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

Augmented Reality Framework for Data Visualization Based on Object Detection and Digital Twins

Hung Pham[1*], Linh Nguyen[2], Nhut Huynh[3], Yong-Ju Lee[4], Man-Woo Park[5]

[1] *Department of Civil and Environmental Engineering, Myongji University, Yongin 17058, Korea,* E-mail address: phamhung0306@mju.ac.kr
[2] *Department of Civil and Environmental Engineering, Myongji University, Yongin 17058, Korea,* E-mail address: linngn@mju.ac.kr
[3] *Department of Civil and Environmental Engineering, Myongji University, Yongin 17058, Korea,* E-mail address: nhutht@mju.ac.kr
[4] *Department of Civil and Environmental Engineering, Myongji University, Yongin 17058, Korea,* E-mail address: leetoday@mju.ac.kr
[5] *Department of Civil and Environmental Engineering, Myongji University, Yongin 17058, Korea,* E-mail address: mwpark@mju.ac.kr

Abstract: While pursuing digitalization and paperless projects, the construction industry needs to settle on how to make the most of digitized data and information. On-site workers, who currently rely on paper documents to check and review design and construction plans, will need alternative ways to efficiently access the information without using any paper. Augmented Reality is a potential solution where the information customized to a user is aligned with the physical world. This paper proposes the Augmented Reality framework to deliver the information on on-site resources (e.g., workers and equipment) using head-mounted devices. The proposed framework was developed by interoperating Augmented Reality-supported devices and a digital twin platform in which all information related to ongoing tasks is accumulated in real-time. On-site resources appearing in the user's field of view are automatically detected by an object detection algorithm and then assigned to the corresponding information by matching the data in the digital twin platform. Preliminary experiments show the feasibility of the proposed framework. Worker detection results can be visualized on HoloLens 2 in near real-time, and the matching process obtained the accuracy greater than 88%.

Keywords: Augmented Reality, Digital Twin, Global Positioning System, Head-mounted device, Object Detection.

1. INTRODUCTION

The construction industry is one of the major fields contributing to national growth. Due to the increasing demand for buildings and infrastructures, the scope and scale of construction sites are becoming extensive. Therefore, enormous technologies are required to obtain effective operation and management. Digitalization offers various digital applications for construction managers and stakeholders to increase their progress, improve overall productivity, and raise their profit. However, the construction field witnesses the slow adoption of digitalization compared with the others. Digital Twin (DT), the virtual model accurately reflecting the physical construction sites, arrives with massive digital tools and aims to provide sustainable solutions for smart construction [1].

Managing a construction site composes a series of processes and massive data that contains essential decision-making information and directly affects the project's outcome. Adequate data improves efficiency and quality, while up-to-date information integrates all on-site aspects and reduces mistakes. Using paper-based documents is the common method for most workers to access on-site information; however, this method is insufficient in the present and near future. Workers may travel between sites or departments to search for a piece of information and carry documents throughout the working time. Consequently, they spend significant time only accessing information and cause considerable delays in on-site processes. This issue would arise as the construction scale expands and human-machine interaction does not seriously improve [2], rising demands for efficient information technologies. One of the information technologies that recently attracted the tremendous attention of construction stakeholders and researchers is Augmented Reality (AR). AR displays virtual information from computers and then aligns to the physical world via AR-supported devices, providing an immersive experience with vital information in real-time [3]. With the core principle to bridge the virtual and physical environment in the real-time process, AR superimposes the virtual model onto the user's field of view, allowing AR users to interact with virtual data and compare it with the physical world [3]. In the construction industry, AR applications are recently studied to improve on-site performance and optimize collaboration based on its real-time process [4, 5]. The outweighed abilities of AR can be shown when information is visualized more intuitively than conventional methods, e.g., paper or screen-based methods. In the near future, enormous AR's abilities could be utilized, and its compound values from integrating with other technologies (e.g., computer vision and DT) can be obtained toward the smart construction's concept.

This paper proposes a framework to visualize real-time information of on-site resources such as workers and equipment via AR head-mounted devices. The framework is mainly based on the You Only Look Once version 4 (YOLOv4) object detection model and DT. On-site resources appearing user's field of view (FOV) are detected by the YOLOv4 model, and their information accessed from DT is then visualized in a virtual pop-up window. Preliminary experiments were implemented and achieved satisfying results, signifying the feasibility of the proposed framework.

2. BACKGROUND

With the core principle to bridge the virtual and physical environment in the real-time process, AR superimposes the virtual model onto the user's field of view, allowing AR users to interact with virtual data and compare it with the physical world. Thus, AR-supported devices require hardware, communicating methods between devices, sensory interface, and battery system to deal with graphic and sound processes. The camera scans the physical environment for essential information to render virtual models, while sensors, e.g., Inertial Measurement Unit (IMU), sense the user's actions in AR [3]. Two common types of devices supporting AR are smartphones/tablets and head-mounted glasses, and AR's principal work in both types is the same. However, head-mounted glasses offer the more comfortable experience when users can observe virtual models in 360-degrees and use both hands to interact. Various AR glasses have been developed, and HoloLens 2 is one of the popular AR wearable devices with 52 diagonal degrees in the field of view [6]. The abilities of this head-mounted device not only provide interaction with AR but also implement complex tasks in many fields, especially construction sites [6].

Several studies utilize AR and its supported device for construction purposes. Hajji et al. proposed a workflow based on Building Information Modeling (BIM) and AR to develop the "EasyBIM" application for accessing and interacting with BIM models through a smartphone [7].

Loporcaro et al. evaluated the performance of HoloLens glasses in general construction checking and recommended this device as the potential checking tool if obtaining more improvements [4]. Most of these studies focus on integrating BIM and AR, which provide a more intuitive visualization of the 3-D/4-D model instead of 2-D drawings on paper or computers. However, the solutions based on AR to display other construction's information, e.g., name, schedules, are still limited even AR solutions can significantly outweigh paper-based methods. Thus, the proposed framework aims to visualize information of on-site resources appearing in the user's FOV in the real-time process of AR. The framework's outcome is expected to save significant time for workers in searching for up-to-date information, leveling up project performance and overall productivity.

3. METHODOLOGY

Figure 1 below demonstrates the framework to visualize information of workers/equipment using HoloLens 2, a local computer (PC), and DT. This framework assumes that DT is a cloud-based information system containing on-site location construction data from GPS (Global Positioning System) and other components such as name, schedule, and experience. On the construction site, the user wears HoloLens 2, which records IMU data, video stream and then transfers them to a local PC. The YOLOv4 model detects workers/equipment appearing in the user's FOV from the video stream. Then, the 2-D image coordinates of detection results are transformed into real-world coordinates according to the user's FOV, which covers the whole user's FOV and remains constant distance, rotation with the user. The user's FOV is mapped based on his

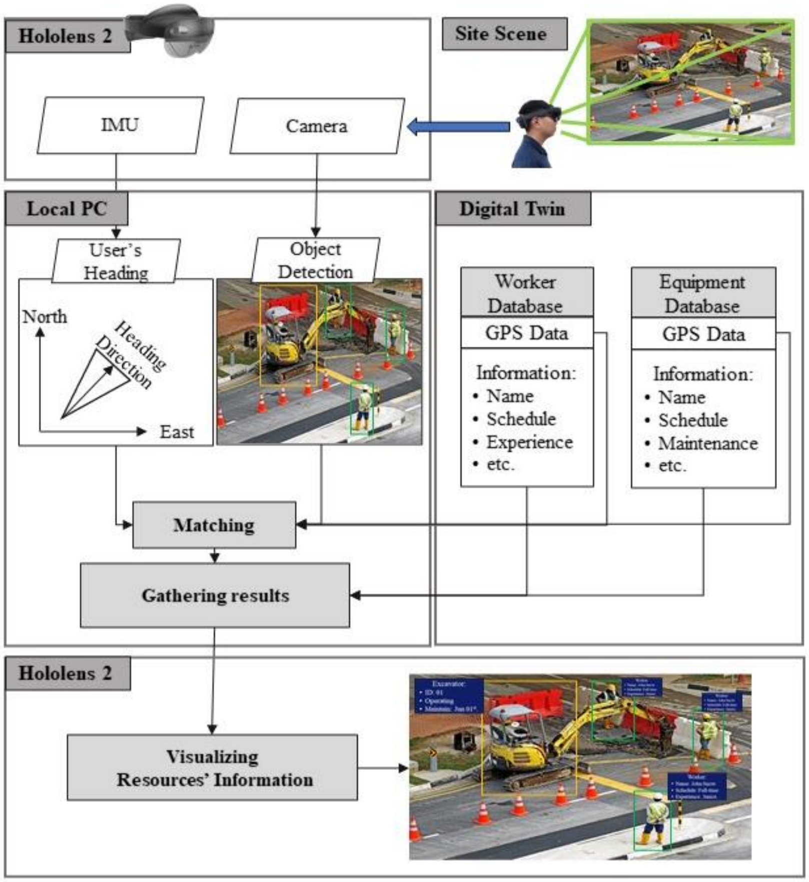

Figure 1. Framework for assigning detection results with corresponding identities

positions from GPS and heading direction computed from IMU data. Next, the identities of detection results are assigned with identities of GPS data that satisfy conditions in the matching process. Once detection results' identities are determined, their information can be accessed from DT. Each detected worker or equipment is visualized in the form of a button, which is located on the mentioned virtual plane with the 2-D coordinates converted onto this plane. Whenever the users press the virtual buttons, information related to the detected object is shown in a pop-up window.

3.1. Detecting workers/objects

Since HoloLens 2 computational power is insufficient for processing the YOLOv4 model, the video stream is transferred to a local computer by Web Real-Time Communication (WebRTC), a protocol optimizing real-time multimedia streaming [8]. The video stream is configured to have 720p30 (1080x720 pixels) resolution and a frame rate of 30 fps. The results from the YOLOv4 model are bounding boxes consisting of five values: object class (e.g., workers, equipment), coordinates in x, y of upper-left corner, width (w), height (h). Detection results are visualized via buttons, which are located on the virtual plane. Thus, the 2-D image coordinate of object detection results are transformed into 2-D local coordinates of the virtual plane to render these buttons on AR as follows:

$$x_{vis} = -\frac{10}{2} + \frac{2\,x_{min}+w}{2*1080} \tag{1}$$

$$y_{vis} = \frac{10}{2} - \frac{2\,y_{min}+h}{2*720} \tag{2}$$

3.2. Matching detection results with satisfied GPS data

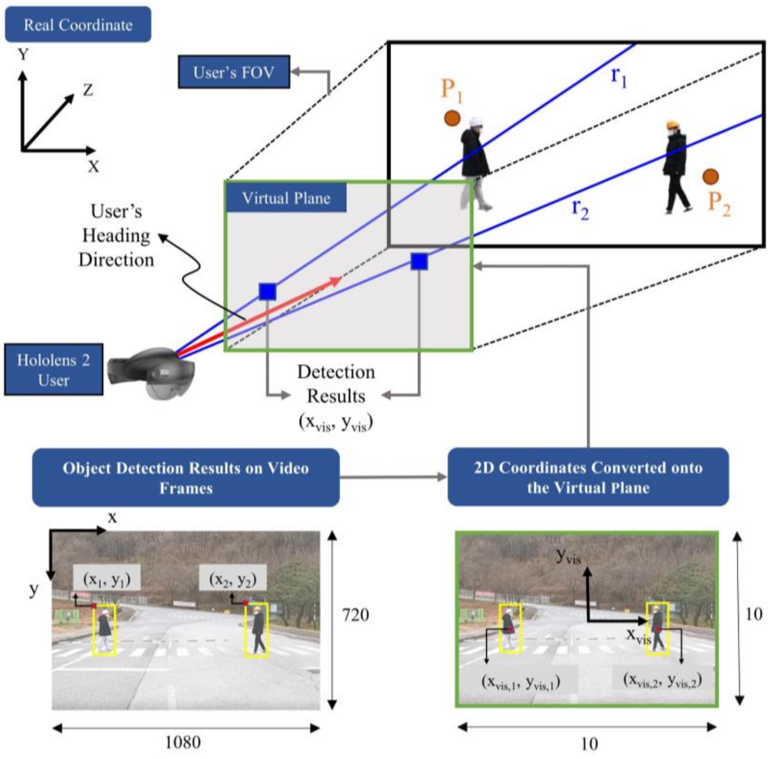
Real Coordinate
Y
Z
X
User's FOV
P_1
r_1
r_2
P_2
Virtual Plane
User's Heading Direction
Hololens 2 User
Detection Results (x_{vis}, y_{vis})
Object Detection Results on Video Frames
2D Coordinates Converted onto the Virtual Plane
x
y
(x_1, y_1)
(x_2, y_2)
720
1080
y_{vis}
x_{vis}
($x_{vis,1}$, $y_{vis,1}$)
($x_{vis,2}$, $y_{vis,2}$)
10
10

Figure 2. Matching between object detection results and GPS data

Figure 2 illustrates the matching process, where detection results in the real coordinate system, user's FOV, and workers/equipment's GPS positions are required. Infinite rays (r_1, r_2) are cast from the user's position to the detection results (x_{vis}, y_{vis}), and the distance between these rays and GPS positions (P_1, P_2) are calculated. The satisfied GPS data that its identity is assigned to the considered detection result is the position having the smallest distance with the infinite ray.

The user's position and heading direction are two elements to map the user's FOV in real coordinates. The user's position is collected from GPS units and smoothed by Kalman Filter [9], and values of IMU's magnetometer are computed to determine the angle between the user's heading and the north direction [10]. Since HoloLens 2's diagonal field of view is 52 degrees [6], the user's FOV is mapped in a real coordinate system. The mentioned virtual plane and buttons are also defined in real coordinates according to the user's FOV.

Worker positions (P_1, P_2) are also collected from GPS units and smoothed by Kalman Filter. After computing the distance between r_1, r_2 and P_1, P_2 sequentially, the distance values are inputted into the Hungarian algorithm to obtain the smallest distance [11] for the matching process, and the identity of detection results is the GPS identity with the smallest distance. Multiple Object Assigning Precision (MOAP) is calculated to determine the average similarity between correct matching results and their ground truth identities (see Equation 3). Once the framework returns assigning results, MOAP is computed from the correct assignment of button *i* with its ground truth identities ($a_{i,t}$) and the number of assignment results (M_t) as:

$$MOAP = \frac{\sum_{t,i} a_{i,t}}{\sum_t M_t} \tag{3}$$

3.3. Visualizing results on HoloLens 2

Results including x_{vis}, y_{vis}, and information of detection results are transferred to HoloLens 2 via Transmission Control Protocol/Internet Protocol as JavaScript Object Notation (JSON) format. x_{vis} and y_{vis} are the local coordinates of buttons in the virtual plane. These buttons update their positions in every frame according to new results from the YOLO detection model and align actual detected objects. User can interact with these buttons by touching or pressing them, and information about detected workers/equipment appears in the user's FOV whenever the buttons are pressed.

4. RESULTS AND DISCUSSION

The proposed framework was tested with preliminary experiments, which consider workers as the main resources to retrieve information. The first part presents the experiments to test the performance of detecting workers from the video streams and visualizing the results on HoloLens 2. The second part evaluates the performance of the matching process with two types of experiments – one with virtual data and the other with real sensing data.

4.1. Visualizing Object Detection Results on HoloLens 2

This section presents the experiments to investigate the performance of the worker detection as well as the lead time taken for i) recording a video frame with the camera on HoloLens 2, ii) transferring the frame to a local PC, iii) detecting workers in the frame, and iv) sending back and visualizing the detection results on HoloLens 2. The application was developed based on Unity Engine and Mixed Reality Toolkits, and the PC that ran the object detection was equipped with the Intel Core i9-10900K processor and the NVIDIA GeForce RTX 3080.

The YOLOv4 model pretrained with the MS COCO dataset [12] was used for detecting workers which were actually identified as the "person" class. The YOLOv4 was tested on the 30-fps videos recorded with the HoloLens 2, in which three people wearing hardhats wandered near the HoloLens

2 user. The tests showed 100.0% precision and 99.5% recall with the YOLOv4 working in real-time. The precision is considered more critical than the recall in the proposed framework to reduce false matchings and avoid providing false information to the user.

Detection results are visualized as the blue buttons on the user's FOV via HoloLens 2 (Figure 3), and the positions of the buttons are updated frame-by-frame according to new detection results. Because the buttons are rendered based on the detection results of the previous frames, the lead time causes displacements from the buttons to the actual appearances of the workers in the user's FOV. Three factors directly affect the lead time: (1) the connection status between the HoloLens 2 and the local PC, (2) the worker's speed, and (3) his distance to the HoloLens 2 user. In the ideal experiments, workers walk around 10~20 away from the user at about 1.2 m/s. The lead time took up to 0.5 seconds throughout the experiments, and the button displacement was trivial as shown in Figure 3.

4.2. Matching between Object Detection Results with GPS Data

The proposed matching method was tested on virtual scenes in which three workers were simulated to walk around within the user's view. The scenes involved several cases of the workers crossing each other closely, which are considered the more challenging scenarios to validate the matching process. The virtual data used for the matching include the 3D locations of the workers as well as the user, and the pixel coordinates of the workers projected on the virtual image plane. All data were generated at 30 Hz. Reflecting the accuracy of portable inexpensive GPS units, random errors generated by the Gaussian distribution were added to 3D location data.

Table 1 summarizes the matching results per various scales of errors imposed on the location data. The accuracy decreases as the magnitude of the errors increases, indicating that the sensors' performance used for acquiring location data is critical to the matching results in real cases. The Kalman filter significantly reduced the effect of the errors, allowing the matching process to achieve the accuracy of over 92% even with the 8 m-errors. The false matchings occurred when workers were positioned close to each other.

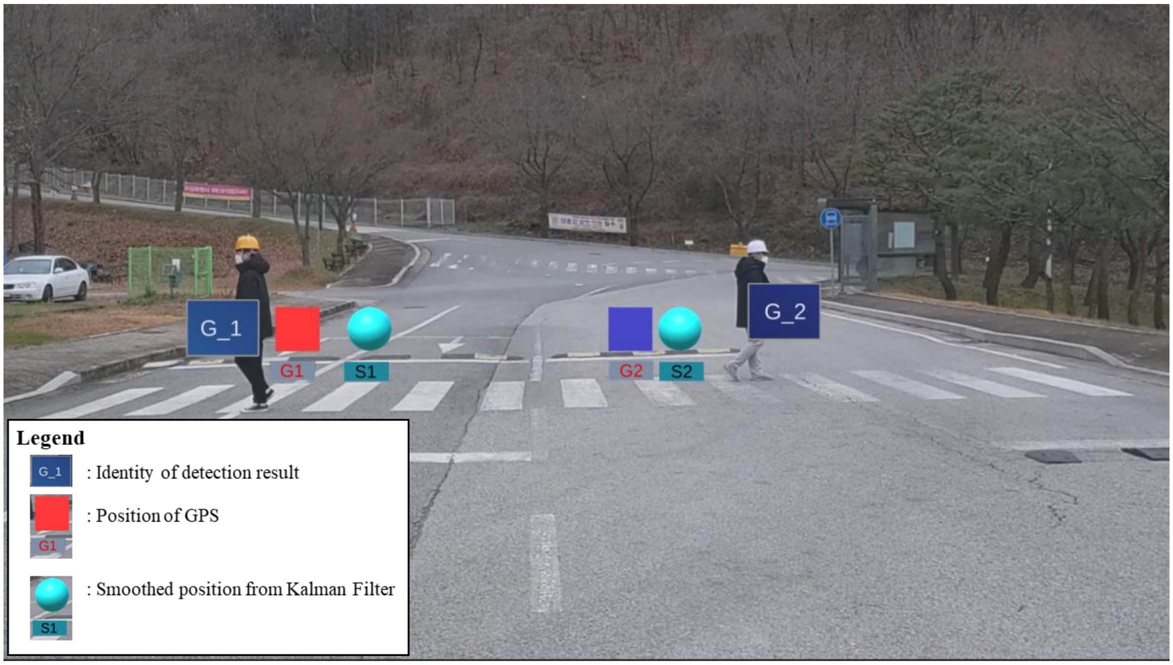
G_1
G1
S1
G2
S2
G_2
Legend
G_1 : Identity of detection result
G1 : Position of GPS
S1 : Smoothed position from Kalman Filter

Figure 3. A sample frame illustrating the button visualization and matching results

Table 1. MOAP with virtual simulated data

Maximum error (m)	Using error-added location data	Using smoothed location data
2	91.79%	94.45%
4	84.03%	93.47%
6	75.32%	92.92%
8	70.27%	92.03%

Experiments using real sensing data had a similar setup to the ones with virtual scenes except that workers' locations, 2D image coordinates, and user's heading direction were collected by using the GPS, YOLOv4, and the IMU sensor on HoloLens 2, respectively. The GPS units used in the experiments (Ascen RCV-3000 [13]) provide 1-Hz data which are reported to contain 3-meter errors in average. It is worth noting that the matching algorithm was post-processed in the experiments, and the real-time processing is left for future works. Also, all types of data were recorded at synchronized time for post-processing.

The data were collected for two scenarios which involved two and three workers moving around, respectively. Table 2 summarizes the matching results and shows that smoothing the GPS data with the Kalman filter did not help the matching performance. It is inferred that the low data rate of the GPS units (1 Hz) was too low to be smoothed well and to provide more accurate locations. As in the experiments with simulated data, the false matchings were associated with the workers in close proximity. The second scenario, which involved three workers in the user's view, created more error-prone cases and led to the lower MOAP than the first scenario. Even with the IMU sensors' errors that were not considered in the simulated data, the proposed method could match the identities of the detected workers with MOAP over 88%.

Table 2. MOAP with real sensing data

Scenario	Using raw GPS data	Using smoothed GPS data
2 workers	92.59%	92.79%
3 workers	88.63%	88.27 %

5. CONCLUSION

The construction industry is facing a new paradigm toward digitalization and paperless projects. This change will demand new ways for on-site workers to access project information without papers. In this aspect, AR technology can be a game-changer to enhance the efficiency of on-site tasks by reducing time spent on searching for data or information. Previous research works on AR in the construction domain have focused on the use of 3D or BIM models as virtual information to overlap onto the users' view. However, there are other types of data or information which are not included in general BIM models.

In this regard, this paper presents the AR framework that realizes accessing the information about on-site work resources such as workers and equipment using a head-mounted device, HoloLens 2. It automatically detects the resource entities from the video frames taken by the HoloLens 2, matches the detected entities with the information accumulated in a sort of a DT system, and visualizes the matched information on the HoloLens 2. This paper detailed the framework and showed preliminary experimental results. Though there are still some gaps to fill to make the proposed framework more practical, the experiments signified the feasibility of the framework. As future works, the matching process needs to be improved to reduce false positive matchings, and real-time processing of the implementation must be validated. Also, the implementation needs to be tested on varied environments and scenes.

ACKNOWLEDGEMENTS

This work was conducted with the support of the National R&D Project for Smart Construction Technology (22SMIP-A158708-03) funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by the Korean Expressway Corporation.

REFERENCES

[1] F. Tao, M. Zhang, and A. Y. C. Nee, "Digital Twin Driven Smart Manufacturing", Academic Press, 2019.
[2] J. Yang, M.-W. Park, P. Vela, and M. Golparvar-Fard, "Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future", Advanced Engineering Informatics, vol. 29, no.2, pp. 211-224, 2015.
[3] A. Oke, C. Aigbavboa, S. Segun, and W. Thwala, "Augmented reality and sustainable construction", in Sustainable Construction in the Era of the Fourth Industrial Revolution, Eds A. Oke, C. Aigbavboa, S. Segun, and W. Thwala, London: Routledge, pp. 21–38, 2021.
[4] G. Loporcaro, L. Bellamy, P. McKenzie, and H. Riley, "Evaluation of Microsoft HoloLens Augmented Reality Technology as a construction checking tool", Proceedings of the 19th International Conference on Construction Applications of Virtual Reality, Bangkok, Thailand, 2019.
[5] R. Tayeh, F. Bademosi, and R. Issa, "BIM-GIS Integration in HoloLens", Proceedings of the 18th International Conference on Computing in Civil and Building Engineering, Paulo, Brazil, pp. 1187–1199, 2021.
[6] Microsoft, "HoloLens 2—Overview, Features, and Specs | Microsoft HoloLens" https://www.microsoft.com/en-us/hololens/hardware.
[7] R. Hajji, A. Kharroubi, Y. B. Brahim, Z. Bahhane, and A. E. Ghazouani, "Integration of BIM and Mobile Augmented Reality in the Aeco Domain", Proceedings of the International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLVI-4/W3-2021, pp. 131–138, 2022.
[8] B. Sredojev, D. Samardzija, and D. Posarac, "WebRTC technology overview and signaling solution design and implementation", Proceedings of the 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1006–1009, 2015.
[9] A. Deep, M. Mittal, and V. Mittal, "Application of Kalman Filter in GPS Position Estimation", Proceedings of the 2018 IEEE 8th Power India International Conference (PIICON), pp. 1–5, 2018.
[10] T. Yoo, S. Hong, H. Yoon, and S. Park, "Gain-Scheduled Complementary Filter Design for a MEMS Based Attitude and Heading Reference System", Sensors (Basel, Switzerland), vol. 11, pp. 3816–30, 2011.
[11] J. Gil-Aluja, "Theoretical Elements of the Hungarian Algorithm", in The Interactive Management of Human Resources in Uncertainty, Eds. J. Gil-Aluja, MA: Springer US, pp. 158–170, 1998.
[12] T. Lin et al., “Microsoft COCO: Common Objects in Context”, arXiv:1405.0312 [cs], 2015.
[13] AscenKorea. https://www.ascenkorea.net/.