Advances in Engineering Research : Possibilities and Challenges
ISSN Print: 3079-5192
ISSN Online: 3079-5206
https://doi.org/10.63313/AERpc.9063

Research on Real time Acquisition System Based on Binocular Stereoscopic Vision

Yuanhang Xuan
Department of Computer Science, Tianjin University of Technology and Education, No. 1310, Dagu South Road, Tianjin, 300222, China.
Email: blistersqizman23@gmail.com

How to cite this paper: Xuan, Y. H. (2025). Research on real time acquisition system based on binocular stereoscopic vision. *Advances in Engineering Research: Possibilities and Challenges*, 3(1), 19-25. ISSN Print: 3079-5192; ISSN Online: 3079-5206.
https://doi.org/10.63313/AERpc.9063
Published: 2025-12-17

Abstract

This paper presents the design and implementation of a real-time 3D acquisition system utilizing binocular stereoscopic vision. The primary objective is to develop a robust and efficient pipeline capable of capturing, processing, and reconstructing three-dimensional geometry of dynamic scenes with low latency. The core methodology integrates synchronized image capture from a calibrated stereo camera pair, followed by real-time stereo rectification and a dense stereo matching algorithm optimized for speed. The system successfully achieves real-time performance, generating dense depth maps at a frame rate sufficient for interactive applications. Experimental results demonstrate the system's accuracy in reconstructing static objects and its capability to track depth variations in moderately dynamic environments. The conclusion highlights that the implemented system provides a practical and effective solution for real-time 3D perception, establishing a reliable foundation for applications in robotics guidance, quality inspection, and augmented reality where immediate spatial feedback is critical.

Keywords

Real-time System; Binocular Stereoscopic Vision; 3D Acquisition; Stereo Matching; Depth Map

1. Introduction

Currently, non-contact ranging technology has become one of the key technologies in fields such as industrial automation, robot navigation, autonomous driving, and virtual reality. Among these, binocular stereoscopic vision ranging technology has attracted significant attention due to its speed and high accuracy. This technique simulates human stereoscopic vision by using two cameras to capture images of the same scene from different angles. By analyzing the disparity between these images, it calculates object distances, providing an effective means for acquiring three-dimensional spatial information.

Although research on binocular stereoscopic vision ranging technology has a history of several years, how to improve the real-time performance and ranging accuracy of systems, along with how to better achieve 3D point cloud visualization, remains a hot topic in current research[1,3]. This paper researches and implements a real-time ranging system based on binocular stereoscopic vision. This system is capable of not only providing high-precision ranging results but also generating 3D point cloud visualizations for key frames of interest. Thereby, it offers users more intuitive depth information and enhances the system's practicality and interactivity.

2. Image acquisition section

2.1. Calibration of binocular cameras

Calibration of binocular cameras is a fundamental step in binocular vision systems, ensuring that the system can accurately recover three-dimensional spatial information from images rest. The calibration process involves determining the internal and external parameters of the camera and correcting lens distortion. The Zhang Zhengyou camera calibration method is adopted, which is a single plane checkerboard camera calibration method proposed by Professor Zhang Zhengyou[1] in 1998. By continuously moving the pre prepared checkerboard positions, multiple sets of checkerboard images are collected using binocular cameras. The stereo Camera Calibrator toolkit in MATLAB is used to calibrate the binocular cameras, export the results, and extract the inner and outer sides and distortion parameters of the binocular cameras.

2.2. Real time image capture and processing

After completing the calibration, the system starts capturing real-time images, and the left and right cameras work synchronously to obtain corresponding scene images[2]. Carry out standing body correction is used to address image distortion caused by differences in viewing angles. Stereoscopic correction improves matching accuracy by transforming images to align corresponding points on the same horizontal line. Binocular cameras are limited by lens assembly, differences in camera hardware, and whether the camera installation is strictly parallel to the optical axis during operation, which can cause distortion in the captured images and misalignment of the left and right images. Based on the internal and external parameters calibrated by the camera, stereo correction is performed on the images captured by the left and right cameras, simplifying disparity calculation and reducing the two-dimensional search problem to a one-dimensional search problem. This greatly reduces computational complexity, improves matching accuracy, and reduces matching errors. Therefore, stereo correction is an essential part of binocular vision systems[3, 5].

3. Stereo Matching and Depth Calculation

Stereo matching, which obtains depth information by calculating the disparity value between the rectified left and right images, is the core step for accurate spatial positioning and is key to binocular ranging and 3D point cloud reconstruction[6,7]. It essentially formulates an optimization problem: an energy cost function is established, and the disparity value for each pixel is estimated by minimizing this function.Stereo matching algorithms can be broadly categorized into three types: local methods, global methods, and deep learning – based methods.

Local Stereo Matching Algorithms work by comparing small windows or regions in the images to find corresponding points[8]. Their advantage lies in fast computational speed, allowing for quick recovery of disparity in texture-rich areas. The disadvantage is that they can cause mismatches in low-texture regions, resulting in a non-dense disparity map that often requires correction via interpolation algorithms in post-processing.

Global Stereo Matching Algorithms consider information from the entire image and solve for the disparity map by optimizing a global energy function[9,10]. These methods typically handle occlusions and textureless areas better but come with a higher computational cost.

Deep Learning – based Stereo Matching Algorithms leverage Convolutional Neural Networks (CNNs) to extract features and compute matching costs. Through adversarial training, they optimize disparity estimation, effectively handling large low-texture areas with high matching accuracy. The drawback is their requirement for large amounts of training data and significant computational resources[11].

The choice of a stereo matching algorithm should consider factors such as the application scenario, image data characteristics, and available computational resources. Each algorithm has its specific suitable scenarios and limitations. Selecting the appropriate algorithm depends on concrete application requirements, available resources, and the needed balance between precision and efficiency[12].

This system employs the Semi-Global Block Matching (SGBM) algorithm provided by OpenCV for stereo matching, optimized with a temporal filter.

After obtaining the disparity map through stereo matching, the process of acquiring the depth map primarily relies on the geometric relationship of binocular vision.

To better understand and visualize depth information, the depth map is typically converted into a pseudo-color image, where different depth values correspond to different colors.

4. Experiment

4.1. Camera calibration results

The production of a checkerboard calibration board requires attention to the number of rows and columns in the checkerboard. When the number of rows and columns is both odd and even, it will result in the edge squares of the last column and the last row not forming a complete detection unit, causing the calibration program to be unable to extract all preset feature points completely. Therefore, the calibration board used in this article is a 9x6 black and white checkerboard with each grid size of 25mm, and each row and column contains 8 and 5 corner points. At the same time, due to the use of A4 paper for printing, in order to reduce errors caused by paper deformation, the printed calibration board is pasted on a hardwood board to reduce paper warping.

After fixing the camera horizontally, you can start capturing calibration board images at different angles and positions. During the calibration process, the aperture and focal length of the camera should be kept constant to ensure consistent light intake and focal length. Any adjustments require recalibration. Each position requires two oblique shots in different directions, and the calibration plate can be fixed and shot from different angles. The imaging area of the calibration board should occupy about 1/4 of the camera's field of view, and overexposure should be avoided to avoid affecting the accuracy of feature contour extraction. Proper lighting and adjusting the distance between the calibration plate and the lens can improve image clarity. Usually, the number of calibration images should be between 15-25 to ensure the accuracy of calibration parameters. The calibration board images captured by the cameras are shown in Figure 1.

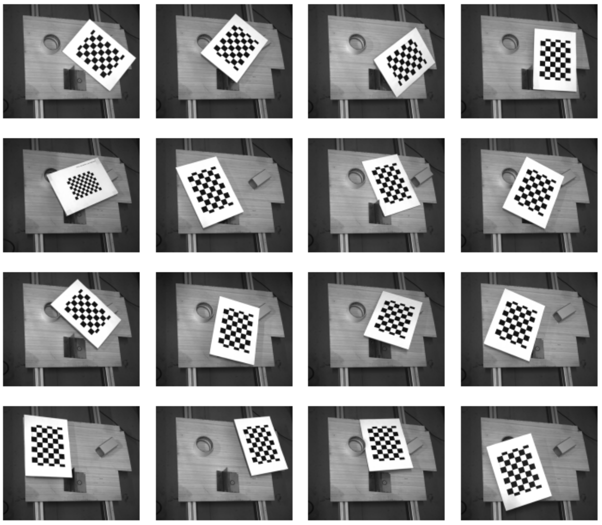

Figure 1. Camera image captured during camera calibration

Open Matlab, this article uses Matlab R2023b. Search for the Stereo Camera Calibrator tool in the App options, click on the Add Images option, set the path to save left and right calibration images, and set the size of each grid. Click the OK button to complete the reading of calibration images and delete samples with large errors.

After importing the image, the software will automatically use sub-pixel corner detection algorithm to extract corner points one by one - the points in the black and white staggered positions of the chessboard. This process can be referred to in the figure. In the diagram, the world coordinate system is constructed with the first detected corner as the origin, where the Z-axis is perpendicular to the surface of the chessboard and faces outward. The origin of the world coordinate system is marked with a yellow box.

Before calibration, it is necessary to select the distortion type of the camera. After the analysis, this experiment considers the radial distortion and tangential distortion of the camera. Since it is not a fisheye camera, only the first two factors are considered for radial distortion. After selecting, you can click the Calibrate button and wait for the camera calibration to complete.

4.2. Measurement error

There are only 14 images in the picture because 2 images with large reprojection errors were excluded, and 14 images were retained to continue the final result calculation. Based on the reprojection error, the accuracy of camera calibration can be preliminarily judged. The reprojection error is the Euclidean distance between the reference point on the calibration board and the corresponding point calculated through projection and distortion correction. From Figure 2, it can be seen that the average reprojection error is only 0.45 pixels, which is less than 1 pixel, indicating that the camera calibration meets the measurement accuracy requirements.

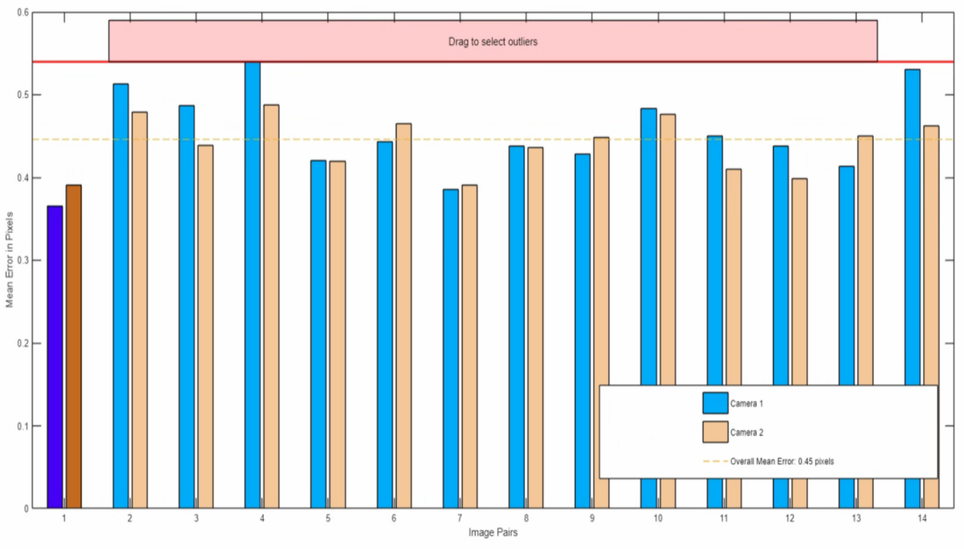

Figure 2. Camera calibration results and errors

5. Conclusion

Image acquisition forms the fundamental and critical first step in the binocular stereo vision pipeline. The quality, synchronization, and geometric fidelity of the captured image pair directly determine the upper limit of performance for all subsequent processes, including camera calibration, stereo rectification, matching, and ultimately, the accuracy of 3D reconstruction.

This study has implemented a systematic image acquisition framework designed for real-time operation. Key considerations included the selection of suitably synchronized cameras, precise control over acquisition parameters (such as exposure and gain to ensure consistent radiometry), and the implementation of hardware or software triggering to guarantee temporal alignment between the left and right frames. The acquisition process was explicitly designed to feed into the calibration and rectification modules, providing the raw data necessary for computing precise geometric transformations.

The success of the subsequent stereo matching and depth calculation stages is fundamentally dependent on the input provided by this acquisition stage. A well-calibrated system capturing synchronized, undistorted image pairs establishes the correct epipolar geometry, without which dense and accurate disparity estimation would be impossible. The real-time capability of the entire system is also primarily constrained by the acquisition frame rate and the efficiency of the initial image preprocessing (e.g., distortion correction).

In conclusion, a robust, calibrated, and synchronized image acquisition subsystem is not merely a preliminary step but the cornerstone of a reliable binocular stereo vision system. The design choices made at this stage—pertaining to hardware selection, synchronization strategy, and initial image conditioning—lay the essential groundwork upon which all advanced algorithms for depth perception are built. Future enhancements to the system's speed or accuracy must, therefore, continually re-evaluate and potentially refine the image acquisition front-end.

References

[1] Zhang, Z. (2000). A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), 1330-1334. https://doi.org/10.1109/34.888718

[2] Hartley, R. I. (1997). In defense of the eight-point algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(6), 580-593. https://doi.org/10.1109/34.601246

[3] Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision, 47(1/3), 7-42. https://doi.org/10.1023/A:1014573219977

[4] Hirschmüller, H. (2008). Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2), 328-341. https://doi.org/10.1109/TPAMI.2007.1166

[5] Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A., & Bry, A. (2017). End-to-end learning of geometry and context for deep stereo regression. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 66-75. https://doi.org/10.1109/ICCV.2017.17

[6] Guo, X., Yang, K., Yang, W., Wang, X., & Li, H. (2019). Group-wise correlation stereo network. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3273-3282. https://doi.org/10.1109/CVPR.2019.00339

[7] Chang, J.-R., & Chen, Y.-S. (2018). Pyramid stereo matching network. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5410-5418. https://doi.org/10.1109/CVPR.2018.00567

[8] Faugeras, O., & Lustman, F. (1988). Motion and structure from motion in a piecewise planar environment. International Journal of Pattern Recognition and Artificial Intelligence, 2(3), 485-508. https://doi.org/10.1142/S0218001488000285

[9] Brown, M. Z., Burschka, D., & Hager, G. D. (2003). Advances in computational stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(8), 993-1008. https://doi.org/10.1109/TPAMI.2003.1217603

[10] Mayer, N., Ilg, E., Häusser, P., Fischer, P., Cremers, D., Dosovitskiy, A., & Brox, T. (2016). A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4040-4048. https://doi.org/10.1109/CVPR.2016.438

[11] Menze, M., & Geiger, A. (2015). Object scene flow for autonomous vehicles. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3061-3070. https://doi.org/10.1109/CVPR.2015.7298925

[12] Lipson, L., Teed, Z., & Deng, J. (2021). RAFT-Stereo: Multilevel recurrent field transforms for stereo matching. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2070-2079. https://doi.org/10.1109/ICCV48922.2021.00208