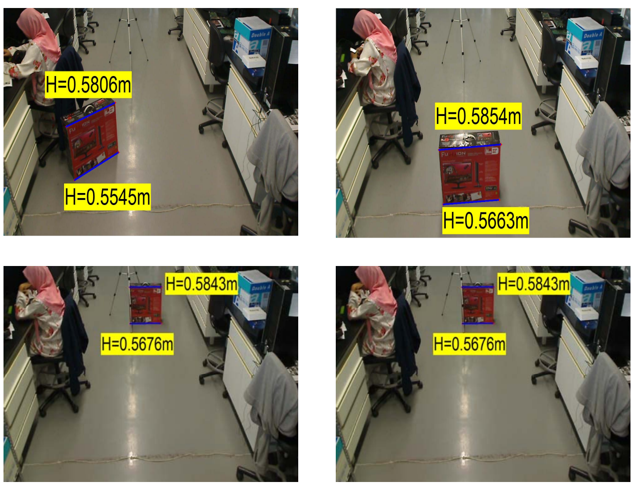

Fig. 7. Measuring the width of a box from single image and at different views.

5. CONCLUSION AND FUTURE WORKS

This paper presented a new method for depth and geometry computation from a single image using triangulation. The proposed methodology can be implemented on any type of image acquisition devices and it can be used for various applications. This method has a very high measurement accuracy and short computational time. The developed method has been tested with various images to measure depth of field, height of a person and width from multiple images and shows good measurement accuracy. Generally the accuracy of the measurement reduces with the distance from the camera because the area covered by a pixel increase.

This method can be improved by equipping it with detection algorithm in order to automate the depth computation process. In addition, this method can be implemented for measuring distance from aerial images and can also be implemented for robot navigation.

6. REFERENCES

[1] A. Saxena, S. H. Chung, and A. Ng, "3-D Depth Reconstruction from a Single Still Image," *International Journal on Computer Vision*, no. 76, pp. 53-69, 2008.

[2] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *in Proceedings of IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV)*, 2001, pp. 131-140.

[3] D.-J. Lee, P. Merrell, and Z. Wei, "Two-frame structure from motion using optical flow probability distributions for unmanned air vehicle obstacle avoidance," *Machine Vision and Applications*, vol. 21, no. 3, pp. 229-240, 2010.

[4] Y. N. Shree K. Nayar, "Shape from Focus," *IEEE Transaction on Pattren Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824-831, 1994.

[5] M. T. Suzuki, Y. Yaginuma, and H. Kodama, "A texture energy measurement technique for 3D volumetric data," *in Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, no. October, pp. 3779-3785, 2009.

[6] R. Zhang, P.-sing Tsai, J. E. Cryer, and M. Shah, "Shape from Shading : A Survey," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 690-706, 1999.

[7] A. Criminisi, I. Reid, and A. Zisserman, "Single View Metrology," International Journal of Computer Vision, vol. 40, no. 2, pp. 123-148, 2000.

[8] D. Rother, K. a Patwardhan, and G. Sapiro, "What Can Casual Walkers Tell Us About A 3D Scene?," *in Proceedings of 11^{th} IEEE International Conference on Computer Vision*, pp. 1-8, 2007.

[9] E. R. Hancock, "Improved pose estimation for texture planes using multiple vanishing points," *in Proceedings of International Conference on Image Processing*, pp. 148-152, 1999.

[10] O. Barinova, V. Konushin, A. Yakubenko, K. Lee, H. Lim, and A. Konushin, "Fast automatic single-view 3-d reconstruction of urban scenes," *in Proceedings of European Conference on Computer Vision*, pp. 1-14, 2008.

[11] K. Peng, L. Hou, R. Ren, X. Ying, and H. Zha, "Single View Metrology Along Orthogonal Directions," *in Proceedings of 20^{th} International Conference on Pattern Recognition*, 2010, pp. 1658-1661.

[12] B. Pribyl and P. Zemcik, "Simple Single View Scene Calibration," *Lecture Notes in Computer Science (Advances Concepts for Intelligent Vision Systems)*, vol. 6915, pp. 748-759, 2011.

[13] G. Wang, Y. Wu, and Z. Hu, "A novel approach for single view based plane metrology," *in Proceedings of International Conference on Pattern Recognition*, pp. 556-559, 2002.

[14] Yasir Salih, Aamir Saeed Malik, "Comparison of Stochastic Filtering Methods for 3D Tracking", *Pattern Recognition*, Vol. 44, Issue 10, pp. 2711-2737, October 2011.
[15] Aamir Saeed Malik, Humaira Nisar, Tae-Sun Choi, "A Fuzzy-Neural Approach for Estimation of Depth Map using Focus", *Applied Soft Computing*, Vol. 11, No. 2, pp. 1837-1850, March 2011.
[16] Roushanak Rahmat, Aamir Saeed Malik, Nidal Kamel, Humaira Nisar, "An Overview of LULU Operators and Discrete Pulse Transform for Image Analysis", *Imaging Science Journal*, 10.1179/1743131X11Y.0000000029, Accepted June 9, 2011.
[17] Seong-O Shim, Aamir Saeed Malik, Tae-Sun Choi, "Noise Reduction using Mean Shift Algorithm for Estimating 3D Shape", *Imaging Science Journal*, DOI: 10.1179/136821910X12867873897553, Vol. 59, No. 5, pp. 267-273, October 2011.
[18] Aamir Saeed Malik, Tae-Sun Choi, "Finding best focused points using intersection of two lines", *Proceedings of IEEE International Conference on Image Processing (ICIP)*, San Diego, California, USA, pp. 1952-1955, 12-15 October, 2008.