2025 IEEE/SICE International Symposium on System Integration (SII)
January 21-24, 2025, Munich, Germany

Accurate Estimation of Fiducial Marker Positions Using Motion Capture System

Matus Tanonwong, Naoya Chiba, and Koichi Hashimoto

Abstract— **In this paper, we present a method for aligning the coordinates of multiple cameras and sensors into a unified coordinate system using a motion capture system. Our simulated convenience store environment includes cameras and sensors with distinct coordinate systems, necessitating coordinate alignment. The motion capture system identifies retroreflective markers, while other cameras detect fiducial markers for position and orientation determination. Three optimization algorithms are experimented with to compute a transformation matrix aligning camera coordinates to motion capture coordinates, with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm achieving the best results (average errors of 1.13 centimeters and 3.90 degrees). Comparisons with fiducial marker pose estimation using an open-source Pupil Core software indicate our method is more robust and consistent, with lower repeatability errors. Additionally, we examine the estimation errors in relation to the distances of the fiducial markers from the camera to minimize these errors, enhancing installation accuracy of cameras and sensors in our simulated environment. This approach enables precise determination of positions and orientations across integrated cameras, consistent with the motion capture system. The findings contribute to our ongoing project, which requires an accurate system integration for customer behavior analysis.**

I. Introduction

In recent years, supermarkets and retail stores have adopted Artificial Intelligence-Powered Closed-Circuit Television (AI-Powered CCTV) systems to analyze visitor behavior, detecting instances such as shoplifting, monitoring shopping trajectories, and offering real-time product recommendations to enhance the shopping experience. Studying customer behaviors during in-store shopping enhances understanding of consumer shopping patterns. This insight can optimize store layout design to create a more convenient, productive, and satisfying shopping experience for consumers. An effective store layout strategically positions merchandise to attract customer attention and encourages exploration of multiple aisles. This increased exposure to merchandise can impact customer decisions, ultimately leading to improved store sales and profitability [1].

Video analytics are utilized on CCTV footage to gain insights into customer shopping behaviors [2], covering tasks like customer detection, shopping item detection, customer identification, tracking, emotion recognition, and action recognition. To generate deep learning models that are capable of these tasks, the models pretrained on large-scale datasets are commonly adopted and finetuned on specific dataset tasks [3], [4]. Due to the privacy issues, it has been challenging to use data in real environments. To collect datasets for finetuning the models and test their applications, the datasets collected in simulated convenience store environments have been widely used by previous studies [5], [6].

We are currently constructing a simulated convenience store environment with the dimension of 4.476 meters by 7.665 meters in our laboratory for data collection, equipped with a range of cameras and sensors. Our setup includes an OptiTrack motion capture system with eight PrimeX 13 cameras, DS-2CD6365G1-IVS HIK 6MP fisheye cameras, RealSense D455 depth cameras, Pupil Core eye trackers [7], GoPro HERO11 Black cameras, and Luxonis OAK-1 OpenCV AI cameras as portrayed in Fig. 1. The images of our environment captured by two surveillance cameras and a fisheye camera are portrayed in Fig. 2.

Our objective is to use the motion capture system to accurately help determine the positions and orientations of all integrated cameras in alignment with its coordinates, enhancing overall system coherence. Synchronizing coordinates and timestamps across multiple devices is crucial for compiling accurate datasets. The motion capture system accurately determines retroreflective marker positions with precision down to less than a millimeter, while other cameras detect fiducial markers to determine their positions and orientations. The fiducial markers in our environment are AprilTag tag36h11 family [8].

The use of accurate positional data obtained from motion capture (e.g., attaching a marker to the human body yields a skeletal model) and image information in the scene obtained from the environmental cameras are valuable in analyzing shopping behavior. However, the coordinate system estimated by the motion capture system is independent of the coordinate system estimated by the marker-based system using the camera. Therefore, a calibration between these coordinate systems is required to bind them.

This paper presents a method to align the motion capture coordinate system with the marker coordinate system by mounting a motion capture tracker on a camera and observing markers in the environment with the camera. The proposed method also estimates the rigid body transformation between the camera origin and the motion capture tracker at the same time as the coordinate transformation estimation, thus achieving coordinate system alignment with a simple experimental setup. The proposed method is a

*This work was supported by JSPS KAKENHI Grant Numbers 21H05298 and 24H01419.

Authors are with the Department of System Information Sciences, Graduate School of Information Sciences, Tohoku University, Aoba-ku, Sendai 980-8579, Japan (tanonwong.matus.q5@dc.tohoku.ac.jp; chiba@nchiba.net; koichi.hashimoto.a8@tohoku.ac.jp)

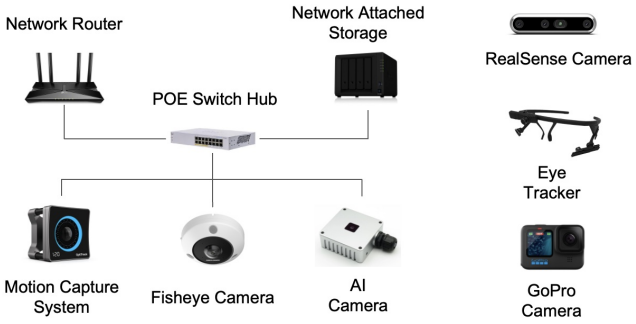
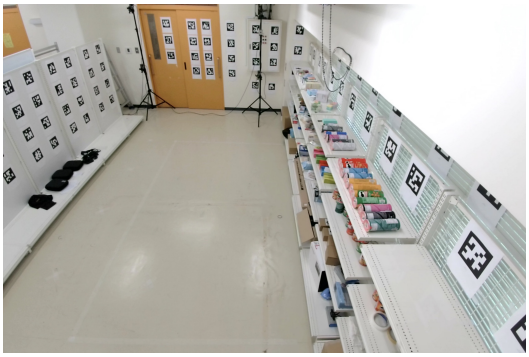


Fig. 1. Details of System Elements.

(a) Surveillance Camera 1



(b) Surveillance Camera 2

(c) Fisheye Camera

Fig. 2. Simulated Convenience Store for Data Collection.

general-purpose technique for scenarios involving multiple cameras and markers in the environment, with a calibration process that easily supports the addition of both static and dynamic cameras.

The key contributions of our work are as follows: 1.) We propose a method to estimate fiducial marker positions using a motion capture system and three optimization algorithms to compute the transformation matrix to align camera coordinates with motion capture coordinates. 2.) Our method is compared with the Pupil Core software by estimating fiducial marker positions in our simulated convenience store environment. 3.) Our investigation explores how estimation errors correlate with the distance of fiducial markers relative to the camera. The results indicate ranges for detecting fiducial markers to minimize errors, enhancing the accuracy of estimating positions and orientations for additional cameras to be installed in the environment.

Our results align with a previous study [9] showing that AprilTag marker position estimation improves as the camera moves closer. Within a 1.5 to 2-meter range, the prior study reports average errors of 6-7 centimeters with a Logitech camera and 3-3.5 centimeters with a Raspberry Pi camera. In comparison, our method achieves lower average errors, around 1 centimeter, ensuring high accuracy for future camera setups in our simulated convenience store environment.

II. Related Work

A. Motion Capture for Human Activity Recognition

A motion capture system has been widely deployed and integrated with other devices to collect human motion datasets in the field of human activity recognition and human motion analysis. The paper [10] introduced a dataset of 3D human poses captured by four camera views and a motion capture system. The study [11] introduced a comprehensive dataset on whole-body human motion, focusing specifically on human-object interactions. The procedures and techniques for capturing of motion, labeling, and organizing the motion capture data were proposed. More recently, many studies provided with large-scale datasets that are beneficial to animation, visualization, and generating training data for deep learning models [12], [13]. The research [14] released an indoor human motion dataset (Oxford-IHM) which includes human-motion trajectories in an indoor environment. The dataset includes ground truth trajectories from motion capture and RGB-D data. A previous work [15] introduced seven datasets collected using inertial-based motion capture technology. The datasets feature professional gestures executed by industrial operators and skilled craftsmen, performed in real-world conditions on-site.

B. Fiducial Markers in Camera Pose Estimation

A fiducial marker is a reference object or pattern for camera pose estimation, which requires a precise localization in the environment. It has been used in various fields, such as computer vision [7], robotics [16], and augmented reality [17], to provide a point of reference for spatial measurements and alignment. The most commonly used fiducial markers

are ARTag [18], AprilTag [8], [19], and ArUco [20]. A previous study [9] provided an overview of these fiducial markers and experimentally compared their accuracy in pose measurements. In our work, we utilize the AprilTag markers because they are supported by Pupil Core software, and they can be detected by any cameras. This allows us to deploy them in the same environment for the camera pose estimation task.

C. Optimization Algorithms

An optimization algorithm is a computational method used to find the minimum or maximum of a function. These algorithms are designed to iteratively improve a solution by adjusting the input parameters to minimize or maximize the objective function, subject to any given constraints [21]. Among available algorithms, we adopt three algorithms to solve our optimization problem. The BFGS algorithm [22] is a quasi-Newton method that approximates the Hessian matrix. The Nelder-Mead algorithm [23] is a simplex method that does not require gradient information, making it suitable for non-smooth functions. Lastly, the Powell algorithm [24] is a direction set method that conducts sequential one-dimensional minimization along each vector of the direction set.

III. Methodology

This section outlines our proposed method for estimating AprilTag marker positions as illustrated in Fig. 3. We begin with the observation model and the objective function of their calibration. Next, we detail the AprilTag sample dataset used to optimize and evaluate the Camera-to-Tracker transformation matrix, the unknown transformation matrix for aligning camera coordinates to motion capture coordinates. After that, we estimate the AprilTag marker positions, comparing our estimation method with the estimation results obtained using the Pupil Core software. Finally, we perform repeatability tests for our proposed method and the method using the Pupil Core software.

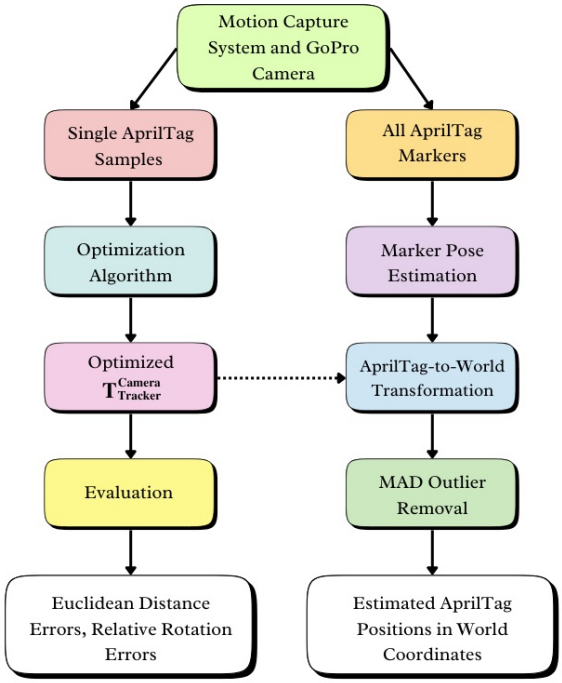
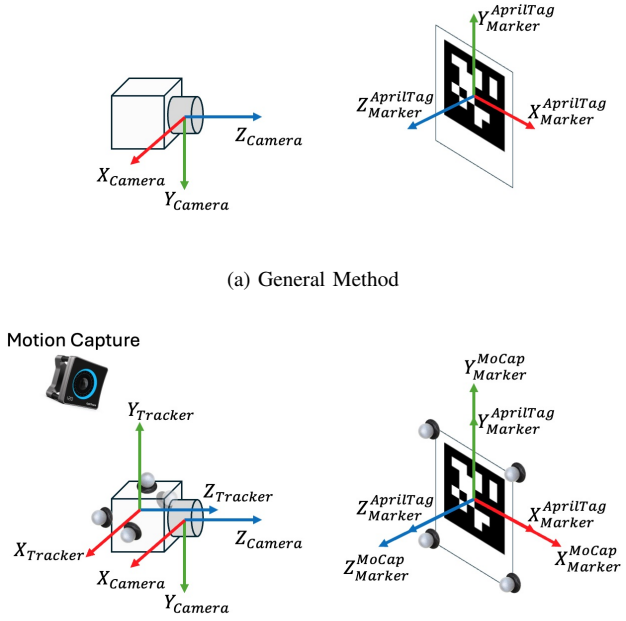


Fig. 3. Workflow of Our Proposed Method.



(a) General Method

(b) Our Proposed Method

Fig. 4. AprilTag Marker Pose Estimation.

A. Proposed Method

We first calibrated the motion capture system and assigned the origin of the motion capture system's coordinates. The motion capture coordinates are designated as the world coordinate system. The comparison of our proposed method and the general method for the AprilTag marker pose estimation is portrayed in Fig. 4. We attached four retroreflective markers to an AprilTag marker, which has a size of 15 cm, and four to a GoPro camera to track their positions and orientations within the motion capture system's coordinates, as shown in Fig. 4b.

The general method for the AprilTag marker pose estimation provides the AprilTag marker coordinate system and the GoPro camera coordinate system as illustrated in Fig. 4a. In contrast, our proposed method introduces two additional coordinate systems: the Tracker coordinate system, generated using four retroreflective markers attached to the GoPro camera, and the AprilTag coordinate system, created using four retroreflective markers attached to the marker. Both of these coordinate systems are obtained from the motion capture system.

Next, we collected the single AprilTag samples to optimize the objective function for calibration and evaluate

the optimized Camera-to-Tracker transformation matrix. We experimented with three optimization algorithms, such as the BFGS algorithm, the Nelder-Mead algorithm, and the Powell algorithm. After determining the optimized Camera-to-Tracker transformation matrix, we utilized the GoPro camera with attached retroreflective markers to estimate the positions of all AprilTag markers in the environment.

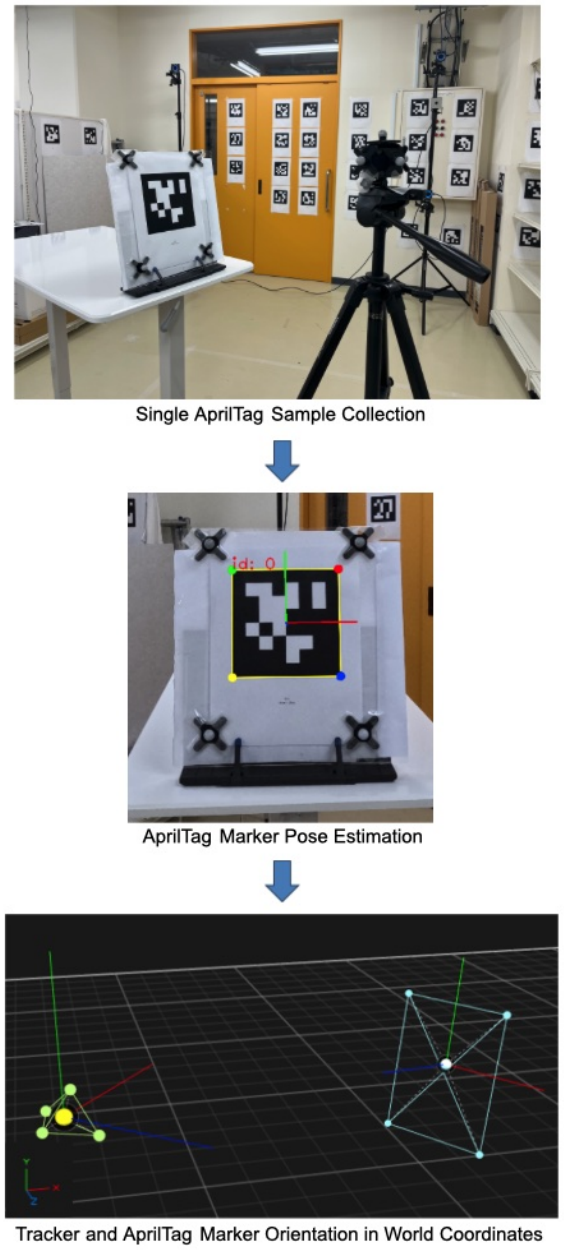


Fig. 5. Setup for Tracking with Motion Capture System and Marker Pose Estimation with GoPro Camera.

B. Observation Model and Optimization Problem for Calibration

We model an equation to transform a point from the AprilTag coordinate system to the world coordinate system as

$$\mathbf{T}^{\text{AprilTag}}_{\text{World}} = \mathbf{T}^{\text{Tracker}}_{\text{World}} \cdot \mathbf{T}^{\text{Camera}}_{\text{Tracker}} \cdot \mathbf{T}^{\text{AprilTag}}_{\text{Camera}}, \tag{1}$$

where $\mathbf{T}^{\text{AprilTag}}_{\text{World}}$, $\mathbf{T}^{\text{Tracker}}_{\text{World}}$, $\mathbf{T}^{\text{Camera}}_{\text{Tracker}}$, $\mathbf{T}^{\text{AprilTag}}_{\text{Camera}}$ in $SE(3)$ are AprilTag-to-World, Tracker-to-World, Camera-to-Tracker, and AprilTag-to-Camera transformation matrices respectively. Here, $SE(3)$ is a special Euclidean group which includes rigid transforms in 3D space. The tracker coordinates are the coordinates of the rigid body created by the retroreflective markers attached to the camera. The unknown transformation matrix is the Camera-to-Tracker transformation matrix as

$$\mathbf{T}^{\text{Camera}}_{\text{Tracker}} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

When the calibration of Camera-to-Tracker transformation matrix $\mathbf{T}^{\text{Camera}}_{\text{Tracker}}$ is done well, eq. (1) should be satisfied; therefore, the objective function for optimization can be designed as the sum of the Frobenius norm [25] as derived in

$$\min_{\mathbf{T}^{\text{Camera}}_{\text{Tracker}}} E = \sum_{i=1}^{N} \left\| (\mathbf{T}^{\text{Tracker}}_{\text{World}})^{-1} \cdot \mathbf{T}^{\text{AprilTag}}_{\text{World}} - \mathbf{T}^{\text{Camera}}_{\text{Tracker}} \cdot \mathbf{T}^{\text{AprilTag}}_{\text{Camera}} \right\|^2_F. \tag{3}$$

C. Single AprilTag Sample Dataset

Our experimental setup explained in Section III-A provides the AprilTag-to-World and Tracker-to-World transformation matrices. By detecting the AprilTag marker and estimating its pose in each image frame, we compute the AprilTag-to-Camera transformation matrix as illustrated in Fig. 5. The relationships among these transformation matrices are described in eq. (1), with the Camera-to-Tracker transformation matrix being unknown.

To solve for the unknown transformation matrix, we collected 24 AprilTag samples from a single marker positioned at eight different locations within the environment. At each location, three samples were captured from different camera angles, as illustrated in Fig. 6.

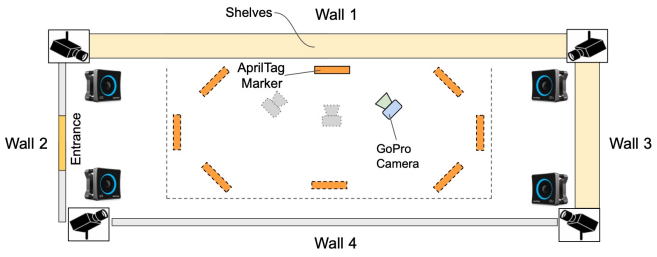


Fig. 6. AprilTag Sample Collection for Optimization.

D. Estimation of AprilTag Marker Positions

After obtaining the optimized Camera-to-Tracker transformation matrix, we used the GoPro camera with attached retroreflective markers to estimate the positions of 157 AprilTag markers in the environment as described in

$$\mathbf{P}_{\text{World}} = \mathbf{T}^{\text{Tracker}}_{\text{World}} \cdot \mathbf{T}^{\text{Camera}}_{\text{Tracker}} \cdot \mathbf{P}_{\text{Camera}}, \tag{4}$$

where $\mathbf{P}_{\text{World}}$ and $\mathbf{P}_{\text{Camera}}$ are a position of the marker in the world coordinates and camera coordinates respectively. The Mean Absolute Deviation method (MAD) [26] is applied to remove outliers. Figure 7 shows examples of estimated AprilTag markers on each side of the wall using our method in the environment.

Additionally, we used a Pupil Core eye tracker to scan and estimate the positions of the AprilTag markers in the environment, allowing us to compare these estimates with those obtained using our method. The Pupil Core eye tracker is not directly involved in solving the primary problem addressed in this study but instead serves as an external device to facilitate data collection by providing reference estimates for comparative analysis. The eye tracker scanned the markers, and the Pupil Core software computed the estimated positions, as shown in Fig. 8. We assigned an AprilTag marker on Wall 1 to be the origin marker in the Pupil Core software.

E. Repeatability and Errors Relative to Marker Distances

We computed Euclidean distance errors between our method's estimated marker positions and those of the Pupil Core software. To assess repeatability, we estimated AprilTag marker positions ten times using our method and the Pupil Core software, sampling four markers on each side of the room. As shown in Fig. 6, marker IDs 25, 26, 36, 37 are on Wall 1; IDs 64, 67, 69, 74 are on Wall 2; IDs 93, 94, 100, 102 are on Wall 3; and IDs 140, 146, 149, 152 are on Wall 4. We lastly calculated repeatability errors relative to marker distances from the GoPro camera using our method to determine the relationship between distance and estimation errors, thereby enhancing the accuracy of future camera installations.

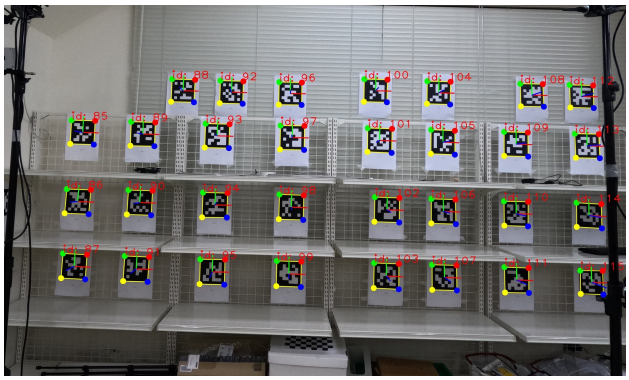
IV. Evaluation

This section presents the evaluation results of the optimization process, compares marker position estimates between our method and the Pupil Core software, analyzes the repeatability errors, and assesses the marker estimation errors of our method based on the distances of the markers from the camera.

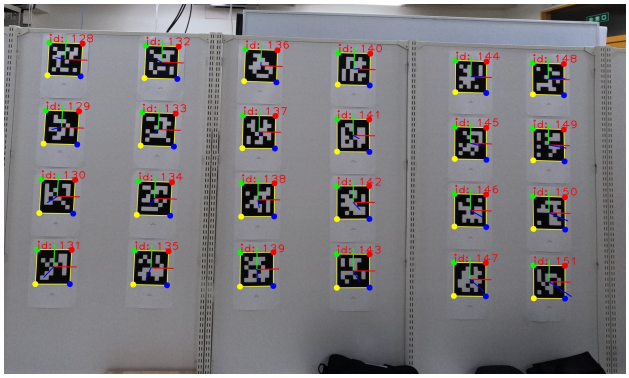
We split the AprilTag samples into optimization and evaluation sets for optimizing the Camera-to-Tracker transformation matrix as 5/10, 10/10, and 14/10 samples respectively. To compare the results of three optimization algorithms, we apply their optimized transformation matrices to the evaluation set to compute $\mathbf{T}_{\text{World}}^{\text{AprilTag}}$ in eq. (1) and compare them with the ground truths from the motion capture system. The evaluation metrics are Euclidean distance error and Relative Rotation error. Table I illustrates that the BFGS algorithm yields the best results on the evaluation set, with average errors of 1.13 centimeters and 3.90 degrees. After the optimized Camera-to-Tracker transformation is obtained, it is used to estimate the positions of all AprilTag markers as described in eq. (4). From this point onward, our method uses the optimized transformation matrix generated by the BFGS algorithm to estimate marker positions.

(a) Markers on Wall 1

(b) Markers on Wall 2



(c) Markers on Wall 3



(d) Markers on Wall 4

Fig. 7. Examples of Estimated AprilTag Markers in Our Environment.

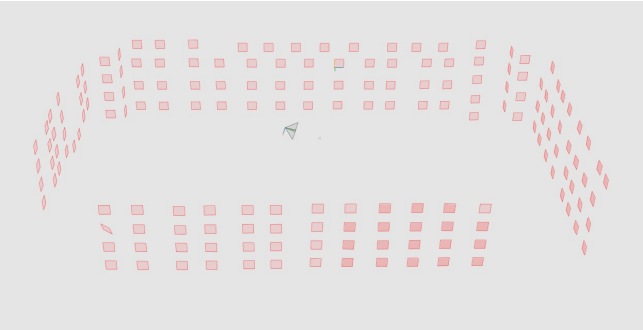


Fig. 8. AprilTag Marker Localization Generated by Pupil Core Software.

TABLE I
EVALUATION RESULTS OF OPTIMIZED CAMERA-TO-TRACKER TRANSFORMATION MATRIX

Method	**Train/Test**	**Euc. Dist.** [cm]	**Rel. Rot.** [deg.]
BFGS	5/10	1.22	5.16
BFGS	10/10	1.21	4.18
BFGS	14/10	**1.13**	**3.90**
Nelder-Mead	5/10	10.13	60.03
Nelder-Mead	10/10	24.31	25.88
Nelder-Mead	14/10	25.98	16.35
Powell	5/10	1.22	5.16
Powell	10/10	1.21	4.25
Powell	14/10	1.13	3.95

Next, our method is compared with the Pupil Core software by estimating AprilTag marker positions in our simulated convenience store in Fig. 2. The differences between estimated marker positions by the two methods are computed as Euclidean distance errors. The graph in Fig. 9 indicates the number of markers in the ranges of Euclidean distance errors. It can be observed that there is a huge difference between the estimated marker positions by our approach and the Pupil Core software. Most of the markers exhibit Euclidean distance errors ranging from 10 to 30 centimeters, with some markers showing errors exceeding 40 centimeters.

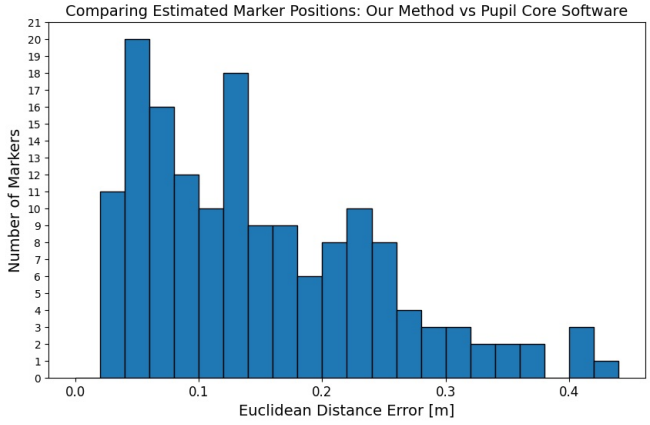


Fig. 9. Distribution of Errors between Marker Positions from Our Method and Pupil Core Software.

To investigate the difference in the estimated marker positions, we estimated AprilTag marker positions ten times using our method and the Pupil Core software, sampling four markers on each side of the room to conduct the repeatability tests as elaborated in Section III-E. The repeatability errors of our method in Fig. 10 show that our method is robust and consistent. The average errors of randomly sampled markers are centered at approximately 1-2 centimeters, which is close to the expected Euclidean distance error from the optimized transformation matrix $\mathbf{T}_{\text{Tracker}}^{\text{Camera}}$.

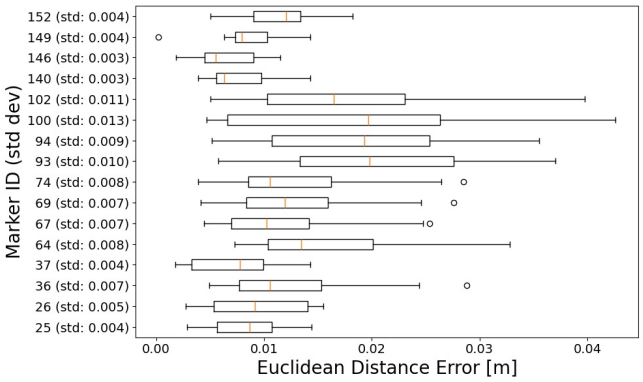


Fig. 10. Repeatability Errors of Our Approach.

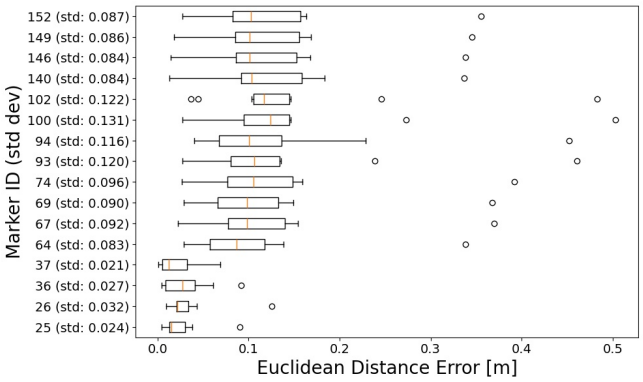


Fig. 11. Repeatability Errors of Pupil Core Software.

In contrast, the Pupil Core software has significant repeatability errors as portrayed in Fig. 11. The markers with ID 25, 26, 36, and 37, which are close to the origin marker in the Pupil Core software explained in Section III-D, have small repeatability errors, while the other markers that are far from the origin marker have high average errors and standard deviations up to over 10 centimeters. This result suggests that the Pupil Core software is not suitable for estimating AprilTag markers in our environment.

As can be seen in Fig. 10, some markers have the average repeatability errors greater than the expected error range such as the markers with ID 93, 94, 100, and 102. We further investigate these errors by examing the relationship between estimation errors and the relative distance of AprilTag markers from the camera as shown in Fig. 12.

Figure 12 indicates that the repeatability errors are proportional to the marker distance relative to the camera.

The errors are centered at around 1-2 centimeters when the marker position relative to the camera is less than 2 meters. On the other hand, the errors become larger when the markers are more than 2 meters away from the camera. Specifically, when the markers are more than 2.5 meters away, the range of repeatability errors become wider. This explains the standard deviation errors of the markers with ID 93, 94, 100, and 102 being larger than other markers in Fig. 10.

Our result is consistent with the findings of a previous study [9] that examined AprilTag marker estimation at varying distances, implying that the closer the camera is to the marker, the more accurate the marker position estimation becomes. In the range of 1.5 to 2 meters, the average errors of the AprilTag marker estimation reported in the previous study are approximately 6-7 centimeters using a Logitech camera and approximately 3-3.5 centimeters using a Raspberry Pi camera. In comparison, our proposed method for estimating the AprilTag marker positions yields lower errors within this distance range, with average errors of approximately 1 centimeter. This allows us to ensure the accuracy of estimating positions and orientations based on the AprilTag markers for additional cameras to be installed in our simulated convenience store environment in the future.

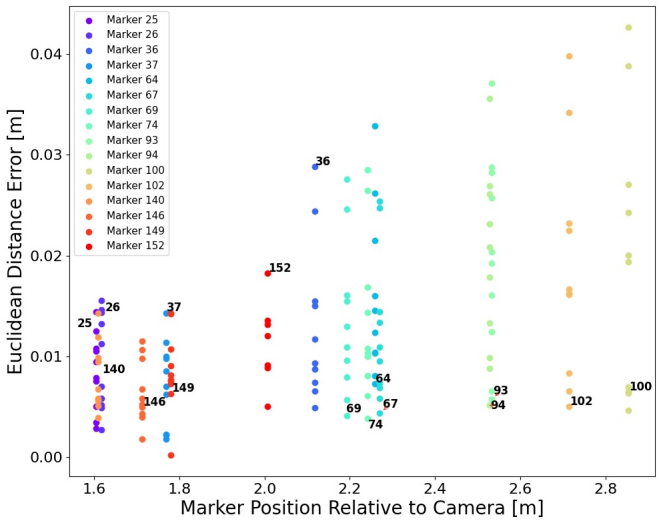


Fig. 12. Relationship between Errors and Marker Position Relative to Camera.

V. Conclusion

In this paper, we presented a method to estimate fiducial marker positions using a motion capture system to enhance the accuracy of marker position estimation. We applied three optimization algorithms to compute the Camera-to-Tracker transformation matrix, achieving average Euclidean distance errors of approximately 1.13 centimeters and average Relative Rotation error of 3.90 degrees. Additionally, we compared our method with the Pupil Core software by estimating the positions of 157 AprilTag markers in our simulated convenience store environment. The repeatability errors suggest that our method is robust and consistent. Finally, we examined the correlation between estimation errors and the positions of markers relative to the camera. This analysis aims to achieve high accuracy in estimating the positions and orientations of additional cameras and sensors to be installed in the environment.

Our future work will explore customer shopping behaviors in our simulated store environment, aiming to attract their attention to advertisements on digital signage. We aim to observe changes in behavior using multiple cameras and sensors to gather various types of data for behavior analysis. This includes skeletons, velocity, and acceleration from the motion capture system; gaze direction and head poses from eye trackers; RGB+D and heatmaps from RealSense cameras; and top-view videos from fisheye cameras. The findings in this paper contribute to our ongoing project, which requires an accurate system integration to precisely capture changes in customer behavior.

References

[1] Kien Nguyen, Minh Le, Brett Martin, Ibrahim Cil, and Clinton Fookes. When ai meets store layout design: a review. *Artificial Intelligence Review*, 55(7):5707–5729, 2022.

[2] Shaphali Gupta and Divya Ramachandran. Emerging market retail: transitioning from a product-centric to a customer-centric approach. *Journal of Retailing*, 97(4):597–620, 2021.

[3] Yuchen Wei, Son Tran, Shuxiang Xu, Byeong Kang, and Matthew Springer. Deep learning for retail product recognition: Challenges and techniques. *Computational intelligence and neuroscience*, 2020(1):8875910, 2020.

[4] Tuan Dinh Nguyen, Keisuke Hihara, Tung Cao Hoang, Yumeka Utada, Akihiko Torii, Naoki Izumi, Nguyen Thanh Thuy, Du Tien Pham, and Long Quoc Tran. Retail store customer behavior analysis system: Design and implementation. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 305–318. Springer, 2024.

[5] Noko Kuratomo, Haruna Miyakawa, Tadashi Ebihara, Naoto Wakatsuki, Koichi Mizutani, and Keiichi Zempo. Attracting effect of pinpoint auditory glimpse on digital signage. *IEEE Access*, 11:42779–42794, 2023.

[6] Matus Tanonwong, Naoya Chiba, and Koichi Hashimoto. Recognition of human relationships using interactions and gazes through video analysis in surveillance footage. In *2023 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1–7. IEEE, 2023.

[7] Moritz Kassner, William Patera, and Andreas Bulling. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14 Adjunct, pages 1151–1160, New York, NY, USA, 2014. ACM.

[8] Edwin Olson. Apriltag: A robust and flexible visual fiducial system. In *2011 IEEE international conference on robotics and automation*, pages 3400–3407. IEEE, 2011.

[9] Michail Kalaitzakis, Brennan Cain, Sabrina Carroll, Anand Ambrosi, Camden Whitehead, and Nikolaos Vitzilaios. Fiducial markers for pose estimation: Overview, applications and experimental comparison of the artag, apriltag, aruco and stag markers. *Journal of Intelligent & Robotic Systems*, 101:1–26, 2021.

[10] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE transactions on pattern analysis and machine intelligence*, 36(7):1325–1339, 2013.

[11] Christian Mandery, Ömer Terlemez, Martin Do, Nikolaus Vahrenkamp, and Tamim Asfour. The kit whole-body human motion database. In *2015 International Conference on Advanced Robotics (ICAR)*, pages 329–336. IEEE, 2015.

[12] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. AMASS: Archive of motion capture as surface shapes. In *International Conference on Computer Vision*, pages 5442–5451, October 2019.

[13] Saeed Ghorbani, Kimia Mahdaviani, Anne Thaler, Konrad Kording, Douglas James Cook, Gunnar Blohm, and Nikolaus F Troje. Movi: A large multi-purpose human motion and video dataset. *Plos one*, 16(6):e0253157, 2021.

[14] Mark Nicholas Finean, Luka Petrović, Wolfgang Merkt, Ivan Marković, and Ioannis Havoutis. Motion planning in dynamic environments using context-aware human trajectory prediction. *Robotics and Autonomous Systems*, 166:104450, 2023.

[15] Brenda Elizabeth Olivas-Padilla, Alina Glushkova, and Sotiris Manitsaris. Motion capture benchmark of real industrial tasks and traditional crafts for human movement analysis. *IEEE Access*, 11:40075–40092, 2023.

[16] Rafael Marques Claro, Diogo Brandão Silva, and Andry Maykol Pinto. Artuga: A novel multimodal fiducial marker for aerial robotics. *Robotics and Autonomous Systems*, 163:104398, 2023.

[17] Mai Trinh, Nikhil V Navkar, and Zhigang Deng. A practical ar-based surgical navigation system using optical see-through head mounted display. In *2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 164–167. IEEE, 2022.

[18] Mark Fiala. Artag, a fiducial marker system using digital techniques. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 590–596. IEEE, 2005.

[19] John Wang and Edwin Olson. Apriltag 2: Efficient and robust fiducial detection. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4193–4198. IEEE, 2016.

[20] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280–2292, 2014.

[21] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

[22] Jorge Nocedal. Updating quasi-newton matrices with limited storage. *Mathematics of computation*, 35(151):773–782, 1980.

[23] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

[24] Michael JD Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The computer journal*, 7(2):155–162, 1964.

[25] Jorge Nocedal and Stephen Wright. *Numerical Optimization*. Springer Science & Business Media, 2006.

[26] Peter J Huber and Elvezio M Ronchetti. *Robust statistics*. John Wiley & Sons, 2011.