3D Scene Understanding using Binocular Images*

Ameya Panchpor
VIT Pune

Jyoti Madake
VIT Pune

***Abstract*—Computer Vision technologies in recent times have improved vastly and usage of such technologies has become an almost integral par our life, right from phone cameras used in daily life to sophisticated military equipment. But as we keep improving in this field, we are yet to beat the ability of our very own eyes. The ability of humans to quickly assess and analyse the surrounding environment through our visual sensors is still no match to our computer vision systems. Hence, understanding the surroundings has been a topic of much research work as its applications are tremendous. In this, we would be exploring the mathematical algorithms for depth measurement for 3D scene understanding to create 3D depth images from binocular images.**

***Index Terms*—computer vision, depth image, scene understanding, binocular images**

I. Introduction

In today's time, advancement in computer vision technologies has allowed humanity to understand the environment and the surrounding in ways never before imagined. We find applications of advanced computer vision systems in everything from smart traffic systems to moon rovers. While there has been an incredible improvement in this domain, the human visual system is still unparalleled in how sharply, accurately and comprehensively it can understand its surrounding environment, Hence, it has been a constant attempt at trying to imitate our human vision for computer vision tasks. Many complex tasks such as depth understanding of the surrounding for rover, plane understanding for augmented reality applications etc. involve obtaining three-dimensional coordinates of an object. This paper works on using mathematical algorithms along with different 3D point cloud mapping software tools to generate a 3D depth image from two 2D images.

II. Theory

The stimulus for working on 3D scene understanding comes from stereopsis, which is how human eyes perceive depth. Between the two right and left eyes, humans view the same object with some slight offset; this happens because of the effect called parallax, where the apparent position of an object appears differently depending on the angle of the line of sight. We can then use concepts of trigonometry to by having the values of distance between the eyes and displacement of the object position known.

In order to generate a 3D image with the use of such method, we would have to operate with two 2D images while also being aware of the caliberation (such as the height, width, displacement, etc.) of the cameras the images were taken from. Both the cameras kept at known distances from each other, are used for imitating our eyes for capturing a particular scene, giving it a some offset because of different line of sight. We then define, the distance of pixels for the same point in the two varying images with the formula: Here Xl and Xr are

$$d = x_l - x_r$$

the x coordinates of the same image point of the left and right images, respectively. If we assume, the location of the object to be P(x,y,z), then we can also find out the depth of the scene, by applying the proportionality formulae to find out the z-coordinate: here b is the displacement between both cameras,

$$x/z = x_l/f$$

$$(x - b)/z = x_r/f$$

and f is their focal length. With the combination both these equations, we can formulate z to be: Now, the most crucial

$$z = (b * f)/(x_l - x_r)$$

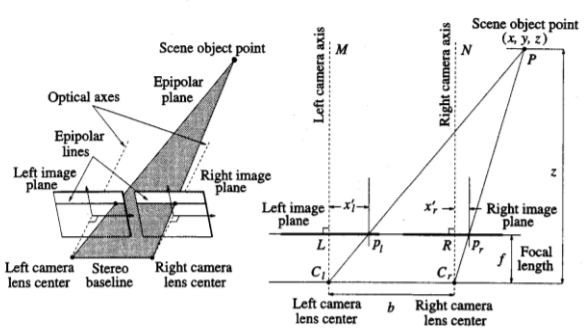
aspect of the project is to find the binocular disparity between each point of the two images. For this, the concept of region matching is of great use to find the binocular disparity. In this process, a window is created around each pixel in the left image, while a similar window is created in the right image around the pixels it is similar most with. In this, the heights of both camera are assumed as same, making the process of comparing the pixels easier. By this, we get a matrix of binocular disparity values which can then be used for a proper 3D reconstruction.

III. Methods And Data

A. Data Description

To access the two 2D images with known calibrations, we used the Middlebury Stereo dataset as it comprises of various high-resolution stereo images along with data regarding their calibration, complex geometry and ground-truth disparity. It acquires the ground-truth disparities using a novel technique

Identify applicable funding agency here. If none, delete this.



which uses structured lighting and does not need the calibration of the light projectors. The images taken had to undergo preprocessing and analysis to find the binocular disparity and then post-processed to get the final results.

B. Pre-processing The Images

The images had to be preprocessed properly, in order to ensure the models to run smoothly and the runtime was not high. The images were coloured and had an original size of 2864 x 1924, which were then converted to grayscale. Also, the images had to be resized to an eighth of the original dimension to keep the runtime low.

Fig. 1. Original image (Left)

C. Finding The Binocular Disparity

The Sum of Absolute Differences approach calculates pixel difference between the pixel window for the left and right images. The right image pixel having the lowest SAD value window is considered as the matching one, and the distance to that pixel is taken as the disparity.

D. Post Processing

Various post-processing procedures were required as the resultant plots still contained noise and discontinuities. In order to denoise the plots, methods like thresholding were used to check if pixel disparity exceeds a threshold (¿25).In case it does exceed the threshold, then calculate mode value in a large window around each pixel. In order to handle discontinuity, we calculate the average in a small window around each pixel and assign the average value if difference between disparity and value exceed a threshold, making the change between regions smoother

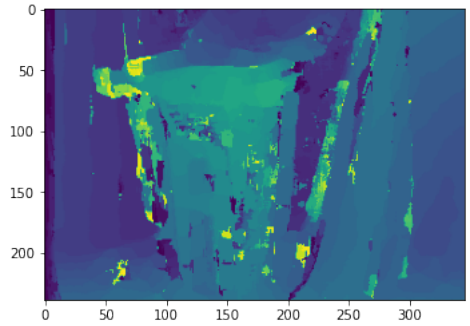


Fig. 2. Disparity Matrix Plot after SAD

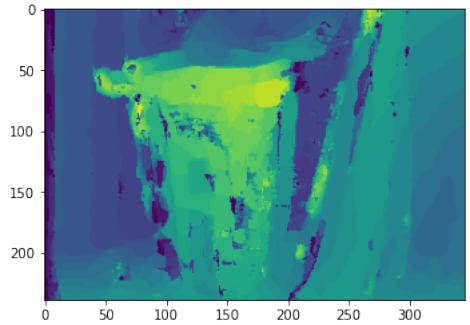


Fig. 3. Original image (Left)

IV. Results

The resulting coordinate values, as well as the RGB values from the original resized image, were collected in a .txt file. The information was entered into Cloud Compare, an open source point cloud reconstruction software.

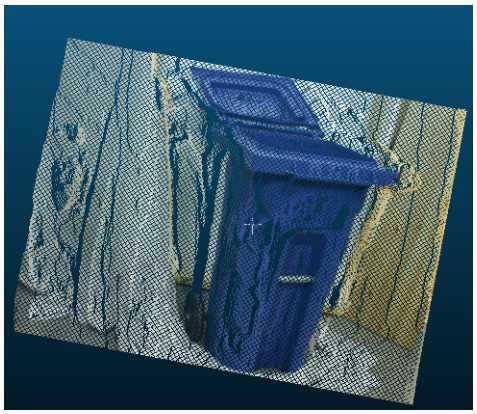


Fig. 4. Original image (Left)

References

[1] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)

[2] Tan M, Le QV (2019) Efficientnet: rethinking model scaling for convolutional neural networks. arXiv:190511946

[3] Chouhan V, Singh SK, Khamparia A, Gupta D, Tiwari P, Moreira C, Dama˘sevi˘cius R, de Albuquerque VHC (2020) A novel transfer learning based approach for pneumonia detection in chest x-ray images. Appl Sci 10(2):559
[4] Zhou B, Khosla A, Lapedriza A, Torralba A, Oliva A (2016) Places: an image database for deep scene understanding. arXiv: 161002055
[5] Sasaki T, Kinoshita K, Kishida S, Hirata Y, Yamada S (2012) Ensemble learning in systems of neural networks for detection of abnormal shadows from x-ray images of lungs. J Signal Process 16(4):343–346
[6] Zhou ZH, Jiang Y, Yang YB, Chen SF (2002) Lung cancer cell identification based on artificial neural network ensembles. Artif Intell Med 4(1):25–36
[7] Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society conference on computer vision and pattern recognition (CVPR), pp 886–893
[8] Loey M, Smarandache F, M Khalifa NE (2020) Within the lack of chest covid-19 x-ray dataset: a novel detection model based on gan and deep transfer learning. Symmetry 12(4):651
[9] Oliva A (2005) Gist of the scene. In: Neurobiology of attention Elsevier, pp 251–256
[10] Khan AI, Shah JL, Bhat MM (2020) Coronet: a deep neural network for detection and diagnosis of COVID-19 from chest X-ray ages. Comput Methods Progr Biomed 196:105581
[11] Panwar H, Gupta P, Siddiqui MK, Morales-Menendez R, Singh V (2020) Application of deep learning for fast detection of covid-19
[12] Civit-Masot J, Luna-Perej´on F, Dom´ınguez Morales M, Civit A (2020) Deep learning system for covid-19 diagnosis aid using x-ray pulmonary images. Appl Sci 10(13):4640
[13] Li C, Zhu G, Wu X, Wang Y (2018) False-positive reduction on lung nodules detection in chest radiographs by ensemble of convolutional neural networks. IEEE Access 6:16:060–16: 067
[14] Qin C, Yao D, Shi Y, Song Z (2018) Computer-aided detection in chest radiography based on artificial intelligence: a survey. Biomed Eng Online 17(1):113
[15] Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
[16] Narin A, Kaya C, Pamuk Z (2020) Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv:200310849
[17] Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic