ICCEPM 2022
The 9th International Conference on Construction Engineering and Project Management
Jun. 20-23, 2022, Las Vegas, NV, USA

AprilTag and Stereo Visual Inertial Odometry (A-SVIO) based Mobile Assets Localization at Indoor Construction Sites

Rabia Khalid[1]*, Muhammad Khan[2], Sharjeel Anjum[3], Junsung Park[4], Doyeop Lee[5], Chansik Park[6]

[1,3,4,5,6]*Construction Technology & Innovation Laboratory (ConTIL), Department of Architectural Engineering, Chung-Ang University, Seoul, South Korea,* E-mail addresses: rabiakhalid@cau.ac.kr[1], sharjeelanjum@cau.ac.kr[3], qkrwnstjd0@cau.ac.kr[4], doyeop@cau.ac.kr[5], cpark@cau.ac.kr[6]
[2]*Construction Innovation Integration Laboratory (CII-Lab), Department of Civil, Construction and Environmental Engineering, University of Alabama, USA,* E-mail address: mkhan13@crimson.au.edu[2]

Abstract: Accurate indoor localization of construction workers and mobile assets is essential in safety management. Existing positioning methods based on GPS, wireless, vision, or sensor based RTLS are erroneous or expensive in large-scale indoor environments. Tightly coupled sensor fusion mitigates these limitations. This research paper proposes a state-of-the-art positioning methodology, addressing the existing limitations, by integrating Stereo Visual Inertial Odometry (SVIO) with fiducial landmarks called AprilTags. SVIO determines the relative position of the moving assets or workers from the initial starting point. This relative position is transformed to an absolute position when AprilTag placed at various entry points is decoded. The proposed solution is tested on the NVIDIA ISAAC SIM virtual environment, where the trajectory of the indoor moving forklift is estimated. The results show accurate localization of the moving asset within any indoor or underground environment. The system can be utilized in various use cases to increase productivity and improve safety at construction sites, contributing towards 1) indoor monitoring of man machinery coactivity for collision avoidance and 2) precise real-time knowledge of who is doing what and where.

Key words: Stereo Visual Inertial Odometry, April Tags, Indoor Positioning, Construction Safety

1. INTRODUCTION

The number of worker safety-related accidents reported from the construction industry has been a global concern [1]. The construction workforce accounts for 7% of the global workforce, which reports 30% of the accidents [2]. The US Bureau of Labor Statistics reported a 2% increase in fatal occupational injuries in 2019, the largest annual number reported since 2017 [3]. Many construction accidents are mainly reported due to falls from height, being struck by an object, electrocution, and struck between or in equipment. Furthermore, approximately 200,000 workers have disabled annually because of work-related injuries at the construction site [4]. Due to the complex, confined, and dynamic working environment at the construction site. As construction work progresses, new hazards occur; manual hazard recognition monitoring is almost impossible for the workers and supervisor on site [5], [6]. Scientists reported that roughly 50% of the hazards at the construction are unrecognized[7], [8]. As a result, the workers are exposed to unseen risks and potential injuries. Safety training programs have been designed and conducted to improve the ability of the works to identify the potential risk at the construction site [9]. Although beneficial, those training methods have not successfully minimized the unseen risk to the desirable levels. Furthermore, according to Cormwell et al. [10], only 10-15% of training investment result in favorable workplace results.

Recent research has shown personalized training approaches to increase risk awareness among construction workers[9], [11]. Martin et al. [12] conducted a five-hour safety training session on fall prevention, dust exposure, leadership, safety planning, and communication for 118 workers. After six months, safety managers conducted an assessment survey, which revealed a considerable improvement in effective communication among workers and greater adherence to safety laws to reduce the risk of accidents. However, there is a significant disparity in the number of safety managers to workers; therefore, educating individual personnel is impractical.

Even if aforementioned challenges are resolved, workers might still fail to identify certain unseen hazards. According to Albert et al. [8] and Jeelani et al. [7] , apart from the safety training, workers still fail to recognize one-fourth of the risk associated with safety. Risk recognition is a visual process that can be affected by factors such as complex and dynamic layout, occlusion, blind zone, and illuminance etc. [13], [14]. Therefore, only training cannot recognize these unseen risks that a worker is exposed to. Hence, there is a need for automation techniques to identify the workers positioning at the complex and dynamic construction site and the risk associated with it.

2. RELATED RESEARCH WORK

Existing research on construction safety has boosted worker safety and visibility for emergency response by deploying wireless intelligence. Indoor and underground working environments are global positioning systems (GPS) [15], [16] denied. This led to the development of wireless technologies like wireless local area networks (WLAN)[17], radio frequency identification (RFID)[18], ultrawideband (UWB)[19], Bluetooth (BLE)[20], Ultrasound[21], etc. However, these require dedicated hardware installation, which is expensive for large-scale environments. Vision-based real-time locating systems (RTLS) are also used in navigation; however, it is inaccurate as it depends on pre-installed cameras that do not handle occlusion and need intensive computations for Structure for Motion (SfM). 2D LIDAR simultaneous localization and mapping (SLAM)[22] and visual SLAM[23], referred to as the process of position and orientation estimation w.r.t. surrounding using LIDAR and camera, simultaneously map the environment. However, 2D LIDAR SLAM needs the input of initial pose estimations and hence is not reliable when there is no prior pose estimation information. Whereas vSLAM is vulnerable to variations in illumination and motion blur when the camera moves too fast.

The infeasibility in the usage of standalone sensors led to the senor fusion methods, which couple multiple sensors to mitigate standalone weaknesses. Georgy et al. [24] integrated MEMS-based inertial measurement units (IMU) with GPS and used Mixture Particle Filter such that the former limits the positional error in the latter during its outage. However, this approach fails when GPS is unavailable in an indoor environment or is jammed. Jiang et al. [25] used cameras in visual odometry (VO) as it is relatively cheaper than GPU and IMU. VO is a part of vSLAM, which analyzes a sequence of images to determine the location and orientation of the camera. Various VO algorithms have been tested on monocular and stereo cameras, where stereo VO reports higher accuracy than monocular VO. To counter the high scene dependence and computation complexity of standalone VO, Li et al. [26] proposed visual-inertial odometry (VIO), which integrates IMU with a monocular camera as visual and inertial odometry complement each other. Inertial Odometry (IO) has high data rate, is environment independent and gravity aware; however, it has high positional drift due to double integration of accelerometer data. This drift is corrected by the camera, where stereo camera corrects the drift better than monocular cameras due to additional depth information.

In recent years, the potential of fiducial landmarks is identified for indoor localization and navigation. Nahangi et al. [27] proposed a method for automated localization of UAVs in GPS-denied environments by employing AprilTags linked to the 3D coordinates of the building. These coordinates are the ground truths known from the building information model (BIM). The UAV localizes itself as long as the AprilTags are within its camera line of sight. Kayhani et al. [28], [29] improved this tag-based localization by using an extended Kalman filter (EKF) and switching to IMU when UAV misses AprilTags. However, being prone to positional drift, IMU doesn't serve as a reliable system, and hence more robust VIO systems need to be integrated. Kayhani et al. [30] integrated a more robust VIO system with AprilTags presenting a system with high accuracy, and root mean square error within 2-5cm. However, in tag-blind zones, the system accuracy relies on odometry-based estimates. Therefore, a more robust, and reliable odometry method of SVIO is proposed to be integrated with AprilTags. Hence, not only providing better accuracy but also decreasing the number of tag installation in the construction site.

3. OBJECTIVE AND SCOPE

This study proposes a conceptual framework of an integrated model of AprilTag fiducial markers and Stereo Visual-Inertial Odometry (A-SVIO) for accurate indoor localization of mobile assets. AprilTag is a 2D bar code style tag allowing complete 6-DOF localization of features from a single image. It also enables the transformation of the SVIO based positional coordinates to the coordinates relative to decoded AprilTag ground truth position. The proposed solution helps the construction sites to prevent collisions, bring visibility into emergency mustering and provide lone worker safety. While the solution prioritizes safety, it also boosts efficiency and productivity at construction sites. The proposed solution is tested on a virtual environment created on NVIDIA ISAAC SIM utilizing ISAAC SDK for moving asset localization and ISAAC Sight to visualize results.

4. PROPOSED FRAMEWORK

The proposed framework of A-SVIO is illustrated in Figure 1. This framework is implemented in the virtual simulation environment of NVIDIA ISAAC SIM. A small warehouse environment is created in this study. AprilTags linked with localization coordinates are assumed to be installed in the entrance points of this indoor working environment. In a practical scenario, these tags can be generated by linking them with the coordinates in the BIM model and then installed on that position as a fiducial landmark. For performing SVIO, an Intel RealSense Stereo Camera 435i, including IMU embedded with NVIDIA Jetson Nano Kit, is installed on the moving asset (such as workers, UAVs, robodog, or machinery like excavators, fork lifters etc.). In this study, the model is tested on simulated virtual indoor warehouse environment containing a moving fork lifter.

For the processing of the proposed solution, NVIDIA ISAAC SDK is utilized. It monitors the pre-calibrated stereo camera and performs SVIO to generate a trajectory of moving asset motion relative to the starting point. AprilTags placed at the starting point are decoded to transform the SVIO computed coordinates to absolute global coordinates.

After accurately localizing the moving asset, the trajectory can be visualized by the site manager on NVIDIA ISAAC Sight or the BIM Model.

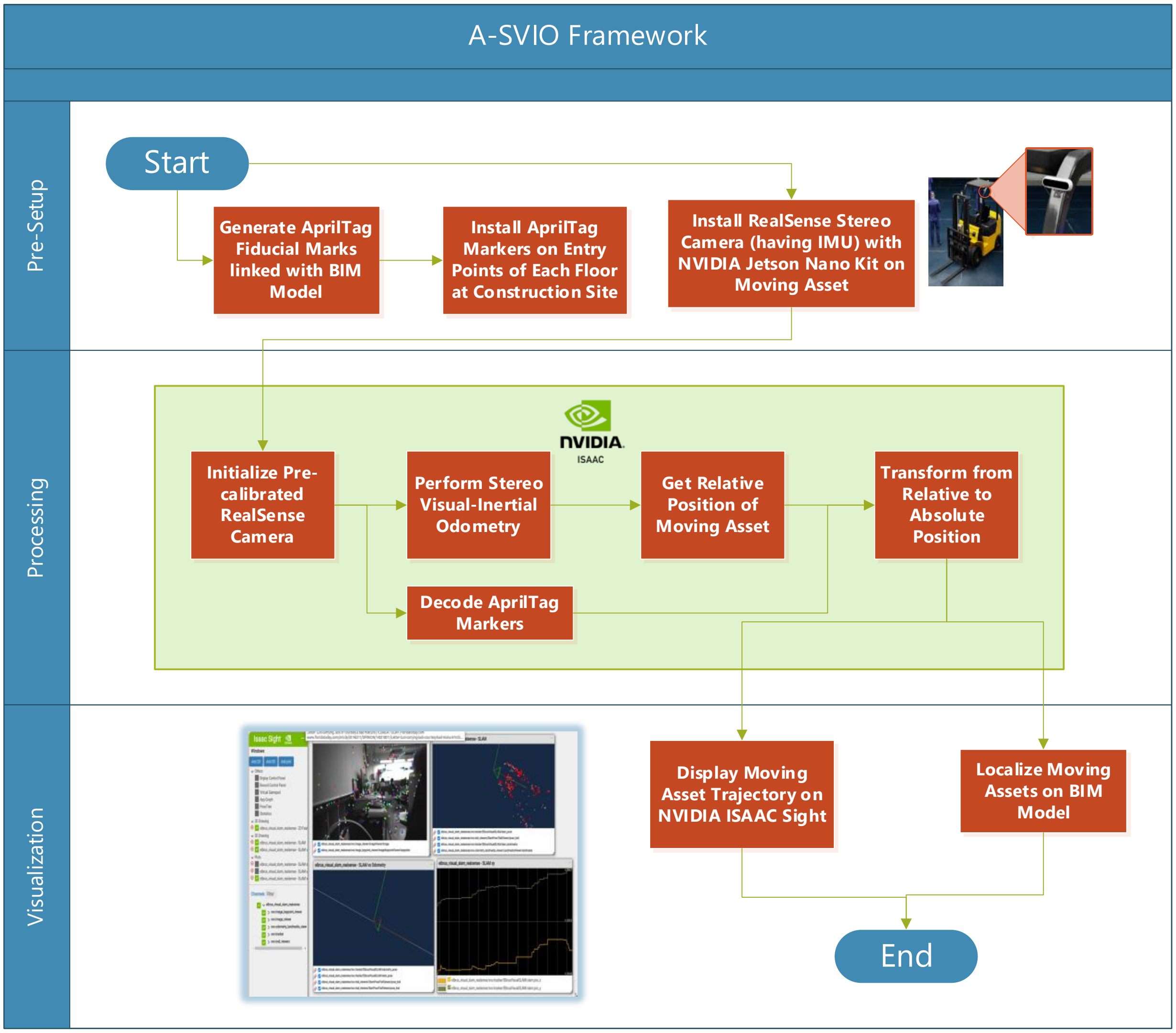

Figure 1. Proposed framework of A-SVIO based positioning system

5. EXPERIMENTATION

The proposed framework is experimented and validated on a virtual indoor warehouse environment, created using NVIDIA ISAAC SIM. The configuration parameters, pre-setup requisites, processing, and visualization of results obtained are explained in this section.

5.1 Configuration Setup

The development environment compatible for this experiment is Ubuntu 18.04 with an NVIDIA Graphic Card version greater than GTX 1080 and driver version greater than 440. The development environment configuration parameters used in this study are shown in Table I.

Table 1. Configuration requirements of development environment

Development Environment	Requirements
Ubuntu OS	18.04
Graphic Driver	NVIDIA GeForce RTX 3080 with NVIDIA driver version 495
SDK	NVIDIA ISAAC

The experimental setup used for the demonstration of this solution is explained in Table II.

Table 2. NVIDIA ISAAC SDK setup tasks

SDK Configuration	Setup Used
NVIDIA ISAAC SIM	• Virtual indoor warehouse environment containing a moving asset (i.e., forklifter). • Intel RealSense 435i (with IMU) with NVIDIA Jetson Nano Kit installed on forklifter. • AprilTag fiducial markers placed on the entry point.
NVIDIA ISAAC Sight	• Elbrus SVIO trajectory on realsense stereo camera feed. • Trajectory transformation to global absolute position from decoded AprilTags.

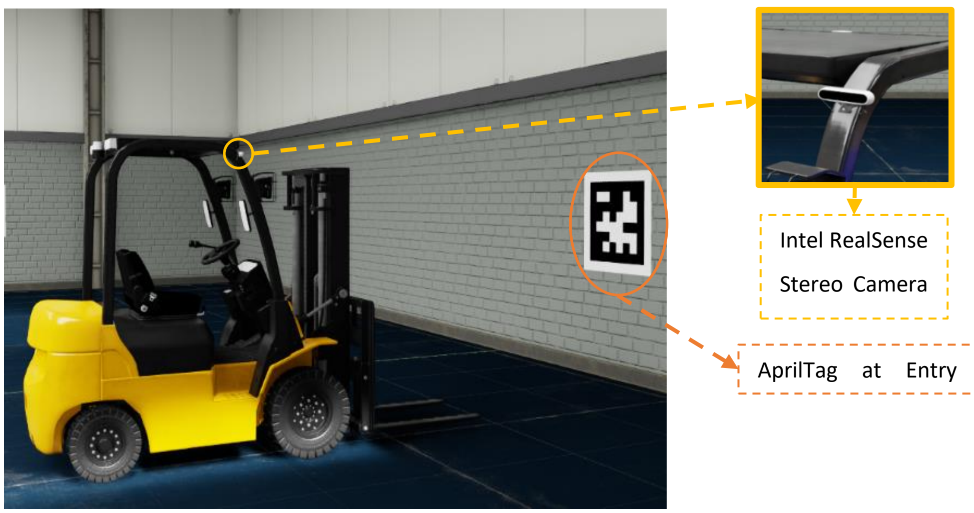

Figure 2. NVIDIA ISAAC SIM virtual indoor warehouse environment for testing of A-SVIO

5.2 Processing

5.2.1 AprilTag Decoding

AprilTags are the most robust and feasible visual fiducial markers proposed by Olson [31] in 2011. These allow 6DOF localization in a single image frame by estimating the camera pose relative to the decoded tag. The tag linked with the BIM model can provide further insights into the indoor environment like floor number, unit number, etc. Installation of tags in the site is prone to errors; however, the installation error can be corrected by utilizing depth information of the stereo camera and the BIM-linked information. Hence, resulting in a more accurate estimate of the camera position and orientation with respect to the tag.

5.2.2 SVIO Processing

NVIDIA ISAAC SDK includes the SVIO codelet that utilizes the built-in Elbrus Visual Odometry library [33], [34] to determine the 3D pose of the moving asset from the video stream of stereo camera and its IMU readings. The library performs real-time with an average detection speed of 30 fps on three cores @ 3.3 GHz and 144 fps on Jetson AGX for 640x480 video resolution. The algorithm reports a drift of ~1% in localization and 0.003 degrees/meter of rotational error in motion on the KITTI benchmark dataset.

Elbrus SVIO utilizes inertial data from IMU embedded inside the stereo camera. If distortion in input images fails visual odometry, the IMU data estimates the positional trajectory for up to a second. The integration of IMU with vision updates pose seamlessly, provided the video interruption lasts for less than a second.

5.2.3 Positional Transformation

Stereo cameras acquire images from two synchronized cameras with known calibration intrinsic. SVIO uses this imaging data to retrieve the 3D pose of the left camera relative to its starting position. For the IMU integration to work with SVIO, it is required that the stereo camera baseline is in a horizontal position at the beginning; otherwise, the initial pose and gravitational acceleration vector are not maintained correctly by Elbrus SVIO.

The 3D trajectory computed by the SVIO is relative to the started point. This is transformed to an absolute global position by utilizing the decoded AprilTag information. It is assumed that the moving asset activates localization computation at the marked entry points, where the embedded stereo camera decodes at least one AprilTag to transform the SVIO position to absolute global coordinates. Every time the moving asset sees an AprilTag, the positional drift (if any) is corrected. Moreover, the most reliable and robust SVIO based localization allows for fewer tag installations in the construction site.

6. RESULTS

The 3D trajectory computed by Elbrus SVIO is visualized on NVIDIA ISAAC Sight. The tag transformed position is the absolute global position, which can be displayed either on ISAAC Sight or on the BIM model of the floor number retrieved from decoded tag. Figure 3 shows the transformed position trajectory of moving asset on the ISAAC Sight.

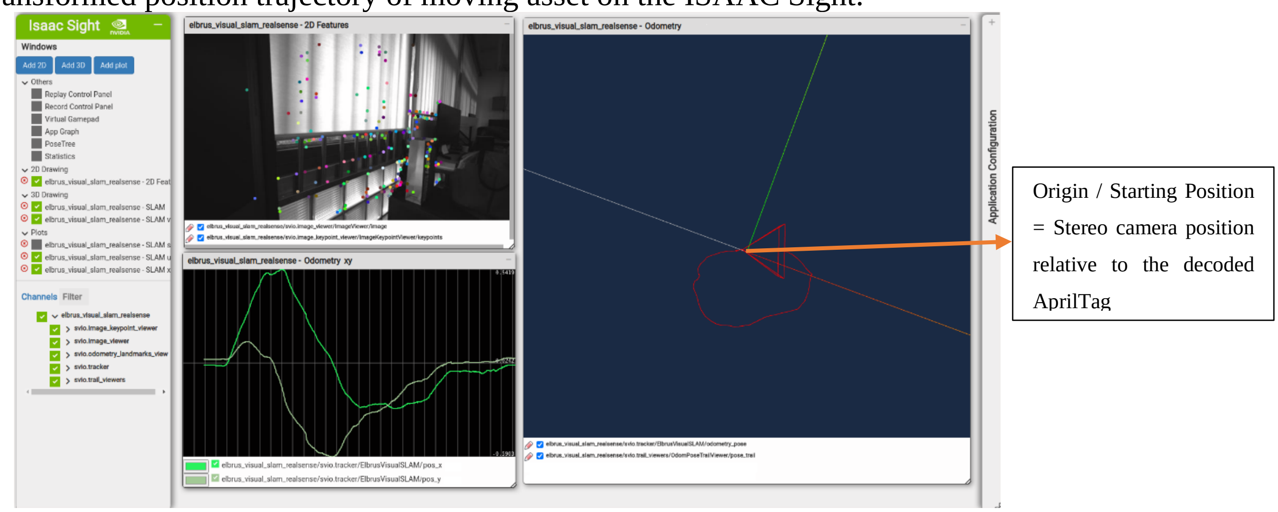

Figure 3. Trajectory of the moving fork lifter visualized on NVIDIA ISAAC Sight

7. CONCLUSTION & FUTURE WORK

In this study, a framework is proposed and experimented with a virtual simulation environment to localize a moving asset in GPS denied indoor environment. Recent research studies on indoor localization integrate tag with the visual-inertial odometry system; however, this odometry does not serve as a reliable approach in the tag-blind zone. Therefore, more reliable stereo visual-inertial odometry is proposed to be utilized. This not only increases the accuracy of localization but also allows for fewer tag installations in the construction site. The proposed methodology is experimented with a simulated environment utilizing open-source NVIDIA ISAAC.

The proposed system will be further tested in the real indoor construction site to validate the performance of the proposed A-SVIO approach. This methodology aims to improve safety and increase productivity at construction sites, contributing towards 1) indoor monitoring of man machinery coactivity for collision avoidance and 2) precise real-time knowledge of who is doing what and where.

ACKNOWLEGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (No. 2021R1A2C2014488).

REFERENCES

[1] B. Manzoor, I. Othman, J. C. Pomares, and H. Y. Chong, "A research framework of mitigating construction accidents in high-rise building projects via integrating building information modeling with emerging digital technologies," *Applied Sciences (Switzerland)*, vol. 11, no. 18, 2021, doi: 10.3390/app11188359.

[2] M. T. Newaz, P. R. Davis, M. Jefferies, and M. Pillay, "Developing a safety climate factor model in construction research and practice: A systematic review identifying future directions for research," *Engineering, Construction and Architectural Management*, vol. 25, no. 6, pp. 738–757, Jul. 2018, doi: 10.1108/ECAM-02-2017-0038/FULL/PDF.

[3] US Bureau of Labor Statistics, "Census of Fatal Occupational Injuries Summary," 2019. https://www.bls.gov/news.release/cfoi.nr0.htm (accessed Dec. 01, 2021).

[4] W. S. A. , S. S. , A. H. Q. , N. S. , M. A. , M. A. Syed Ammad, "Personal Protective Equipment In Construction, Accidents Involved In Construction Infrastructure Projects," *Solid State Technology*, vol. 63, no. 6, pp. 4147–4159, Oct. 2020.

[5] M. Khan, R. Khalid, S. Anjum, N. Khan, S. Cho, and C. Park, "Tag and IoT based safety hook monitoring for prevention of falls from height," *Automation in Construction*, vol. 136, no. December 2021, p. 104153, 2022, doi: 10.1016/j.autcon.2022.104153.

[6] S. Anjum, N. Khan, R. Khalid, M. Khan, D. Lee, and C. Park, "Fall Prevention From Ladders Utilizing a Deep Learning-Based Height Assessment Method," *IEEE Access*, vol. 10, pp. 36725–36742, 2022, doi: 10.1109/ACCESS.2022.3164676.

[7] I. Jeelani, K. Han, and A. Albert, "Development of Immersive Personalized Training Environment for Construction Workers," *Congress on Computing in Civil Engineering, Proceedings*, vol. 2017-June, pp. 407–415, 2017, doi: 10.1061/9780784480830.050.

[8] A. Albert, M. R. Hallowell, and B. M. Kleiner, "Enhancing Construction Hazard Recognition and Communication with Energy-Based Cognitive Mnemonics and Safety Meeting Maturity Model: Multiple Baseline Study," *Journal of Construction Engineering and Management*, vol. 140, no. 2, p. 04013042, 2014, doi: 10.1061/(asce)co.1943-7862.0000790.

[9] V. Kaskutas, A. M. Dale, H. Lipscomb, and B. Evanoff, "Fall prevention and safety communication training for foremen: Report of a pilot project designed to improve residential construction safety," *Journal of Safety Research*, vol. 44, no. 1, pp. 111–118, 2013, doi: 10.1016/j.jsr.2012.08.020.

[10] S. E. Cromwell and J. A. Kolb, "An examination of work-environment support factors affecting transfer of supervisory skills training to the workplace," *Human Resource Development Quarterly*, vol. 15, no. 4, pp. 449–471, 2004, doi: 10.1002/hrdq.1115.

[11] I. Awolusi, E. Marks, and M. Hallowell, "Wearable technology for personalized construction safety monitoring and trending: Review of applicable devices," *Automation in Construction*, vol. 85, no. October 2017, pp. 96–106, 2018, doi: 10.1016/j.autcon.2017.10.010.

[12] L. S. Marín and C. Roelofs, "Promoting Construction Supervisors' Safety-Efficacy to Improve Safety Climate: Training Intervention Trial," *Journal of Construction Engineering and Management*, vol. 143, no. 8, p. 04017037, 2017, doi: 10.1061/(asce)co.1943-7862.0001330.
[13] W. Fang, P. E. D. Love, H. Luo, and L. Ding, "Computer vision for behaviour-based safety in construction: A review and future directions," *Advanced Engineering Informatics*, vol. 43, no. November 2019, p. 100980, 2020, doi: 10.1016/j.aei.2019.100980.
[14] W. Zhang, S. Zhu, X. Zhang, and T. Zhao, "Identification of critical causes of construction accidents in China using a model based on system thinking and case analysis," *Safety Science*, vol. 121, no. April 2019, pp. 606–618, 2020, doi: 10.1016/j.ssci.2019.04.038.
[15] S. N. Razavi and O. Moselhi, "GPS-less indoor construction location sensing," *Automation in Construction*, vol. 28, pp. 128–136, Dec. 2012, doi: 10.1016/J.AUTCON.2012.05.015.
[16] C. H. Caldas, M. Asce, ; David, G. Torrent, and C. T. Haas, "Using Global Positioning System to Improve Materials-Locating Processes on Industrial Projects," *Journal of Construction Engineering and Management*, vol. 132, no. 7, pp. 741–749, Jul. 2006, doi: 10.1061/(ASCE)0733-9364(2006)132:7(741).
[17] W. S. Jang and M. J. Skibniewski, "A wireless network system for automated tracking of construction materials on project sites," *Journal of Civil Engineering and Management*, vol. 14, no. 1, pp. 11–19, Mar. 2008, doi: 10.3846/1392-3730.2008.14.11-19.
[18] P. M. Goodrum, M. A. McLaren, and A. Durfee, "The application of active radio frequency identification technology for tool tracking on construction job sites," *Automation in Construction*, vol. 15, no. 3, pp. 292–302, May 2006, doi: 10.1016/J.AUTCON.2005.06.004.
[19] J. Park, Y. K. Cho, and D. Martinez, "A BIM and UWB integrated Mobile Robot Navigation System for Indoor Position Tracking Applications," *Journal of Construction Engineering and Project Management*, vol. 6, no. 2, pp. 30–39, Jun. 2016, doi: 10.6106/JCEPM.2016.6.2.030.
[20] J. Park, S. M. Asce, ; Kyungki Kim, Y. K. Cho, and A. M. Asce, "Framework of Automated Construction-Safety Monitoring Using Cloud-Enabled BIM and BLE Mobile Tracking Sensors," *Journal of Construction Engineering and Management*, vol. 143, no. 2, p. 05016019, Aug. 2016, doi: 10.1061/(ASCE)CO.1943-7862.0001223.
[21] C. Medina, J. C. Segura, and Á. de la Torre, "Ultrasound Indoor Positioning System Based on a Low-Power Wireless Sensor Network Providing Sub-Centimeter Accuracy," *Sensors 2013, Vol. 13, Pages 3501-3526*, vol. 13, no. 3, pp. 3501–3526, Mar. 2013, doi: 10.3390/S130303501.
[22] Y. T. Wang, C. C. Peng, A. A. Ravankar, and A. Ravankar, "A Single LiDAR-Based Feature Fusion Indoor Localization Algorithm," *Sensors 2018, Vol. 18, Page 1294*, vol. 18, no. 4, p. 1294, Apr. 2018, doi: 10.3390/S18041294.
[23] M. Salas, W. Hussain, A. Concha, L. Montano, J. Civera, and J. M. M. Montiel, "Layout aware visual tracking and mapping," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2015-December, pp. 149–156, Dec. 2015, doi: 10.1109/IROS.2015.7353367.
[24] J. Georgy, T. Karamat, U. Iqbal, and A. Noureldin, "Enhanced MEMS-IMU/odometer/GPS integration using mixture particle filter," *GPS Solutions*, vol. 15, no. 3, pp. 239–252, Sep. 2011, doi: 10.1007/S10291-010-0186-4/TABLES/6.

[25] R. Jiang, R. Klette, and S. Wang, “Modeling of unbounded long-range drift in visual odometry,” *Proceedings - 4th Pacific-Rim Symposium on Image and Video Technology, PSIVT 2010*, pp. 121–126, 2010, doi: 10.1109/PSIVT.2010.27.
[26] M. Li and A. I. Mourikis, “High-precision, consistent EKF-based visual-inertial odometry,” *The International Journal of Robotics Research*, vol. 32, no. 6, pp. 690–711, May 2013, doi: 10.1177/0278364913481251.
[27] M. Nahangi, A. Heins, B. McCabe, and A. Schoellig, “Automated localization of UAVs in GPS-denied indoor construction environments using fiducial markers,” *ISARC 2018 - 35th International Symposium on Automation and Robotics in Construction and International AEC/FM Hackathon: The Future of Building Things*, 2018, doi: 10.22260/ISARC2018/0012.
[28] N. Kayhani, A. Heins, W. D. Zhao, M. Nahangi, B. McCabe, and A. P. Schoellig, “Improved tag-based indoor localization of UAVs using extended Kalman Filter,” *Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019*, pp. 624–631, 2019, doi: 10.22260/ISARC2019/0083.
[29] N. Kayhani, B. McCabe, A. Abdelaal, A. Heins, and A. P. Schoellig, “Tag-Based Indoor Localization of UAVs in Construction Environments: Opportunities and Challenges in Practice,” *Construction Research Congress 2020: Computer Applications - Selected Papers from the Construction Research Congress 2020*, pp. 226–235, 2020, doi: 10.1061/9780784482865.025.
[30] N. Kayhani, W. Zhao, B. McCabe, and A. P. Schoellig, “Tag-based visual-inertial localization of unmanned aerial vehicles in indoor construction environments using an on-manifold extended Kalman filter,” *Automation in Construction*, vol. 135, p. 104112, Mar. 2022, doi: 10.1016/j.autcon.2021.104112.
[31] E. Olson, “AprilTag: A robust and flexible visual fiducial system,” *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 3400–3407, 2011, doi: 10.1109/ICRA.2011.5979561.