ISSN 2278-3091
Volume 8, No.5, September - October 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse112852019.pdf
https://doi.org/10.30534/ijatcse/2019/112852019

Depth Analysis of Single View Image Objects based on Object Detection and Focus Measure

Jyoti B. Kulkarni[1], Manna Sheela Rani Chetty[2]
[1] Research Scholar, Department Of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India – 522502,
mailjyotikulkarni@gmail.com
[2] Professor, Department Of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India – 522502,
sheelarani_cse@kluniversity.in

ABSTRACT

In today's Computer Vision era, one important step is 3D scenes generation from 2D images. One of the challenging and difficult tasks is generation of Depth Map from Single View Image. Depth Map gives depth information of pixels in 2D form. Image may have many objects in it and focus measure is one of the important depth cues. In this paper, indoor images with different objects have considered and the focus measures of various objects in the images have calculated using Discrete Wavelet Transform (DWT) by considering it as one of the evaluation metrics. The results show that if the particular object is near, far from the camera, or from one another, the focus parameter changes with respect to that. The object that is near to the camera has greater focus measure as compared to the objects, which are relatively farer. In addition to that, the entropies of different objects computed as another evaluation metrics. Deep Learning used here for object detection from image and the depth map generated.

Key words: Focus Measure, Discrete Wavelet Transform **(**DWT), Object Detection, Entropy, Deep Learning, Depth Map.

1. INTRODUCTION

In this paper, the relative depth of various objects in a single image, is estimated. Due to single image as an input, the job is somewhat complex. The image does not provide any depth information. In this paper, this issue is addressed by computing focus measures of different objects and by comparing those measures. Indoor images with different objects are considered here.

From single view image, object detection is done using deep learning. Object detection is more complicated task rather than object classification. YOLO framework (You Only Look Once) is smart neural network that detects objects from images online. The entire image is taken at once and YOLO does the prediction of the bounding box coordinates and class probabilities for these boxes**.** The most important advantage of it is the faster performance. YOLO understands generalized object representation also.

In paper [1], the authors proposed an innovative depth estimation method that uses object classification based on the Bayesian Learning algorithm and used Mean Shift Algorithm [1]. Training dataset of six attributes is used in [1] and classified the objects from the single image into four different types. To each object, a relative depth value is assigned depending on the type and a simple 3D model is generated [1].

In paper [2], for depth estimation, the authors used an aperture camera that is phase-coded, and it also consisted of an optical phase mask providing depth-related color features for the captured image [2]. The scene depth map is estimated using a fully convolutional neural network [2]. The authors, in paper [3], proposed a new quality criteria set, that allows thorough analysis by focusing on particular characteristics of depth maps [3]. They have also used a Digital Single Lens Reflex camera along with a laser scanner to obtain high-resolution images and very accurate depth maps [3].

The authors, in paper [4], proposed a novel method in [4] that uses two deep network stacks. The authors, in paper [4], collected various cues based on deep network and combined local and global information. Two stacks are used in such a way that global stack processes coarse prediction of entire image and local stack does the prediction refinement [4].

In paper [5], the authors deliberated about the combination of Neural Network and Wavelet Transform. In this paper, mixture of two techniques, i.e. DWT and NN on IRS-1D images, used and improved results obtained.

In paper [6], the authors proposed a novel method to measure a focus of an image based on wavelet transform applied on it.

The authors decomposed the image up to third level of wavelet transform and the wavelet coefficient computed. The coefficient is nothing but a high pass band and low pass band norms' ratio. It is proved, in paper [6], that the measure is directly proportional to the amount of defocus or blur in the image.

In paper [7], the authors designed a fully convolutional architecture. Here, a ResNet-50 network used. It has fully trained prior and image features retrieved then. The authors further replaced ResNet-50 network layer by the blocks of upsampling and the size of the depth map recovered in paper [7]. This model further compared with CNN network and demonstrated the efficiency of it in paper [7].

The authors, in paper [8], discussed about focus measure using Discrete Wavelet Transform and image reconstruction techniques from blur and noisy images. Further, image reconstruction time analyzed [8].

The authors, in paper [9], have employed an approach based on a learning that is supervised, to generate a disparity map i.e. depth map from an image that is single view. They collected some monocular images as a training set. Their model customs a Markov Random Field, which is trained, that includes image features that are local as well as global [9].

In paper [10], the authors proposed scene segmentation based on semantic approach and used labels for directing the three dimensional reconstruction. The authors designed to extend their system that learnt from various strong data sources consisting of synthetic data like ray-traced scenes and poorly labelled images i.e. with the height of some classified objects [10].

The authors, in paper [11], proposed a method to estimate depth. The system implemented for a mobile platform. Monocular vision used here. The system developed in the paper [11], required a calibration. The data gained from calibration, can be used to generate the depth estimation equation. This task is only limited to mobile platform and not for general purpose.

In paper [12], a queue of images considered for which the depth is known, as there is absence of depth cue in single image. In paper [12], depth estimation of monocular images is done. The issue tackled as a discrete and continuous optimization problem. The variables formulates the depth of the super-pixels in the input image in paper [12].

In paper [13], the author developed a filter that estimates the depth of an object in image. A blur image considered here, in paper [13]. The proposed model is based on an assumption that, an image of an object, which is defocused, is a convolved image of a sharp image with 2D Gaussian function. The object depth is related to it's spread parameter.

In paper [14], the authors developed a monocular image depth estimation model using Bayesian approach. Three dimensional scene structures recovered here [14]. From single monocular images; robust, reliable, depth related features retrieved. The Natural Scene Statistics model used in [14], illustrates the relationships between two dimensional images and their corresponding depth maps [14].

In paper [15]; the authors used object recognition techniques and a novel technique for multiple object detection in an image. They have used template matching technique. The technique needs huge amount of image templates for accurate recognition of objects from images [15].

The authors, in paper [16], developed a new depth image coding system grounded on Block Compressive Sensing method. From the encoder side, the entropy of pixels in each block implemented, in paper [16]. Due to that, the sparsity of depth signals is denoted.

In paper [17], the authors employed the object detection that is done for detection and tracking of objects in the video.

The authors, in paper [18], tells about Discrete Wavelet Transform (DWT) and Intensity-Hue Saturation (IHS) using fusing criteria for selecting rough and detailed sub-images. When DWT has mixed with any other technique like HIS, it will give well results.

The authors, in paper [19], used a new method called Conditional Adversarial Networks (CAN) and used it for converting the clouds of sparse points into dense maps. The method is effective to estimate depth map. The paper emphases on sparse points only [19].

In paper [20], the authors studied the edge detection methods based on intensity levels that are not continuous. The conclusion is made that the Canny result is superior for a particular image, because various edge detection algorithms work good for different situations. The authors in paper [20] said that it is a challenging job for detecting the predicting same image without noise from the original image. In paper [21], the authors proposed an image retrieval based on Cubic Bezier Curve. Nonlinear relationship has discussed between query image and database image.

The remaining part of this paper has arranged as follows:
Section 1 explains about Introduction, Section 2 explains about Wavelet Focus Measure by DWT, Section 3 explains about proposed method, section 4 explains about Experimental Results and Discussion and Section 5 describes conclusion.

2. WAVELET TRANSFORM BASED FOCUS MEASURE: DISCRETE WAVELET TRANSFORM (DWT)

When an image is considered, then its Wavelet Transform has explained as a ratio of high pass and low pass norms. Discrete Wavelet Transform and Discrete Cosine Transform are same when there comes a conversion of spatial to frequency domain

[5]. In Discrete Wavelet Transform, the image splits in 4 sub images. It is shown in Figure l (a). In it, a_1 is original image’s approximation. b_1, b_2 and b_3 are the sub images that consist of diagonal, horizontal and vertical components of the original input image [5]. Further, DWT can be applied on a_1 and a 2 level DWT breakdown is obtained that result in 7 sub images which is shown in Figure. 1 (b). The procedure may continue to produce decomposition up to nth level [5].

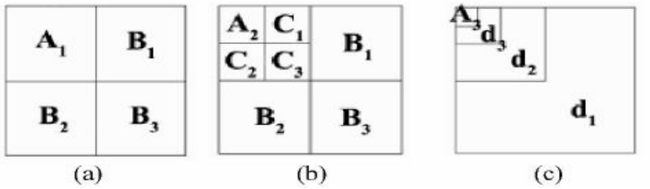


Figure 1: A wavelet coefficients map after **-** a) at first level, b) at second level, c) at third level decomposition.

In this paper [5], at jth level of the Wavelet Transform decomposition, 3 sub images are merged into a single image and called as $d_j(p,q)$. 3rd level decomposition and the resulting images at every level of decomposition, are shown in Figure. 1 (c). The focus measure i.e. wavelet coefficient *w* is calculated depending on Discrete Wavelet Transform.

Assume an image *I*. Coefficient *w* is wavelet coefficient and it is computed as a ratio of high pass to low pass band.

$w = \frac{\|h(I)\|}{\|l(I)\|}$; where, $h(I)$ and $l(I)$ are high pass band and low pass band respectively……………………… (1)

The coefficient *w* is not varying as per the measure of the image. Due to the preserving norms and the blur, the energy reduces in high pass bands for the image and at the same time, the energy of the image rises in low pass bands. So, the difference or variation of the coefficient gets increased.

3. PROPOSED METHOD:

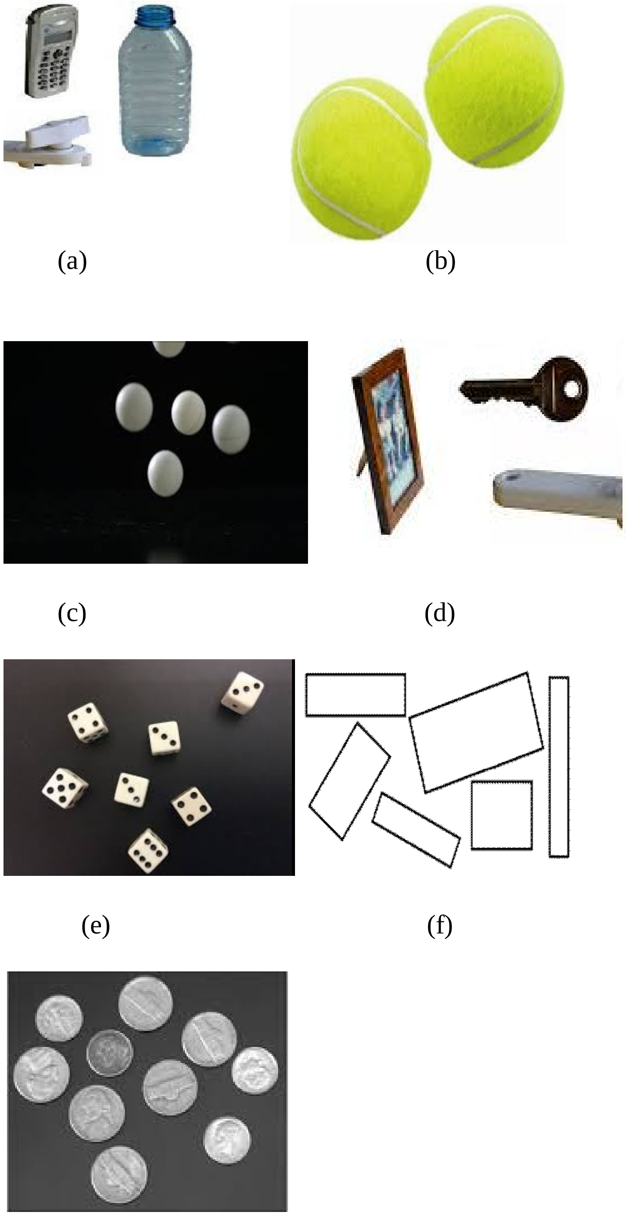
Algorithm:

a. Acquire an image.
b. Detect various objects by Deep Learning using YOLO3 framework.
c. Compute wavelet coefficient *w* i.e. focus measure of objects in the image.
d. Compute entropies of objects in the image.
e. Compare wavelet coefficients and entropies.
f. Find near and far objects in the image.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Firstly, the single view image objects are detected using deep learning. YOLO3 framework has used for object detection. In this paper, indoor images with various objects that have not overlapped on each other are considered e.g. images of all balls, images with various objects like bottles, calculator, balls, etc. Once all objects have detected, their focus measures and entropies have computed. Discrete Wavelet Transform (DWT) has used to measure the focus of objects in the image. Wavelet-based measure *w* shows good robustness. When depth of the wavelet decomposition increases, the robustness increases. In the experiments, level 2 produces sufficient robustness [5]. Mostly the images with same objects of roughly same size is considered so that the computed focus measures are justified e.g. the images of balls, coins, bats, etc. However, this approach also worked for some images with different objects.

4.1 Input sample images:



(a) (b)

(c) (d)

(e) (f)

(g)

Figure 2: Input test images

4.2 Results

Table 1:Focus measures calculated for objects in Figure 2. (a)

Objects	Focus Measure (*w*)	Entropy
object2 (far)	0.5438	3.63
object3 (middle)	0.42	4.7
object1 (near)	0.4164	6.51

Table 2:Focus measures calculated for objects in Figure 2. (b)

Objects	Focus Measure (*w*)	Entropy
object2 (far)	0.2091	5.5493
object1 (near)	0.1945	5.6798

Table 3: Focus measures calculated for objects in Figure 2. (c)

Objects	Focus Measure (*w*)	Entropy
Object1 (nearest)	0.4773	5.78
Object2 (far)	0.4633	4.41
Object3 (far from object1)	0.4569	5.61
Object4 (far from object1)	0.4535	5.87
Object5 (far from object1)	0.4162	5.77

Table 4:Focus measures calculated for objects in Figure 2. (d)

Objects	Focus Measure (*w*)	Entropy
Object1 (nearest)	0.92	4.77
Object2 (near)	0.90	4.61
Object3 (near, but far from object1)	0.88	4.12
Object4 (far from object3)	0.84	4.01
Object5 (very far)	0.80	3.81

The results in the table1, table 2, table 3 and table 4 show that the focus measures are less for the objects which are relatively far from the objects that are nearer (to camera). the objects are that are near have greater focus measures [5].

The entropies are also affecting on the relative depths of the objects in the image. The entropies are relatively decreased as the distance of the object increases from camera.

4.3 Graphical representation

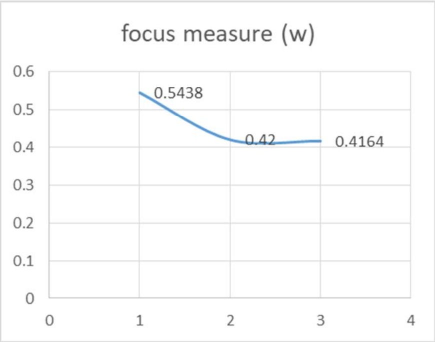


Figure 3: Graphical representation of Figure 2. (a)

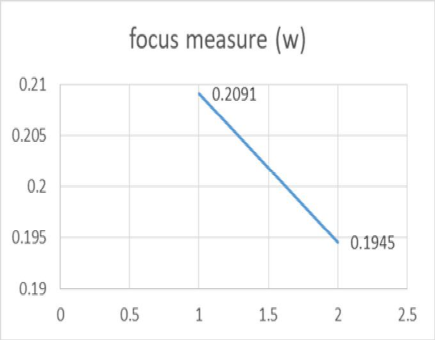


Figure 4: Graphical representation of Figure 2. (b)

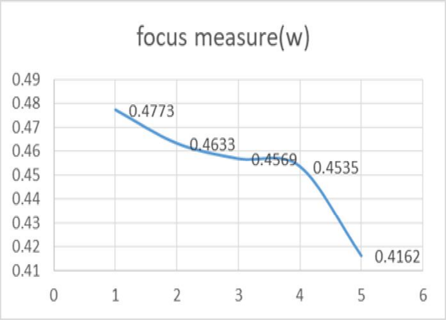


Figure 5:Graphical representation of Figure 2. (c)

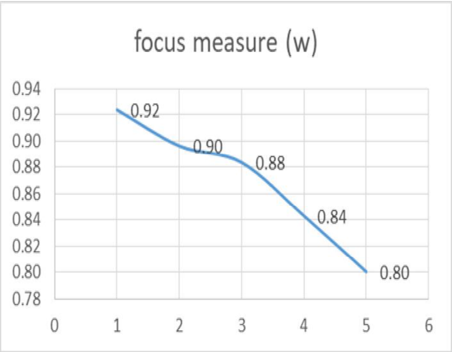


Figure 6. Graphical representation of Figure 2. (e)

The graphical representation in Figure 3, 4, 5, 6 show that the descending values represent the lower focus values as the distance of objects goes on increasing.

5. CONCLUSION

In this paper, the issue of depth estimation from single view image has addressed. Here, the focus measures of different objects in the single view image have calculated using Discrete Wavelet Transform. Two level Discrete Wavelet Transform has used to calculate focus measure. The conclusion has made that if a particular object is near, far from the camera, or from one another, the focus measure changes with respect to that. The results show that the object near to the camera has greater focus measure as compared to the object that is relatively farer. In addition, the entropies of different objects have computed as another evaluation metrics and the values relatively decreased as the distance of the objects increase from camera. The object detection from the image has done by Deep Learning.

REFERENCES

[1] Jae I Jung, Yo Sung Ho, "Depth Map Estimation from Single View Image using Object Classification based on Bayesian Learning", IEEE Explore, 2010.
https://doi.org/10.1109/3DTV.2010.5506603

[2] Harel Haim, Shay Elmalem, Raja Giryes, "Depth Estimation from a Single Image using Deep Learned Phase Coded Mask", IEEE Transactions on Computational Imaging, 2018.

[3] Tobias Koch, Lukas Liebel, Friedrich Fraundorfer, Marco Körner, "Evaluation of CNN-based Single-Image Depth Estimation Methods," May 2018.

[4] David Eigen, Christian Puhrsch, Rob Fergus, "Depth Map Prediction from a Single Image using a Multi-ScaleDeepNetwork", Dept. of Computer Science, Courant Institute, New York University, 2015.

[5] C M Sheela Rani, V VijayaKumar, B Sujatha, "An Efficient Block based Feature Level Image Fusion Technique using Wavelet Transform and Neural Network", International Journal of Computer Applications (0975 – 8887), Volume 52– No.12, August 2012.
https://doi.org/10.5120/8253-1780

[6] Jaroslav Kautsky, Jan Flusser, Barbara Zitova, Stanislava Simberova, "A new wavelet-based measure of image focus", Pattern Recognition Letters 23 (2002) 1785–1794, Elsevier Science Direct Journal.

[7] Xiaobai Ma, Zenglin Geng, Zhi Bie, "Depth Estimation from Single Image Using CNN-Residual Network".

[8] Jyoti B Kulkarni and Dr Manna Sheela Rani Chetty, "Comparative Analysis of Image Restoration Techniques by measuring Focus Parameter using Discrete Wavelet Transform", Journal of Advanced Research in Dynamical & Control Systems, Vol. 10, Special Issue, 2018.

[9] Ashutosh Saxena, Sung H Chung, and Andrew Y Ng, "Learning Depth from Single Monocular Images", Computer Science Department, Stanford University, 2006.

[10] Beyang Liu, Stephen Gould, Daphne Koller "Single Image Depth Estimation From Predicted Semantic Labels", 2010.
https://doi.org/10.1109/CVPR.2010.5539823

[11] Z. Said, K Sundarajand, M N A Wahab, "Depth Estimation for a Mobile Platform Using Monocular Vision", Elsevier, Science Direct Journal, Procedia , IRIS 2012.

[12] Miaomiao Liu, Mathieu Salzmann, Xuming He, "Discrete-Continuous Depth Estimation from a Single Image", CVPR, 2015, IEEE open access online.

[13] V Aslantas, "A depth estimation algorithm with a single image", OCIS codes: (100.2000) Digital Image Processing; (150.6910) Three Dimensional Sensing, Vol. 15, No. 8, published 10 Apr 2007.
https://doi.org/10.1364/OE.15.005024

[14] Che Chun Su, Lawrence K Cormack, Alan C Bovik, "Bayesian depth estimation from monocular natural images", Journal of Vision (2017) 17(5):22, 1–29

[15] Khushboo Khurana, Reetu Awasthi, "Techniques for Object Recognition in Images and Multi-Object Detection", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 2, Issue 4, April 2013, ISSN: 2278 – 1323.

[16] Huihui Bai, Mengmeng Zhang, Meiqin Liu, Anhong Wang and Yao Zhao, "Depth Image Coding Using Entropy-Based Adaptive Measurement Allocation", ISSN 1099-4300, Open access, 6590-6601.
2014.

[17] Suraj Pramod Patil, "Techniques and Methods for Detection and Tracking of Moving Object in a Video", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 5, May 2016.

[18] Prajakta patil, C M Sheela Rani, Meenakshi Arya, "A hybrid Pan-sharpening approach using maximum local extrema", International Journal of Electrical and Computer Engineering (IJECE), Vol. 9, No. 1, February 2019, pp. 255~263, ISSN: 2088-8708, DOI: 10.11591/ijece.v9i1.pp255-263.

[19] Yaoxin Li, Keyuan Qian, Tao Huang and Jingkun Zhou, "Depth Estimation from Monocular Image and Coarse Depth Points based on Conditional GAN", MATEC Web of Conferences 175, (2018), IFCAE-IOT 2018.

[20] Muthukrishnan R. and M Radha, "EDGE DETECTION TECHNIQUES FOR IMAGE SEGMENTATION", International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 6, Dec 2011.
https://doi.org/10.5121/ijcsit.2011.3620

[21] Sandeep Pande, Manna Sheela Rani Chetty, "Position Invariant Spline Curve Based Image Retrieval Using Control Points", International Journal of Intelligent Engineering and Systems, Volume 12, January 8, 2019.
https://doi.org/10.22266/ijies2019.0831.17

-