



Research Article

Depth perception in single rgb camera system using lens aperture and object size: a geometrical approach for depth estimation

P. J. A. Alphonse¹ · K. V. Sriharsha¹

Received: 26 June 2020 / Accepted: 12 January 2021

Published online: 01 May 2021

© The Author(s) 2021 **OPEN**

Abstract

In recent years, with increase in concern about public safety and security, human movements or action sequences are highly valued when dealing with suspicious and criminal activities. In order to estimate the position and orientation related to human movements, depth information is needed. This is obtained by fusing data obtained from multiple cameras at different viewpoints. In practice, whenever occlusion occurs in a surveillance environment, there may be a pixel-to-pixel correspondence between two images captured from two cameras and, as a result, depth information may not be accurate. Moreover use of more than one camera exclusively adds burden to the surveillance infrastructure. In this study, we present a mathematical model for acquiring object depth information using single camera by capturing the in focused portion of an object from a single image. When camera is in-focus, with the reference to camera lens center, for a fixed focal length for each aperture setting, the object distance is varied. For each aperture reading, for the corresponding distance, the object distance (or depth) is estimated by relating the three parameters namely lens aperture radius, object distance and object size in image plane. The results show that the distance computed from the relationship approximates actual with a standard error estimate of 2.39 to 2.54, when tested on Nikon and Cannon versions with an accuracy of 98.1% at 95% confidence level.

Keywords Focal length · De-focus in-focus · Aperture number · f_{stop} · Film speed · Exposure time · Auto focusing

1 Introduction

1.1 Motivation

Depth recovery is pre-requisite for identifying human actions in surveillance and this information can be attained from sequence of images taken in multiple views using more than one camera. But the real challenge lies in dealing with conventional photography that captures only two dimensional projection of a 3d scene. As conventional cameras [1] are capable of capturing in-focused portions of object(s) which are away from the plane of focus, focusing is considered to be one of the contributing factors for retrieving coarse depth information. When image is in-focus, knowledge of camera parameters help us in estimating range of an object with a typical experimental set up as

shown in Fig. 4. When an object is near to the camera, the image tends to grow larger and tends to diminish gradually while moving away from the camera. This inverse relationship is studied in detail by relating three parameters lens aperture radius, object size in image plane and the object distance along camera axial line.

1.2 Research contributions

- A relations is established between the object distance (or depth), object size and lens aperture radius and observations are taken in auto focusing mode.
- A multiplying factor is introduced to compensate the effect of object lightning on camera exposure settings while estimating depth of an object.

✉ P. J. A. Alphonse, pjaalphonse@gmail.com | ¹Department of Computer Applications, NIT, Trichy, Tamil Nadu, India.



- A curve estimation regression analysis is made for validating the relation between camera to object distance (or depth) and object size.

Literature Survey is presented in Section 2, proposed work in section 3 with subsections Prerequisites, Experimental set up and Camera Calibration, Database creation for Depth Estimation, Results & interpretation in Section 4 with subsections dealing with the influence of object lighting and Camera exposure settings during Depth Estimation, Finding Real height from Depth estimated, Method validation, Model accuracy with reference to the existing works. The rest of paper is concluded in Section 5.

2 Literature survey

Depth Estimation or Extraction refers a set of algorithms or techniques aiming at reconstructing a spatial structure of a 3D scene. In order to obtain depth of an object in a 3D scene, several approaches have been in existence. Electronically, all these approaches can be categorized in two ways namely active and passive. Among the active methods so far proposed, light is the first kind of energy to measure distance. Incandescent light produced from high temperature of a coil is used for distance measuring. But the system is sensitive to color of illuminated object and hence it fails.

Time of Flight depth [2–4] using phase delay of the reflected IR light, estimates the depth directly without the help of conventional computer vision algorithms. But as the principle is based on phase shift calculation, only a range of distances with in one unambiguous range $[0, 2\pi]$ can be retrieved. And also calculations made from the phase shift θ gives possible measurement errors. Pulse modulation approach is an alternative ToF, where depth of object is associated not only with duration of light pulse but also with duration of camera shutter. This solves the range ambiguity but at the same time suffers from calibration issues.

Nonsystematic depth errors [5], caused by light scattering and multi path propagation effects in 3D scene reconstruction makes ToF cameras inefficient for practical use. Another critical problem seen in ToF depth image is motion blur [6] occurs when camera or target is in motion. The error in phase measurements induces overshoot or undershoots in depth transition regions with in the bounded integration time limiting the depth accuracy and frame rate.

Triangulation methods offer better solution in estimating depth by capturing a scene taken from multiple views. Here we are restricting our discussion to binocular

vision (left and right views). As discussed in [7] differences between two images, give depth information and these differences are known as disparities. In other words shift between stereocorresponding points is termed as disparity.

However estimation of disparity map is considered to be the fundamental problem in computer vision. As mentioned in [8] stereo correspondence points are determined either with maximum correlation or with minimum sum of squared differences (SAD) or absolute intensity Differences (AID) and hence disparity map is constructed. Furthermore efforts are continued in order to increase accuracy in depth estimation and in this process depth map merging approaches [9] multi view stereo came into existence. In this process, depth map is computed for each view point using binocular stereo, synthesized according to path based normalized cross correlation metric and are merged to produce 3D models. However, the disparity map construction needs knowledge of camera configuration with epi-polar geometry constraint. And also occlusions in the monitoring environment creates mismatch in pixel to pixel correspondence which in turn leads to ambiguity in disparity map computation.

Power consumption, camera mounting space [10] and memory stack space for computation process [11] are also the factors that alarmed the researchers to go with monocular vision, i.e. information from single image) in order to obtain depth information. A single view point of ToF depth sensor [12, 13] offers significant benefit in providing accurate depth measurements without being affected by ambient lightning, shadows and occlusions. Here the design itself provides illumination and also phase measurement is taken as criteria for depth measurement but not the intensity.

As on Today CW-ToF (Continuous Wave ToF) sensors are dominating the consumer electronics and low end robotics space with certain limitations. As stated in [14] the design mainly suffers from three fundamental issues. One is the range which is limited by power consumption and eye safety considerations. Second is accuracy which is adversely affected by illumination affects. And finally the interference comes in to picture when they start operating in bulk amount in indoor and outdoor environments. Further in order to evaluate depth range accuracy, Plenoptic cameras [15–17] design came in to existence. This camera uses not only the intensity of light in the scene but also the directional information of light distribution in scene for retrieving depth of the surface. But with poor reconstruction depth quality and large storage requirements, human action recognition in surveillance systems is impossible and hence the design is not recommended.

Considering the economic and technical issues in all the above mentioned approaches, we proposed a theory

relating three fundamental parameters Object size, lens aperture radius and object depth. This theory can be well established in any conventional surveillance cameras without incorporating any additional hardware or software and even it works in both indoor and outdoor environments under any circumstances.

3 Proposed work

3.1 Prerequisites

3.1.1 Focal length

Consider an object 'O' at infinity focus, photographed by rectilinear convex lens with a focal length 'f' and aperture radius 'a_R' to form an image at a distance equal to 'f' from the image plane.

As shown in Fig. 1 lens with aperture of radius 'a_R' project incoming rays on to image plane and hence image distance for object distance at infinity defines the focal point 'F'. In this case, an object is at infinite focus appears sharp with an image distance $l_{inf} = f$

As shown in Fig. 2, the point 'F' is focused at a distance 'l_{of}' behind the lens but in front of the image plane. As receiving plane is not at the Image focus, blurriness appears in the image.

'O_d' - Object distance from lens

'l_{inf}' - Distance between lens and image plane when image is in perfect focus

'l_{inf} - l_{of}' - Distance between lens and image plane when image is de-focused

As shown in Fig. 2 Light rays emanating from an object 'O' fall on lens and converges at distance 'l_{of}' on sensor side of the lens.

For a thin lens [18] with focal length 'f', the relation between 'O_d' and 'l_{of}' is stated as

$$\frac{1}{O_d} + \frac{1}{l_{of}} = \frac{1}{f} \tag{1}$$

$$O_d = \frac{f \times l_{of}}{l_{of} - f} \tag{2}$$

We now explore what happens if an object is at a distance greater than or less than O_d is imaged. As show in Fig. 2, object at infinity focus, would be focused at distance l_{of} behind the lens but in front of the sensor creating a sharp image.

From the Geometry of Fig. 3, we obtain

$$\frac{a_R}{l_{of}} = \frac{O_s}{l_{inf} - l_{of}} \tag{3}$$

$$l_{of} = a_R \times \frac{l_{inf} - l_{of}}{O_s} \tag{4}$$

From the Geometry of Fig. 3, we also obtain

Fig. 1 Image is in Focus

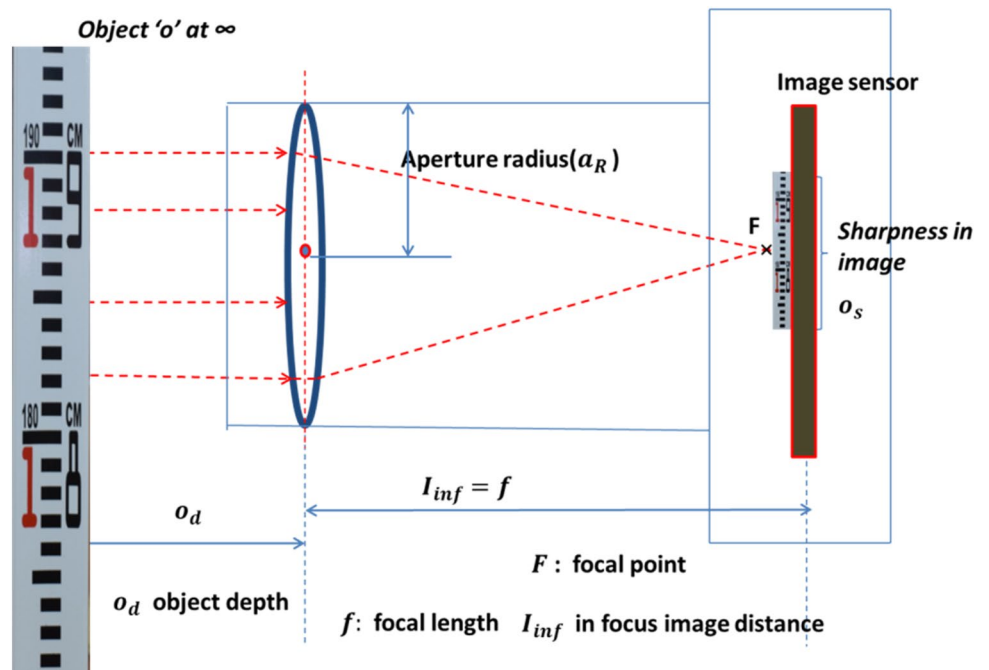


Fig. 2 A Fusion of in-Focus and Out Focus Scenarios

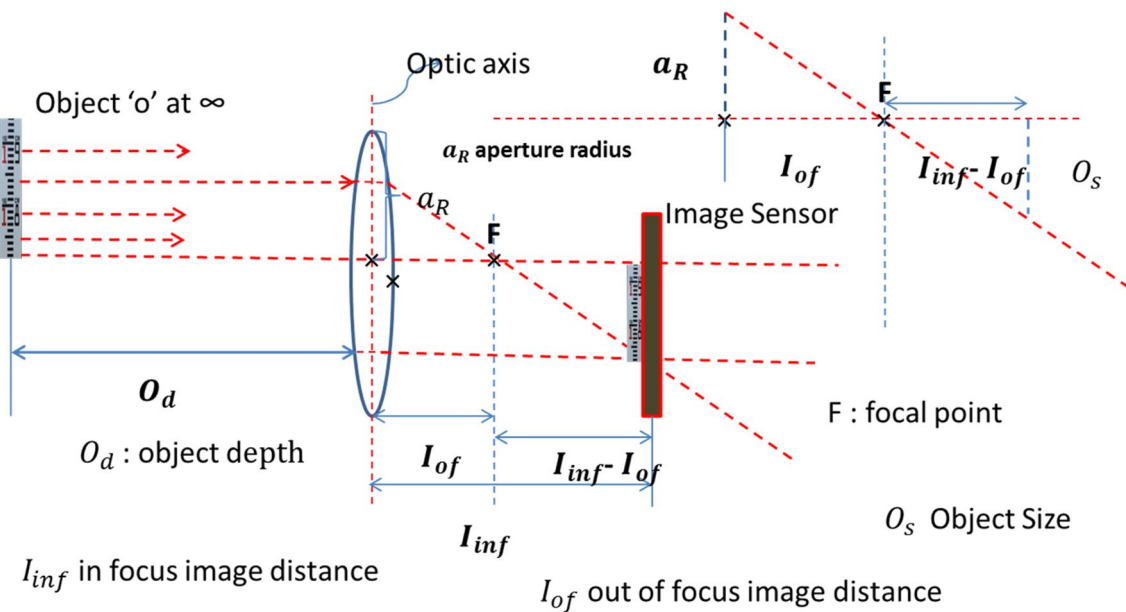
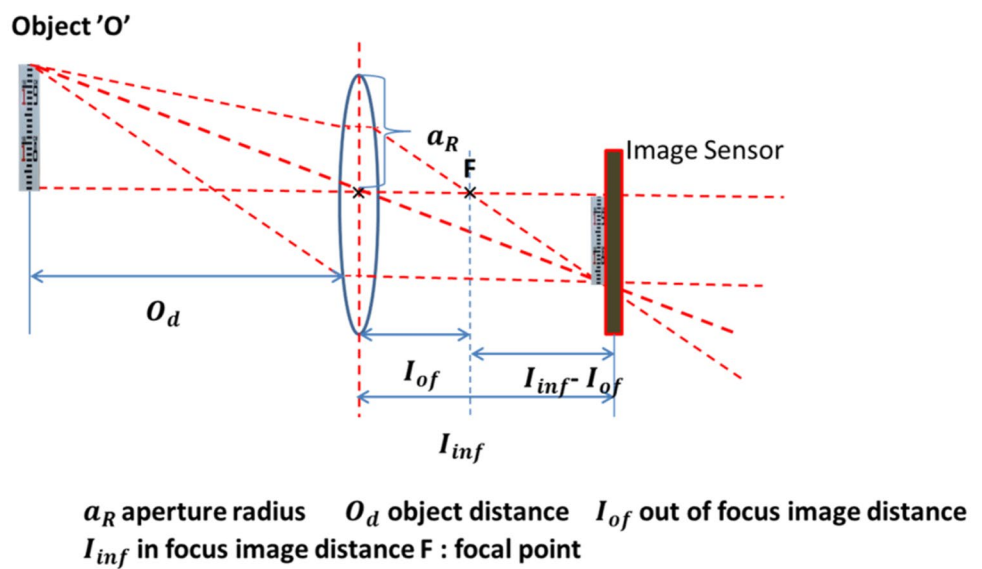


Fig. 3 Geometry of Imaging Relating Aperture Radius a_R and Object Size O_s

$$\frac{I_{inf} - I_{of}}{I_{of}} = \frac{O_s}{a_R} \tag{5} \quad I_{of} = I_{inf} \times \frac{a_R}{O_s + a_R} \tag{8}$$

$$\frac{I_{inf}}{I_{of}} = 1 + \frac{O_s}{a_R} \tag{6}$$

Substituting Eq. (8) in Eq. (2), we obtain

$$\frac{I_{inf}}{I_{of}} = \frac{O_s + a_R}{a_R} \tag{7} \quad O_d = \frac{\left(f \times \left(I_{inf} \times \left(\frac{a_R}{O_s + a_R} \right) \right) \right)}{\left(I_{inf} \times \left(\frac{a_R}{O_s + a_R} \right) - f \right)} \tag{9}$$

$$O_d = \frac{(f \times I_{inf} \times a_R)}{(I_{inf} \times a_R) - (f \times (O_s + a_R))} \quad (10)$$

where

$$a_R = (f/f_{stop})/2$$

Knowing the object distance from camera O_d , object real height can be determined from the relation

$$\frac{\text{Real Height}}{\text{Image height}} = \frac{O_d}{\text{Image distance}} \quad (11)$$

In many imaging applications, always object distance from lens is considerably greater than image distance. Hence the Eq. (1) can be approximated as

$$\frac{1}{f} \cong \frac{1}{\text{Image distance}} \quad (12)$$

By substituting Eq. (12) in Eq. (11), we obtain

$$\frac{\text{Real Height}}{\text{Image height}} = \frac{\text{Object distance } (O_d)}{\text{focal length } (f)} \quad (13)$$

$$\text{Real Height} = \frac{\text{Image height} \times \text{Object distance}}{f} \quad (14)$$

3.2 Experimental set-up and camera calibration

Our research aims at measuring the depth of a stationary object using a single RGB camera by establishing a relationship among focal length (f), lens aperture (a_R), object size (O_s) and camera to object distance (or depth) (O_d) parameters. In this perspective, a 50 mm Nikkor prime lens (known as conventional lenses with fixed angle of view) is used to focus objects at different working distances starting from 10 cm by successively altering the opening f_{stop} for repeated collection of readings. ISO and shutter speed parameters are tuned in required ratios for sensor to render the sharp image. The device and its exposure are mentioned in Tables 1 and 2.

Table 1 Experimental Device Details for Photogrammetry Experiment

Devices	Values
Camera model	Nikon D5300
Image sensor	23.5 × 15.6 mm CMOS sensor
Image size (pixels)	6000 × 4000
Pixel size	3.9 micron
Lenses	Nikon AF-S Nikkor 50 mm f/1.8G Prime Lens for Nikon DSLR Camera

Table 2 Parameter Influencing the Camera Exposure








Exposure settings	Values
ISO	1000
Aperture (f_{stop})	1.8 mm to 16 mm
Exposure time (shutter speed)	1/100 seconds
View Finder size	0.82 × (50 mm f/1.4 lens at infinity)

The experiment is carried out by setting the camera in auto focus mode. When testing the auto focus accuracy of a camera or lens, it is always essential to have exposure consistency. In our Nikon D5300 with Nikkor 50 mm prime lens, to maintain auto focus consistency, we ensure that camera is set at AF-S/Single Servo mode. Internally the camera relies on contract-detect to acquire focus. Using this focus method, camera forces the lens to shift back and forth anywhere in the frame by adjusting the focal point 'F' to get a sharp image. This focus confirmation happens electronically through the camera sensor and hence it will always be accurate. Ensure that camera is placed on a stable tripod and an object of interest; say aluminum leveling staff is placed in front of the camera at a distance of 10 cm with 1.8 mm f_{stop} , 1000 ISO and 1/100 s exposure setting as shown in Fig. 4.



Fig. 4 Experimental Set-Up

Table 3 Sample Photographs with f-Stop=1.8mm at Focus = ∞ f = l_{inf} 5cm, Aperture Radius (a_R) = $(f/f_{stop})/2$ = 7.81 cm, k = 0.45

S.No	b_s (pixels)	O_s (cm)	Est_{dist} (cm)	Act_{dist} (cm)	Figure
1	1519.25	0.59	53.27	50	
2	1285.32	0.50	62.97	60	
3	1081.04	0.42	74.87	70	
4	924.96	0.36	87.50	80	
5	835.04	0.32	96.92	90	
6	778	0.30	104.03	100	
7	718.03	0.28	112.72	110	

4.1 Influence of object lighting on camera exposure settings during depth estimation

It is evident from [19]; each pixel N_d in the image is proportional to the scene luminance I_s and also dependent on camera settings like Film speed (ISO), exposure time (t) and aperture number f_{stop} , where








$$N_d = K_c \times (\text{Exposure time} \times \text{ISO}) / f_{stop}^2 \times I_s$$

$$I_s = 0.2162 \times R + 0.7152 \times G + 0.0722 \times B \quad (15)$$

(R, G and B are red, green and blue values of i^{th} pixel.)

The range of K_c as recommended by ANSI/ISO 2720-1974 [20] varies from 10.6 to 13.4. In practice, value 12.5 is recommended for Canon and Nikon camera. It is understood from our observation that changes in aperture settings affects $\frac{I_s}{f_{stop}^2}$ ratio. This in turn affects the brightness and also the subject distance from the lens. In order to

Table 4 Sample photographs with f-stop=2.8 mm at Focus = ∞ f = l_{inf} = 5cm, Aperture Radius (a_R) = $(f/f_{stop})/2$ = 8.928cm, k = 0.66








S.No	b_s (pixels)	O_s (cm)	Est_{dist} (cm)	Act_{dist} (cm)	Figure
1	1522.25	0.59	50.08	50	
2	1242.62	0.48	61.35	60	
3	1062.77	0.41	71.73	70	
4	1141.14	0.36	81.97	80	
5	820.03	0.31	96.92	90	
6	759.99	0.29	100.31	100	
7	696.67	0.27	109.42	110	

compensate the subject lighting affects induced by exposure settings, a multiplying factor 'k' is included in Eq (10). The modified equation with respect to object lightning on film exposure settings is written as

$$O_d = k \times \frac{(f \times l_{inf} \times a_R)}{(l_{inf} \times a_R) - (f \times (O_s + a_R))} \tag{16}$$

It is observed that in most of the cases, the value 'k' varies in accordance with aperture reading except for $f_{stop} = 1.8$ to 2 mm, 2.2 mm to 2.5 mm, 4 to 4.5 mm and 8 to 9 mm. Table 8 discloses the following information.

Table 5 Sample Photographs with f-Stop=3.2 mm at Focus = ∞ $f = l_{inf} = 5$ cm, Aperture Radius (a_R) = $(f/f_{stop})/2 = 7.81$ cm, $k = 0.76$

S.No	b_s (pixels)	O_s (cm)	Est_{dist} (cm)	Act_{dist} (cm)	Figure
1	1518.9	0.59	50.70	50	
2	1248.81	0.48	61.67	60	
3	1062.77	0.41	71.73	70	
4	936.97	0.36	82.20	80	
5	840.93	0.32	91.59	90	
6	759.99	0.29	100.31	100	
7	754.03	0.29	102.14	110	

4.1.1 Finding optimum value of 'k'








But our objective is to find the optimum 'k' value for which the estimated camera to object depth approximate the actual value. It is inferred from our observations, by applying Generalized reduced gradient method that at $k=0.666$, we attain minimum MSE of 2.566 with in the defined constraints for k.

$$k \geq 0.45 \text{ and } k \leq 3.2$$

It is concluded from our observation that for achieving a closes approximation in camera to object depth, an $f_{stop} = 2.8$ mm and k value = 0.666 is chosen in surveillance camera design.

In a similar fashion , the proposed equation Eq. (16) also tested on cannon cameras for the f_{stop} ranging from 5 to 32mm and found that camera to object depth approximating to actual value for lens $f_{stop} = 7.1$ mm with optimum 'k' value = 1.44 .

Table 6 Sample Photographs with f-Stop=3.5 mm at Focus = ∞ $f = l_{inf} = 5\text{cm}$ Aperture Radius (a_R) = $(f/f_{stop})/2 = 7.81\text{cm}$, $k = 0.85$

S.No	$b_s(\text{pixels})$	$O_s(\text{cm})$	$Est_{dist}(\text{cm})$	$Act_{dist}(\text{cm})$	Figure
1	1507.25	.58	51.64	50	
2	1276.13	0.49	60.99	60	
3	1111	0.43	70.06	70	
4	937.08	0.36	83.06	80	
5	846.97	0.33	91.90	90	
6	793	0.30	98.15	100	
7	714.86	0.25	108.88	110	






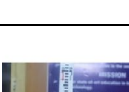

4.2 Finding real height of an object from depth (O_d) estimated

Real height can be obtained from an object centered at distinct operating distances provided that object distance (O_d) from the lens, image height and focal length are known. The object depth is calculated from Eq. (16) and the image height is computed by cropping the object

of interest for a specified size over each image. As height is obtained in pixels, it is multiplied with pixel size as mentioned in NIKON D5300, for getting measurement in SI units.

On substituting the image height, object depth and focal length in Eq. (16), object real height is obtained. For all the f_{stop} readings ranging from 1.8 to 16 mm, object is placed at different working distances starting from 50

Table 7 Sample Photographs with f-stop=4 mm at Focus = ∞ f = l_{inf} = 5cm, Aperture Radius (a_R) = $(f/f_{stop})/2$ = 6.25cm, k = 0.95

S.No	<i>bs</i> (pixels)	<i>Os</i> (cm)	<i>Estdist</i> (cm)	<i>Act_{dist}</i> (cm)	<i>Figure</i>
1	1507.27	0.58	50.50	50	
2	1261.16	0.49	60.35	60	
3	1069.03	0.41	71.20	70	
4	912.46	0.35	83.42	80	
5	852.72	0.33	89.26	90	
6	771.99	0.30	98.60	100	
7	666.36	0.25	114.23	110	

to 110 cm and, photographs are taken. And real height is extracted from each photographic object and the results are tabulated in the Tables 9 and 10.

4.3 Method validation

The independent variable is object size (O_s)

The Table 11 provides the R , R^2 , adjusted R^2 and std error of estimate determines how well a inverse model fits the data.

The R column represents multi correlation coefficient, which is considered to be measure quality of prediction of dependent variable. A value of a 0.990 indicates good level of prediction. The R^2 column represents the value of proportion of variance in dependent variable. From our value of 0.990, we can see that 99% of the variation of our dependent variable is explained by our autonomous variable 'objects size'.

The F-ratio in Table 12 tests whether the reverse model fits the information well. The Table 12 demonstrates that the independent variable predicts the dependent variable

Table 8 Variation in k Value Subject to Effect of Exposure on Subject Lightning

f_{stop} range(mm)	k	f_{stop} range(mm)	k
1.8–2	0.45	2.2–2.5	0.5
2.8	0.66	3.2	0.76
3.5	0.85	4–4.5	0.95
5.6	1.12	6.3	1.33
7.1	1.43	8–9	1.818
10	2	11	2.2
13	2.7	14	2.9
16	3.2		

statistically and substantially, $F(1, 74) = 2294.695, p < 0.0005$ (i.e. the inverse model is a good fit for the data).

As shown in ‘Sig’ column of Table 13 all independent variable coefficients are statistically and significantly different from 0 (zero), and $p < 0.05$ concludes that the coefficients are statistically and significantly different to 0 (zero). Hence from general form of an equation to predict estimated distance from object size is expressed as :

$$\text{Pred.Estdist}(O_d) = 1.221 + 24.910 \times 1/\text{Object size} \tag{17}$$

4.4 Comparison of our work with existing works

The Table 14 illustrates the fitting process of the model in one of the validation folds. It is found from our observations that our model is competing Yujiao Chen [21] depth estimation technique by showing high consistency between the predicted (estimated) and reference (actual) depth values with a Pearson’s correlation coefficient of $r = 0.99213$. Besides, variables estimated distance and object size are also proved to be highly consistent with the Pearson’s correlation coefficient of $r = 0.86$ (95% confidence level) as shown in Table 14. The negative

value indicates that for every positive increase in distance there is negative decrease of fixed proportion in object size. Hence the absolute value is taken. It is also observed that our depth estimation technique is outperforming the Mylene C.Q.Farias [22] stereo vision system based distance measurement in single camera system by giving a good quality of curve fit with RMSE of 0.041356 on Nikon model data set acquisition and 0.056232 on canon model. The details are furnished in the Table 15.

5 Conclusions

This work provided a theory related to object size, lens aperture radius and object depth. For this purpose, near range photography is used for this purpose to explore the impact of object size and lens aperture radius on depth (or distance) in auto focusing modes. It is also found that change in aperture settings induces a variation in luminance of an object which is to be calculated. This variation is also taken into account when estimating an object’s depth from the center of the lens. For experimentation, Nikon D5300 with 50 mm prime lens is used. By placing the lens in infinite focus, the distance measurements are calculated for different apertures.

This research created an inverse model to object distances with a permissible standard error estimate of 3.012, $R = 0.990$ and $R^2 = 0.981$ regulated by object size values with camera. It is noted that when the lens aperture changes, object lightning also affects the camera to object distance (or depth). This change in exposure settings affects $\frac{f_s}{f_{stop}^2}$ ratio. We considered this ratio as ‘k’. We are able to achieve an optimum ‘k’ value = 0.45 for Nikon model with $f_{stop} = 2.8$ mm and $k = 1.44$ for canon model with $f_{stop} = 7.1$ mm. When compared with the Mylene C.Q.Farias [22], Yujiao Chen [21] stereo vision set-up and Said Pertuz et al. [17] Palmieri et al. [23] light field imaging experimental set up, the proposed techniques used single

Table 9 Sample Photographs with Object Distance and Real Height Measurements $f = 5\text{cm}$












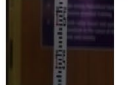
S. No	Figure	$f_{stop}(mm)$	$Img_{ht}(pixels)$	$Img_{ht}(cms)$	$O_d(cm)$ estimate d	$Real_{ht}(Act)$	$Real_{ht}$ (Estimate d)
1		1.8	3981.86	1.55	74.86	22	23.25
2		1.8	3989.27	1.55	104.02	29	32.3
3		2	3969.13	1.54	63.69	17	17.92
4		2	3977.52	1.55	109.35	31.5	30.84
5		2.2	3976.06	1.55	83.05	23	25.74
6		2.2	3957.98	1.54	112.48	32	34.72
7		2.5	4000	1.56	47.02	13.5	14.67
8		2.5	3988.34	1.55	50.08	26	28.49
9		2.8	3976	1.55	109.42	13	15.54
10		2.8	3987.86	1.55 69	.69.30	33	34.07
11		3.2	4000.02	1.56	50.70	13	15.82
12		3.2	3991.37	1.55	91.62	26	28.52

Table 10 Sample Photographs with Object Distance and Real Height Measurements (contd) $f = 5\text{cm}$











S. No	Figure	$f_{stop}(mm)$	$Img_{ht}(pixels)$	$Img_{ht}(cms)$	$O_d(cm)$ estimate	$Real_{ht}(Act)$	$Real_{ht}(Estimate)$
13		3.5	3992.87	1.55	70.06	20	21.82
14		3.5	3995.02	1.55	108.88	32	33.93
15		4	3976.3	1.55	83.46	23	25.96
16		4	3965.08	1.54	102.85	29	32
17		4.5	3961.98	1.54	49.21	13.5	15
18		4.5	3983.02	1.55	72.71	20	22.59
19		5	3976.14	1.55	80.85	23	25.24
20		5	3990.01	1.55	112.34	32	34.96
21		5.6	3989.01	1.55	57.90	16	17.85
22		5.6	4009.99	1.56	79.46	23	24.63

Table 11 Determining How Well Model Fits: Model Summary

R	RSquare	Adjusted R Square	Std. Error
0.990	0.981	0.981	3.012

Table 12 ANOVA Table-Statistical Significance

	Sum of squares	df	Mean square	F	Sig.
Regression	34796.441	1	34796.441	3835.039	0.000
Residual	671.424	74	9.073		
Total	35467.865	75			

Table 13 Estimated Model Coefficients -Statistical Significance of Independent Variables

	B	Std E	Beta	t	Sig.
1/ <i>Objectsize</i>	24.910	0.990	-2.965	61.928	0.000
(constant)	1.221	1.333		0.916	0.363

Table 14 Correlation between Actual Versus Estimated Depth, Object Size Versus Estimated Distance Values Obtained from Different Camera Models

Camera model	Act range(cm)	Observations	Act vs Est(r1)	Object size vsEst.dist (r2)	Confid.level
NikonD5300	50–110	119	0.99213	0.86	95%
Canon	50–110	136	0.971356	0.92036	95%

Table 15 Model Accuracy: Evaluating Prediction Accuracy with Reference to Existing Works

Ref Document	Experimental set-up	Technique	Act(Range incms)	Performance metric
Chen et al. [21]	Considerations: Central plane of an object. Micro lens array. An elemental image array captured from plenoptic camera	SSD for computing disparity	not specified	Pearson.s correlation coefficient(r)=0.999
Farias et al. [22]	Considerations: Parallel stereo system mount setup with two identical cameras. Optical table for calibration	Triangulation method	27.9–81.3	RMSE=0.667
Said Pertuz et al. [17] & Palmieri et al. [23]	Considerations: Target, main lens, Micro lens array and Sensor	Light Field Refocusing	20–160	Pearsons correlation coefficient(r)=0.99
Proposed technique Using Nikon	Considerations: NikonD5300,Nikor AF-S 50 mm prime lens, Tripod	For each fixed lens aperture radius successive variation in object size is taken for Consideration.	50–110	RMSE=0.0413 r=0.99213
Proposed technique Using Canon	Considerations: Nikon D5300,Nikor AF-S 50 mm prime lens, Tripod	For each fixed lens aperture radius successive variation in object size is taken for consideration.	50–110	RMSE=0.056232 r=0.971356

lens arrangement for finding depth estimates. The depth estimates obtained from the proposed model approximate the ground truth with RMSE of 0.05, when compared with the recent works on stereo and light imaging by exhibiting around 98.1% correlation.

In our current study we are confined to finding the depth estimate only when object is on camera axial line. But in future the work is likely to be extended in such a way that depth of person is estimated when person is in slanting position. And also finding the depth of person in long range, say up to 40 mts is also a task which is under execution.

Acknowledgement The research is supported by Viswesvaraya Ph.D. scheme for Electronics & IT [Order No: PhD-MLA/4(16)/2014], a division of the Ministry of Electronics & IT, Govt of India. We gratefully acknowledge the association of SICA (Southern Indian Cinematographers), Chennai, for creating the Dataset to validate the proposed theory. The author would also like to thank the Student society, Department of Computer Applications who have been with us and supported us in Dataset creation.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Chaudhuri S, Rajagopalan AN (2012) Depth from defocus: a real aperture imaging approach. Springer, Germany
2. Hansard M, Lee S, Choi O, Horaud RP (2012) Time-of-flight cameras: principles, methods and applications. Springer, Germany
3. Lefloch D, Nair R, Lenzen F, Schäfer H, Streeter L, Cree MJ, Kolb A (2013) Technical foundation and calibration methods for time-of-flight cameras. In Grzegorzec M, Theobalt C, Koch R, Kolb A (eds) Time-of-flight and depth imaging. Sensors, algorithms, and applications. Springer, Berlin, Heidelberg, pp 3–24
4. Li L (2014) Time-of-flight camera—an introduction. *Technical white paper*, (SLOA190B)
5. Fuchs S (2010). Multipath interference compensation in time-of-flight camera images. In *2010 20th International Conference on Pattern Recognition* (pp. 3583–3586). IEEE
6. Whyte O, Sivic J, Zisserman A, Ponce J (2012) Non-uniform deblurring for shaken images. *Int J Comput Vision* 98(2):168–186
7. Cabezas I, Padilla V, Trujillo M (2011) A measure for accuracy disparity maps evaluation. In San Martin C, Kim S-W (eds) *Iberoamerican congress on pattern recognition*. Springer, Berlin, Heidelberg, pp 223–231
8. Scharstein D, Szeliski R (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int J Comput Vision* 47(1–3):7–42
9. Liu Y, Cao X, Dai Q, Xu W (2009) Continuous depth estimation for multi-view stereo. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2121–2128). IEEE
10. Wahab MNA, Sivadev N, Sundaraj K (2011) Development of monocular vision system for depth estimation in mobile robot—Robot soccer. In *2011 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (STUDENT)* (pp. 36–41). IEEE
11. Shan-shan C, Wu-heng Z, Zhi-lin F (2011) Depth estimation via stereo vision using Birchfield's algorithm. In *2011 IEEE 3rd International Conference on Communication Software and Networks* (pp. 403–407). IEEE
12. Langmann B (2014) Depth Camera Assessment. In *Wide Area 2D/3D Imaging* (pp. 5–19) Springer Vieweg, Wiesbaden
13. Sarbolandi H, Lefloch D, Kolb A (2015) Kinect range sensing: Structured-light versus Time-of-Flight Kinect. *Comput Vision Image Underst* 139:1–20
14. Achar S, Bartels JR, Whittaker WLR, Kutulakos KN, Narasimhan SG (2017) Epipolar time-of-flight imaging. *ACM Transactions Gr (ToG)* 36(4):1–8
15. Wang TC, Efros AA, Ramamoorthi R (2016) Depth estimation with occlusion modeling using light-field cameras. *IEEE Transactions Pattern Anal Mach Intell* 38(11):2170–2181
16. Monteiro NB, Marto S, Barreto JP, Gaspar J (2018) Depth range accuracy for plenoptic cameras. *Comput Vision Image Underst* 168:104–117
17. Pertuz S, Pulido-Herrera E, Kamarainen JK (2018) Focus model for metric depth estimation in standard plenoptic cameras. *ISPRS J Photogramm Remote Sens* 144:38–47
18. Kingslake R, Johnson RB (2009) *Lens design fundamentals*. Academic press, Cambridge
19. Hiscocks PD, Eng P (2011) Measuring luminance with a digital camera. *Syscomp electronic design limited*, 686
20. Conrad J (2007) Exposure metering relating subject lighting to film exposure
21. Chen Y, Wang X, Zhang Q (2016) Depth extraction method based on the regional feature points in integral imaging. *Optik* 127(2):763–768
22. Sánchez-Ferreira C, Mori JY, Farias MC, Llanos CH (2016) A real-time stereo vision system for distance measurement and underwater image restoration. *J BrazSocMechSciEng* 38(7):2039–2049
23. Palmieri L, Scrofani G, Incardona N, Saavedra G, Martínez-Corral M, Koch R (2019) Robust depth estimation for light field microscopy. *Sensors* 19(3):500

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.