IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 37, No. 2, April, 2020, pp. 333-340
Journal homepage: http://iieta.org/journals/ts

Depth Perception in a Single RGB Camera Using Body Dimensions and Centroid Property

P.J.A. Alphonse*, K.V. Sriharsha

Department of Computer Applications, National Institute of Technology, Tiruchirappalli 620015, India

Corresponding Author Email: alphonse@nitt.edu

https://doi.org/10.18280/ts.370220

Received: 18 November 2019
Accepted: 20 February 2020

Keywords:
stereo imaging, anthropometric, perspective errors, body dimensions, centroid, surveillance, vision

ABSTRACT

Infrastructure Supervision is a compelling need for buildings and open areas. It is facilitated through the joint use of stereo vision cameras, techniques and algorithms. This Stereoscopic assessment helps monitoring systems to reconstruct people's visible surface and also provides a robust estimation of the position and posture of the person that allows 3D scene activities and interactions. In practice, in occluded fields, the correspondence between pixels and pixels interferes with the flow of data in surveillance. Structured light ToF imaging and Light Field imaging sensors came into being considering the restriction. These techniques, however failed in addressing the inaccuracies and noise introduced in the phase of profound capture. Based on the Human Anthropometric research, we suggested a technique for estimating the depth of an individual from a single RGB camera. As we deal with moving objects in a scene, also consideration is given to centroid ownership. The system is trained by feeding stature, body width and centroid as inputs to estimate a person's actual height using gradient boosting model. And a person's further anticipated height and actual height are used to predict distance. After taking actual depth (camera to person distance) and real height as ground truth, the suggested model is validated and it is inferred that the camera to person distance anticipated (Pred$_{dist}$) from estimated Real height is 95% correlated with actual Camera to Person distance (or depth) at a confidence level of **99.9%** with RMSE of **0.092.**

1. INTRODUCTION

1.1 Motivation

Human beings, with their Binocular vision capacities and Biological Neural Network tend to perceive target depth. In reality, the same idea is used in Machine Vision technologies such as surveillance systems to discriminate sequences of human action in 3d environments. Given better accuracy requirements, the use of more than two cameras captures multiple views of the scene. Over moment, correspondence issues in stereo imaging systems are realized and addressed in Shao et al. [1]. Structured Light imaging systems by calibrating the disparity map from the distortions acquired with reference to known target structures in the projected models.

Smooth areas and also regions containing depth discontinuities are the fundamental reasons for occurrence of potential disparities. In such cases these can be minimized but the issue is grid lines width must be sufficient for obtaining the required accuracy in depth. The above issue is resolved in Microsoft's latest version of Kinect, time of flight principle [2]. This theory unlike the previous version, uses single shot photography and capture the scene from a single view point. But while in motion time of flight camera requires multiple shots. In such circumstances, camera may shake in motion, induces a blur in depth map and as a result, motion artifacts get corrupted. Considering all the implications chosen for their practicality and ability to act, a person's real height is estimated from an image captured by RGB camera and then used to predict the distance (or) depth of a camera to object. Neither extra hardware nor special software is required to communicate with the current underlying monitoring infrastructure. This invention is applicable in distributed surveillance system, where in order to determine human action sequences, depth information is essential. Knowing the depth, position and orientation of person relative to the camera is known. Without relying on existing triangulation and other existing depth estimation techniques, the proposed model can give depth information for static as well as moving objects with utmost accuracy.

1.2 Research contributions

- ✓ Data set comprising 22 males and 11 females including adults and children with heights ranging from 110cm to 190cm is created using Nikon D5300 DSLR camera with 18-55mm Nikor lens.
- ✓ Anthropometric based Feature Extraction is introduced for Real height estimation.
- ✓ A Standard Error is estimated in order to deal with Perspective Errors that arise in lens properties.

Related Work is presented in Section 2, Methodology including Influencing Factors, Creating Datasets, Feature Extraction for Real Height Prediction is presented in Section 3, Results and Discussion including Estimation of Measurement Errors Method Validation and Comparison with reference to existing works is discussed in Section 4. Rest of the paper is concluded in Section 5.

2. RELATED WORK

It is well recognized that traditional 2D methods are no longer appropriate and the capacity to achieve the necessary level of precision in 3D imaging is also lacking. However, in today's consumer landscape, three techniques are accessible for 3D imaging, including [3] Stereo Triangulation, Time of Flight and Structured Light. As pioneered in Marr [4], D Scharstein et al. [5-7], Disparity map estimation strategies are used for depth retrieval. But in the presence of noise, lot of irregularities are seen in depth retrieval. At the same time Anandan [8] and Hannah [9] used correlation and Sum of Squared Differences for determining pixel to pixel correspondence. But there is a mismatch in pixel to pixel correspondence whenever occlusions are encountered in surveillance environment. Selection of an appropriate window size is another problem observed while using correlation technique. And there is also a possibility that information flow under surveillance may break when any discontinuity arises at boundaries. However, Qayyum et al. [10] these uncertainties are removed at boundaries using Bayesian approach. This approach shows poor results when subjected to changes in illumination and contrast. Fradi et al. [11] Bidirectional Matching is well applied in occluded and low textured areas and low pass filtering technique is used as a pre processing step to fill the missing information whenever there is break in information flow under surveillance at the boundaries. On the other side this technique may fail whenever there is high variation and distortion in information inside a window area. In order to overcome the illumination affects, rather than dealing with image intensities symbolic features are considered to serve the purpose. In this context, Feature Based Strategies fail in selecting an appropriate interpolation method for non-featured areas. Touzene and Larabi [12] trained a neural correlation network with data comprising hundred pairs of matched and unmatched pixels. This training cannot be accommodated in resource constrained surveillance cameras. Structured Light imaging based camera eliminates correspondence problem by constructing disparity map from distortions obtained in projected patterns with reference to known pattern pertaining to subject of interest. But the limitation is grid line width must be high enough to obtain required accuracy in depth values. Microsoft’s latest version of Kinect uses single view point for computing depth by utilizing Time of Flight principle. But to operate in dynamic scenes, it requires multiple shots. In such circumstances, camera shake in motion and induces a blurr in depth map and as a result motion artifact may get corrupted. Fusion of ToF and Stereo imaging also introduced but du e to correspondence problems, the technique failed in giving desired depth estimates. Plenoptic cameras [13-17] 4D structures encoding angular information provide better solutions for vision and scene understanding. This technique is capable enough to deal with video stabilization, object detection, tracking and recognition problems but due to Poor reconstruction depth quality, large storage requirements and high investment it is not in use. Owing to economic and technical feasibility addressed in the literature, we have proposed a theory by relating Real Height and Camera to Object Distance (or Depth). This theory can be well established in all existing surveillance camera systems without incorporating any additional hardware or software.

3. METHODOLOGY

3.1 Influencing factors

3.1.1 Finding real height

Since the primary objective is to discover the Camera to Person distance, consideration is given to human Anthropometric research. As referred in the study [18], among 9 Anthropometric measurements namely, Stature, Neck height, Acrominal height, Head length, Mouth to top of Head distance, Forehead to Chin distance, Sellion to Chin distance, Biocular distance and Bitragion distance, Stature (or Body height), and width are taken as criterion for human height estimation. As we are dealing with the body motions in a scene, centroid of the person in an image is also considered for better results. Distance (or depth) is estimated with the outcomes acquired pertaining to actual person height. The process is illustrated in subsequent sections.

3.2 Experimental study

The experiment is conducted at Sensors and Security Lab, Department of Computer Applications, National Institute of Technology, Tamil Nadu state, India. The research seeks to measure a Person's depth using a single RGB camera by interpreting the measurements of the human body based on human anthropometric research. In this perspective, a DSLR Nikon camera with 18-55mm zoom lens is used for capturing the persons with different heights with in 381cm to 681cm range. Parameters like aperture, shutter speed and ISO are tuned in required ratios to achieve the precise light on sensor. The device and its exposure are mentioned in Table 1 & Table 2.

Table 1. Experimental device details for photogrammetry experiment

Devices	Values
Camera Model	NikonD5300
Image sensor	23.5 × 15.6mm CMOS sensor
Image size(pixels)	6000 ×4000
Pixel size	3.9 micron
Lenses	Nikon AF-S Nikkor 18-55mm

Table 2. Parameter influencing the camera exposure

Exposure settings	Values
ISO	2000
Aperture (f-stop)	f/16
Exposure time (shutter speed)	1/5 s

3.2.1 Creating Image Dataset for real height estimation

As mentioned in Sriharsha et al. [19], we invited 33 distinct subjects (22 males and 11 females) for our data collection and captured 264 image samples with a fixed camera view point at different focal lengths under the supervision of under the supervision of Sica Southern Indian Cinematographers Association, Chennai. In order to show the variation and diversity of the data set used in our experiment(s), the age, gender and height are taken for consideration. The dataset is created for real height estimation using anthropometrics and later used in this work for retrieving depth from Real height estimated. Apart from the training data taken from data set comprising 22 males and 11 females including adults and children with heights ranging from 110cm to 190cm, additionally 524 images are created from existing training data using Data augmentation. Brightness width shift and scaling characteristics are applied on each training sample and as a result different version of transformed images are reproduced. Though dataset is restricted to indoor environments, we provide the ambiance continuity by capturing the same conditions because of the operational limitation of the Acquisitions sensor. The Augmentation process is mentioned in Algorithm 1.

Algorithm 1: Augment_Image_Bright _shift Width_Scale
Input(s): Image(s)
Output: Image
Initalize Params: None
Steps:

```
1. Store the image as an array
   // Applying the scaling
2. Scale the image to the required dimension
   // Applying brightness and width shift properties
3. Create an instance of ImageDataGenerator (on 4
   Brightness ranges)
4. Extract this array of image with an instance and sample
   and convert back to image
5. Create an instance of ImageDataGenerator (on required
   width shift ranges)
6. Extract this array of image with an instance and sample
   and convert back to image
```

End algorithm.

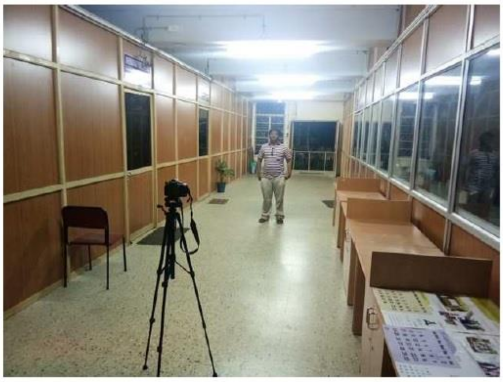


Figure 1. Standing posture of person along camera axial line

As shown in Figure 1, starting at a distance of 381 cm from the camera axial line, you keep moving away from the camera until you reach 681cm. Two hundred and Sixty four (264) photographs are continuously shot on standing positions at an interval of 30cm by varying focal length. All these images are pre-processed and used for forecast of actual height and distance. For removal of blur Sriharsha et al. [20] in image due to linear motion or unfocussed optics, filtering operation is applied. As it is known that Wiener filter is suitable for reconstruction of original from the noisy image and hence it is chosen for image filtering operation. Finally dilation and erosion operations are applied for removal of image imperfections.

3.3 Feature extraction for real height prediction

Unlike in the work of Sriharsha et al. [19], out of 9 anthropometric measurements, only two metrics namely stature (head to foot range), body width and centroid property are used for Real height estimation. Initially YOlO object detector issued in the video to detect objects (say persons) in each frame. The YOLO, divides the system the incoming frame of a video into an S×S grid. If the object center drops into a grid cell, the grid cell detects the object. Each grid cell predicts B boundary boxes and probability scores for these boxes. These scores show that the model is confident that the box contains an object (say person) and how exact it thinks the box is that it predicts the stature, width and centroid properties are obtained based on the greatest probability score. Here the probability score is normalized to [0, 1] range. The procedure for feature extraction is mentioned in Algorithm 2. For the corresponding real height of the person in the image, person height (Body Height), and width (Body Width), and centroid location (CenterX, CenterY) are extracted in pixels. Pixel measurements are listed in Table 3.

Table 3. Extraction of body height, width and centroid measurements in (pixels)

SNo	Real_{ht}	D1	D2	D3	D4
1	110	1029	300	972	2517
2	146	2062	681	1050	2647
3	147	1404	375	962	2423
4	151	1927	633	963	2423
5	152	2992	757	1022	2235
6	157.1	2366	739	1135	2107
7	159	2310	763	943	2065
8	162	1791	536	1159	1916
9	163	2268	623	972	1982
10	164	3238	921	976	2221
11	167	2261	774	1065	1974
12	168	1442	449	904	1776
13	169.5	1767	490	1119	1867
14	170	1899	567	1140	1782
15	171	2151	648	1120	1864
16	172	1744	514	1087	1783
17	176	2227	546	995	1823
18	180	1498	404	974	2275
19	184	2264	571	1203	1802
20	190	2003	549	1059	1718

D1: Body Height (pixels); D2: Body Width(pixels); D3: CenterX (pixels); D4: CenterY (pixels)

Algorithm 2: Extract_Body_ht_wt_loc _Measurements
Input(s): Image(s), W: Weight, H: Height
Output: Face height, Neck height, Mouth to Forehead, Eyes to Chin, binocular
Initalize Params:

```
1. Assign paths to yolo-weights file, yolo-names file and yolo-config file
2. Load YOLO:
get an instance of Dense Neural Network of Darknet
provide config and weights to them
Define the labels with names file
```

Steps:

```
  1. Get the shape(W,H) of an image
  2. Take an instance from the blob from image by Dense Neural Network
  3. Get a list of layer _Output of all outputs by (net & blob instance)
For output in Layer_Output:
        For detection in output:
            get the classID and scores of detection
                IF classID is for Human:
                    IF scores are satisfactory:
```

```
                get Height,Width and Centroid of Object
                get box for each detected object
            EndIf
        EndIf
    EndFor
EndFor
(ids)=filter the boxes with thresold of min probability of Human
IF Detected:
    For each id in (ids):
        return height,width and centroid
    EndFor
Else
    return None
```

3.4 Privileged real height and distance prediction

Table 4. Sample photographs with predicted real height and distance measurements

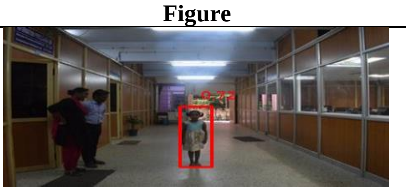
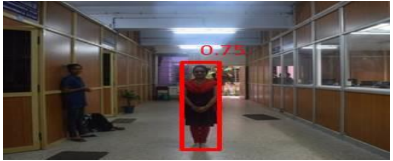
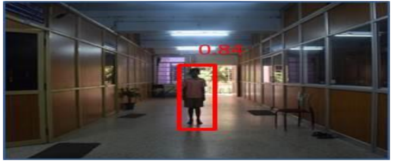
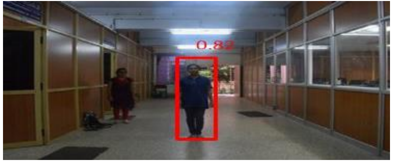
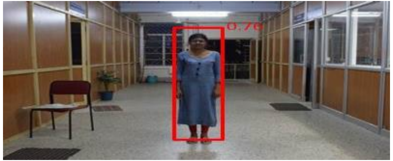
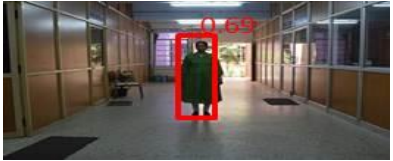
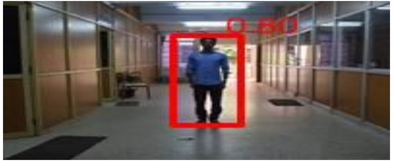
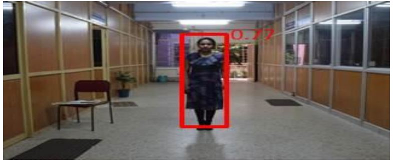
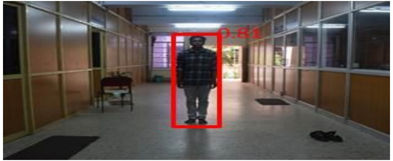
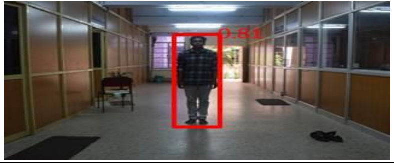
Real_{ht}	Pred_{ht}	Act_{dist}	Pred_{dist}	Figure
110	0.4013	531	527.44	
146	144.52	381	384.11	
147	147.52	531	537.73	
151	148.70	381	384.15	
152	154.29	381	378.64	
157.1	159.46	411	415.14	
162	162.94	561	561.46	
163	165.91	441	451.88	
164	162.66	381	379.98	
170	169.09	591	598.35	

Table 5. Sample photographs with predicted real height and distance measurements (continuation)

Real_{ht}	Pred_{ht}	Act_{dist}	Pred_{dist}	Figure
170	171.91	681	663.14	
171	170.48	591	577.48	
172	172.46	681	663.22	
177.69	176.62	561	566.53	
180	170.80	621	623.65	
184	170.80	621	623.65	
190	180.28	651	642.23	
159	162.85	441	422.40	
167	166.34	411	419.72	
168	168.38	681	670.23	

A Regressor (GBR) gradient [21, 22] is trained with three characteristics: stature, person width and centroid, as mentioned in Section 3.3 in order to obtain Pred_{ht}. The Pred_{ht} prediction rate is evaluated using RMSE and Pearson's correlation coefficient(r) for a test_size=0.2. Once real height is expected, the expected distance Pred_{dist} is acquired for the respective test and train samples. The procedure is stated in Algorithm 3 and results are tabulated as shown in Table 4 and Table 5.

Algorithm 3: Predict_Height_Distance
Inputs: Act_{ht}, $D_{i=1....4}$, pixel _size

```
Outputs: Pred_ht (cms), Pred_dist (cms)
InitalizeTrainingParams: test_size=0.2, iteration(k)=200, batch_size=788
hiddenlayer_size=200, activaltion=relu, learningrate_init=0.001, early_stopping=true, random _state=66
Steps:
1.$X \leftarrow$ Extract_Features_ from _Dataset ($D_{i=1 \text{ to } 4}$),
2. $y \leftarrow$ Extract_Label_ from _Dataset ($ht_{i=110 \text{ to } 190}$)
//splitting of data
3.$Train_X$,$Test_X$,$Train_y$,$Test_y$
        $\leftarrow$ split ($X$, y, test_size=0.3, random _state=66)
        //Training of model
4.For i=1 to K do
        grb $\leftarrow$ GBR model trained on ($Train_X$, $Test_y$)
5.EndFor
6.predictions $\leftarrow$ predict ($Train_X$)
//Evaluate RMSE,r
//Distance Prediction
7.height _pred=predictions
8.$X_d$=concat (height _pred,X)
9.$Y_d$=$Act_{dist}$
10$X_d$_train, $X_d$_test, $Y_d$_train, $Y_d$_test=train_test_split ($X_d$, $Y_d$, test _size, random _state)
11.Pred $\leftarrow$ predict (Xd_test)
12.Evaluate rmseD, CorrD
End algorithm
```

4. RESULTS AND DISCUSSION

4.1 Estimation of measurement errors

Overall, when captured by the camera using standard lenses, perspective errors [21] are seen compressed or extended around the middle of a picture. As shown in Figure 2, lens are subjected to Perspective Errors due to variation in camera view (treated as α) and lateral displacement of an object from the point where it is placed. As a consequence, these errors influence the characteristics of lens magnification. This also in turn shows an impact on precision of object distance measurement. The corrected distance measurements are furnished in Table 6. From Figure 2, the Perspective Error δx [22] computed is as follows:

$$\text{Perspective Error } (\delta x)= \delta z \times \tan \alpha \qquad (1)$$
$$\delta z \text{ - out of plane displacement}$$

Table 6. Sample distance measurements corrected with δx

SNO	Act $_{dist}$	Pred $_{dist}$	Pred$_{dist}$ corrected
1	531	527.44	527.44
2	381	384.11	384.11
3	531	537.73	530.08
4	381	384.15	384.15
5	381	378.64	378.64
6	411	415.14	415.14
7	561	557.28	557.28
8	441	451.88	444.23
9	381	379.98	379.98
10	591	598.35	590.70
11	441	450.72	443.06
12	681	670.23	677.88
13	591	607.68	600.02
14	651	639.68	643.34
15	501	485.24	493.34

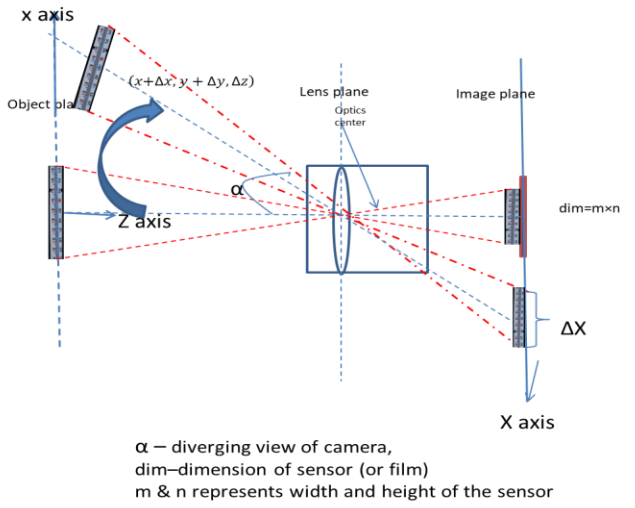

x axis
Object pla
$(x+\Delta x, y+\Delta y, \Delta z)$
Lens plane
Optics center
Image plane
α
Z axis
dim=m×n
ΔX
X axis
α – diverging view of camera,
dim–dimension of sensor (or film)
m & n represents width and height of the sensor

Figure 2. Perspective errors caused by out of plane displacement

The camera to Person distance estimated (Estimated$_{dist}$cm)) is corrected with a factor ± δx

$$\begin{aligned} \text{Pred}_{dist} \text{ corrected} \\ &= \text{Pred}_{dist} + \delta x \ \text{ if } \ \text{Pred}_{dist} < \text{Act}_{dist} \\ &= \text{Pred}_{dist} - \delta x \ \text{ if } \text{Pred}_{dist} > \text{Act}_{dist} \end{aligned} \qquad (2)$$

where, $\delta x = \sigma / \sqrt{n} = 7.6582$
n: number of samples =159

4.1.1 Method validation

Considering null hypothesis as no significance between current and suggested model and, alternatively as a significant difference in proposed and actual model, the procedure delineated is as follows. To consolidate the efficiency of the suggested model, 159 samples are taken into consideration. As stated in Table 6, difference in Pred$_{dist}$ corrected from Actd$_{ist}$ is selected as test variable for carrying out one sample t-test.

Calculating $\sum d^2 = (\sum d^2)$ gives $\sum d^2 = 83561.99758$, $(\sum d^2) = 6910119.268$.

$$t_{cal} = \sum d \,/\, \sqrt{(n \times \sum d^2 - (\sum d)^2/(n-1))} \qquad (3)$$
$$= 2628.71/2636.91 = 0.9968 \text{ on } 158 \text{ df}$$
n-1: degrees of freedom

Looking this up in tables gives p = 0.001. Therefore, there is strong evidence that, on average, the module does lead to improvements.

For hypothesis testing, calculated t_{cal} is compared against tabled value (inferred from values of t-distribution) at (α)=99.99. As t_{cal}<3.390 at confidence level (α) =**99.99%,** null hypothesis is accepted and hence it is found that there is no significance between the predicted and standard observations. It is inferred from the discussion that the model correctly fitting the data with confidence level 99.99%.

4.2 Model accuracy with reference to the existing works

The comparison is produced on the level of experimental configuration, technology and precision. The details are furnished in Table 7. It is found from Table 7 that our model performs equally well with the Chen [14] depth estimation technique by demonstrating elevated consistency between the expected (estimated) and reference (real) depth values (r) with a Pearson's correlation coefficient of r=0.95. Unlike the micro-lens and primary lens arrangement, as mentioned in the works [13, 14, 17, 23], for 2D images, we used single DSLR nikkor 18-55 mm zoom lens. Unlike the disparity and distance between the image planes and the micro lens, we used body dimension and centroid as characteristics for true height forecast. No where binocular vision techniques are used in the suggested technique and a single RGB camera is used to measure depth.

Table 7. Model assessment: Referring design, technique and performance of existing works

Sno	Ref. doc	Experimental set-up	Technique	Performance Metric
1	Wang [13]	Considerations: ✓ central plane of an object is taken ✓ Micro lens array ✓ elemental image array captured from plenoptic camera	SSD for computing disparity Calculating micro lens pitch Distance between image plane and micro lens array, D = (p*g)/d	Pearson's correlation coefficient(r)=0.999
2	Farias [23]	Considerations: ✓ Parallel stereo system mount setup with two identical cameras ✓ Optical table for calibration	**Background subtraction** for determining position (center of mass) in both cameras. Compute the disparity between the two points. Apply the triangulation method for distance measurement	experimented with in 27.9-81.3 range and RMSE=0.667
3	Said Pertuz [14, 17]	Considerations: ✓ Target, Main lens ✓ Micro Lens array, sensor	✓ Find the pixel pitch ✓ Finding the focal length of microlens ✓ Finding micro lens diameter ✓ Finding the separation between real and synthetic focal plane ✓ computing refocusing parameter ✓ computing focusing distance(or)depth	experimented with in 0.2-1.6m range and r=0.99
4	Proposed Technique	Considerations: ✓ DSLR nikkor 50mm Prime lens (DSLR nikkor 18-55mm zoom lens (**single RGB camera**)) ✓ Camera mounted on Tripoid	**Relating real height of a person and camera to object distance** ✓ Initially real height is inferred from body dimensions and centroid property ✓ Person distance is predicted from Real height using Machine Learning Model	experimented with in **3.81m-6.81m range** and pearson's correlation coefficient(r)=**0.95**

5. CONCLUSION

This work presented a theory in relating Real height of a person to the camera to object distance. For this purpose, close range photography was used to investigate the impact of human anthropometric study on camera to object distance (or) depth and also the impact of perspective errors while estimating object distance from the center of the lens. A Nikkon DSLR camera model with 18-55 mm zoom AF-P Nikkor is used to capture still images of people in a standing position. The suggested model was tested on dataset in 3.81m-6.81m distance range in indoor environment. 22 males and 11 females are photographed at various distances starting from 381cm with in an interval of 30cm along the camera axial line with variable focal lengths in order to estimate actual height of individual with each of 11 instances. Considering anthropometric human body, actual height is estimated and the person distance (or depth) values are subsequently acquired from the expected height. The technique is validated using one sample T-test on 159 samples and it is discovered that the depth values acquired from the suggested theory show 95% confidence level correlation with ground truth at 99.9% confidence level.

In future, this work is extended to outdoor environment up to 40 mts distance. Within the same range, we are also expected to find the slanting distance of a person (i.e. person moving away from camera axial line). And also using depth estimated from the data, we further try to estimate the position and orientation relative to camera. This would help in discriminating human action sequences from their movements relative to camera.

ACKNOWLEDGEMENT(S)

This work is supported under 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. The writers also gratefully acknowledge the association of SICA (Southern Indian Cinematographers Association), 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. The Author is also grateful to the In-charge Faculty, Sensors & Security Lab for offering experimental validation facilities.

REFERENCES

[1] Shao, L., Han, J., Kohli, P., Zhang, Z. (2014). Computer vision and machine learning with RGB-D sensors. Part of the Advances in Computer Vision and Pattern Recognition book series (ACVPR), 3-26. https://doi.org/10.1007/978-3-319-08651-4

[2] Sarbolandi, H., Lefloch, D., Kolb, A. (2015). Kinect range sensing: Structured-light versus Time-of-Flight Kinect. Computer Vision and Image Understanding, 139: 1-20. https://doi.org/10.1016/j.cviu.2015.05.006

[3] Liu, Y.B., Fang, L., Gutierrez, D., Wang, Q., Yu, J.Y., Wu, F. (2017). Introduction to the issue on light field image processing. Pattern Recognition Letters, 11(7): 923-925. https://doi.org/10.1109/JSTSP.2017.2759458

[4] Marr, D.C. (1982). A Computational Investigation into the Human Representation and Processing of Visual Information. The MIT Press.

[5] Scharstein, D. (1999). View Synthesis Using Stereo Vision. Springer-Verlag.

[6] Scharstein, D., Szeliski, R. (1998). Stereo matching with nonlinear diffusion. International Journal of Computer Vision, 28: 155-174. https://doi.org/10.1023/A:1008015117424

[7] Scharstein, D., Szeliski, R., Zabih, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001), Kauai, HI, USA, pp. 131-140.https://doi.org/10.1109/SMBV.2001.988771

[8] Anandan, P. (1989). A computational framework and an algorithm for the measurement of visual motion. International Journal of Computer Vision, 2: 283-310. https://doi.org/10.1007/BF00158167

[9] Hannah, M.J. (1974). Computer Matching of Areas in Stereo Images. Stanford Univ Ca Dept of Computer Science.

[10] Qayyum, A., Malik, A.S., Saad, M.N.B.M., Abdullah, F., Iqbal, M. (2015). Disparity map estimation based on optimization algorithms using satellite stereo imagery. In 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 127-132. https://doi.org/10.1109/ICSIPA.2015.7412176

[11] Fradi, H., Dugelay, J. (2011). Improved depth map estimation in stereo vision. Stereoscopic Displays and Applications XXII, 7863: 78631U. https://doi.org/10.1117/12.872544

[12] Touzene, N.B., Larabi, S. (2010). Neural disparity map estimation from stereo vision. International Arab Journal of Information Technology, 9(3).

[13] Wang, T.C., Efros, A.A., Ramamoorthi, R. (2016). Depth estimation with occlusion modeling using light-field cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(11): 2170-2181. https://doi.org/10.1109/TPAMI.2016.2515615

[14] Chen, Y., Wang, X., Zhang, Q. (2016). Depth extraction method based on the regional feature points in integral imaging. Optik, 127(2): 763-768. https://doi.org/10.1016/j.ijleo.2015.10.171

[15] Pertuz, S., Pulido-Herrera, E., Kamarainen, J. (2018). Focus model for metric depth estimation in standard plenoptic cameras. ISPRS Journal of Photogrammetry and Remote Sensing, 144: 38-47. https://doi.org/10.1016/j.isprsjprs.2018.06.020

[16] Marin, G., Agresti, G., Minto, L., Zanuttigh, P. (2019). A multi-camera dataset for depth estimation in an indoor scenario. Data in Brief, 27*:* 104619. https://doi.org/10.1016/j.dib.2019.104619

[17] 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. https://doi.org/10.3390/s19030500

[18] Bailar, I.I.I., Meyer, E.A., Pool, R. (2007). Assessment of the NIOSH head-and-face anthropometric survey of US respirator users. The National Academies Press. https://doi.org/10.17226/11815

[19] Sriharsha, K.V., Alphonse, P.J.A. (2019). Anthropometric based real height estimation using multi layer peceptron ANN architecture in surveillance areas. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-6. https://doi.org/10.1109/ICCCNT45670.2019.8944862

[20] Sriharsha, K.V., Rao, N.V. (2015). Dynamic scene analysis using Kalman filter and mean shift tracking algorithms. In 2015 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-8. https://doi.org/10.1109/ICCCNT.2015.7395214

[21] Sciammarella, C., Considine, J., Gloeckner, P. (2016). Experimental and Applied Mechanic. CRC Press. https://doi.org/10.1007/978-1-4419-9792-0

[22] Chen, L., Wei, H., Ferryman, J. (2013). A survey of human motion analysis using depth imagery. Pattern Recognition Letters, 34(15): 1995-2006. https://doi.org/10.1016/j.patrec.2013.02.006

[23] Sánchez-Ferreira, C., Mori, J.Y., Farias, M.C.Q., Llanos, C.H. (2016). A real-time stereo vision system for distance measurement and underwater image restoration. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 38: 2039-2049. https://doi.org/10.1007/s40430-016-0596-5