I-014

AR マーカーを用いた姿勢推定におけるローカル手法とグローバル手法の評価
Assessment of Local and Global Methods for AR Marker-Based Pose Estimation

エムディ ヌルル　イスラム†　　オキ ディッキ アルディアンシャー　プリマ†
MD Nurul Islam　　Oky Dicky Ardiansyah Prima

1. Introduction

Augmented Reality (AR) markers offer a simple and effective solution for Six-Degree-of-Freedom (6DoF) pose estimation, enabling the precise spatial placement of virtual 3D objects. Utilizing multiple markers enhances pose accuracy by mitigating occlusion-related issues and providing redundant geometric information. However, inaccuracies in corner detection—particularly for markers viewed at oblique angles—can degrade overall estimation performance.

This study investigates and compares two pose estimation strategies—local and global—within a multi-marker configuration. The local method selects the most reliable marker, specifically the one most closely aligned with the camera's optical axis, to minimize distortion in pose computation. In contrast, the global method aggregates data from multiple markers, excluding those captured at extreme viewing angles. The proposed approach includes techniques for adaptive marker selection and noise reduction during image capture. Quantitative evaluation is conducted by comparing the estimated poses with ground truth data.

2. Related Work

Pose estimation using fiducial markers has been extensively investigated due to its computational efficiency and suitability for real-time augmented reality (AR) applications. Existing approaches are broadly categorized into local methods, which estimate pose using individual markers, and global methods, which utilize multiple markers arranged in a known spatial configuration.

Local methods compute the camera pose based on the detection of a single marker, typically employing algorithms such as Perspective-n-Point (PnP) or POSIT [1]. Prominent systems such as ARToolKit [2], AprilTag [3], and ArUco [4] utilize binary square markers with distinctive patterns to ensure reliable detection and identification. These methods are computationally lightweight and easy to deploy; however, their accuracy tends to degrade in scenarios involving partial occlusion, oblique viewing angles, or suboptimal lighting conditions.

In contrast, global methods aim to improve robustness and accuracy by exploiting multiple spatially co-located markers with predefined geometric relationships—often referred to as marker maps or bundles. When at least one marker is visible, the camera pose is estimated with respect to the global configuration and further refined using optimization techniques such as graph-based optimization or bundle adjustment [5]. Systems like ARToolKitPlus [6] and Chilitags [7] exemplify this approach by supporting structured marker arrangements for enhanced stability and precision in tracking. While global methods exhibit superior performance in complex or dynamic environments, they generally require increased computational resources and more elaborate setup procedures. Furthermore, erroneous detections or inaccurate localization of some markers within the configuration can adversely affect the overall pose estimation accuracy.

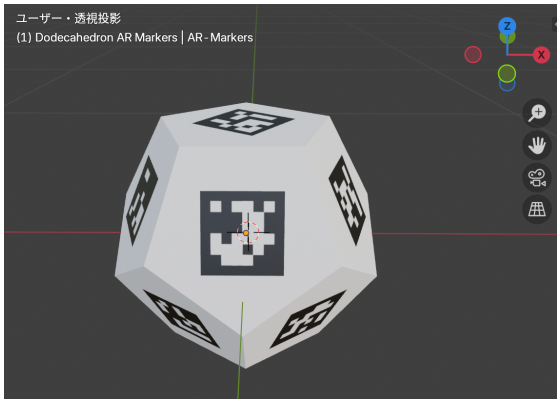


Figure 1. A 12-faced dodecahedron with ArUco markers affixed to its surfaces for pose estimation.

3. Methods

3.1 Experimental Setup

This study employs a rigid 3D-printed dodecahedral object with twelve 16 × 16 mm ArUco markers affixed to its faces as shown in Figure 1. The 3D coordinates of each marker's four corners are defined in the object-centric coordinate frame. A calibrated USB webcam captures at 30 frames per second and 1080p resolution, during which the object undergoes a smooth yaw rotation from 0° to 120°. The manually recorded yaw angle at each frame serves as the ground-truth reference.

3.2 Adaptive Marker Selection

An adaptive marker selection strategy is applied to identify and exclude markers with unreliable detection. Two criteria are used:

a. Corner-Angle Deviation

Internal angles at each corner are calculated based on their appearance in the camera image. If any angle deviates by more than ±20°, the corresponding marker is excluded from use in the global pose estimation method.

† 岩手県立大学大学院ソフトウェア情報学研究科
The Graduate School of Software and Information Science, Iwate Prefectural University.

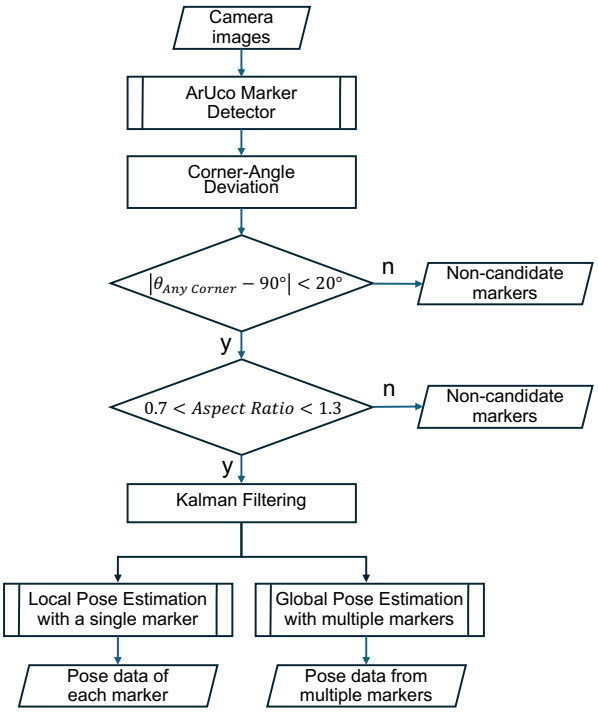


Figure 2. A diagram illustrating our strategies for assessing pose estimation results from the local and global methods.

b. Aspect Ratio

From the detected bounding rectangle around the marker's corners as seen by the camera, the width-to-height ratio is calculated. Markers are rejected if this ratio falls below 0.7 or exceeds 1.3.

3.3 Kalman Filtering

The Kalman Filter is employed to estimate the state of a dynamic system based on noisy measurements of marker corner detections. It effectively reduces noise while preserving the underlying motion or structure. The refined corner positions are subsequently used for both local and global pose estimation.

3.4 Pose Estimation

Markers that pass the adaptive selection criteria are used for pose estimation. In the local method, the solvePnP function [8] from the OpenCV library is applied individually to each valid marker to solve the Perspective-n-Point (PnP) problem and estimate its pose. In contrast, the global method aggregates the corner points from all valid markers, treating them as a composite marker to estimate a single, unified pose. Figure 2 illustrates the strategies used to assess pose estimation results from both the local and global methods.

4. Experiments and Results

An experiment was conducted to evaluate the effects of the proposed adaptive marker selection and Kalman filtering on marker selection consistency and pose estimation accuracy. Eleven ArUco markers were attached to the dodecahedron, excluding the bottom face. Markers with IDs 1 and 11 were occluded or outside the camera's field of view. Figure 3 illustrates the experimental setup. The dodecahedron was moved from left to right while being recorded by a camera. An angle indicator was used to collect ground truth data corresponding to each frame.

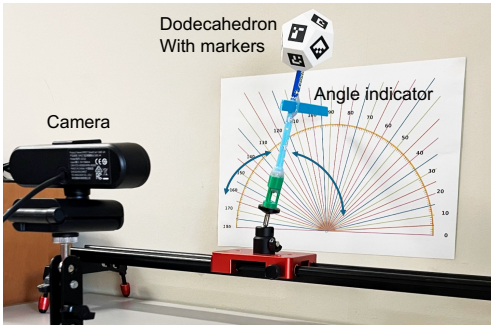
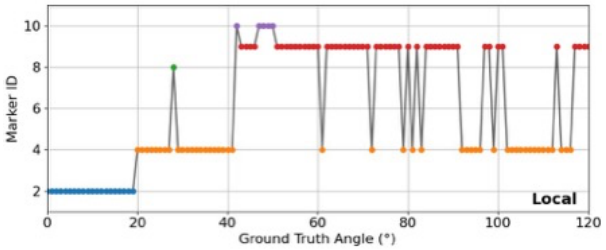
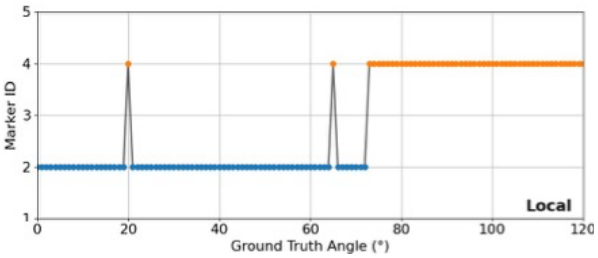


Figure 3. Experimental setup



(a) Without adaptive marker selection



(b) With adaptive marker selection

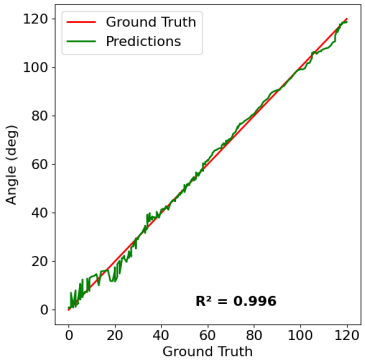
Figure 4. Markers selected for optimal angle estimation in the local pose method.

4.1 The Effects of the Adaptive Marker Selection

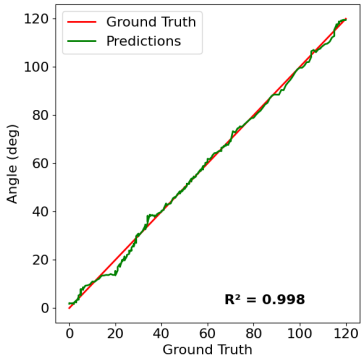
Pose estimation was performed individually for each marker based on the local pose method, followed by regression analysis against the ground truth data. The error between each marker's estimated pose and the corresponding ground truth was calculated, and the marker with the smallest error was selected. Figure 4(a) presents the results without adaptive marker selection, where five different markers (IDs 2, 4, 8, 9, and 10) were frequently switched to represent the tilt of the dodecahedron from 0 to 120 degrees. In contrast, as shown in Figure 4(b), the application of adaptive marker selection resulted in the use of only three markers, with marker IDs 2 and 4 handling most of the tracking.

4.2 Pose Estimation Accuracy

Pose estimation was performed using all valid markers based on the global pose method, and the results were validated through regression analysis against the ground truth data. A Kalman Filter was applied to the detected corners of each marker. Figure 5



(a) Without adaptive marker selection and Kalman filtering



(b) With adaptive marker selection and Kalman filtering

Figure 5. Angle estimation using the global pose method.

presents the angle estimation results with and without the use of adaptive marker selection and Kalman filtering. The regression analysis indicates an improvement in the coefficient of determination (R^2) from 0.996 to 0.998 when both techniques were applied. Additionally, the graph shows reduced fluctuation, reflecting the enhanced stability associated with the increased R^2 value.

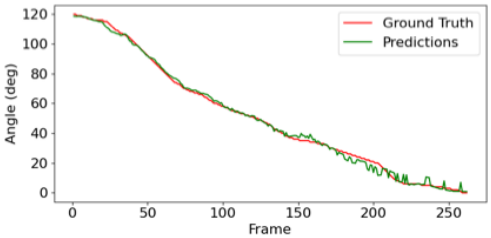
4.3 Object Tracking Smoothness

Figure 6 shows the per-frame angle estimation using the global pose method. The continuous angle estimation, enhanced by adaptive marker selection and Kalman filtering, results in a smoother trajectory that more closely aligns with the ground truth data.

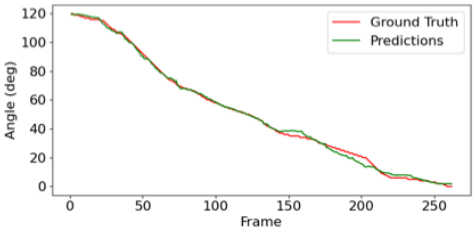
5. Discussion

The improvements resulting from the use of adaptive marker selection and Kalman filtering have demonstrated enhanced pose estimation performance using ArUco markers. The adaptive marker selection effectively eliminates poorly detected markers, while Kalman filtering reduces the influence of noise during image capture. Together, these techniques contribute to improved pose estimation accuracy and smoother object tracking.

However, there remains potential for further refinement. Marker corner detection could be enhanced by re-evaluating detected corners using Harris corner detection. If a detected corner significantly deviates from the actual corner position, the corresponding marker could be excluded from the pose estimation process to maintain accuracy.



(a) Without adaptive marker selection



(b) With adaptive marker selection

Figure 6. Per-frame angle estimation using the global pose method.

6. Conclusion

In this study, we experimentally evaluated both local and global pose estimation methods using ArUco markers and proposed an approach to enhance estimation accuracy. By utilizing a dodecahedron as the experimental platform, we were able to detect multiple markers from various viewing angles. Although up to five markers could be simultaneously detected, some were inaccurately identified due to distortion or poor visibility. The proposed adaptive marker selection method effectively excluded such unreliable markers. Furthermore, the integration of Kalman filtering contributed to improved pose accuracy and resulted in smoother, more stable object tracking.

References

[1] Dementhon, D.F. and Davis, L.S., “Model-based object pose in 25 lines of code,” International Journal of Computer Vision, Vol. 15, pp. 123–141 (1995).
[2] Kato, H. and Billinghurst, M., “Marker tracking and HMD calibration for a video-based augmented reality conferencing system,” Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR’99), pp. 1-10, (1999).
[3] Olson, E., “AprilTag: A robust and flexible visual fiducial system,” 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, pp. 3400-3407, (2011).
[4] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., and Marín-Jiménez, M. J., “Automatic generation and detection of highly reliable fiducial markers under occlusion,” Pattern Recognition, Vol. 47, No. 6, pp. 2280–2292, (2014).
[5] R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision,” Cambridge Univ. Press, (2004).
[6] Wagner, D. and Schmalstieg, D., “ARToolKitPlus for pose tracking on mobile devices,” Computer Vision Winter Workshop, pp. 1-8, (2007).
[7] Jurado-Rodriguez, D., Muñoz-Salinas, R., Garrido-Jurado, S. et al. Planar fiducial markers: a comparative study. Virtual Reality, Vol. 27, pp. 1733–1749, (2023).
[8] OpenCV, https://docs.opencv.org/3.4/d5/d1f/calib3d_solvePnP.html. [retrieved at June 12, 2025]