UNB EST. 1785 UNIVERSITY OF NEW BRUNSWICK | Off-site Construction Research Centre

Northumbria University NEWCASTLE

Transforming Construction with Off-site Methods and Technologies (TCOT) Conference: Designing Tomorrow's Construction, Today

August 20-22, 2024, Fredericton, New Brunswick, Canada

ENHANCING BUILT ENVIRONMENT MANAGEMENT: A VISION-BASED APPROACH UNIFIED WITH FIDUCIAL MARKERS AND OMNIDIRECTIONAL CAMERA POSE ESTIMATION

Taherian, Gelare[1*], Rezazadeh Azar, Ehsan[1]
[1] Department of Architectural Science, Toronto Metropolitan University, Canada [*]
gelare.taherian@torontomu.ca

Abstract: Construction jobsites are characterized by their dynamically changing environment driving the need for improved managerial methods, such as automated progress tracking and quality inspection. These applications rely on detecting physical changes, which necessitate accurate data capturing technologies and analysis of as-is against the as-planned status. Vision-based methods have emerged as promising tools for localizing the real-time query thanks to the availability of low-cost data capturing, transferring, storing, and computing systems. However, they should overcome certain challenges to reliably detect and locate construction resources and building elements in frames, especially in complex indoor environments. Despite significant advancements, recent approaches still require considerable effort to perform reliably, such as creating comprehensive 3D point clouds through frequent 3D reconstruction using overlapping images. This paper proposes a new approach inspired by recent innovations in computer science which can reduce the data capturing efforts needed in other approaches. The method proposes the use of omnidirectional images and detecting the fiducial markers in the generated visualizations to capture the asis status and further retrieve the relative pose in the as-planned status through Building Information Modeling (BIM). This approach facilitates comparative analysis through state-of-the-art computer visionbased object detection and classification methods for change detection between the as-is query and the as-planned status across a wide view. Additionally, it moderates camera pose estimation efforts, enhancing efficiency for various built environment management applications, including construction progress tracking.

Keywords: Vision-based Localization, Camera Pose Estimation, Marker-based, Omnidirectional View, Indoor Environment**.**

1 INTRODUCTION

Detecting changes in the constantly evolving built environment, especially within Architectural, Engineering, and Construction (AEC) contexts, is crucial for various operational and managerial purposes, such as construction progress monitoring and documentation. Indoor built environment management (IBEM) encompasses the concept of visual-comparative analysis of the as-is vs as-planned status, for which the estimation of the user's location within the indoor environment is an integral element to accurately capture

the as-is status and correctly detect the objects of interests. Indoor localization refers to identifying the precise location and orientation of objects or individuals in an indoor environment. The as-planned information is typically derived from the building information modelling (BIM). The as-is status is mainly captured manually and there are some ongoing research and development to capture it by the help of ground robots and UAVs. Despite advancements in vision-based camera pose estimation (CPE) technologies, existing methods often require high amount of data and efforts and still struggle with accurately localizing the camera in complex indoor environments. Marker-based CPE methods are viable solutions to the current challenges. Fiducial markers are the reference points, creating an artificial landmark to recognise, detect and localize different objects within the environment (Yu et al. 2019). These markers typically come in various forms, with known size and shapes, ranging from simple patterns to more complex designs that include encoded IDs or messages (Kalaitzakis et al. 2021). Marker-based CPE methods have received recognition in recent years owing to their ease of use and reliable accuracy. They alleviate the CPE process by only requiring the detection of the four corners of a single square-shaped or the center of a circular-shaped fiducial marker. The main advantage of the fiducials is undoubtedly their extremely low cost and the need for only a calibrated camera (Ulrich et al. 2022). However, their exploration and evaluation have been largely centred around their application as 2D planar and non-planar artificial landmarks in environments captured in 2D imagery, which requires a considerable number of images with markers present in them for pose estimation. Though limited research studies (Hajjami et al. 2020; Adapa et al. 2023) have explored non-planar marker detection in 2D/3D images, the proposed methods are developed based on employing specific devices or equipment for holding non-planar markers. Additionally, to the best of authors' knowledge, no research work has focused on implementing omnidirectional imaging for CPE specifically in dynamic and comprehensive AEC industry. Consequently, there remains a significant opportunity for further investigation into fiducial markers detection as 2D planar markers within the context of omnidirectional images, also known as 360-degree images, enabling a comprehensive view of the environment through a single capture. This research aims to enhance built environment visual assessment by streamlining the capture of the environment's current status through the novel integration of fiducial markers with omnidirectional CPE. The objectives are to: (1) reduce data capturing efforts, (2) improve camera localization efficiency in indoor settings. This research avenue is essential for addressing the persistent challenges and limitations encountered in existing approaches to marker-based CPE, including issues with detection under occlusion and constraints related to the detectable distance range, even in conventional scenarios involving flat markers in 2D images.

2 MARKER-BASED CPE WITH OMNIDIRECTIONAL VIEWS FOR IBEM

The position and orientation of the camera can be estimated using the pose of a detected marker within an image with respect to the marker’s reference system. This representation involves six degrees of freedom and employs diverse formats to communicate alterations in both position and orientation, covering translation and rotation (Xu et al. 2023). Therefore, the pose can be expressed as a combination of two components:

1) Translation vector: a tuple of three elements that identifies the absolute coordinates x, y, and z in a reference space (R)

2) Rotation matrix: a 3*3 matrix that is represented using Euler angles, where the tuple contains three values corresponding to the rotations around the three principal axes (e.g., roll, pitch, and yaw) in a reference space (R)

2.1 Marker-based CPE

The typical process for pose estimation using square-based markers involves embedding any geometric shape or pattern within a square border, which serves as the marker's ID. This distinctive pattern aids in distinguishing between multiple markers. Square-shaped fiducial markers, primarily established on the foundation of ARToolkit (Kato and Billinghurst 1999), stand as a prevalent choice in marker-based CPE.

ARToolkit employs a global threshold to identify regions that can be delineated by four-line segments. Evolved from ARToolkit, ARTag (Fiala 2005) incorporates digital coding theory to enhance marker reliability, facilitating detections under conditions of partial occlusion. ArUco (Garrido-Jurado et al. 2014), a derivative of ARTag, introduces a significant advancement by allowing users to generate customized marker libraries tailored to specific application requirements. This customization capability has led to the development of smaller marker libraries and reduced computational overhead, enhancing efficiency and flexibility in marker-based applications. In addition to these established frameworks, methodologies such as DeepTag (Zhang et al. 2023) have emerged to broaden the capabilities of markers in CPE, supporting the detection of a wide variety of existing marker families and designing new marker families with customized local patterns. Once a marker is successfully detected within an image, its pose relative to the camera reference system can be estimated, followed by subsequent computational steps to determine the pose of the camera with respect to the marker's reference system.

2.2 The Contribution of Omnidirectional View to IBEM and CPE

State-of-the-art methods for CPE with fiducial markers rely on cameras with a limited Field of View (FOV), leading to tracking interruptions when markers become occluded or go out of sight. In contrast, 360-degree cameras offer distinct advantages over regular perspective cameras by providing an omnidirectional view of a scene rather than a narrow FOV. This omnidirectional view facilitates the extraction of comprehensive information from the entire indoor environment in a single capture, demonstrating invaluable for systems that require a detailed understanding of their surroundings, such as analyzing construction progress for IBEM. However, the adoption of 360-degree cameras presents challenges due to distortions inherent in the captured images, impeding accurate marker detection compared to images captured from a perspective view. Therefore, there is active research about employing cameras with an enhanced FOV for improved marker detection. Hajjami et al. (2020) investigated the detection of non-planar (polygonal) ArUco markers in omnidirectional images, employing a spherical representation of the image data. Images were captured using a 3D rig of markers to create a dataset. In contrast, our research focuses specifically on detecting and identifying 2D planar markers in omnidirectional images captured within indoor environments, eliminating the need for additional equipment and setups for marker installation.

3 METHODOLOGY

The methodology revolves around a structured experiment where an omnidirectional camera was used to capture an indoor built environment. Fiducial markers (ArUco and DeepTag) were strategically placed to facilitate the detection and localization processes. Accordingly, the empirical test setup involved the installation of six 2D planar square-shaped markers, comprising four ArUco and two DeepTag markers, in different locations within an office environment, covering an area of 6.5m × 8.5m. The markers were generated with random IDs including 16, 23, 45, and 68 from the ArUco and 323 from the DeepTag libraries. Considering that marker detection is influenced by the distance of the camera from the markers, the camera was held within a fix height as 150 cm and positioned relatively close to the markers to capture a 360-video, ensuring that the distance parameter, which is not under the scope of current research, was adequately fulfilled. This research used the Insta360 ONE X2 camera and Insta360 Studio software, which facilitated the extraction of snapshots with different FOVs from a specific frame. In response to the need for distortion coefficient estimation in CPE, the camera was calibrated using the OpenCV pipeline. The calibration was performed under standard conditions; however, the calibration was uniformly applied across different FOVs due to practical constraints in dynamically adjusting distortion coefficients.

Employing 360-degree visualizations, various perspectives within the environment were examined, including snapshots with different FOVs (Narrow, Linear, Linearplus, Wide, and Ultrawide) and images extracted from the video recording of the environment in forms of Flat and 360-degree images. To facilitate frame extraction from the video, the 360-degree video was first extracted into Flat and 360-degree video formats. Subsequently, frames were extracted from each exported video at a rate of 10 frames per second

(fps). These visualizations were rigorously evaluated using the ArUco marker detection pipeline in OpenCV. Marker detection was performed with the detectmarker() function, while the camera's position relative to the marker was determined through a three-dimensional transformation with solvePnP() method encompassing rotation and translation vectors (OpenCV 2015).

4 EXPERIMENTAL RESULTS

It is important to note that while ArUco is primarily designed for ArUco marker detection, it also exhibits robustness in detecting other marker packages, including DeepTag. This versatility is evident where successful detections of both ArUco and DeepTag markers are showcased in Linearplus image type, as outlined in Table 1.

Table 1: the results of marker detection and estimated poses in different image types

Data		Initial detection		Error						
Image Type	FOV	Marker	Detection	ID	tx	ty	tz	Rx	Ry	Rz
Video -to-image	360	DeepTag	0%	343	No marker detected					
		ArUco	75%	23	No marker detected					
				45	CPE failed					
				16	CPE failed					
				68	CPE failed					
Video -to-image	Flat	DeepTag	0%	343	No marker detected					
		ArUco	100 %	23	CPE failed					
				45	0.10	0.31	0.81	2.11	0.85	0.59
				16	0.09	0.87	0.74	2.11	3.90	0.27
				68	0.35	0.12	0.26	2.71	3.08	0.31
Snapshot	Narrow	ArUco	100 %	23	CPE failed					
				45	0.04	0.22	0.87	3.11	2.12	0.61
				16	0.05	0.38	1.01	2.10	3.92	0.26
				68	0.29	0.04	0.31	5.86	0.51	0.36
Snapshot	Linear	ArUco	100 %	23	0.20	0.04	1.36	1.03	3.54	0.13
				45	0.17	0.45	0.68	3.17	2.06	0.60
				16	0.08	0.69	0.84	2.10	3.89	0.26
				68	0.40	0.23	0.14	5.87	0.51	0.34
Snapshot	Linearplus	DeepTag	100%	343	CPE failed					
		ArUco	100 %	23	0.19	0.04	1.31	1.03	3.56	0.15
				45	0.23	0.56	0.59	3.12	2.11	0.65
				16	0.06	0.81	0.77	2.14	3.89	0.22
				68	0.45	0.33	0.06	5.87	0.53	0.29
Snapshot	Wide	ArUco	100 %	23	CPE failed					
				45	0.14	0.40	0.72	3.11	2.13	0.61
				16	0.04	0.62	0.88	2.10	2.09	0.24
				68	0.38	0.19	0.18	5.14	0.52	0.65

Snapshot	Ultrawide	DeepTag	0%	343	No marker detected					
		ArUco	100 %	23	0.19	0.04	1.30	1.02	3.19	0.07
				45	0.24	0.59	0.56	3.08	2.12	0.62
				16	0.06	0.90	0.72	2.09	2.12	0.24
				68	0.47	0.36	0.03	5.09	0.52	0.68

* tx, ty, tz values represent the position error in meters and Rx, Ry, Rz values represent the rotation error in Euler angels.

Although DeepTag markers were not consistently detected in all image types, their detection, particularly in certain FOVs such as Linearplus, underscores the significant impact of FOVs types extracted from 360degree images on detection results as well as the position of the maker within the images. In contrast, ArUco markers exhibited consistent detection across all image types. Namely, image types with wider FOVs, like Flat video-to-image extraction, Linearplus, Wide, and Ultrawide, consistently resulted in successful detections for ArUco markers. However, the detection rate for these markers in 360 video-toimage extraction format was 75%, indicating the potential of failure in this FOV. These results highlight the provided chance of marker detection in images with broader FOVs compared to regular perspective images. However, there remains room for further exploration and improvement to enhance the robustness to detection of all types of marker packages in addition to increasing the rate of detection in image types with the wider view, such as 360 video-to-image extracted image, in which the markers may not entirely be detected due to high rate of distortions.

The scene featured four ArUco markers, each positioned on separate walls parallel to the camera lens. According to Table 1, the estimated pose errors represent consistent results across different image types with minor variations. Marker 16 exhibited minor variations in all image types except for orientation errors in Wide and Ultrawide images. Marker 23 showed the most consistency in estimated poses across various image types. Lastly, markers 45 and 68 displayed reasonable consistency in estimations among all image types, though with variations in rotation errors in flat images. Discrepancies in estimation errors, especially in rotation, may be attributed to marker position in the images and varying conditions in the environment, such as the lighting or the camera distance from the markers. Table *2* illustrates the average errors for CPE in each image type in which the markers were detected and CPE was successful. The errors are ranging from 0.13 to 0.75 meters for position and 0.33 to 3.69 degrees for rotation. While mean errors suggest minor variations overall, errors in Rx and Ry are significantly higher, which make them worth to be explored further. Additionally, CPE with the 360-degree image type was unsuccessful for all markers, indicating the need to address distortions in wide visual representations to enhance marker detection and CPE accuracy.

Table 2: The average pose estimation errors (in meter for position and Euler angels for rotation) according to the detected markers in different image types

Image Type	|Δtx|	|Δty|	|Δtz|	|ΔRx|	|ΔRy|	|ΔRz|
Flat	0.18	0.43	0.60	2.31	2.61	0.39
Linear	0.13	0.21	0.73	3.69	2.18	0.41
Linearplus	0.21	0.35	0.75	3.04	2.50	0.33
Narrow	0.23	0.43	0.69	3.04	2.52	0.33
Ultrawide	0.18	0.40	0.59	3.45	1.58	0.50
Wide	0.24	0.47	0.66	2.82	1.99	0.40
Average	0.20	0.38	0.67	3.06	2.23	0.39

5 CONCLUSION

The experiments yielded results in two key aspects: marker detection rate and estimated camera pose for the extracted frames. Preliminary findings indicated promising marker detection rates with the provided omnidirectional input. Despite the overall success of marker detection, challenges persist, particularly in

image types with extensive distortions, such as 360 video-to-image extracted images. In these cases, markers are not consistently and accurately detected due to high levels of distortion, necessitating further exploration of 360-degree image rectification methods. The consistent results of CPE in this research represent a significant contribution to the comprehensive and automated visual assessment of built environments, particularly within the field of construction management. Key tasks such as progress monitoring and quality control require an accurate understanding of the project's as-is status. The proposed method provides a broad visualization of the actual environment, eliminating the high amount of data capturing and efforts for CPE to acquire the as-is status, such as the need for multiple camera setups. Accordingly, it facilitates direct visual comparisons with the as-planned status using the advanced computer vision techniques. While minor discrepancies are reflected in CPE results, future research on developing adaptive calibration methods to encounter practical uniform implementation of calibration results across different FOVs can enhance the CPE accuracy. Additionally, investigating CPE for sequences from the video that do not include markers will be a focal point for further exploration. These explorations are especially crucial in indoor environments, which have more challenges compared to outdoor settings, including clutter, similar patterns, and limitations in lighting conditions. Addressing these challenges will be pivotal in enhancing the robustness and accuracy of marker detection in omnidirectional images, thereby advancing the efficacy of built indoor environments management.

6 ACKNOWLEDGEMENTS

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of a funding agency.

7 REFERENCES

Adapa, D, VS Shekhawat, A Gautam, and S Mohan. 2023. Autonomous Mapping and Navigation Using Fiducial Markers and Pan-Tilt Camera for Assisting Indoor Mobility of Blind and Visually Impaired People, October.

Fiala, M. 2005. ARTag, a Fiducial Marker System Using Digital Techniques. In *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*. Vol. II. doi:10.1109/CVPR.2005.74.

Garrido-Jurado, S, R Muñoz-Salinas, FJ Madrid-Cuevas, and MJ Marín-Jiménez. 2014. Automatic Generation and Detection of Highly Reliable Fiducial Markers under Occlusion. *Pattern Recognition* 47 (6). doi:10.1016/j.patcog.2014.01.005.

Hajjami, J, J Caracotte, G Caron, and T Napoleon. 2020. ArUcOmni: Detection of Highly Reliable Fiducial Markers in Panoramic Images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Vol. 2020-June. doi:10.1109/CVPRW50498.2020.00325.

Kalaitzakis, M, B Cain, S Carroll, A Ambrosi, C Whitehead, and N Vitzilaios. 2021. Fiducial Markers for Pose Estimation: Overview, Applications and Experimental Comparison of the ARTag, AprilTag, ArUco and STag Markers. *Journal of Intelligent and Robotic Systems: Theory and Applications* 101 (4). doi:10.1007/s10846-020-01307-9.

Kato, H, and M Billinghurst. 1999. Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System. In *Proceedings - 2nd IEEE and ACM International Workshop on Augmented Reality, IWAR 1999*. doi:10.1109/IWAR.1999.803809.

OpenCV. 2015. Detection of ArUco Markers. Open Source Computer Vision Library. https://docs.opencv.org/4.x/d5/dae/tutorial_aruco_detection.html.

Ulrich, J, A Alsayed, F Arvin, and T Krajník. 2022. Towards Fast Fiducial Marker with Full 6 DOF Pose Estimation. In *Proceedings of the ACM Symposium on Applied Computing*. doi:10.1145/3477314.3507043.

Xu, M, Y Wang, B Xu, J Zhang, J Ren, Z Huang, S Poslad, and P Xu. 2023. A Critical Analysis of Image-Based Camera Pose Estimation Techniques. *Neurocomputing*, 127125. doi:10.1016/J.NEUCOM.2023.127125.

Yu, G, Y Liu, X Han, and C Zhang. 2019. Objects Grasping of Robotic Arm with Compliant Grasper Based on Vision. In *ACM International Conference Proceeding Series*. doi:10.1145/3351917.3351958.

Zhang, Z, Y Hu, G Yu, and J Dai. 2023. DeepTag: A General Framework for Fiducial Marker Design and Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45 (3). doi:10.1109/TPAMI.2022.3174603.