

## Depth Estimation from Defocused Images: A Survey

Jyoti Kulkarni and Dr. Manna Sheela Rani Chetty

Department Of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,  
Guntur, 522502 Andhra Pradesh, India

**Abstract:** An important step in 3D data generation is the generation of depth map. Depth map is a black and white image which has exactly the same size of the original captured 2D image that indicates the relative distance of each pixel from the observer to the objects in the real world. This study presents a survey of depth perception from defocused or blurs images as well as image from motion. The change of distance of the object from the camera has direct relation with the amount of blurring of object in the image. The amount of blurring will be calculated with a comparison in front of the camera directly and can be seen with the changes at gray level around the edges of objects.

**Key words:** Depth map, defocusing or blurring, gray level image, depth perception, triangulation, stereo vision

### INTRODUCTION

There are many evolutions, from black-and-white television to colored television and high-definition television is now the most popular television in the market from the birth of television. The development of 3D television system has grabbed more and more attention after that.

Depth estimation or extraction is a measure of the distance of, ideally each point of the seen scene to the Stereo vision research field. In absence of the information for absolute depth measurements such as motion, defocus, binocular disparity, the absolute distance between the observer or camera and the objects in the scene or image cannot be measured. The depth estimation is very important aspect in various applications. So, when the distance of the cameras and some camera parameters when changed, then the depth estimation is useful to calculate exact distance of camera from image in both the cases as discussed in Tiwari (2010). The information of shading, edges and junctions may provide 3-D Model but it will not give the exact scale of the space.

There are many ways to retrieve 3D information. The most common and straightforward method is to use “active device”, i.e., active cameras which can capture the original image of the scene and detect the depth of each of pixels of the scene simultaneously. The examples are infrared cameras, sonar cameras, etc. However, most of the active devices need additional sensor to obtain depth information, so, they are generally more expensive

compared to common cameras. Whereas, since, they have specific designed sensors most of the times the detected depth maps are more accurate. In addition to active device, the “passive device” can also be used which requires some post processing after the acquisition of the scene of interest such as binocular cameras and common cameras are with focus tuning functionality. Compared to active devices, passive devices are much cheaper and without the limitation of sensor energy issue because of the active devices, passive devices can generate a higher resolution of depth map.

In the area of passive methods, single common camera is the most preferred camera, since, it can be acquired from general markets and it is easy to use for most of the people. However, the lack of information from additional sensors like the active devices or different angles of view from additional camera, the binocular cameras make the calculation of depth map from a single image very difficult. To compensate this defect, a series of images of a scene can be captured on different focus planes to gather more information than just a single image alone. The result in the previous works turns out to be better than just a single image alone and many algorithms are even suitable for future hardware implementation for real time application.

Also, depth perception from stereo vision is based on the triangulation principle (Anonymous, 2012). Two cameras can be used with projective optics and arrange them side by side such that their view fields overlap at the desired object distance. By taking a picture with each

camera, the scene can be captured from two different viewpoints. The above scope considers focused images. Depth from defocus is the challenge in perception of accurate depth as all the objects in the scene are not focused always (Beck and Kubota, 2013). The camera positions, light intensity and focal lengths of camera may vary which yield in blur images. So, in this study, survey on depth from defocus is done from the above motivation.

**Literature review:** In a natural image, objects on the focus plane are sharper than those out of focus due to the depth of the objects. Those works or algorithms that analyze the sharpness or blur of an object as their information of depth are categorized as “Depth From Focusing” or “Depth From Defocusing” algorithms.

Either “Depth From Focusing” or “Depth From Defocusing” algorithm needs focus measure which finds out the sharpness of an object to determine its depth. There are plenty of ways to measure the sharpness, either in spatial domain or frequency domain. All of them are just different aspects of some sort of high pass filters in different point of view.

Rajabzadeh *et al.* (2010) study were introduced a new method which used similar characteristics of defocus blur. It was found out in study of Rajabzadeh *et al.* (2010) that the change in object distance from the camera has direct relation with the amount of defocus blur in the image. The proposed method (Rajabzadeh *et al.*, 2010) compared to conventional defocus and other methods was shown to be a blind method, i.e., no focus state object image is required. Another advantage of this method (Rajabzadeh *et al.*, 2010) is that, it is independent from the 3-D attributes of objects or the scene. The complexity is also less compared to similar methods..

The researchers Swain *et al.* (1994) improved the accuracy of depth from defocus using fuzzy logic technique. Fuzzy logic in Rajabzadeh *et al.* (2010) is combined with a depth from Defocus technique to correct for uncertainty and imprecision in depth estimation.

Swain *et al.* (1994) have given two inputs to fuzzy algorithm are focus quality and focal error. Focus quality is a measure of the amount of defocus in the image. Experimental results in Swain *et al.* (1994) show that fuzzy logic significantly improves depth estimation compared to nonfuzzy depth from defocus.

Yang and Schonfeld (2010) presented a novel method for virtual focus and object depth estimation from defocused video captured by a moving camera. They used the term virtual focus to refer to a new approach in

Yang and Schonfeld (2010) for producing in-focus image sequences by processing blurred videos captured by out-of-focus cameras. The method used in study of Yang and Schonfeld (2010) relies on the concept of Depth-from-Defocus (DFD). The researchers explored several blur models which can be used to recover arbitrary transfer functions.

Kim *et al.* (2012) research work uses a novel approach to depth estimation using a Multiple Color-filter Aperture (MCA) camera and its application to multifocusing. An image acquired by the MCA camera in Kim *et al.* (2012) contains spatially varying misalignment among RGB color channels, where the direction and length of the misalignment is a function of the distance of an object from the plane of focus. Therefore, if the misalignment is estimated from the MCA output image in Kim *et al.* (2012) multifocusing and depth estimation become possible using a set of image processing algorithms.

The MCA camera in Kim *et al.* (2012) with proposed image processing algorithms enables automatic, computationally efficient multifocusing using a three-step process that involves: image segmentation for classifying clusters, color shift modelbased registration and fusion, and image restoration. More specifically, an image acquired by the MCA camera is first segmented into multiple clusters each of which has the uniform color and then the corresponding rectangular region is generated that encloses each cluster. The MCA camera significantly enhances the visual quality of an image containing multiple objects of different distances (Kim *et al.*, 2012).

## MATERIALS AND METHODS

In the study of Miled *et al.* (2009) they presented a new method for addressing robust depth estimation from a stereo pair under varying illumination conditions. First, a spatially varying multiplicative model is developed in Miled *et al.* (2009) to account for brightness changes induced between left and right views. The depth estimation problem, based on this model (Miled *et al.*, 2009) is then formulated as a constrained optimization problem in which an appropriate convex objective function is minimized under various convex constraints modeling prior knowledge and observed information.

The resulting multiconstrained optimization problem in Miled *et al.* (2009) is finally solved via a parallel block iterative algorithm which offers great flexibility in the incorporation of several constraints. Experimental results

on both synthetic and real stereo pairs, Miled *et al.* (2009) demonstrate the good performance of our method to efficiently recover depth and illumination variation fields, simultaneously. The researchers in study of Miled *et al.* (2009) have proposed a convex programming approach for the problem of disparity estimation in the presence of illumination variations.

In the study written by Paramanand and Rajagopalan (2012), the objective was to recover the 3D structure of a scene from motion blur or optical defocus. In the proposed approach, the difference of blur between two observations is used as a cue for recovering depth, within a recursive state estimation framework.

For motion blur in Paramanand and Rajagopalan (2012) the researchers used an unblurred (focused) blurred image pair. Since, the relationship between the observation and the scale factor of the point spread function associated with the depth at a point is nonlinear, they proposed and developed a formulation of unscented Kalman filter for depth estimation.

Depth estimation from a single image is a challenging problem in computer vision research. By analyzing the defocus cues produced by the depth of field of lens, the information of depth can be determined. Chan *et al.* (2011) in their research employed reverse heat equation which is simple and effective for this analysis.

Because the depth map is required to be smooth in many applications, a mean shift segmentation and graph cut based method is proposed in to infer the depth map of the scene. The confidence of depth estimation is incorporated into the energy function of graph cut to preserve details of the depth map. Experimental results show that the proposed method can produce a good depth map even from a single image.

## RESULTS AND DISCUSSION

**ALARM:** Yang and Schonfeld (2010) original image blurred image, focused image reconstruction using Restoration from Magnitude (RFM), focused image reconstruction using blur function magnitude estimation, focused image reconstruction using blur function magnitude and phase estimation.

The researchers in study of Yang and Schonfeld (2010) test PTF estimation with a sequence ALARM as shown in Fig. 1a shows the first frame and Fig. 1b shows the blurred first frame as a result of a synthetic OTF consisting of a Gaussian MTF and an arbitrary PTF. Figure 1c shows the reconstruction result using Restoration from Magnitude (RFM). It is a technique (Yang and Schonfeld, 2010) based on projection onto convex set (POCS) while the two convex sets are the set of space-limited functions and the set of all functions that

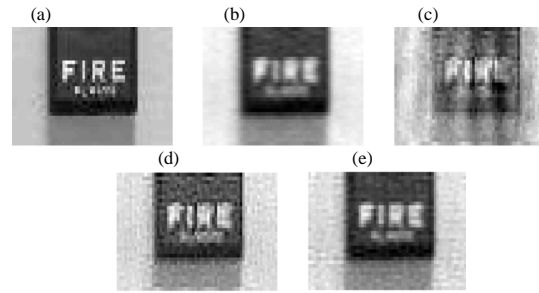


Fig. 1: Comparison of focused image reconstruction methods for video

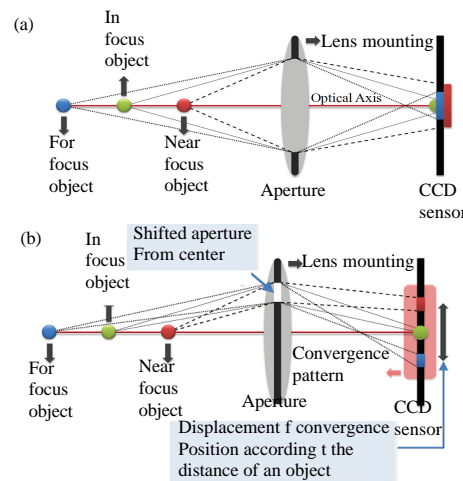


Fig. 2: a) Two different single-aperture models (Kim *et al.*, 2012) and b) Aperture is aligned on the optical axis of the camera

Aperture is shifted away from the optical axis which produces various convergence positions according to the distance of an object (Fig. 2).

As shown in Fig. 3b, the single-eye range method mostly failed to estimate depth of the defocused scene background by Kim *et al.* (2012). On the other hand, the proposed depth estimation method produces the best result as shown in Fig. 3c.

From Fig. 4, it is noticed that local methods give noisy results and are very sensitive to illumination changes (Miled *et al.*, 2009) while the SGM algorithm and the proposed method in allow to obtain a smooth disparity map with sharp depth discontinuities. Both GC and MRF algorithms combined with a histogram and rank transform, respectively by Miled *et al.* (2009) also show good performance for this stereo pair due to the presence of large homogeneous regions.

In Fig. 5, the method from study of Paramanand and Rajagopalan (2012) has the additional advantage that, it can be applied to DFD without constraining the PSF to be Gaussian. By enforcing sparsity constraint, the

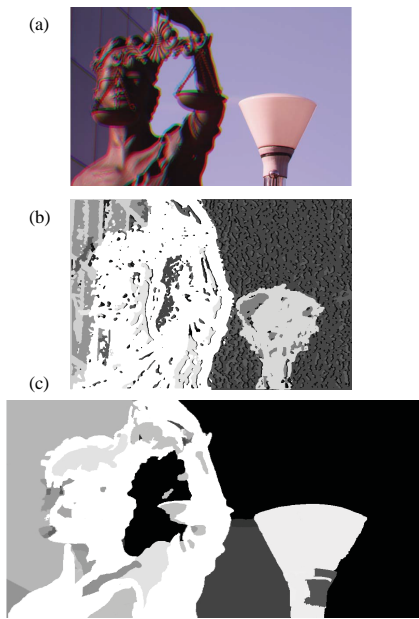


Fig. 3: Comparison of depth estimation of three different algorithms (Kim *et al.*, 2012); a) Input image acquired by the MCA camera; b) Result of depth estimation using the single-eye range method and c) Result of the proposed depth estimation algorithm

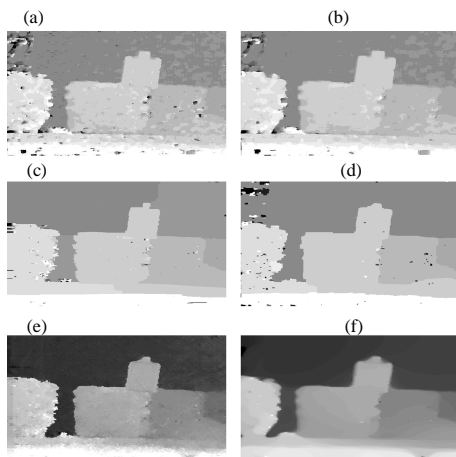


Fig. 4: Results on the Shrub stereo pair. Estimated disparity map using (Miled *et al.*, 2009); a) SD with affine illumination; b) Normalized cross correlation; c) GC with histogram transform; d) MRF with rank transform; e) SGM algorithm and f) Proposed method researchers also addressed the problem of depth estimation when an observation undergoes simultaneous motion and optical blur

have a Fourier transform magnitude equal to a prescribed function. Figure 1d shows the reconstruction result using only proposed OTF magnitude estimation (Yang and Schonfeld, 2010) and PTF is considered to be zero. Figure 1e shows the reconstruction result using both proposed OTF magnitude estimation and phase estimation (Yang and Schonfeld, 2010). It can be seen that the restoration including PTF estimation performs better than the restoration without phase and restoration using RFM (Yang and Schonfeld, 2010).

Aperture is shifted away from the optical axis which produces various convergence positions according to the distance of an object (Fig. 2).

As shown in Fig. 3b, the single-eye range method mostly failed to estimate depth of the defocused scene background by Kim *et al.* (2012). On the other hand, the proposed depth estimation method produces the best result as shown in Fig. 3c.

From Fig. 4, it is noticed that local methods give noisy results and are very sensitive to illumination changes (Miled *et al.*, 2009) while the SGM algorithm and the proposed method in allow to obtain a smooth disparity map with sharp depth discontinuities. Both GC and MRF algorithms combined with a histogram and rank transform, respectively by Miled *et al.* (2009) also show good performance for this stereo pair due to the presence of large homogeneous regions.

In Fig. 5, the method from study of Paramanand and Rajagopalan (2012) has the additional advantage that, it can be applied to DFD without constraining the PSF to be Gaussian. By enforcing sparsity constraint, the researchers also addressed the problem of depth estimation when an observation undergoes simultaneous motion and optical blur.

The experimental results in the study in Fig. 6 show that this depth estimation technique is reliable. In this study, researchers illustrate passive depth estimation method to extract depth map from a single image captured in a narrow depth of field setting. This method employs reverse heat equation for pre-process and use the proposed hierarchy mean shift segmentation and graph cut with a confidence to infer the depth map.

From above literature survey, it has been observed that the defocused or blur image and the Depth estimated from it, is a big challenge. But if focal length and focal error of the lens are known and the intensity of light is also known, then by various approaches, depth map can be estimated to reconstruct 3D view from that.

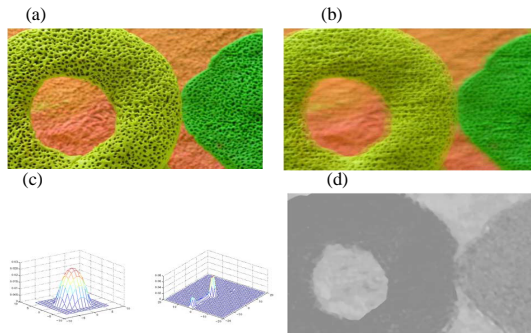


Fig. 5: a) Reference image; b) Simultaneously blurred observation; c) Known optical blur; d) Estimated motion kernel and e) Estimated depth by proposed method (Paramanand and Rajagopalan, 2012)

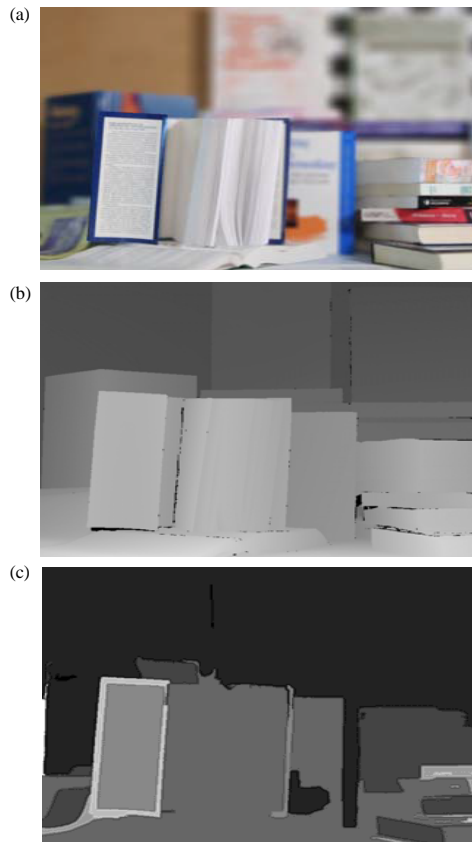


Fig. 6: a) The images of Middlebury Stereo datasets; b) The ground truth depth map and c) The depth map estimated by proposed method

## CONCLUSION

Focus quality is a measure of the amount of defocus in the image. Depth from defocus or depth from motion blur is a challenge.

## REFERENCES

- Anonymous, 2012. Obtaining depth information from stereo images. IDS Imaging Development Systems, Obersulm, Germany. [https://en.ids-imaging.com/tl\\_files/downloads/whitepaper/IDS\\_Whitepaper\\_3\\_D\\_Stereo\\_Vision.pdf](https://en.ids-imaging.com/tl_files/downloads/whitepaper/IDS_Whitepaper_3_D_Stereo_Vision.pdf)
- Beck, L. and T. Kubota, 2013. Accuracy of visual depth perception in an open field. *J. Math. Biol. Sci.*, 1: 1-17.
- Chan, P.P., B.Z. Jing, W.W. Ng and D.S. Yeung, 2011. Depth estimation from a single image using defocus cues. *Proceedings of the 2011 International Conference on Machine Learning and Cybernetics (ICMLC'11) Vol. 4, July 10-13, 2011, July 10-13, 2011*, pp: 1732-1738.
- Kim, S., E. Lee, M.H. Hayes and J. Paik, 2012. Multifocusing and depth estimation using a color shift model-based computational camera. *IEEE. Trans. Image Process.*, 21: 4152-4166.
- Miled, W., J.C. Pesquet and M. Parent, 2009. A convex optimization approach for depth estimation under illumination variation. *IEEE. Trans. Image Process.*, 18: 813-830.
- Paramanand, C. and A.N. Rajagopalan, 2012. Depth from motion and optical blur with an unscented Kalman filter. *IEEE. Trans. Image Process.*, 21: 2798-2811.
- Rajabzadeh, T., A. Vahedian and H. Pourreza, 2010. Static object depth estimation using defocus blur levels features. *Proceedings of the 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM'10), September 23-25, 2010, IEEE, Chengdu, China*, ISBN:978-1-4244-3708-5, pp: 1-4.
- Swain, C., A. Peters and K. Kawamura, 1994. Depth estimation from image defocus using fuzzy logic. *Proceedings of the 3rd IEEE Conference on Fuzzy Systems and the World Congress on Computational Intelligence, June 26-29, 1994, IEEE, Orlando, Florida*, ISBN:0-7803-1896-X, pp: 94-99.
- Tiwari, K.K., 2010. A polynomial based depth estimation from a single image. Master Thesis, Cornell University, Ithaca, New York, USA.
- Yang, J. and D. Schonfeld, 2010. Virtual focus and depth estimation from defocused video sequences. *IEEE. Trans. Image Process.*, 19: 668-679.