

Integrating Stereo Vision with Robotic Arm for Unstructured Pick-and-Place Tasks

Mariusz Wisniewski, Michał Nawotka, Tomasz Lubowicki
Central Office of Measures, Ul. Elektoralna 2, 00-139 Warsaw, Poland

A photogrammetric system was designed for EPM 23IND08 DI-Vision project to precisely determine the position and orientation of objects within a workspace. In an automated environment, the system identifies objects in random locations and poses, then directs a robotic arm to grasp, lift, and place them in a predefined location. In the current phase of this work, ArUco coded targets were utilized to determine the positions of both the robot arm and the objects to be manipulated. A series of tests were conducted to evaluate the overall system performance. As part of these evaluations, the measurement uncertainty of the vision system and the repeatability of the robot's movements were determined, confirming its capability for dynamic task execution.

1. INTRODUCTION

The rise of Industry 4.0 has made the automation of complex and dynamic tasks a necessity. While traditional industrial robots excel at repetitive tasks in highly structured environments, their application in unstructured settings—where the position and orientation of objects are unknown—remains a significant challenge. Vision-based robotic systems address this challenge by providing the robot with the ability to perceive its environment and adapt its actions accordingly. This paper presents a novel integrated system developed as part of the EPM 23IND08 DI-Vision project [1], which combines a photogrammetric vision system with a robotic arm to perform accurate and repeatable pick-and-place operations on randomly located objects.

Documentary standards, such as, ISO 10360-13 [2], VDI/VDE 2634-1 [3] provide detailed description for the placement of reference lengths in the measurement volume of the stereo vision system under test to clearly capture systematic errors in these systems.

2. METHODS AND SYSTEM COMPONENTS

The system is composed of two primary components: the vision system and the robotic manipulator.

The multicamera photogrammetric system was prepared for the EMPIR 20IND02 DynaMITE project [4] which scope is focused on traceable dynamic measurements. The vision system uses images from one or more cameras to obtain precise geometric data. The measurement system consists of a twin camera with lens and set of coded targets. Each element requires calibration or dimensioning with an accurate measuring device. Both cameras are mounted above the workspace, capturing the entire operational area.

The vision system was previously used for validation of the Pozyx system [5], a commercial indoor positioning system based on the Ultra-Wide Band (UWB) technology and IMUMETER aircraft performance monitoring device [6]. In both tasks the working area was very large and ranged from 15 cubic meters of volume for Pozyx to over 200,000 cubic meters for the airport measurements.



Figure 1. Measurement campaign to determine aircraft landing positions for the validation of the IMMUTER system.

To simplify object localization and robot tracking, for initial phase of project, the system uses ArUco markers [7]. These square, black-and-white fiducial markers are highly robust to lighting changes, and their unique patterns allow for fast and reliable identification. The accuracy of ArUco marker pose estimation is influenced by factors such as camera resolution, distance from the camera, and the size of the marker itself.

The core of the vision system's functionality is built using the OpenCV library (Open Source Computer Vision Library) [8]. OpenCV provides a comprehensive set of tools for real-time image processing, including functions for camera calibration, marker detection, and 3D pose estimation.



Figure 2. Measurement campaign to determine aircraft landing positions for the validation of the IMMUTER system.

For this project, the Dobot Magician [9] was selected as the robotic platform. It is a compact, affordable, and versatile four-axis robotic arm. Its key features include a modular design, multiple end-effector options (including a gripper), and an open-source development environment, making it an ideal choice for research purposes. The Dobot's compact size is well-suited for laboratory-scale experiments, allowing for a controlled setup to validate the vision-guided manipulation concept.

The accuracy of a vision-based robotic system is fundamentally dependent on precise calibration. The stereo vision system was calibrated using implementation of the Zhang method [10] for the pin-hole model. Two phases of calibration were performed: intrinsic and extrinsic.

Intrinsic camera calibration determines the internal parameters of the camera, such as focal length, optical centre, and lens distortion coefficients. This process is crucial for converting 2D pixel coordinates from an image into accurate 3D coordinates in the real world. We followed a standard procedure using a chessboard pattern to compute the camera matrix and distortion coefficients.

Extrinsic calibration, establishes the spatial relationship between the camera's coordinate system and 3D position of objects. The calibration of the multi-camera system was carried out placing a chessboard in the field of view of both cameras. During calibration, several dozen positions of the reference board are used. Chessboard size is used as a reference to scale scene. To control robot we need to know spatial relationship between the camera's coordinate system and the robot's coordinate system. This step is essential for translating the camera's pose estimations into commands that the robot can execute. Using a known calibration object (e.g., an ArUco marker) placed at several known locations, the transformation matrix between the two coordinate frames was calculated. This allows the system to accurately command the robot to move to a position observed by the camera.

Extensive testing has shown that the estimation error for ArUco markers can be as low as a fraction of a millimetre in position. Our system's performance is sufficient for the pick-and-place task.

3. TESTING SYSTEM PERFORMANCE

To evaluate the system's performance, three testing scenarios were completed. In first test a set of reference blocks with ArUco markers were placed at different, known locations within the workspace. To facilitate the positioning of the blocks, they were placed on grid paper, with the outermost points marked with ArUco codes. For each location, the system determines the marker's position and pose. In this test, we checked the positioning errors of the vision system and the accuracy of the robot's movement to the indicated position.

In second test we evaluate the system's repeatability of ability to perform the pick-and-place task. The robot's task was to pick up a block, move it to a random position, and then return to the same place and put the block back.

In the third test, the robot was given four blocks and asked to move them to any desired location. The algorithm checked to see if a block already existed at the target location. This approach verified the accuracy of movements across the entire space using randomized control point filling. Simultaneously, the obstacle avoidance algorithm was validated.

The tests carried out showed the accuracy and repeatability of movements at a level below 1 mm.

4. CONCLUSION AND FUTURE WORK

The developed photogrammetric system integrated with the Dobot Magician robotic arm has demonstrated its capability to perform dynamic pick-and-place tasks with high precision and repeatability. The use of ArUco markers and the OpenCV library has provided a robust and reliable foundation for object localization and robot guidance.

Future work will involve transitioning from marker-based to markerless vision systems. This will require the implementation of more advanced machine learning and computer vision algorithms, to identify and localize objects. This will significantly increase the system's flexibility and applicability to a wider range of industrial tasks.

FUNDING STATEMENT

The project 23IND08 DI-Vision has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

REFERENCES

- [1] EPM 23IND08 DI-Vision project, available on: <https://projects.lne.eu/jrp-di-vision/>
- [2] ISO 10360-13:2021, Geometrical product specifications (GPS) — Acceptance and reverification tests for coordinate measuring machines (CMM) — Part 13: Optical 3D CMS.
- [3] VDI/VDE 2634-1, 2002, Optical 3D measuring systems – Imaging systems with point-by-point probing.
- [4] EMPIR 20IND02 DynaMITE project, available on: <https://empir.npl.co.uk/dynamite/>
- [5] Pozyx Accurate positioning site, available on: www.Pozyx.io.
- [6] Wiśniewski M., Kaźmierczak A., Pytka J., "Observation of aircraft ground performance with multicamera video monitoring system", 2025, Measurement: Sensors, Volume 38, Supplement, 101844, DOI: 10.1016/j.measen.2025.101844
- [7] Garrido-Jurado S, Muñoz-Salinas R, Madrid-Cuevas FJ, Marín-Jiménez MJ (2014) Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognition 47(6):2280–2292
- [8] OpenCV library (Open Source Computer Vision Library) , available on: <https://opencv.org/>
- [9] Dobot Magician, available on: <https://www.dobot-robots.com/>
- [10] Z. Zhang, A flexible new technique for camera calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), 2000, pp. 1330-1334.