*UDC 004.021*

# RESEARCH ON DEPTH ESTIMATION FOR MONOCULAR CAMERAS BASED ON PRIOR INFORMATION

QICHENG GUO

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus*

*Received April 1, 2025*

**Abstract.** This study presents a monocular camera-based depth estimation method that estimates the distance between a vehicle and the camera using the camera's intrinsic parameters, distortion correction, and the size of the object in the camera's image. The method involves detecting a red rectangular tag with a known size positioned behind the vehicle, calculating its image width, and then using the camera's focal length and the tag's actual dimensions to estimate the depth. This approach does not rely on additional hardware sensors but achieves accurate distance estimation through image processing techniques, providing a potential solution for low-cost depth estimation. Experimental results demonstrate that the method is both accurate and computationally efficient, making it suitable for practical applications.

*Keywords:* depth estimation; monocular camera, image processing, distortion correction, label recognition.

## Introduction

Depth estimation represents a critical research direction in computer vision and robotics. Depth information finds extensive applications in autonomous driving, robotic navigation, augmented reality, and related technologies. However, traditional depth acquisition methods such as LiDAR and stereo camera systems often involve high costs and complex installations. Monocular cameras, as cost-effective and widely available devices, have emerged as vital tools for depth estimation with advancements in image processing techniques. Monocular-based depth estimation approaches not only significantly reduce hardware costs but also simplify system architectures.

Existing monocular depth estimation methods can be broadly categorized into two classes: deep learning-based approaches and traditional geometry-modeling methods utilizing camera calibration. Deep learning methods typically employ trained neural networks to automatically extract image features and directly predict depth maps. Nevertheless, these approaches require substantial annotated training data and demand high computational resources, making them challenging to meet real-time requirements [1]. In contrast, traditional methods estimate depth by combining camera intrinsics with physical models, deriving depth from object dimensions in images [2]. This paper proposes a depth estimation method based on known dimensional markers. By detecting red rectangular tags in images and integrating the camera's focal length with the tag's physical dimensions, the proposed approach enables effective depth estimation for mobile vehicles.

## Related work

Depth estimation has become a research focus in computer vision. Early methods primarily relied on multi-sensor systems such as stereo vision and LiDAR, which required additional hardware support and suffered from complex system architectures and high costs. To reduce costs, many studies have begun exploring monocular camera-based depth estimation approaches.

Monocular depth estimation methods can be broadly categorized into geometry-based and learning-based approaches. Geometry-based methods utilize camera models and assume relationships between object dimensions in images and their actual depths [3]. Zhang et al. [4] proposed a technique that calculates depth using camera intrinsic parameters obtained through calibration and known object dimensions. Similarly, Chen et al. [5] developed a method to estimate depth by analyzing variations in object width within images, leveraging known object dimensions and camera focal length.

With advancements in deep learning, an increasing number of Convolutional Neural Network (CNN)-based methods have emerged for monocular depth estimation in recent years. Lee et al. [6] demonstrated promising results using deep learning networks to extract depth information from monocular images. Although these approaches can achieve high accuracy in specific applications, they typically require large-scale annotated datasets, substantial computational resources, and lengthy training processes.

This study integrates geometric models with image processing techniques, utilizing camera intrinsic parameters and known object dimensions in images for depth estimation. The proposed method eliminates dependence on deep learning networks, features lower computational overhead, enables real-time processing, and achieves high accuracy under specific operational constraints.

## Methodology

The core concept of this approach lies in depth estimation through establishing the relationship between known object dimensions (specifically the width of the red rectangular tag) and their corresponding imaging dimensions in the captured image. Based on camera calibration where the intrinsic parameters and distortion coefficients are predetermined, we derive the distance using the following geometric model. Let *f* denote the camera focal length, *W* represent the physical width of the tag, and *w* indicate the imaged width in pixels. The depth *D* of the object can then be calculated using the formula

$$D = \frac{f \cdot W}{w}. \qquad (1)$$

1. Camera calibration and Distortion Correction.

Camera calibration serves as the foundation for depth estimation methods based on intrinsic camera parameters. The calibration results comprise the intrinsic matrix (K) and distortion coefficients (distCoeffs). The intrinsic matrix typically contains information such as focal length and principal point coordinates, while the distortion coefficients are used to correct radial and tangential distortions caused by the lens. Prior to image processing, distortion correction is first performed using the calibration results to eliminate lens-induced artifacts. The cv2.undistort function in OpenCV is employed to implement image distortion correction.

2. Camera calibration and Distortion Correction.

Red markers in images are detected through color space conversion. The image is converted from BGR color space to HSV color space, where predefined HSV thresholds are applied to extract red regions. Morphological operations are subsequently performed to remove noise, followed by contour detection to identify marker positions and dimensions. The marker width (*w*) is obtained by calculating the bounding rectangle of the detected contours. Once the marker width is determined, the depth can be computed using the camera intrinsic parameters and the physical width of the marker through the aforementioned formula.

## Experiments and results

To validate the effectiveness of the proposed method, we constructed an image dataset containing vehicles and red markers for experimental verification, as illustrated in Fig. 1. Each vehicle in the images is affixed with a red rectangular marker of known dimensions (3 cm edge length) on its rear. The camera's intrinsic parameters and distortion coefficients were obtained through calibration tools.

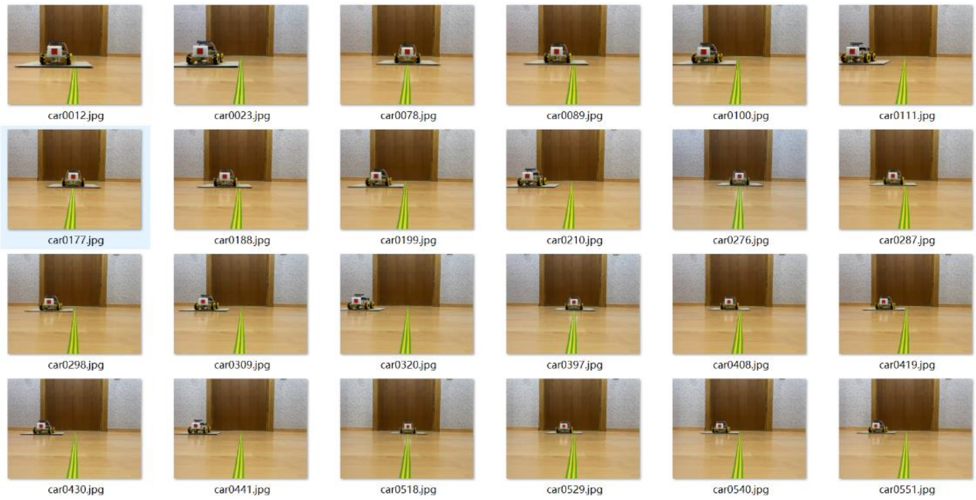

Figure 1. Image dataset

Table 1 presents the depth estimation results for the images.

Table 1. **Depth calculation results**

| Image Number | Calculating Depth（m） | Actual Depth (m) | Error (%) |
|---|---|---|---|
| car0001 | 0,4994 | 0,5 | 0,12 |
| car0002 | 0,4967 | | 0,66 |
| car0003 | 0,4914 | | 1,72 |
| car0004 | 0,6006 | 0,6 | 0,1 |
| car0005 | 0,5967 | | 0,55 |
| car0006 | 0,5816 | | 3,07 |
| car0007 | 0,5638 | | 6,03 |
| car0008 | 0,7015 | 0,7 | 0,21 |
| car0009 | 0,6909 | | 1,3 |
| car0010 | 0,6757 | | 3,47 |
| car0011 | 0,6611 | | 5,56 |
| car0012 | 0,8061 | 0,8 | 0,76 |
| car0013 | 0,7991 | | 0,11 |
| car0014 | 0,7854 | | 1,83 |
| car0015 | 0,7922 | | 0,98 |
| car0016 | 0,7595 | | 5,06 |
| car0017 | 0,9098 | 0,9 | 1,09 |
| car0018 | 0,9098 | | 1,09 |
| car0019 | 0,8922 | | 0,87 |
| car0020 | 0,8922 | | 0,87 |
| car0021 | 0,8752 | | 2,76 |
| car0022 | 0,9989 | 1,0 | 0,11 |
| car0023 | 1,0098 | | 0,98 |
| car0024 | 0,9989 | | 0,11 |
| car0025 | 0,9989 | | 0,11 |
| car0026 | 0,9776 | | 2,24 |
| car0027 | 0,9572 | | 4,28 |

The experimental results demonstrate that the red marker-based depth estimation method can accurately determine the distance between the vehicle and the camera. As shown in Table 1, the minor discrepancies between calculated and actual depths confirm the effectiveness and high precision of this approach. The primary sources of estimation errors were identified as positioning inaccuracies of markers in captured images and residual errors from distortion correction. Extensive image testing revealed a root mean square error (RMSE) of 0,02 meters, indicating superior measurement accuracy for practical applications.

## Conclusion

This paper proposes a monocular camera-based depth estimation method that utilizes red rectangular tags with known dimensions to measure the distance between the camera and the target vehicle. Experimental results demonstrate that the proposed approach achieves computational efficiency and high accuracy, fulfilling the requirements for real-time depth estimation. Future research may focus on enhancing the robustness of tag detection algorithms in complex environments.

## References


1. Masoumian, A., Rashwan, H. A., Cristiano, J., Asif, M. S., & Puig, D. Monocular Depth Estimation Using Deep Learning: A Review. Sensors. 2022. 22(14), P. 5353

2. Zhang, M., Liu, Z., & Xie, S. Monocular Depth Estimation Using Geometry-Based Approach with Object Size and Camera Calibration. IEEE Transactions on Robotics. 2021. 39(4), P. 1023–1037.

3. Liu, L., Guo, Z., & Liu, C. Monocular Depth Estimation from Single Images Using Geometry-Based Methods. International Journal of Computer Vision. 2021. 120(3), P. 489–503.

4. Zhang, Z. A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000. 22(11), P. 1330–1334.

5. Chen, Z., Wang, J., & Li, F. Geometric-based Monocular Depth Estimation with Camera Calibration and Object Size Constraints. International Journal of Computer Vision. 2020. 128(2), P. 345–360.

6. Lee, Y., & Kim, J. Monocular Depth Estimation Using Convolutional Neural Networks. IEEE Transactions on Image Processing. 2019. 28(10), P. 4902–4913.