Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Occlusion Robust Depth Estimation from Binocular Stereo Video

Ahmed Bin Zubair [1], Dr. Maznu Shaik [2]

Vidya Jyothi Institute of Technology, Telangana-500075

Article Info
Page Number: 5451 - 5463
Publication Issue:
Vol 71 No. 4 (2022)

Article History
Article Received*: 25 March 2022*
Revised*: 30 April 2022*
Accepted*: 15 June 2022*
Publication*: 19 August 2022*

Abstract

Binocular stereo vision is used for the reconstruction of depth information and it is sensitive to senses with significant occlusion. Light field imaging is an emerging computational photography method to passive depth perception by storing many angular perspectives in a single exposure. It provides a novel solution. This work investigates binocular SV and LF imaging in order to develop the binocular LF imaging system. Based on geometrical optics theory, imaging theory is created by modelling the imaging process and examining disparity properties. Using the proposed method, we can find a depth estimation for the stereo video. Video is nothing but the sequence of frames. For the video, the structure is the same as the depth map calculation applied to the image.

Keywords: Binocular stereo vision, Light field imaging (LF), stereo vision (SV),

I. Introduction

Calculating image distance is a crucial task. The goal of image classification research is to identify visual representations that could be automatically used to categories photos into a small number of groups. We frequently measure distances in our daily lives, such as when determining the height of things or the distance to the ground.

It is the ability to determine the distance between two objects only based on their relative positions in the two eyes.Stereovision uses two cameras to view the same thing. The2 cameras are separated known as baseline, the distance between which is considered to be precise. At a same time, two images are captured by the two cameras. The two photos are evaluated to determine the differences.

Depth Estimation is the process of calculating the distance between each pixel and the camera. To determine the link between the images, traditional methods utilize multi-view geometry.Depth is

calculated from stereo images.Aim of depth estimation is to recover the 3-D shape and appearance of objects in imaging by obtaining a representation of the spatial structure of a scene.An occlusion is the region between two overlapping objects that move in opposite directions. Occlusion means that any blur video or some noise is there.

Depth estimation is most typically done in computer vision and robotics using stereo vision, which uses images from two cameras to triangulate and calculate distances

In computer visionstereo matching attracts great interest and has different usesin the area of depth estimation. In Stereo algorithms, according to the classification, presents 4 steps is the computation of cost, cost aggregation, disparity computation, and disparity refinement.

First, input the binocular LF image pair, then analysis of the ray tracing disparity in that if any noise or blur part is there it detects after that we have to get a disparity map means the output. It is shown below.

II. Literature survey

There is minimal work done on the performance of efficiency. In that paper, they discuss that dense classification of two-frame stereo methods. The type is created to access the various components and design decisions made in stereo algorithms separately.[1] Using this classification, we area classify the stereo methods and present an experiment to find the multiple variants. We construct a flexible C++ to establish a software platform and collect a data set for the easy evaluation that C++ enables the find the single induvial component, which is easily extended to include a new algorithm. We produce different new multi-frame stereo sets of data with ground truth .and make both code and data set available on the web. Moreover; we compare many of the best stereo algorithms.

In this paper, we see the classification of the dense two-frame algorithm. We use this classification to highlight the feature of the stereo algorithm and component in isolation. We are implementing a set of stereo matching algorithms and creating a harmless test that combines these to change the algorithm in a controlled way on exciting data set to test the algorithm's performance.

[2] The research on the existing disparity map algorithm. In 2002 It concentrated on four main stages of processing. In taxonomy, scharstain &szeliski proposed an evaluation of a dense two-frame stereo algorithm to help the researchers develop their stereo matching algorithm. For every stage of

processing, a summary of the existing algorithm is created, which are also providedthe survey notes the implementation of prior software-based and hardware-based algorithm.

The software-based central processing module is implemented using the central processingunit. The hardware-based performance requires more than one processor for the processing module, like a graphical processing unit. The stereo vision disparity mapping is familiar with the state of the art and is a time-consuming task. In this survey, there is software development of the stereo component algorithm, and hardware-based implementation is there to help its purpose.

The stereo correspondence matching cot combination is one of the oldest and most popular combination methods. However, practical and efficient, the matching cost is typically averaged over a locally defined support zone. That is clearly only optimal locally, and the computing difficulty of the entire kernel implementation is frequently proportional to the size of the region. The aggregation of the cost problem is revisited in this study, and non-local solutions are proposed. Based on a pixel, the corresponding cost values are combined favorably.[3]

Similarly, to preserve a depth edge on tree structure derived from stereo image pair. The nodes of that tree are image pixels at the border—edges formed by the closest neighboring pixel. We can find a similarity between any two pixels by their minimum distance on the tree. The suggested solution is non-local since each node in the tree receives assistance from the other nodes on the standard benchmark. The proposed non-local solution assumes all predicted local cost aggregation approaches. With extremely low computational complexity, it is of great use. Characteristic box filter is one of the forest local cost combination methods, but at the time of edge, depth is a blur**.**
we jointly proposed an inclined plane model that recovered an image segmentation. From a static, we estimate a sense of dense depth and boundary labels Two frames of a stereo pair captured from a moving car are provided. Towards this aim for our SLIE-like objective, a new optimization algorithm is proposed that preserves a connection. In the form of border length, image segments and shape regularization are applied.[4] We demonstrate our technique's efficiency in challenging the KITTI dataset, showing state-of-the-art results can achieve a faster recovery of dense depth and motion from the stereo video, assuming that the scene is constant. Our approach is current inclined plane methods by order of magnitude. Currently, we are checking the parallel implementation of our approach that can run in real-time.

In stereo matching, two techniques are considered: the cost-filtering method and energy minimization algorithms. However, it tends to fail in the occluded region; we get a good result for

the cost-filtering approaches. This paper aims to improve the stereo matching result.[5-8] By using the filtering method, we show that optimization of the fully connected model can be solved from this algorithm. The contribution of this paper is two-step global optimization.

An efficient stereo matching algorithm that imposes adaptive smoothness restrictions utilizing texture and edge information. First, we find the non-texture and edge regions on the input image income flat pixel values. Second, we identify denoised edges that with high probability correlate to depth discontinuities by combing two edges maps from the input image and a pre-estimated disparity map. The proposed algorithm significantly outperforms the conventical method based on these adaptive smoothness limitations, it delivers cutting-edge performance on the Middlebury stereo benchmark. On the denoised edges, it adds a significant penalty for even slight variations in nearby descriptions. [6] The new NIDE algorithm score first in the Middlebury stereo benchmark according to experimental data, which shows it gelati outperforms the traditional stereo approaches.

III.Proposed Method

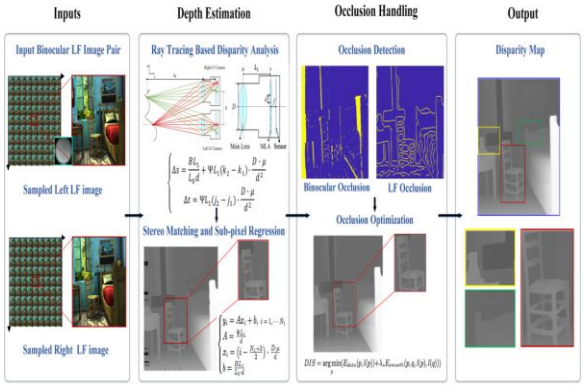
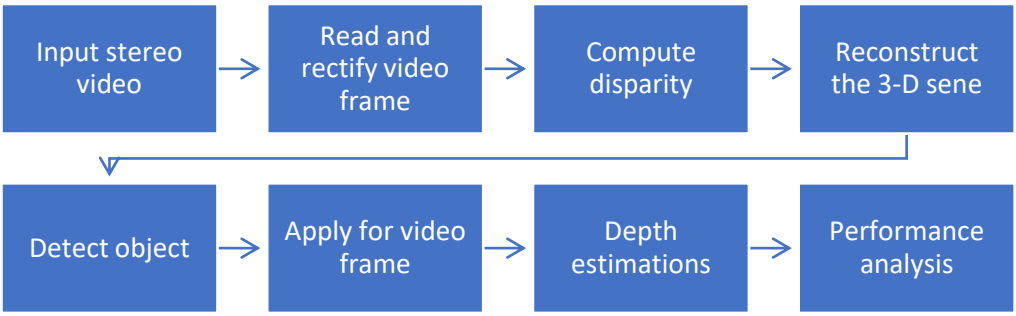


Fig.:1 The Framework of The Proposed Algorithm.



Input Stereo Video

A stereo camera features two or more lenses, each with its own image sensor or film frame. Here we considered for binocular images means right side and left side view of a same image. ‘Vision.VideoFileReader’this command is used for loading both left and right sequences of same video from left side and right side.

Both videos of light field.

Read and Rectify video frame

If we want to compute a disparity to reconstruct a 3D scene, the camera's frame from left and right should have been correct. The rectified images have a horizontal epipolar line and a row aligned. That simplifier disparity computation reduces the space in search for the point matching one dimension. Fixed images can be combined with the anaglyph that is seen in 3D with stereo red cyan glasses.

Compute Disparity

In a corrected stereo image, any pair of related points is placed on the same pixel row. Determine the distance between each pixel in the left image and its equivalent in the right image. The disparity is the distance between the camera and the related world point, and it is proportional to it.

Reconstruct the 3-D Scene

Recreate the 3-D world coordinates of the disparity map's points corresponding to each pixel. It consists of three coordinates: x, y, and z. Z represents spatial variances. Make a point cloud viewer that streams.

Detect Object

To detect a people, use the |vision. People Detector| system (Vision.ObjectDetector). Make a people detector object fora speedLimit the low object size.

Detect Object from Video

Video is nothing but sequence of frames. Detecting object from image can easily then converted to detecting object from video.

Depth estimation

Find the distance between the persons from the camera and the 3D world coordinate of the centroid for each overserved person also the distance between centroid and the camera in meters. Determine the distances in meters from the camera.

Occlusion -Robust Disparity Estimation

Inn that we see from the binocular -LF imaging system occlusion robust disparity estimation. The disparity map is generated on using the stereo matching and refined using linear regression across various baseline angular views. In order to remove the matching paradoxesand outliersfollowing that, occlusions induced by binocular and single LF imaging are discovered and addressed. Finally, using the global disparity optimization to obtain smooth surfaces and exact shape structures method is developed

1) Initial disparity estimation

The LF row data is decoded from a succession of angular views as the first input. In that study, we used the decoding and calibration approach to obtain left and right 4D LF data**.**

2) Occlusion detection

Occlusion means there is any blur video or noise is there. In the stereo matching the occlusion is a challenging problem. Without priors, occluded region matching ambiguities are frequently avoided n filled with disparity values of surrounding or background pixels. Due to the wide and narrow baseline for LF -binocular has strong occlusion over the scene's left edge and bounds.

Types of occlusions

1) The Binocular Occlusion
2) The LF Occlusion
3) Occlusion-handling Disparity Refinement

IV. Result

Following are the results obtained by proposed methodology and it shows the superior performance of our proposed model. Proposed model works on binocular videos for finding depth ofobjects.

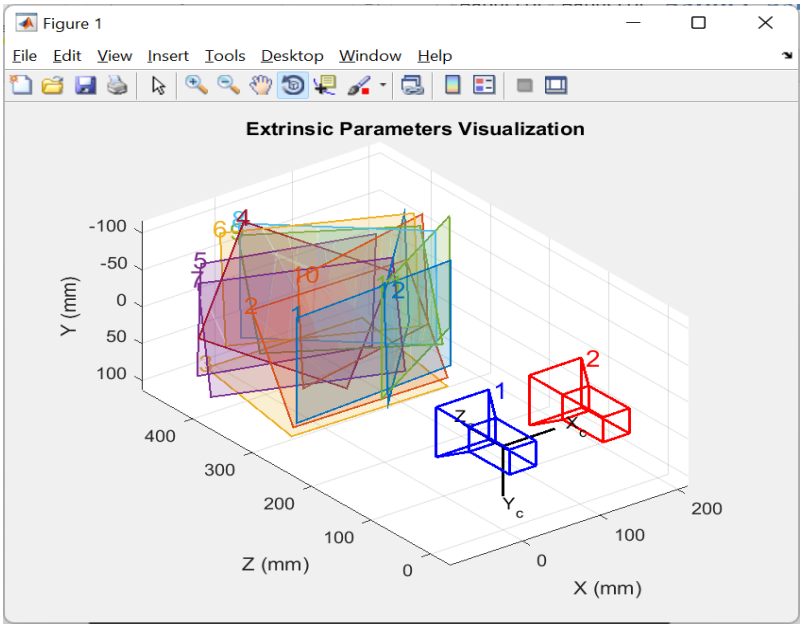


Fig:2 Extrinsic Parameter Visualization

Extrinsic parameter visualization is shown in above figure. Which shows the three dimensions X (in mm), Y (in mm) and Z (in mm). Two cameras are shown as 1 and 2 which indicates there are 2 binocular cameras.

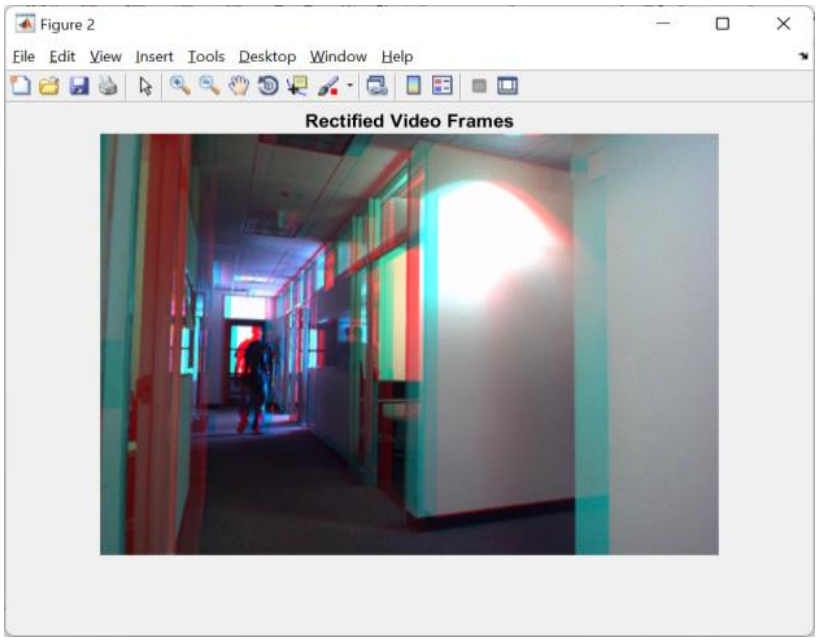


Fig:3 Rectified Video Frames

Video is rectified to get change is video framerate as there is need of slower video framerate to get detailed analysis visible.

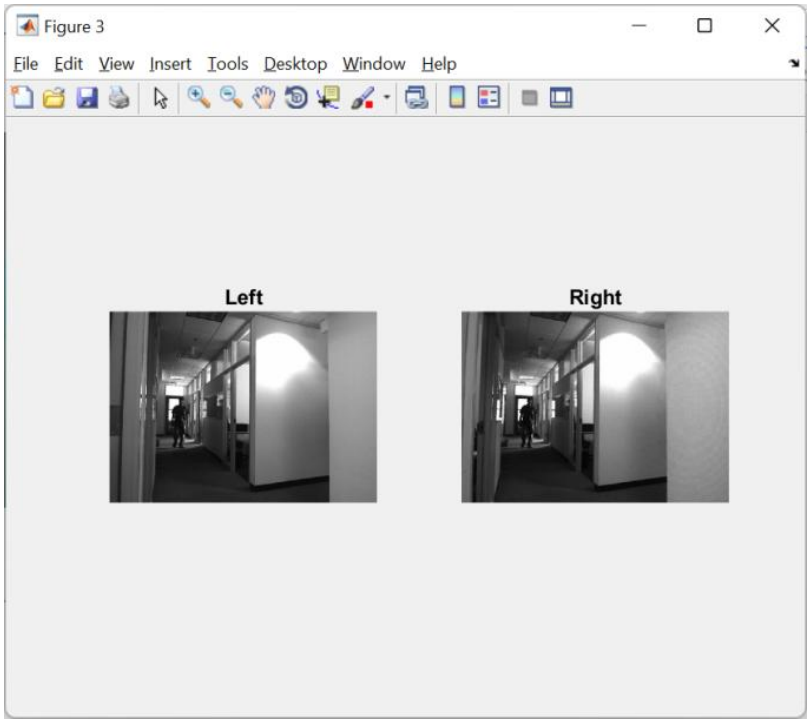


Fig:4 Left Stereo Image and Right Stereo Image

As the video is from binocular camera the resultant video has left side view and right-side view, both views are shown in above figure with gray scale image.

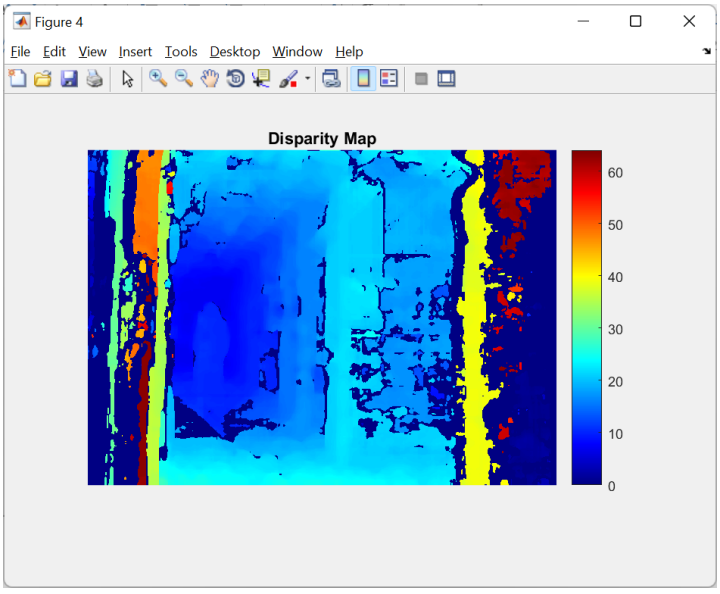


Fig:5Disparity Map generated

Disparity map is generated as shown in above figure. The disparity is the distance between the camera and the related world point, and it is proportional to it.

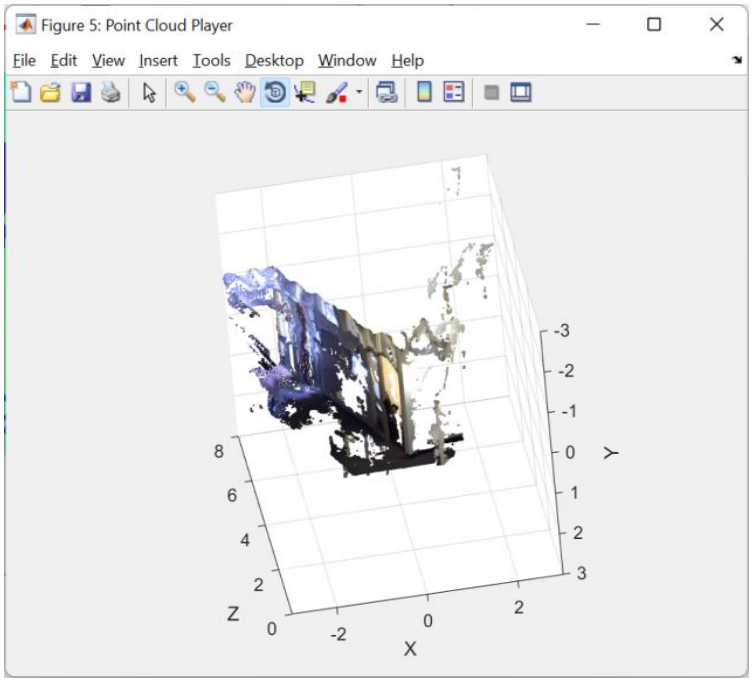


Fig:6 Point Cloud Player

Above figure shows point cloud player which indicates color, location and points of the objects in the frame of video.

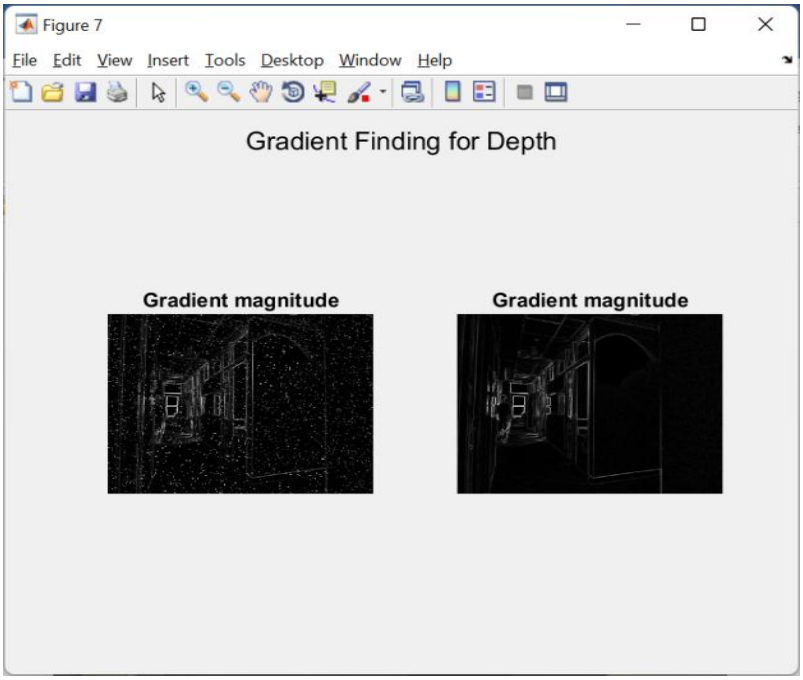


Fig. Gradient magnitude for both left and right stereo data

In above figure gradient magnitude is applied for both binocular videos and shown the outcomes.

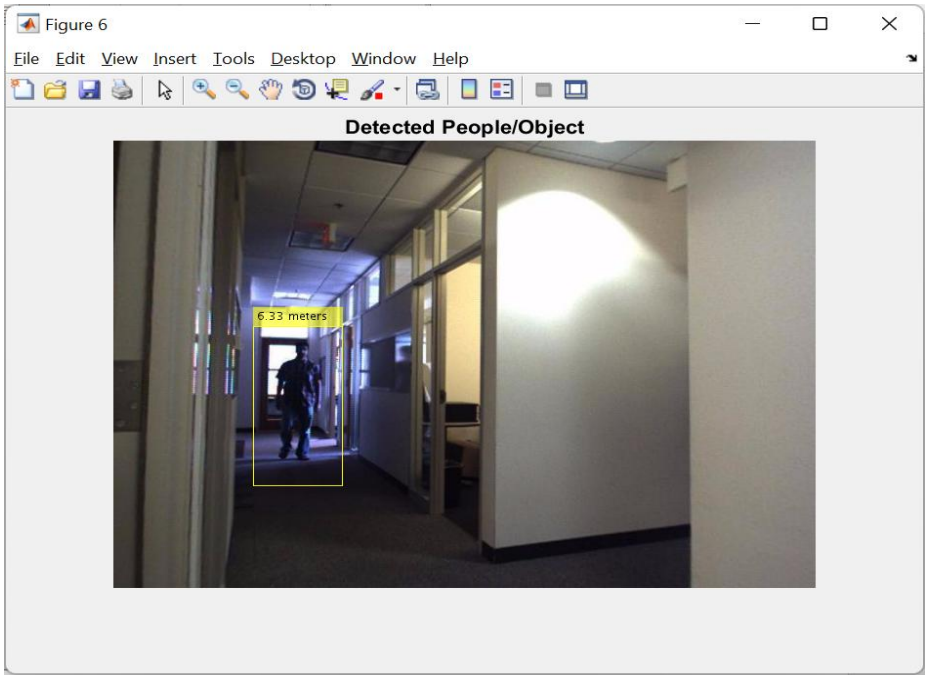


Fig:7 First frame details checked for object and distance

First, we are detecting objects/ people for finding its distance. Then the depth calculation will be applied for object to find its distance from camera.

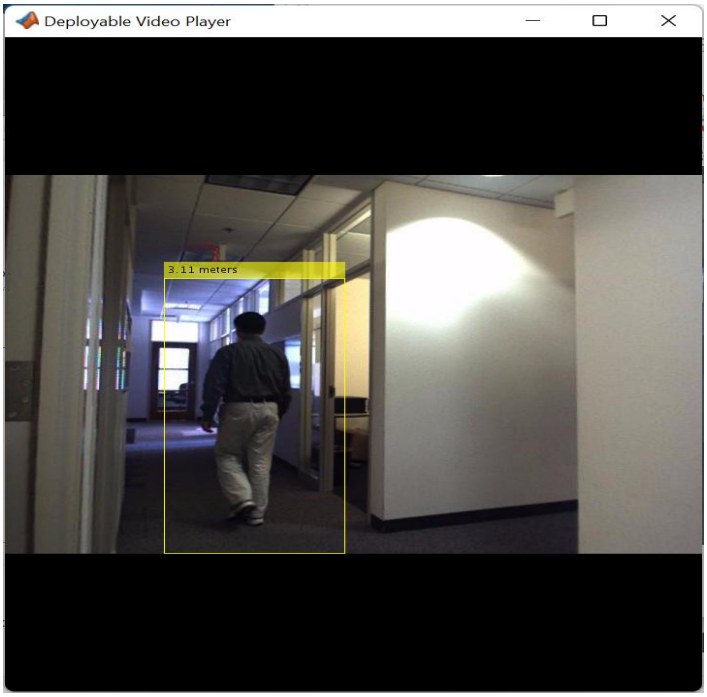


Fig:8 After verification, depth calculation started for video frames

Depth calculation is done by using proposed model which gives depth distance on bounding box for moving object.

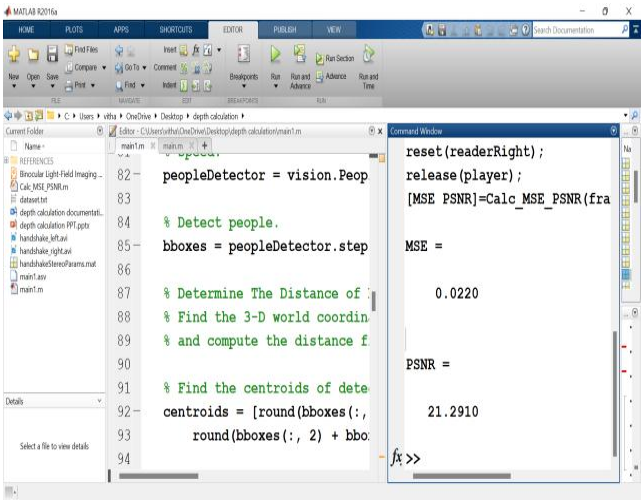


Fig. Performance Evaluation of depth estimation

Performance parameters such as MSE (Mean Square Error) and PSNR (Peak Signal to Noise Ratio) are calculated for this depth calculation by proposed work and it is found that MSE is very less for the proposed model and PSNR found high for proposed model which shows superior performance of proposed model.

V. Conclusion

In this project, depth estimation from the binocular stereo video is successfully found by using the proposed model. In this research, we use ray tracing to simulate binocular LF depth estimation approach that takes. The experimental result demonstrates that the proposed algorithm in the occlusion region preserves excellent shape information and is an incentive for texture changes. In that paper, we are seen that from the LF imaging, the depth estimation can improve more accurate occlusion detection like a multi occlude method. To prevent the distortion, we consider the small binocular baseline.

References

[1] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” International journal of computer vision (IJCV), vol. 47, no. 1-3, pp. 7–42, 2002.

[2] R. A. Hamzah and H. Ibrahim, “Literature survey on stereo vision disparity map algorithms,” Journal of Sensors, no. 2, pp. 1–23, 2016.

[3] Q. Yang, “A non-local cost aggregation method for stereo matching,” Computer Vision and Pattern Recognition (CVPR), pp. 1402–1409, 2012.

[4] K. Yamaguchi, D. McAllester, and R. Urtasun, “Efficient joint segmentation, occlusion labeling, stereo and flow estimation,” European Conference on Computer Vision (ECCV), pp. 756–771, 2014.

[5] M. G. Mozerov and J. van de Weijer, “Accurate stereo matching by two-step energy minimization,” IEEE Transactions on Image Processing (TIP), vol. 24, no. 3, pp. 1153–1163, 2015.

[6] K. R. Kim and C. S. Kim, “Adaptive smoothness constraints for efficient stereo matching using texture and edge information,” International Conference on Image Processing (ICIP), pp. 3429–3433, 2016.

[7] K. Zhang, Y. Fang, D. Min, L. Sun, S. Yang, S. Yan, and Q. Tian, “Cross-scale cost aggregation for stereo matching,” Computer Vision and Pattern Recognition (CVPR), pp. 1590–1597, 2014

[8] K. Zhang, Y. Fang, D. Min, L. Sun, S. Yang, and S. Yan, “Cross-scale cost aggregation for stereo matching,” IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), vol. 27, no. 5, pp. 965– 976, 2017.

[9] V. Q. Dinh, C. C. Pham, and J. W. Jeon, “Robust adaptive normalized cross-correlation for stereo matching cost computation,” IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), vol. 27, no. 7, pp. 1421–1434, 2017.

[10] L. Li, X. Yu, S. Zhang, X. Zhao, and L. Zhang, “3d cost aggregation with multiple minimum spanning trees for stereo matching,” Applied optics, vol. 56, no. 12, pp. 3411–3420, 2017.

[11] L. Sun, K. Chen, M. Song, D. Tao, G. Chen, and C. Chen, “Robust, efficient depth reconstruction with hierarchical confidence-based matching,” IEEE Transactions on Image Processing (TIP), vol. 26, no. 7, pp. 3331–3343, 2017.

[12] S. Kim, D. Min, S. Kim, and K. Sohn, “Feature augmentation for learning confidence measure in stereo matching,” IEEE Transactions on Image Processing (TIP), vol. 26, no. 12, pp. 6019–6033, 2017.

[13] C.-C. Yang, S.-K. Huang, K.-T. Shih, and H. H. Chen, “Analysis of disparity error for stereo autofocus,” IEEE Transactions on Image Processing (TIP), vol. 27, no. 4, pp. 1575–1585, 2018.

[14] J. Zbontar and Y. LeCun, “Computing the stereo matching cost with a convolutional neural network,” Computer Vision and Pattern Recognition (CVPR), pp. 1592–1599, 2015.

[15] W. Luo, A. G. Schwing, and R. Urtasun, “Efficient deep learning for stereo matching,” Computer Vision and Pattern Recognition (CVPR), pp. 5695–5703, 2016.

[16] F. Zhang and B. W. Wah, “Fundamental principles on learning new features for effective dense matching,” IEEE Transactions on Image Processing (TIP), vol. 27, no. 2, pp. 822–836, 2018.

[17] T. Taniai, Y. Matsushita, and T. Naemura, “Continuous Stereo Matching using Locally Shared Labels,” Computer Vision and Pattern Recognition (CVPR), pp. 1613–1620, 2014.

[18] T. Taniai, Y. Matsushita, Y. Sato, and T. Naemura, “Continuous 3d label stereo matching using local expansion moves,” IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. PP, no. 99, pp. 1–1, 2017.

[19] M. Levoy and P. Hanrahan, “Light field rendering,” Computer Graphics and Interactive Techniques, pp. 31–42, 1996.

[20] N. Ren, L. Marc, B. Mathieu, D. Gene, H. Mark, and P. Hanrahan, “Light field photography with a hand-held plenoptic camera,” Computer Science Technical Report CSTR, vol. 2, no. 11, pp. 1–11, 2005.