UDC 004.021

RESEARCH ON RELATIVE LOCALIZATION OF MOBILE ROBOTS BASED ON MONOCULAR VISION AND KNOWN TAGS

QICHENG GUO

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received April 1, 2025

Abstract. This paper proposes a monocular vision-based method for relative pose estimation utilizing known-dimension tags. By equipping vehicle B with a rectangular red tag of precisely calibrated dimensions, vehicle A acquires visual data through a monocular camera. Through image processing and geometric model derivation, the proposed approach achieves accurate localization of the target vehicle. The method demonstrates advantages including low computational complexity, cost-effectiveness, and real-time performance, making it suitable for applications in autonomous vehicles and robotic navigation systems.

Keywords: monocular camera, pose estimation, fiducial marker, image processing, relative localization.

Introduction

With the continuous advancement of autonomous driving technology, visual perception, as one of its core technologies, plays a vital role in autonomous navigation and localization. Common visual sensors include monocular cameras, stereo cameras, and LiDAR. While stereo vision-based localization methods can calculate depth information of target objects through binocular disparity, they suffer from high hardware costs and significant computational complexity [1]. In contrast, monocular cameras have been widely adopted in mobile robotics and related fields due to their advantages such as low cost, lightweight design, and ease of installation [2].

In monocular visual localization, several studies have utilized reference objects with known dimensions to facilitate relative position computation. Chen et al. [3] proposed a calibration pattern-based monocular localization method that achieves positioning through geometric feature detection of predefined markers. The vehicle localization problem addressed in this study operates within a monocular vision framework. By attaching a rectangular tag of known dimensions to vehicle B, vehicle A can determine the relative position of vehicle B through visual analysis of the captured tag image. This approach provides a simple yet effective solution for collaborative robotics applications, demonstrating significant potential in mobile robot coordination scenarios.

Methodology

1. Camera calibration.

The proposed approach is based on monocular camera-captured images of vehicle B, utilizing the relationship between the known physical dimensions of the tag and its pixel coordinates in the image. Assuming the rectangular tag attached to vehicle B has predefined dimensions and occupies a defined pixel area in the image, the relative pose can be calculated through the following steps using perspective geometry principles.

To enable the transformation from pixel coordinates in images to real-world coordinates, camera calibration is first required. By capturing a series of calibration target images with precisely known dimensions, both the intrinsic parameters (e.g., focal length, principal point coordinates) and extrinsic parameters (i.e., transformation matrices between the camera and world coordinate systems) of the camera are derived.

2. Tag feature extraction.

Vehicle A captures images of Vehicle B and its attached tag using a monocular camera. Due to the superior performance of the HSV color space in red object detection [4], the tag typically manifests as a rectangular red region in the acquired images. The system first extracts the tag's contour through image processing techniques (e.g., color threshold segmentation and contour detection). Subsequently, dimensional attributes (e.g., length and width) and pixel coordinates of the extracted tag are acquired, which serve as fundamental parameters for subsequent pose estimation calculations.

3. Pose estimation based on the EPnP algorithm.

The Efficient Perspective-n-Point (EPnP) algorithm is designed to compute the camera pose (rotation matrix R and translation vector t) from n pairs of 3D spatial points and their corresponding 2D image projections. These correspondences are governed by the perspective projection relationship

$$p_i = K\left[R \mid \mathrm{t}\right] X_i \,, \tag{1}$$

where $p_i = \left[\mathrm{u}_i, v_i\right]^T$ denotes the pixel coordinates of the *i*-th point in the image; $X_i = \left[X_i, Y_i, Z_i\right]^T$ represents the 3D coordinates of the *i*-th point in the world frame; *K* is the camera intrinsic matrix, which includes the focal length, principal point, and other calibration parameters; *R* is the rotation matrix describing the orientation of the camera; *t* is the translation vector representing the positional offset of the camera.

Experiments and results analysis

To validate the effectiveness of the proposed method, the following experiments were conducted: A red rectangular tag of known dimensions was affixed to the rear of vehicle B, while vehicle A captured images of vehicle B using a monocular camera. The tag contour was extracted through color threshold segmentation processing, and the relative pose of vehicle B with respect to vehicle A was computed using the aforementioned methodology. During the experiments, multiple image sets were acquired under varying distances and angles. The pixel dimensions of the tag were measured and compared with actual distances. Figure 1 illustrates the image dataset captured by vehicle A's camera.

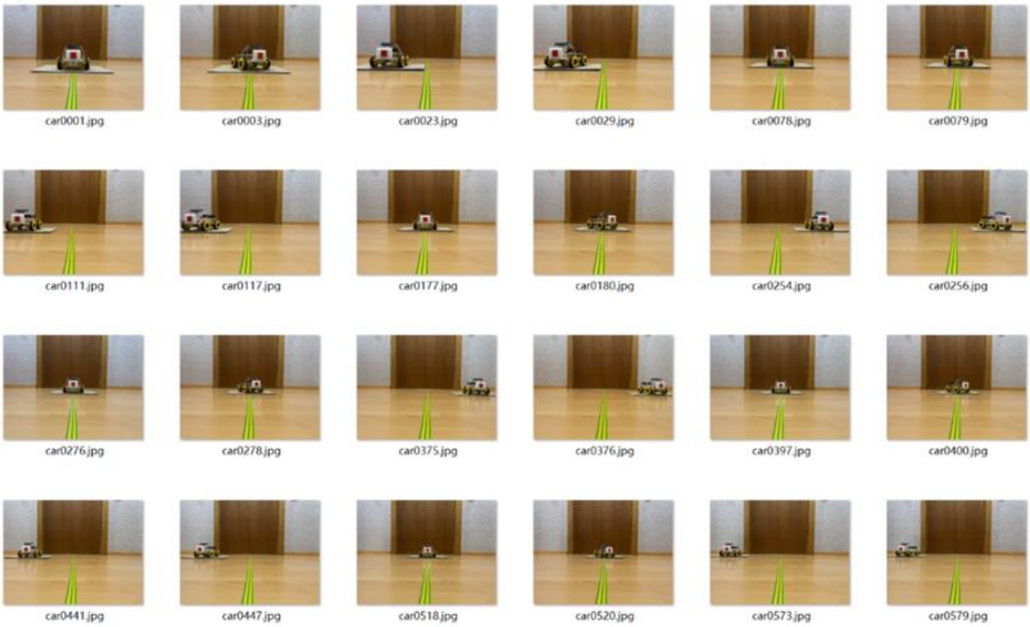

Figure 1. The relationship between the four coordinate systems

To comprehensively evaluate the accuracy of the proposed method, this paper conducts an analysis of reprojection errors. The reprojection error quantifies the discrepancy between the projected positions of reconstructed 3D world coordinates on the image plane and their corresponding actual positions in the acquired images. This metric directly impacts positioning accuracy, particularly when calculating the relative position of Robot B with respect to Robot A, where the magnitude of these errors significantly impacts the accuracy of the final position estimation. The calculated mean and maximum errors are presented in Table 1.

Table 1. **Average error and maximum error**

Number	1-4	5-8	9-12	13-16	17-20	21-24	Average
Average error	0,92 0,43 1,37 1,14	0,84 1,05 0,64 0,68	0,74 1,12 0,83 0,91	0,58 1,02 0,53 0,36	0,58 1,02 0,53 0,36	0,97 0,82 0,79 0,59	0,7633
Maximum error	1,04 0,46 1,44 1,22	0,95 1,13 0,69 0,71	0,81 1,17 0,87 0,94	0,64 1,04 0,56 0,38	0,07 0,53 0,87 0,6	1,06 0,84 0,82 0,61	0,8104

The experimental results demonstrate that the calculated position of vehicle B exhibits minimal deviation from the ground truth, with a mean average error of 0,7633 pixels and a mean maximum error of 0,8104 pixels. These findings indicate that when the reprojection error remains controlled within 1 pixel, the positioning accuracy sufficiently satisfies practical application requirements.

Conclusion

This paper presents a monocular vision-based relative pose estimation method utilizing known-size markers, with experimental validation confirming its feasibility and effectiveness. The proposed approach demonstrates computational efficiency and cost-effectiveness, making it applicable to relative positioning tasks in automated systems such as mobile robots and vehicles.

While the method achieves computational simplicity and satisfactory real-time performance, several limitations persist. For instance, the accuracy of marker extraction degrades significantly when the marker occupies a small area in the image or when subjected to excessive viewing angle deviation. Furthermore, environmental factors such as image noise and illumination variations may further increase computational errors. Future work will focus on optimizing image preprocessing techniques and enhancing marker detection algorithms to improve system robustness.

References

1. Huang, Y., Zhang, L., & Li, H. (2018). A stereo vision-based positioning system for mobile robots. Journal of Robotics and Automation, 5(1), P. 23–34.
2. Zhang, X., Li, J., & Wang, H. (2019). Real-time monocular localization using feature matching and pose estimation. International Journal of Advanced Robotics, 10(3), P. 45–59.
3. Chen, Y., Wang, D., & Zhang, X. (2017). A monocular camera-based localization approach using fiducial markers. Robotics and Autonomous Systems, 89(4), P. 125–134.
4. Smith, S. (2017). Real-time color segmentation for robotics applications. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) P. 567–574.