# Aerial Robotics Kharagpur

Divyansh Sharma

***Abstract*— This project implements a depth detection algorithm using stereo images. By calculating disparities between corresponding points in left and right images, the algorithm determines depth information. The approach uses block matching to compute disparity matrices and can be applied in fields like autonomous driving, robotics, and 3D reconstruction.**

## I. INTRODUCTION

Depth perception is critical for understanding 3D scene geometry in computer vision systems. This project addresses the challenge of estimating depth from stereo images captured by horizontally displaced cameras. The initial implementation explored basic block matching with fixed window sizes, revealing sensitivity to illumination variations and repetitive patterns. Subsequent optimizations incorporated search range constraints and parameter tuning to improve computational efficiency and result quality. Key technical hurdles included managing computational complexity ($O(n^2)$ for image resolution) and handling ambiguous matches in low-texture regions.

## II. PROBLEM STATEMENT

The objective is to calculate depth information from stereo images by identifying disparities between corresponding points in two images captured from slightly different perspectives. The problem involves:

- Reading and preprocessing stereo image pairs.
- Computing disparity matrices using a block-matching approach.
- Visualizing the depth map.

Depth estimation from stereo images involves calculating disparities, which are the horizontal shifts between corresponding points in the left and right images. These disparities arise because of the slightly different perspectives of the two cameras. The relationship between disparity and depth is governed by the principle of triangulation.

The depth Z of a point in the scene can be calculated using the formula:

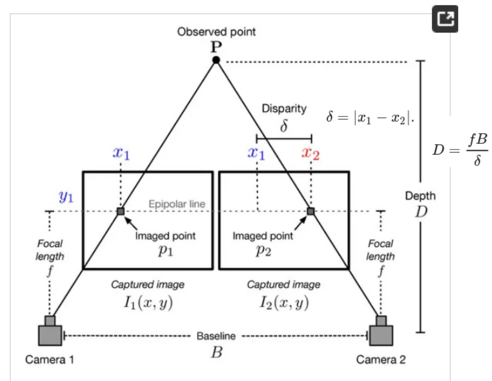
Z=f*B/d

Where:

- Z: Depth of the point from the camera plane.
- f: Focal length of the camera (in pixels).
- B: Baseline distance between the two cameras (in meters).
- d: Disparity, which is the difference in horizontal pixel coordinates of a point between the left and right images.

For two pixels pL (left image) and pR (right image), the disparity d is given by:

d=mod (xL -xR) where xL and xR are the horizontal coordinates of pL and pR respectively.




Relationship between Depth Estimation and Disparity in case of Stereo Image setup

Fig. 1.

## III. RELATED WORK

Deep Learning: End-to-end disparity prediction with convolutional networks

## IV. INITIAL ATTEMPTS

We began with the simplest method locating a pixel within the image. However, given the potential for numerous similar pixels, we opted to use a block of pixels. This block was then searched for along the horizontal axis at the same vertical level in another image, since the shift between the two images occurs solely in the horizontal direction.Using this approach, we successfully generated a disparity matrix for the image. However, the computational cost associated with the process was significant, posing challenges for efficient implementation in resource-constrained environments.

## V. FINAL APPROACH

In this study, we initiated the process by reducing the resolution of the input image, thereby minimizing the number of pixels and improving the computational efficiency of disparity matrix generation. Additionally, converting the image to grayscale significantly reduced computational complexity, as calculations were performed on a single channel instead of three RGB channels, reducing computation by approximately one-third. To further optimize the process, we implemented a search radius strategy. During block matching, we restricted the search for corresponding blocks to the vicinity of the coordinates from which the block was selected, substantially enhancing the speed of block matching and disparity matrix computation. We experimented with varying block sizes and search radii to identify the optimal parameters. After determining the optimal block size and search radius, we computed the disparity matrix and visualized the results through a heatmap. The heatmap effectively represented the depth information of the image, demonstrating the efficiency and accuracy of our approach.

## VI. Results and Observation

A depth map was successfully generated for the given image, providing a clear visualization of depth variations between objects. However, certain irregularities were observed in the depth map, primarily in regions with minimal texture. These irregularities are attributed to improper block matching, which highlights a limitation of this method. Specifically, the approach does not perform well in low-texture areas, where accurate block correspondence becomes challenging

## VII. Future Work

The algorithm encountered two primary challenges: incorrect depth detection in texture-less regions and high computational demands. The inaccuracies in texture-less areas arise from the limitations of the current block matching technique, which struggles to establish reliable correspondences in regions with minimal or no texture. Furthermore, the computational cost of the block search process remains a significant bottleneck in the implementation. To address these challenges, future work could explore alternative block matching techniques that are better suited for texture-less regions. Additionally, optimizing the block search process could further enhance computational efficiency, enabling faster and more accurate disparity matrix generation.

## Conclusion

This implementation successfully demonstrates fundamental stereo depth estimation principles using block matching.Despite challenges like repetitive textures and computationally intensive, the approach provides a foundational framework for applications in robotics and autonomous systems.

## References


[1] https://medium.com/analytics-vidhya/distance-estimation-cf2f2fd709d8

[2] https://medium.com/@satya15july$_1 1937/depth - estimation - from - stereo - images - using - deep - learning - 314952b8eaf9$