buildings

MDPI

Article

Augmented Reality Framework for Retrieving Information of Moving Objects on Construction Sites

Linh Nguyen, Htoo Thiri Htet, Yong-Ju Lee * and Man-Woo Park

Department of Civil and Environmental Engineering, Myongji University, Yongin 17058, Gyeonggi-do, Republic of Korea; linngn@mju.ac.kr (L.N.); htoothiri@mju.ac.kr (H.T.H.); mwpark@mju.ac.kr (M.-W.P.)
* Correspondence: leetoday@mju.ac.kr

Abstract: The construction industry is undergoing a digital transformation, with the digital twin serving as a core system for project information. This digital twin provides an opportunity to utilize AR technology for real-time verification of on-site project information. Although many AR developments for construction sites have been attempted, they have been limited to accessing information on stationary components via Building Information Models. There have been no attempts to access information on dynamically changing resources, such as personnel and equipment. This paper addresses this gap by presenting an AR framework that enables site managers to verify real-time information on specific personnel or equipment. It introduces a matching algorithm for retrieving the necessary information from the digital twin. This algorithm is pivotal in identifying and retrieving the specific information needed from the vast dataset within the digital twin. The matching process integrates object detection and tracking algorithms applied to video frames from AR devices, along with GPS and IMU sensor data. Experimental results demonstrate the potential of this matching algorithm to streamline on-site management and reduce the effort required to interact with digital twin information. This paper highlights the transformative potential of AR and digital twin technologies in revolutionizing construction site operations.

Keywords: augmented reality; computer vision; digital twin; global positioning system; head-mounted device; information access

check for updates

Citation: Nguyen, L.; Htet, H.T.; Lee, Y.-J.; Park, M.-W. Augmented Reality Framework for Retrieving Information of Moving Objects on Construction Sites. *Buildings* **2024**, *14*, 2089. https://doi.org/10.3390/buildings14072089

Academic Editors: Eugeniusz Koda and Irem Dikmen

Received: 15 April 2024
Revised: 23 June 2024
Accepted: 2 July 2024
Published: 8 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the realm of construction management and on-site inspections, traditional practices have long relied on the tangible comfort of paper-based documents. While these practices have served their purpose over the years, they are not without their limitations. The cumbersome nature of paper-based processes often translates into inefficiencies, delays, and an increased risk of errors. Construction site managers, tasked with overseeing complex projects, are burdened by stacks of documents, drawings, and plans [1], making real-time decision making a daunting challenge [2]. This reliance on manual record keeping has underscored the urgent need for innovation within the construction industry.

Contemporary trends within the construction sector are unmistakably steering projects toward a paperless future. With advancements in technology and a growing emphasis on sustainability, the industry is adopting digital tools and processes at an unprecedented pace [3]. The transition to paperless projects is not merely a convenience but a necessity, promising to revolutionize how construction sites are managed and inspected. At the heart of this transformative journey lies the concept of the digital twin. This virtual representation of a physical construction site, complete with all its components and complexities, has emerged as a linchpin in the pursuit of paperless projects [4]. Being more advanced than the Building Information Model (BIM), the digital twin includes sensors and a networking system to collect data and simulate construction sites [5]. This technology serves as a dynamic, real-time mirror of the site, generating a wealth of information for decision

making, tracking progress, and safety management [6,7]. The simulated sites receive the collected data to update the transformation in their physical twins and then return necessary adjustments to the actual sites [8].

In tandem with the digital twin, Augmented Reality (AR) applications hold the promise of elevating construction management to unprecedented levels of efficiency and effectiveness [9]. AR creates a virtual environment that aligns with the real environment and displays extra information via AR devices [10]. By harnessing the power of AR, construction site managers can access the wealth of information residing within the digital twin in real time, overlaying vital data onto their physical surroundings [11,12]. This fusion of the digital and physical environments empowers on-site managers with instant insights, enhancing their ability to make informed decisions, improve worker safety, and streamline operations [13].

The efforts made for on-site AR applications have all focused on retrieving information about stationary components which can be accessed through the overlaid BIMs. In addition to the information available in the BIM, data about site resources that are constantly in motion, such as construction personnel and equipment, are also crucial for understanding and managing the site situation. If this information can be received and checked in real time from the digital twin, it would be highly beneficial for site management. While AR has seen varying degrees of exploration and implementation across the construction sector, there has been an absence of comprehensive research dedicated to the real-time access of information pertaining to workers and equipment.

Aiming to bridge this critical gap, this paper proposes a new AR framework that allows a user to easily access information about any workers or equipment being observed. The framework basically connects the user's observation with the data being collected on the digital twin to provide the user with the information corresponding to the user's observation. The object detection and tracking algorithms applied to the video frames streamed from the AR device summarize the user's observation by informing on which regions of the user's view workers are present. On the other hand, all data about workers retained in the digital twin are represented by their GPS data. The focus of this paper is on the matching process between the object detection and tracking results with the GPS data so that each worker appearing in the user's view is correctly linked to their corresponding information in the digital twin. Once each of the detection and tracking result is matched with GPS data, the framework is ready to deliver and visualize any information linked to the GPS data whenever the user queries. Therefore, by using the proposed method, AR users are expected to access digital twin information online with minimal effort during interactions with the digital twin.

The matching process involves the 3D geometry composed of (i) the user's position and orientation, (ii) the image plane, (iii) object detection and tracking results (rectangular bounding boxes) on the image plane, and (iv) the workers' positions. The process primarily relies on the distance between lines and points—the former being vectors connecting the user's position to the bounding boxes and the latter being the workers' positions. Given that the GPS and IMU data used for acquiring the positions and orientation contain significant errors, this paper details the method for isolating the matchings with high confidence scores. The paper presents experimental results to evaluate the matching process and discusses the feasibility of the AR framework.

2. Literature Review

2.1. AR Device

Two common AR-supported devices are smartphones/tablets or head-mounted devices. These devices are equipped with sensors to recognize the surroundings and track user behaviors. Sensors such as cameras, IMUs, and those of the GPS, along with communication capabilities like Wi-Fi or Bluetooth, enable these devices to perceive the user's surrounding environment and receive relevant information. Although the implementation of AR is similar in both types of devices, the user experience may differ. The AR experience

via smartphones/tablets involves a see-through video where virtual objects are visualized on the screens and overlaid onto the video frames. In contrast, head-mounted devices support hands-free interaction with virtual objects and provide a more immersive environment. Currently, there are various head-mounted devices on the market, with HoloLens 2 being one of the most prominent [14].

2.2. General AR Applications

AR applications for smartphones/tablets and head-mounted devices are developed and released in various fields. Generally, these applications allow users to observe digital information in 3D models from a 360-degree perspective and interact with the models via gestures. Additionally, some AR practices use data from real environments to determine the users' requests and facilitate interaction with physical objects. Most AR applications are limited to providing users with an immersive experience without any relation to the real environments surrounding them [15], from which many fields, such as training and education can still benefit thanks to intuitive visualization [16–18].

Reflecting the user's surroundings and providing associated information in real time can truly be considered genuine AR. In this context, a widely utilized AR application is location-based information delivery, where relevant information is provided based on the user's position and viewing direction. When the position and orientation of virtual objects are determined in the physical environment, AR applications can accurately place the virtual information at their desired locations [19]. Examples include applications that display nearby restaurants or gas stations on the camera screen of a mobile device. Games like Pokémon GO also fall under location-based AR applications [20]. For example, Han et al. [21] created an AR system to visualize cultural heritage sites in outdoor environments, where the users have no difficulties accessing the information on these sites. This point gives the users a more data-rich perspective, which significantly improves their engagement with the received information.

Another method of reflecting the user's surroundings is the view-based AR that utilizes the camera video frames [22–26]. This involves recognizing specific objects in the camera image and providing associated information based on the recognized objects such as vehicle parts [22] and medical specimens [24]. This approach incorporates image object detection algorithms, and recent advancements in deep learning-based algorithms are significantly enhancing its utility. A mobile application that searches using a camera and all applications utilizing QR codes can be considered examples of view-based AR. Additionally, an application that recognizes pictures in books for educational purposes and displays related 3D models [27] or an application that recognizes a billiard table and balls to provide appropriate guidance [28], can also be considered examples of view-based AR.

2.3. Construction AR Applications

In the construction sector, various studies have been undertaken regarding the application of AR, largely underpinned by the introduction of the BIM, which can encompass diverse information. The BIM plays a crucial role in augmenting reality by providing virtual information. Therefore, AR in construction predominantly focuses on visualizing the BIM. Similar to the introduction in Section 2.2, BIM-based construction AR can also be categorized into applications independent of the user's environment, location-based AR, and view-based AR.

Augmented Reality (AR) applications that are independent of the user's surroundings primarily leverage the intuitive and immersive characteristics of AR. The advantages of these characteristics are maximized when users visualize and interact with the BIM in the form of 3D models. Several studies have enabled stakeholders to visualize and share BIM data from construction sites to facilitate decision-making processes [29,30]. Immersive indirect experiences have been utilized for construction safety training [31]. However, these methods are functionally closer to VR technology. To maximize the advantages from

visualizing BIM and digital twin information, it is essential to accurately place models to enable user interaction, which requires recognizing the user's surroundings.

In location-based AR, various positioning methods have been employed to recognize the user's surroundings, with the GPS being the most widely used. Since the GPS only measures location, additional sensors are necessary to obtain orientation information. Kamat and Behzadan [32] demonstrated the utility of the GPS for overlaying 3D models via Head-Mounted Display (HMD) AR by developing hardware that integrates magnetometers and inertial measurement units for attitude measurement. Zollmann et al. [33] developed a mobile AR system using drones for construction site monitoring, successfully overlaying 3D models onto the drone's camera view using the GPS and IMU mounted on the drone. However, these methods rely heavily on sensor performance, and high overlap accuracy cannot be guaranteed without high-performance sensors, including those of the GPS. To overcome these limitations, studies have utilized high-precision sensors like Real-Time Kinematic (RTK) GPS [34]. Commercial solutions using RTK GPS are available to accurately overlay the BIM of terrestrial structures or even subsurface infrastructure [35]. However, these devices are large, costly, and still dependent on satellite signals, which limits their effectiveness for indoor use.

View-based AR, as mentioned in Section 2.2, recognizes the user's surroundings through a camera to visualize and display the BIM accordingly. Markers are one of the most commonly utilized methods for this purpose. There is a study where a 2D drawing was used as a marker to overlay and display a 3D model on top of the drawing, aiding in the user's understanding of the 2D drawings [36]. Additionally, fiducial markers like QR codes or ArUco have been employed to overlay full-scale 3D models from the user's view. Hübner et al. [37] achieved an overlay of indoor environments with virtual room-scale model data with a spatial accuracy of a few centimeters using markers. However, marker-based AR is not free from occlusion issues, and errors accumulate over time when the marker is out of the user's view. Markerless view-based AR has been researched to overcome these limitations. This approach requires the use of various computer vision algorithms to recognize objects, locations, and orientations instead of markers.

Despite continuous efforts in AR, it is difficult to find applications that are practically used on construction sites. Moreover, there are none related to accessing real-time information about workers and equipment. Unlike the previously introduced AR studies, this research distinguishes itself significantly in that the target of the information to be acquired is not a static object like structural components, but rather a moving object. In BIM-based AR, since the target object is static, once the BIM is overlaid and positioned, only the user's location and line of sight need to be tracked to access the information on the already overlaid BIM. However, for moving objects such as personnel or equipment, both the user and the target object must be tracked in real time and continuously associated with each other. Accordingly, it necessitates an entirely different framework from existing BIM-based AR technology.

3. Research Problem and Objective

To address the knowledge gaps identified earlier, this study introduces an AR framework designed for site managers to access real-time information about on-site workers who move within the user's field of view. The framework emphasizes the interaction between the AR user and the digital twin, where the user acts as an agent seeking information about on-site workers, while the digital twin serves as a repository of comprehensive project data. This interaction includes an automated data query process. Unlike BIM-based AR applications, fiducial markers are impractical in this context due to size constraints that limit their detectability and distance from the user. Moreover, markers on moving objects could often be occluded from view due to their dynamic poses.

The data querying process proposed in this study relies on the 3D locations of the user and the target object, as well as the user's view. The study assumes that location data are gathered using GPS sensors and stored in the digital twin along with other relevant

project information. The user's view information is obtained through the camera and the magnetometer embedded in the AR device. Figure 1 illustrates the proposed AR framework. The magnetometer determines the user's facing direction. In addition, object detection and tracking algorithms analyze video frames from the camera informing the direction of each image object within the user's field of view (FOV). Since the user's observation does not include workers' identities, the framework necessitates further processing to match image objects with worker information in the digital twin. As mentioned, the workers' location data play a key role in the matching process. Upon successful matching, the relevant information of the worker is retrieved from the server and transmitted back to the AR device for display.

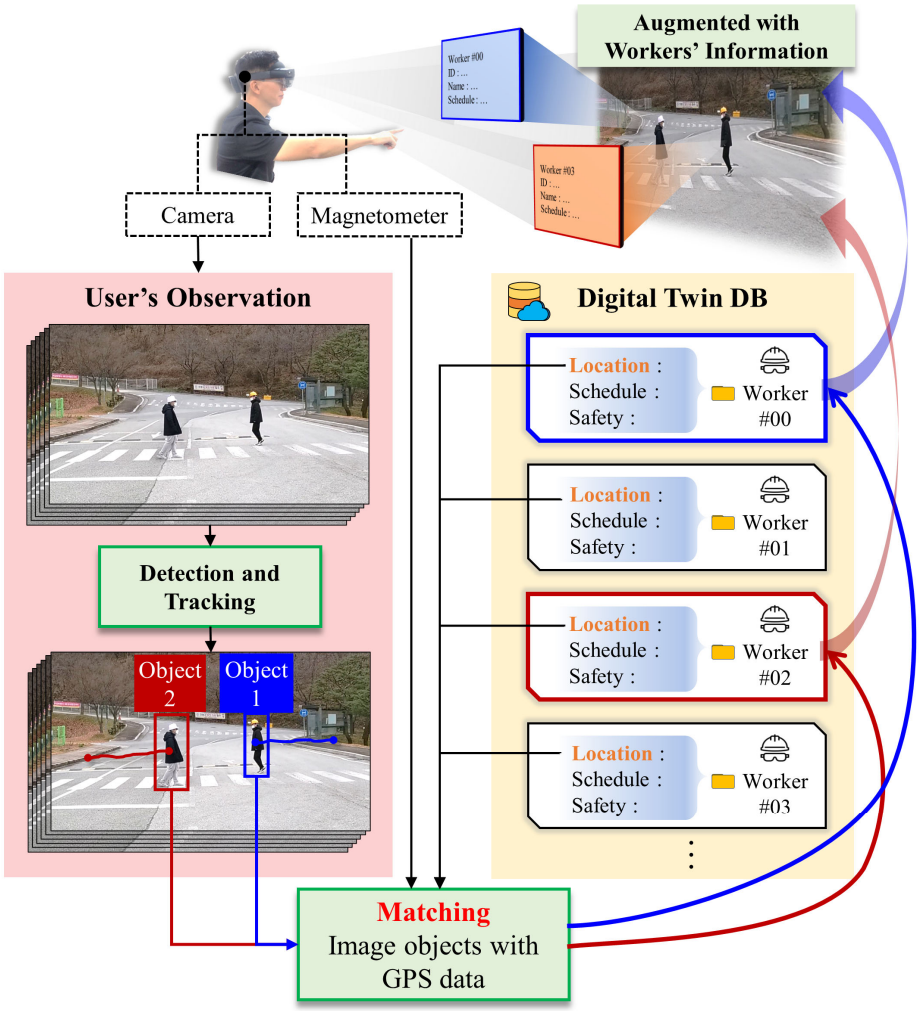


Figure 1. Overall framework of the proposed AR application.

While proposing the AR framework for retrieving on-site dynamic objects, this paper specifically focuses on the matching process as a critical step in the framework. This paper develops the detailed matching algorithm and evaluates its performances and feasibility of the AR framework through experiments.

4. Methodology

This section presents the matching process shown in Figure 2, which is a key component to realize the proposed AR framework. Each step of the matching process is detailed in the subsections.

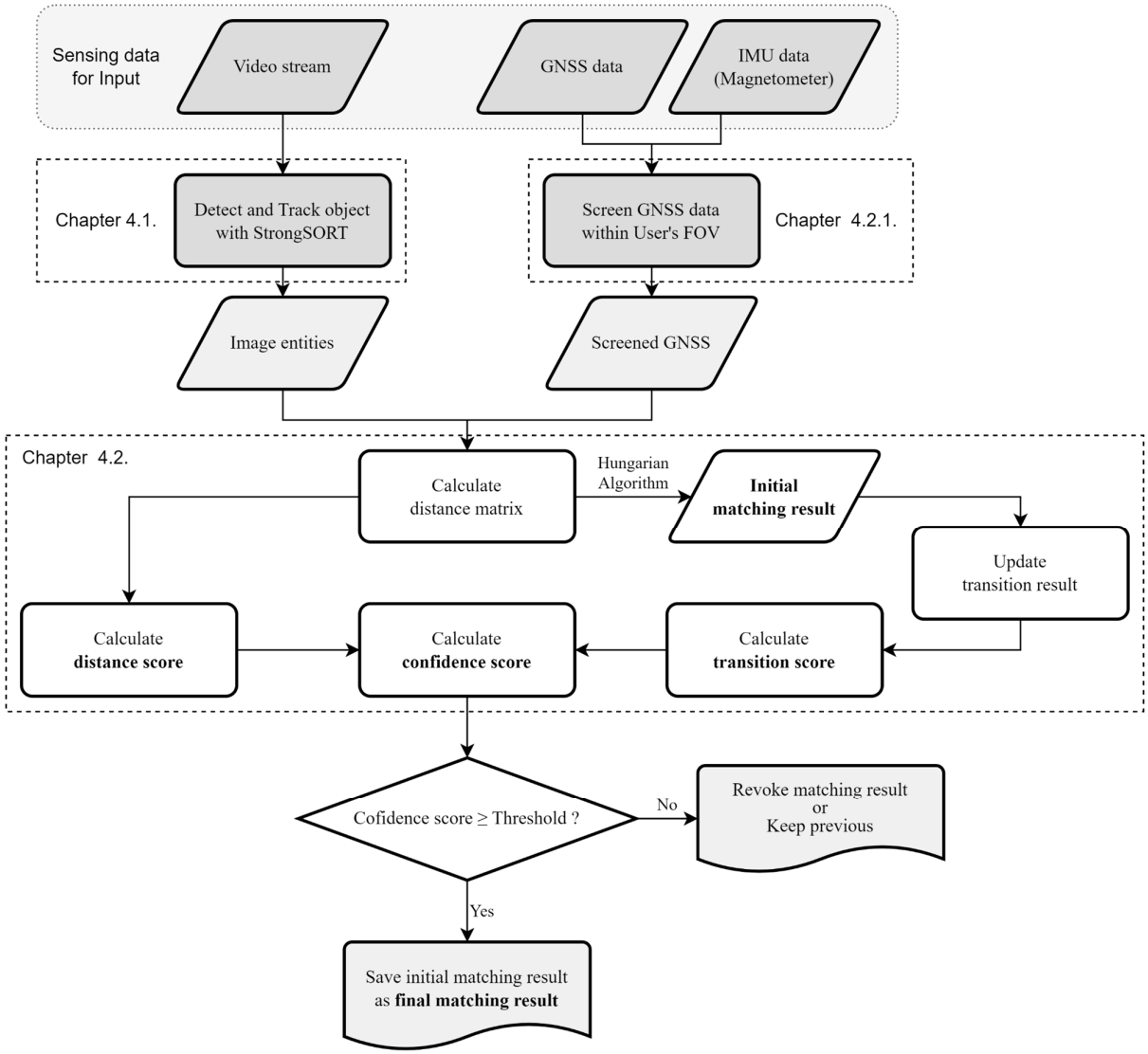


Figure 2. Matching algorithm.

4.1. Localization of Image Objects through Detection and Tracking

Workers within the user's FOV are detected and tracked in the video frame by YOLOv5 [38] and StrongSORT [39]. Each video frame captured by the head-mounted AR device is analyzed to identify and track image objects of workers. The localization of the objects on the image plane involves two steps. Firstly, detection using the YOLOv5 algorithm recognizes these image objects and returns their image coordinates frame-by-frame as bounding boxes [38]. However, detection results distinguish objects up to classes, providing a list of bounding boxes per class in each frame (Figure 3a). In other words, the result of each frame is independent of other frames, and no relation across the frames is provided. Hence, detection alone cannot identify and track the movement paths of each object over frames. As shown in Figure 1, the proposed AR framework requires a matching process between image objects and GPS data. If only detection results are used as image object information, accurate matching results must be obtained for each frame. Additionally, there is no opportunity to utilize the matching history of each object, which is actually a critical factor of the transition score in the proposed matching process (Figure 2).

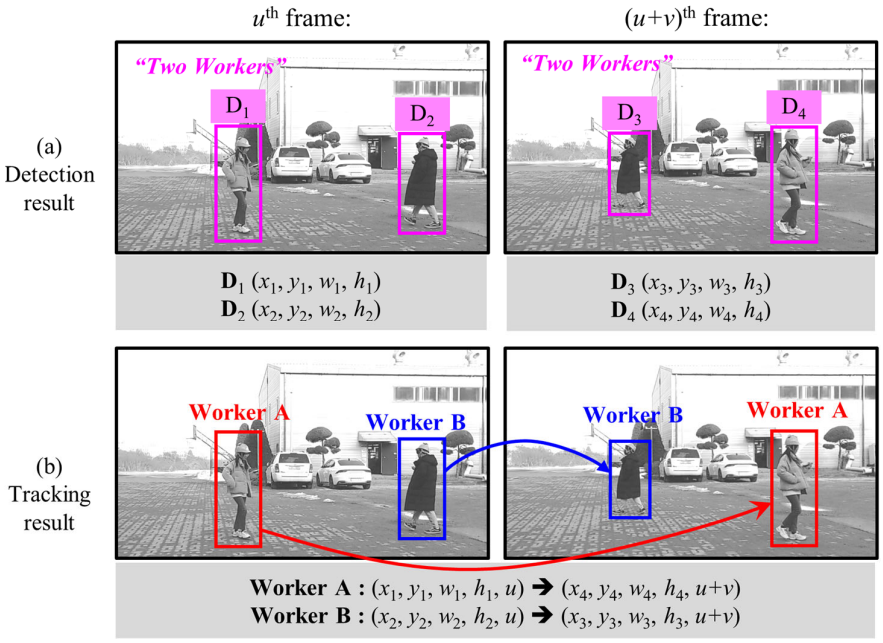


Figure 3. (**a**) Detection and (**b**) tracking results in the video frames.

Therefore, the StrongSORT tracking algorithm is utilized to identify workers by unique IDs [39]. This enables monitoring their position changes in the video sequence and maintaining the IDs consistently as long as the objects are present in the user's FOV. Moreover, the IDs can be retained for short durations of disappearance due to occlusion or leaving the FOV. The tracking result per worker can be represented as a time series of bounding box data composed of the centroid coordinate (x and y), the width (w), and the height (h) (Figure 3b).

To relate the tracking results on the 2D image plane with GPS data, the 2D image data need to be transformed into 3D data. For this purpose, a virtual 3D image plane is placed centrally in front of the user, onto which the tracking results are projected. The virtual image plane is dynamically positioned at a fixed distance from the user, aligning with the user's heading direction. The user's heading direction is determined by the magnetometer, a sensor included in the IMU unit of the AR device. The size of the plane and its distance from the user are calibrated based on the camera resolution and the FOV. Accordingly, the coordinates of the image objects can be conceptualized in three dimensions. The raycasts from the user's position to the image objects on the virtual image plane are then calculated, which is utilized in the next stage, the matching process.

4.2. Matching Algorithm

The tracked image objects from the previous step are matched with the on-site workers' GPS data in the following process (Figure 2). The GPS data, stored in the digital twin, represent the locations of the workers in the construction site. The GPS data are crucial for matching the tracked objects with their real-world counterparts, enabling the system to provide accurate and relevant information to the user. To begin the matching process, n GPS data points within the user's FOV are identified. These are then matched with the m image objects described in Section 4.1. The distances between the raycasts to the image objects and the GPS points are calculated to form an $m \times n$ distance matrix. The Hungarian Algorithm is then used to return the initial matching results for the m image objects. To compensate errors from inaccurate GPS data, two scoring methods—Distance Score and Transition Score—are employed to verify and compute confidence scores for the initial matching results. The overall process is detailed below.

4.2.1. Screen GPS Data within the User's FOV

Since there may be a significant number of workers on-site, processing GPS data for all workers during the matching process is time-consuming and inefficient. To improve this inefficiency, a data screening process is implemented. The user's FOV region can be defined as the yellow area in Figure 4, using the heading direction from the magnetometer, AR user's GPS data, and the camera's intrinsic parameters. The matching process only uses GPS data within the user's FOV (the green positions in Figure 4). GPS data outside this area are excluded from the matching process, reducing unnecessary calculations. The tracked objects will be matched with the screened GPS data in the next step. The workers' GPS positions obtained in this process follow the East–North coordinate system of UTM.

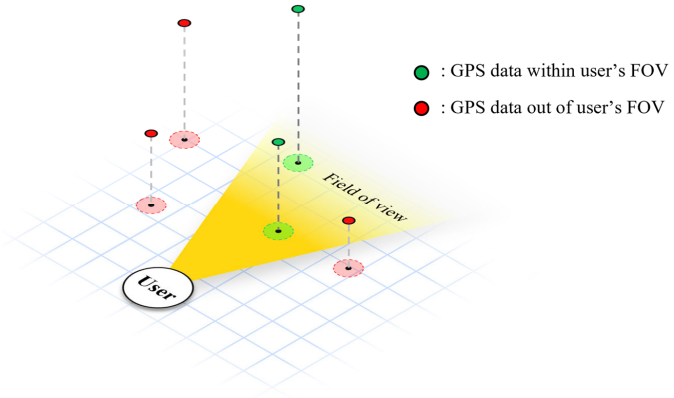


Figure 4. User's FOV region and available GPS for the matching process.

4.2.2. Match Image Objects to GPS Data

With the known AR user position, heading directions of AR user, and camera parameters from camera calibration, it is feasible to estimate the relative positioning between the user and the virtual image plane. This estimation enables the generation of vectors, referred to as raycasts ($\mathbf{r}_1$ and $\mathbf{r}_2$ in Figure 5), extending from the user toward the image coordinates of the tracked objects. These vectors represent sets of potential locations where the corresponding on-site workers may be situated. Ideally, the GPS data of an on-site worker would align precisely with these vectors. However, due to inherent errors from sensor data, there may exist a discrepancy in the distance (d in Figure 5) between the GPS data and raycasts. Therefore, a logical approach is to identify the most likely matching candidate by calculating the distance from a raycast to the GPS data and selecting the nearest one. For example, in Figure 5, to identify image object 1, a corresponding vector $\mathbf{r}_1$ is generated, and the distances between $\mathbf{r}_1$ and GPS_1, GPS_2 are calculated, resulting in distances d_{11} and d_{12}, respectively. Similarly, for image object 2, the distances between $\mathbf{r}_2$ and GPS_1, GPS_2 are calculated, yielding distances d_{21} and d_{22}. These calculated distances form the distance matrix $\mathbf{D}$ depicted in Figure 5.

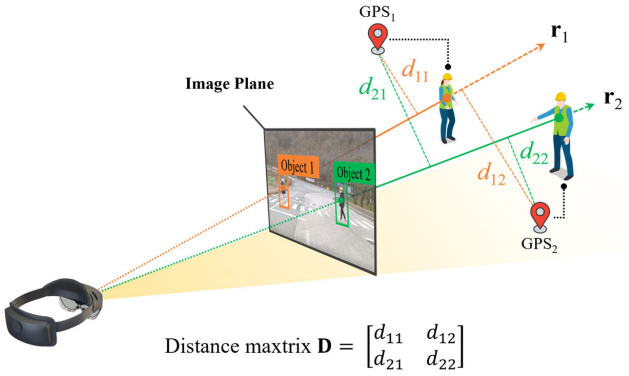


Figure 5. Scenario of distance calculation in the matching process.

In situations where workers are in close proximity, the Hungarian Algorithm, employed for the matching process based on distance calculations, can yield inaccurate results, thereby reducing both precision and recall values. To mitigate this issue and enhance accuracy, the algorithm incorporates a grouping mechanism for workers. Specifically, when the distance between the image coordinates of two or more tracked objects falls below a predefined grouping threshold, these objects are grouped together. By forming these groups, the matching results consider all matched GPS data corresponding to the grouped objects, thereby minimizing false matching results and improving the algorithm's overall accuracy.

4.2.3. Evaluate Matching Results by Confidence Score

Matching results are significantly dependent on sensor performance, leading to potential calculation errors. To minimize these errors, various scoring methods have been implemented to evaluate the matching results. The scoring methods, i.e., distance score (S_D) and transition score (S_T), were developed to skip the false matching assignments at the current frame and retain the correct ones from the previous frames. Distance score (S_D) assesses the confidence level based on the calculated distance matrix, while transition score (S_T) evaluates the consistency of matching results across consecutive frames, promoting stability and reducing the impact of temporary errors. To comprehensively assess the matching results between image objects and GPS data, a confidence score (S_C) has been implemented. A matching result's confidence score (S_C) is a weighted sum of S_D and S_T, reflecting the confidence in both distance and time series. The matching result having S_C higher than the threshold will be confirmed and finalized as the matched GPS for the objects. Conversely, the matching results of objects will remain the previous result or no matching result if the S_C is lower than the threshold. The details of each score are discussed below.

Distance score (S_D) indicates the confidence of matching results based on the distance between the GPS data and raycasts to image objects. It is calculated using a modified SoftMax equation that limit the range from 0 to 100. This equation considers the distances between each object's raycast and all available GPS data to determine the likelihood of each GPS being a match for a specific object. S_D is defined as:

$$S_{D,ij} = \frac{e^{-d_{ij}}}{\sum e^{-d_{ij}}} \times 100\% \quad (1)$$

where:

$i = 0, 1, 2, \ldots, m-1$
$j = 0, 1, 2, \ldots, n-1$
$S_{D,ij} =$ Distance score for image object i, w.r.t GPS j
$d_{ij} =$ Distance between raycast of image object i and GPS j

Due to inherent latency and distance errors, GPS data may inaccurately follow the workers' trajectories, leading to improper assessment of matching results. To address this issue, the Transition score (S_T) is employed to review matching results in terms of time series. It considers the maximum number of consecutive matches from the historical data of all GPS, thereby evaluating the accuracy of the match.

S_T is calculated as:

$$S_{T,\,ij} = \frac{T_{ij}}{\sum T_{ij}} \times 100\% \quad (2)$$

where:

$i = 0, 1, 2, \ldots, m-1$
$j = 0, 1, 2, \ldots, n-1$
$S_{T,ij} =$ Transition score for image object i, w.r.t GPS j
$T_{ij} =$ Maximum number of consecutive frames that image object i is matched with GPS j

Prior to computing S_T, a matrix **T** for transition frames' result is constructed with dimension $i \times j$, where i is the identity of the object and j is that of the GPS. Each entry in matrix **T** records the number of consecutive frames that object i is matched with GPS j. If the object is not matched with GPS j, or if the number of current consecutive frames is not greater than the number of consecutive frames results from the previous frame, the entry for the object and GPS remains unchanged. To explain using Figure 6 as an example, let us assume that there are two GPS data for object 1, GPS_1 and GPS_2. T_{11} and T_{12} represent the transitions between object 1 and GPS_1, and object 1 and GPS_2, respectively. Until the 25th frame, only the signal from GPS_1 is present, so only T_{11} has increased. From then on, only the signal from GPS_2 is present, so the value of T_{12} gradually increases as the frames progress, while the value of T_{11} remains unchanged. Starting from the 60th frame, the signal from GPS_1 occurs, but until the 85th frame, the number of consecutive frames for GPS_1 does not exceed the previous count of 25, so both T_{11} and T_{12} remain unchanged. From the 86th frame onwards, the value of T_{11} gradually increases with frame progression. Later, at the 105th frame, only the signal from GPS_2 occurs again, but since there are only 20 consecutive frames until the final frame, T_{12} retains its previous value. This ensures that S_T reflects the longest streak of consecutive frames in which each object is correctly matched with a particular GPS, enhancing the overall accuracy and reliability of the matching process. By combining these two scores, S_C is computed as in Equation (3), with the weight λ ranging from 0 to 1. This framework credits $\lambda \times 100\%$ from S_D and $(1 - \lambda) \times 100\%$ from S_T, and the λ value of 0.5 was chosen in this paper. The matching result with S_C greater than the threshold value is confirmed as the final matching result.

$$S_{C,ij} = \lambda S_{D,ij} + (1 - \lambda) S_{T,ij} \tag{3}$$

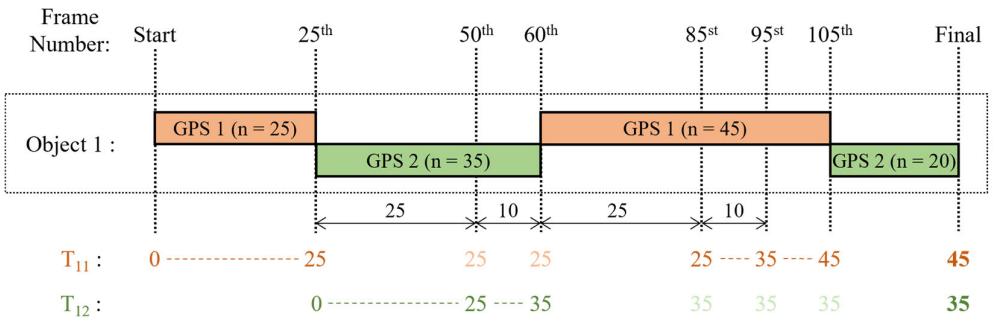


Figure 6. Example of transition results calculation.

5. Experiments and Results

Experiments were conducted with workers to examine the feasibility and performance of the proposed framework. The evaluation of the matching framework was performed by post-processing matching with the sensor data from the digital twin. This study used HoloLens 2 as the head-mounted AR device, portable GPS units, and a local PC equipped with an NVIDIA Geforce RTX 3080 Ti GPU.

We used commercially available, general-purpose portable GPS units. Although the accuracy specified on the specification sheets provided by each manufacturer was approximately 2–3 m, we found it necessary to verify this through experimentation. We assessed the positional accuracy of each device and selected those suitable for the purposes of this study. Similarly, we conducted performance verification of the magnetometer integrated into the HoloLens 2. Both portable GPS units and the HoloLens 2 were used without any additional hardware modifications. Detailed information on these aspects is discussed further below.

The detection and tracking processes utilized the YOLOv5 and StrongSORT algorithms, selecting one of the pre-trained models provided with the network by default. The model used for YOLOv5 was YOLOv5l, and the model used for StrongSORT was

osnet_x0_25_msmt17. We used the basic models without additional training. Additionally, all parameters required for the application of StrongSORT were set to their default values as configured in the network [39].

5.1. Performance of Sensors

GPS and magnetometers are critical sensors that significantly influence experimental outcomes. Thus, it is imperative to evaluate their performance prior to the main experiments. The performance of the GPS was assessed using four distinct devices labeled A, B, C, and D to ascertain their effect on matching accuracy. The magnetometer, integrated within the HoloLens 2, could not be substituted with alternative devices; therefore, its performance was rigorously evaluated to ensure it met the study's objectives.

5.1.1. GPS Sensor

The performance of the GPS is evaluated by two criteria: the error between the GPS data and the ground truth data, and the trajectories of the GPS compared to the expected route. Firstly, five reference points (Figure 7) were marked by the Trimble R8s GNSS system, and their positions were used as ground truth data. Then, GPS sensors were placed at the reference points to collect data for comparison against the established ground truth values. The errors of GPS sensors are shown in Table 1. Notably, device B exhibited the lowest mean error, indicating superior average accuracy compared to that of other GPS sensors. However, the high standard deviation of device B showed significant variability in accuracy across measurements. Device D, conversely, had the highest average error but maintained more consistent performance. Furthermore, examining the maximal and minimal errors of each GPS type provided insights into the range of accuracy and highlights instances of poor performance. In the second experiment to evaluate the GPS trajectory, a route from P3 to P2, P1, P4, and P5 was predefined (Figure 7). The GPS sensors were carried along the route, recording data simultaneously. Upon comparing the GPS trajectories to the predefined route, device A displayed the most accurate trajectory, closely following the predefined route, while other devices showed greater deviations. The evaluation of GPS performance led to the conclusion that the devices could not provide highly accurate data.

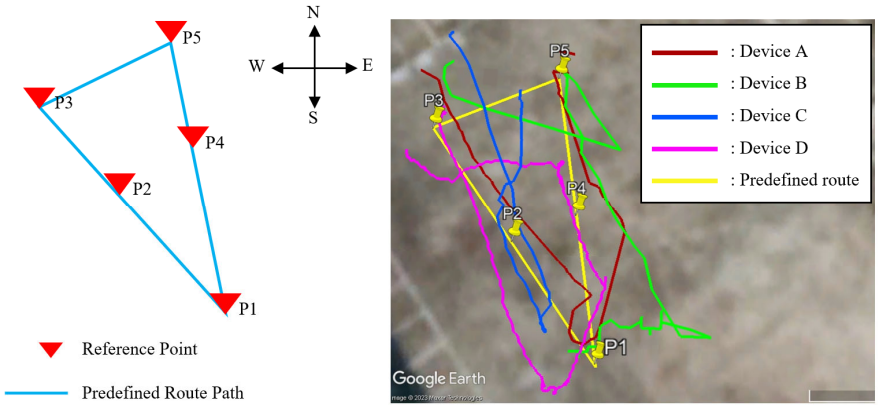


Figure 7. Reference points and GPS trajectories along the predefined route.

Table 1. GPS sensor accuracy.

GPS Device	Error (m)			
	Mean	Max	Min	Standard Deviation
A	4.78	9.91	1.41	1.01
B	2.99	11.70	0.00	2.58
C	3.42	8.98	0.83	1.92
D	6.73	13.41	1.97	1.02

5.1.2. Magnetometer

Similar to the evaluation of GPS performance, the magnetometer's performance was assessed based on the error of the heading angle measured by the sensor compared to ground truth data. The predefined heading direction was established as the user standing at P1 and facing P3, where P1 and P3 were reference points measured by the Trimble R8s GNSS system in the GPS evaluation test. The ground truth heading angle was calculated using the coordinates of P1 and P3. Subsequently, magnetometer data were recorded by the user wearing the HoloLens 2, positioned at P1, and following the predefined heading direction. The sensor errors relative to the ground truth heading angle are tabulated in Table 2. The evaluation of the magnetometer's performance yielded an average error of approximately 3 degrees, with a standard deviation of about 1 degree, indicating moderate consistency in the sensor's measurements. Despite not exhibiting as significant performance issues as GPS sensors, the data from the magnetometer sensor were not entirely accurate.

Table 2. The magnetometer's accuracy.

Error (Degree)	Mean	Max	Min	Standard Deviation
	2.768	6.697	0.728	1.055

5.2. Performance of the Matching Framework

The experiments were designed to evaluate the matching algorithm under different scenarios, representing real-world conditions in construction sites. Table 3 summarizes five experiments based on the scenes, number of workers involved, type of GPS used, and number of crossing events observed. Experiment 1 established a baseline for the matching algorithm's performance with all objects using the same GPS device. Experiments 2 and 3, with setups like that of Experiment 1, evaluated the impact of GPS device variation on matching results and how the algorithm handled matching when the objects were equipped with different GPS devices. Experiments 4 and 5, featuring an additional worker, increased the complexity of the matching process, which allowed for the evaluation of the algorithm's scalability and performance with more objects.

Table 3. Configuration of experiments for the matching framework.

Experiment No.	Scene No.	Number of Frames	Number of Workers	GPS Type	Crossing Events
1	1	3173	2	A	5
2	2	4752	2	B	8
3		4752	2	C	8
4	3	11,149	3	A	7
5	4	3323	3	D	13

5.2.1. Performance Evaluation of Tracking Algorithm

By processing the above scenes into the YOLOv5 and StrongSORT algorithms, the results of detection and tracking image objects were obtained and examined via precision, recall, and MOTA (Multi-Object Tracking Accuracy) metrics as follows:

- Precision $= \mathrm{TP}/(\mathrm{TP}+\mathrm{FP})$
- Recall $= \mathrm{TP}/(\mathrm{TP}+\mathrm{FN})$
- MOTA $= 1 - \frac{\sum(\mathrm{FN}+\mathrm{FP}+\mathrm{SW})}{\sum \mathrm{GT}}$
- TP (True Positive) = number of correctly identified and tracked objects
- FP (False Positive) = number of incorrectly identified and tracked objects
- FN (False Negative) = number of undetected and untracked objects

- SW = number of swapped tracking results among ground truth
- GT = number of ground truth detection

This framework aims at detecting and tracking results as precisely as possible. Therefore, the StrongSORT algorithm took into consideration two key parameters: the tracking score, which served as a minimal score that the tracking result was predicted as true, and the maximal age, which defined the maximum duration that an unmatched tracking result could exist. In the experiments, the tracking score threshold was set to 85% for, and the maximal age was 10 s for tracking objects. FP and SW did not appear in all experiments; however, FN increased as only the objects with tracking scores above the threshold were tracked. When image objects crossed paths, the tracking result of the occluded object was lost. Hence, the identity of the object could remain the same or change to a new identity based on the duration of the occlusion. Additionally, a new identity of the object would occur when the object disappeared from the camera's view and came into the view again. The values of precision, recall, and MOTA are shown in Table 4. The algorithm performed exceptionally well in Scene 1, achieving high accuracy in a simple scenario with two objects crossing five times. The algorithm maintained high precision and MOTA scores in Scenes 2, 3, and 4 despite an increase in the number of objects and crossing events.

Table 4. Accuracy of detection and tracking results.

Scene No.	Number of Frames	Number of Objects	Crossing Events	Number of Object Identities	Precision	Recall	MOTA
1	3173	2	5	2	100.00%	99.79%	99.79%
2	4752	2	8	5	100.00%	99.20%	99.20%
3	11,149	3	7	3	100.00%	98.72%	98.72%
4	3323	3	13	6	100.00%	99.69%	99.69%

5.2.2. Performance Evaluation of Matching Algorithm

The tracking data of all the processed scenes and the corresponding sensor data, including those of GPS and magnetometer, were collected at the synchronized time. The average processing time for one matching iteration was 15 to 120 milliseconds. Figure 8 provides a visual representation of the framework's matching results. The ground truth matchings are demonstrated with colored lines by connecting the image objects, which referred to workers, and the GPS (indicated by red and blue boxes in Figure 8). The matching result of each worker is displayed with its ground truth-matched GPS and confidence score. In the 224th frame, Worker 1 and Worker 2 were grouped together since the distance between their image coordinates was less than the threshold, 15% of the image frame size in the experiments. As a result, their matching results are shown as a grouped matching, including the matched GPS candidates for both workers.

The performance of the matching algorithm was evaluated based on precision and recall, defined in Equations (4) and (5). In this framework, precision indicates the proportion of correct matching results ($\sum a_i$) among the total matching results ($\sum M_i$) of all tracked workers, and recall is the proportion of correct matching results ($\sum a_i$) among the total tracking results of workers ($\sum T_i$). The precision and recall of the matching results using two S_C thresholds, 80% and 70%, in the experiments are presented in Table 5.

$$\text{Precision} = \frac{\sum_i a_i}{\sum_i M_i} \tag{4}$$

$$\text{Recall} = \frac{\sum_i a_i}{\sum_i T_i} \tag{5}$$

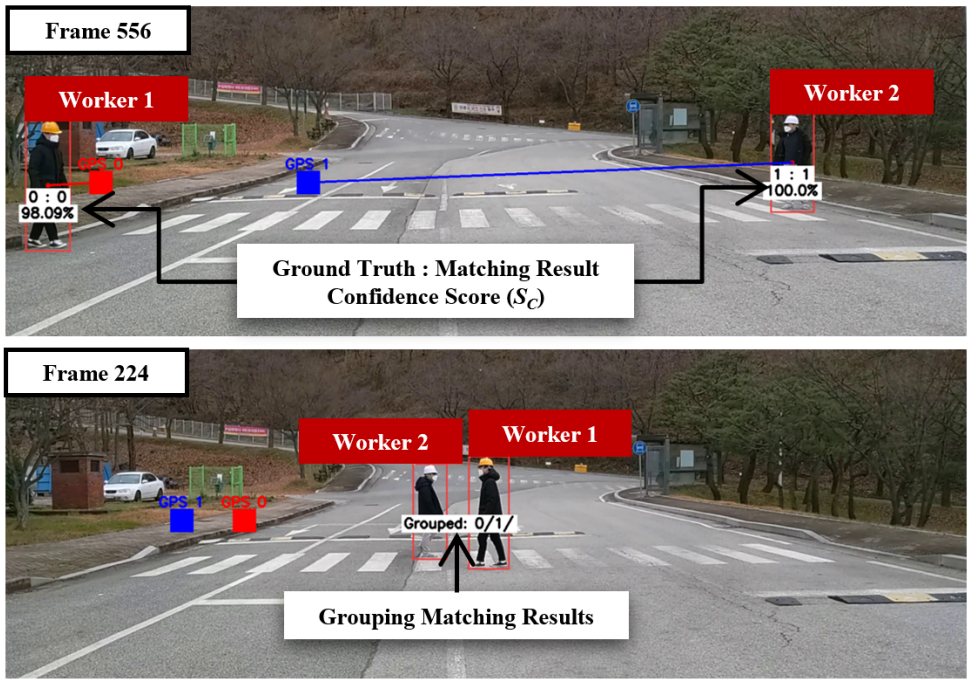


Figure 8. Matching results of the framework for visualization.

Table 5. Accuracy of matching results.

Exp No.	$\sum T_i$	$S_C \geq 80\%$				$S_C \geq 70\%$			
		$\sum a_i$	$\sum M_i$	Precision	Recall	$\sum a_i$	$\sum M_i$	Precision	Recall
1	5847	5787	5787	100.00	98.97	5787	5787	100.00	98.97
2	7206	6644	6644	100.00	92.20	6226	6917	90.01	86.40
3	7206	6261	6580	95.15	86.89	6063	6580	92.14	84.14
4	25,164	21,219	21,219	100.00	84.32	20,738	21,219	97.73	82.41
5	9247	9125	9125	100.00	98.68	8462	9125	92.73	91.51

When the threshold for S_C was 80%, the precision of the matching results was 100%, except for Experiment 3, which consisted of two workers using GPS device C with 8 crossings.

The accuracy scores from Table 5 can help evaluate how different thresholds impact the accuracy of matching results. The S_C threshold plays a crucial role in retrieving the correct matching results and avoiding non-matching results. When the threshold is too high, the matching results from the initial frames may fail to meet the threshold and are not verified as final matching results, leading to a lower recall percentage. Lowering the threshold allows more matching results to meet the threshold, thus avoiding non-matching results. However, this adjustment may affect both precision and recall due to false matching results. False matching results can occur when workers are in close proximity or cross paths. For example, in Figure 9, the 289th frame of Experiment 4 displayed a false matching result of Worker 2. GPS data (GPS 0, GPS 1 and GPS 2) and raycasts to each image object of workers (referred to as Raycast 1, 2, and 3) can also be seen in top view of Figure 9. Ground truth matching is indicated by the same color for both GPS and Raycasts. Raycast 2 closely aligned with both GPS 0 and GPS 1, resulting in S_C of 70.19%. This score exceeded the threshold of 70%, leading to an incorrect update of the initial matching result.

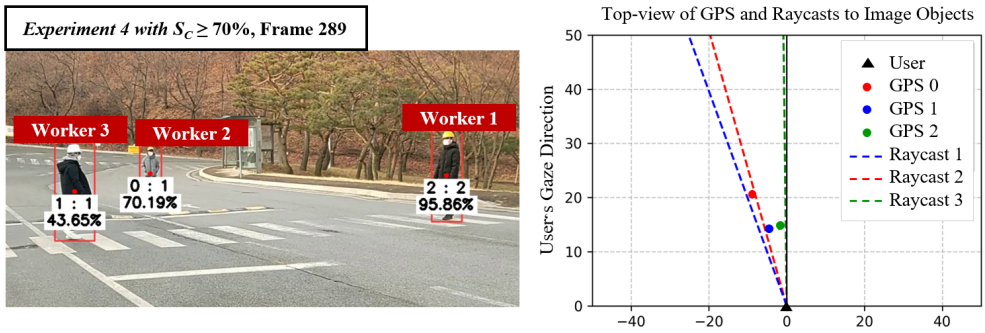


Figure 9. False matching result in Experiment 4.

Similarly, crossed paths of workers in Experiment 5 could be found through the top view from the 224th frame to the 626th frame (Figure 10). The plots indicate a delay in GPS data update for subsequent frames, causing it to lag behind the actual movement of the workers, depicted by raycasts to image objects. False matching of Worker 3 occurred when the initial matching result was promptly updated upon meeting the S_C threshold, which was 73.49%. These observations underscore the importance of GPS data accuracy in close proximity scenarios and highlight the significance of S_C in such contexts.

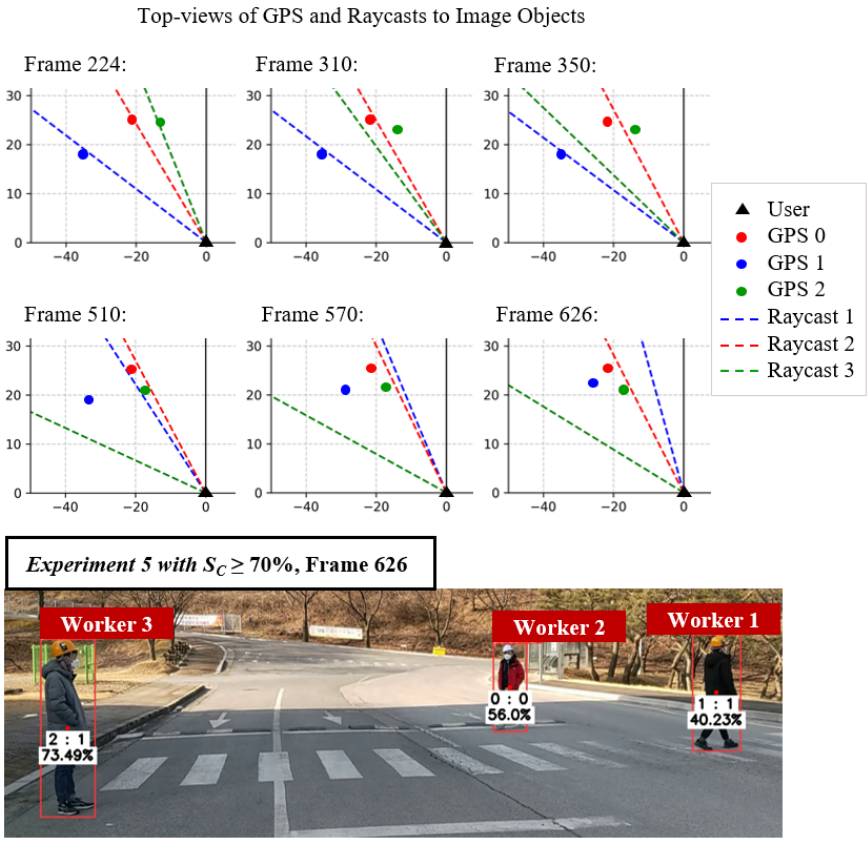


Figure 10. Top-view plots and matching result of the 626th frame in Experiment 5.

Verification of the matching results begins after accumulating 30 initial matching results for each image object, ensuring sufficient data for reliable transition scores. When a worker moved out of the user's view for a certain number of frames and reappeared or when a worker crossed paths with others, the identity of the worker could switch into a new identity. In those cases, the algorithm re-evaluated the matching process for that new identity, starting the matching process anew. This approach aimed to maintain a high level of precision, ensuring that each object was correctly matched with its corresponding GPS

data. Despite the algorithm's precision remaining at 100%, there may be instances where non-matching results happened due to the reevaluation process triggered by changes in object identity, leading to a lower recall rate (Figures 11 and 12). For instance, as depicted in Figure 11, Worker 6 exited the user's view in the 1783rd frame and re-enters in the 1973rd frame with a new identity as Worker 8, resulting in no matching result. Likewise, in the crossing event of Worker 1 and Worker 3 in Figure 12, the identity change occurred from Worker 3 in the 933rd frame to Worker 7 in the 935th frame, resulting in no matching result. Afterwards, Worker 7 could regain a new matching result in the 991st frame (Figure 12).

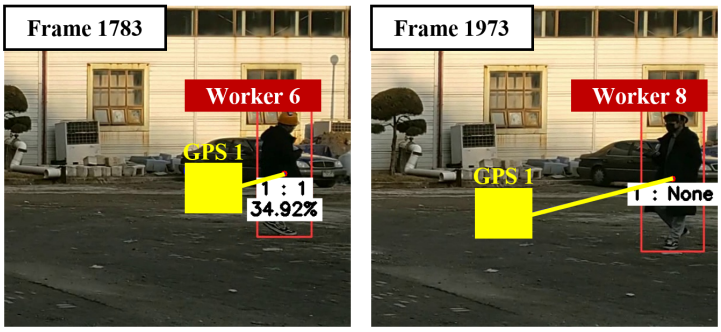


Figure 11. Worker's identity switch in out-of-view event in Experiment 2.

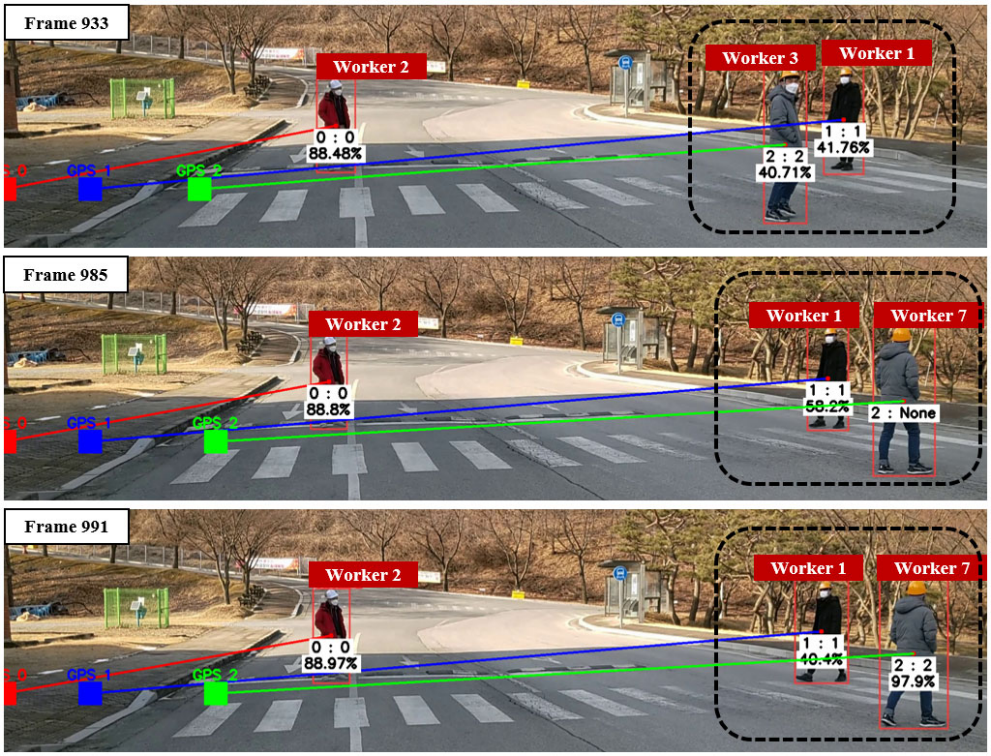


Figure 12. Worker's identity switch in crossing event in Experiment 5.

Experiments 2 and 3 emphasized the importance of GPS data accuracy by revealing specific errors in top views (GPS 0 and GPS 1, Figure 13) from the same scene and the 1459th frame when using different GPS devices, B and C. Specifically, in Experiment 3, using device C for Worker 6 resulted in an S_C of 33.45%, below the 70% threshold, retaining the previous correct matching result. Conversely, in Experiment 2, device B yielded a new, incorrect matching result for Worker 6, as its S_C was 72.3%, above 70%. This highlights the crucial role of GPS accuracy in ensuring the matching algorithm's effectiveness.

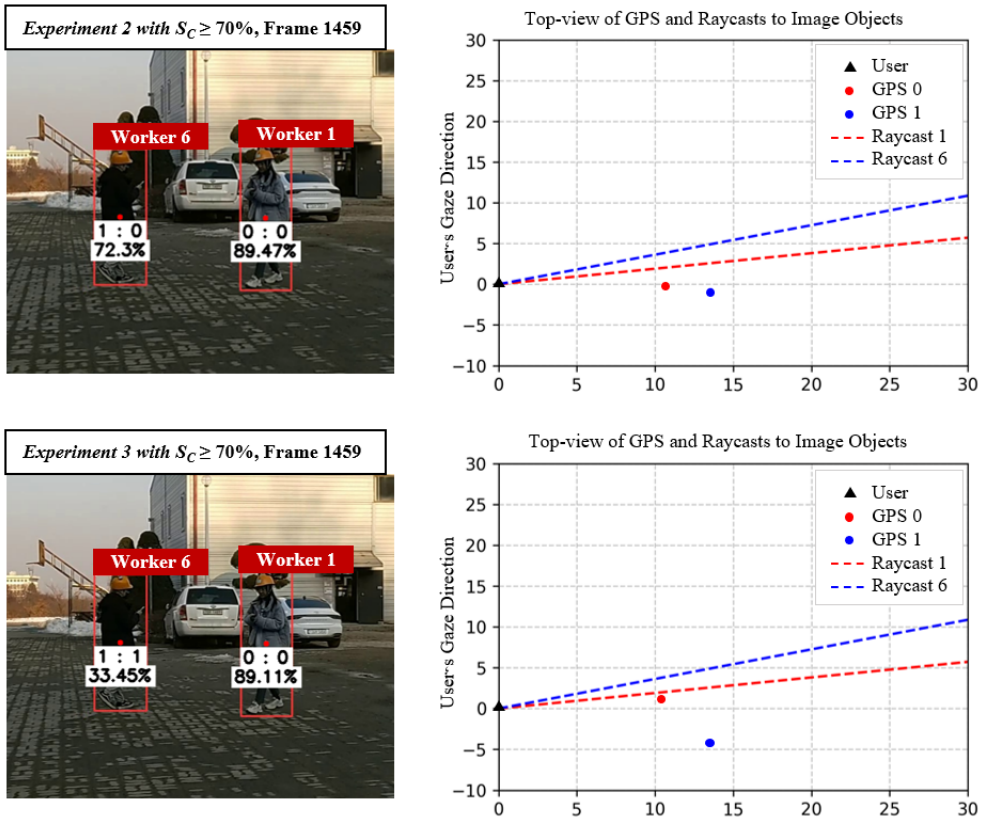


Figure 13. Comparison of GPS data and matching results using different GPS for workers in Experiments 2 and 3.

5.3. Discussion

The performance of the matching algorithm was validated through experiments, and the results showed nearly 100% precision and over 80% recall. This confirmed the feasibility of implementing an AR framework capable of verifying real-time information about personnel or equipment on-site. Introducing a confidence score that reflects cumulative matching results and applying a high threshold maximized precision and minimized the possibility of providing incorrect information. Additionally, applying tracking alongside object detection minimized the reduction in recall caused by the high threshold. The 80% recall rate meant that data matching failed for only one frame out of five on average. For a video running at 10 fps, this implies that more than eight matches are successfully made per second. This level of accuracy is sufficient for the objective of this study.

Although this paper deals only with workers as the target objects, it is expected that the proposed method would yield equivalent or better results for construction equipment, which is relatively easier to distinguish and has a lower clustering density. Furthermore, it is promising that this framework could be extended to apply to any object if the following two conditions are met: (i) its real-time location is available, and (ii) it is feasible to detect it in video frames. The proposed AR framework based on the matching algorithms is anticipated to ultimately assist in assessing site conditions and facilitating rational decision making.

It should be noted that matching performance is highly dependent on the accuracy of the sensors, particularly GPS and magnetometers. Commercial grade GPS trackers on the market or those on mobile devices, as well as the magnetometer on HoloLens 2, were not reliable enough to directly match with image objects without further considerations. Accordingly, various GPS trackers were tested in the experiments to verify their performance and develop a proper matching method that suits the proposed AR framework. Given that the sensors used in the experiments were not high-performance components, the matching accuracy demonstrated in the experiments is very encouraging. It is expected

that the potential for utilizing the proposed AR framework will increase even further as the accuracy of sensors in mobile and AR devices improves in the future.

The YOLOv5 and StrongSORT algorithms have demonstrated robust performance in maintaining object identities even when objects momentarily left the user's FOV. This capability is crucial for the practical implementation of AR systems in construction environments, which often involve occlusions and rapid movements. The framework could benefit from robust re-identification mechanisms integrated into the tracking process to ensure continuous identity verification, thereby minimizing errors during occlusions. However, real-world environments are more intricate than controlled experiments, and unforeseen scenarios can occur. Accordingly, while the proposed matching algorithm operates efficiently in the presented experiments, further research may be needed to evaluate its performance in more complex scenarios involving higher object density and multiple object types. The case of significant altitude differences between objects is another environmental factor not considered in this paper. It is believed that various validations in actual field environments will need to be conducted in the future.

6. Conclusions

The evolving landscape of the construction industry, with its shift towards digitalization and paperless operations, demands efficient methods for accessing the project. This is particularly critical on construction sites where the use of personal computers is limited. AR technology emerges as a valuable solution to facilitate seamless information access in such environments. This paper focuses on an AR framework tailored for retrieving information about dynamic on-site objects, enhancing project management efficiency. The framework outlined herein involves a matching process that queries information pertaining to the object a user is observing from a digital twin that serves as a repository for all project-related data. Detailed within are the sensors utilized in this matching process, along with key components for processing the sensor data, such as object detection and tracking algorithms. Experimental evaluations of the proposed matching process demonstrate remarkable precision, even in the presence of significant sensor errors, affirming the viability of the AR framework in providing instantaneous access to information regarding moving on-site objects.

Furthermore, it is envisaged that the proposed framework and its outcomes will augment the functionalities of the digital twin concept. While this study primarily focuses on workers, the applicability of the framework extends to other mobile objects, such as equipment, provided their location data are accessible. It is pertinent to acknowledge that the experimental results presented in this paper are based on post-processing, thus warranting further research for real-time implementation. Future research will encompass enhancing networking capabilities, synchronizing data in real time, and optimizing user interface design to realize the full potential of the AR framework.

Author Contributions: Conceptualization, M.-W.P. and L.N.; methodology and data curation, L.N. and H.T.H.; validation and formal analysis, L.N., H.T.H. and Y.-J.L.; writing—original draft preparation, L.N., H.T.H. and Y.-J.L.; writing—review and editing, Y.-J.L. and M.-W.P.; supervision, M.-W.P.; project administration, M.-W.P.; funding acquisition, M.-W.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was conducted with the support of the "National R&D Project for Smart Construction Technology (No. RS-2020-KA158708)" funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by the Korea Expressway Corporation.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rezgui, Y.; Boddy, S.; Wetherill, M.; Cooper, G. Past, Present and Future of Information and Knowledge Sharing in the Construction Industry: Towards Semantic Service-Based e-Construction? *Comput.-Aided Des.* **2011**, *43*, 502–515. [CrossRef]
2. Turner, C.J.; Oyekan, J.; Stergioulas, L.; Griffin, D. Utilizing Industry 4.0 on the Construction Site: Challenges and Opportunities. *IEEE Trans. Ind. Inform.* **2021**, *17*, 746–756. [CrossRef]
3. Ofori, G. Construction Industry Development: Role of Technology Transfer. *Constr. Manag. Econ.* **1994**, *12*, 379–392. [CrossRef]
4. Boje, C.; Guerriero, A.; Kubicki, S.; Rezgui, Y. Towards a Semantic Construction Digital Twin: Directions for Future Research. *Autom. Constr.* **2020**, *114*, 103179. [CrossRef]
5. Sacks, R.; Brilakis, I.; Pikas, E.; Xie, H.S.; Girolami, M. Construction with Digital Twin Information Systems. *Data-Centric Eng.* **2020**, *1*, e14. [CrossRef]
6. Han, M.; Baek, K.; Lee, K.-T.; Ko, S.; Kim, J.-H. A Study on Supporting Design Decision Making in Office Building Remodeling Projects by Introducing Mixed Reality. *Korean J. Constr. Eng. Manag.* **2021**, *22*, 3–12. [CrossRef]
7. Jiang, Y.; Li, M.; Guo, D.; Wu, W.; Zhong, R.Y.; Huang, G.Q. Digital Twin-Enabled Smart Modular Integrated Construction System for on-Site Assembly. *Comput. Ind.* **2022**, *136*, 103594. [CrossRef]
8. Lee, D.; Lee, S. Digital Twin for Supply Chain Coordination in Modular Construction. *Appl. Sci.* **2021**, *11*, 5909. [CrossRef]
9. Kikuchi, N.; Fukuda, T.; Yabuki, N. Future Landscape Visualization Using a City Digital Twin: Integration of Augmented Reality and Drones with Implementation of 3D Model-Based Occlusion Handling. *J. Comput. Des. Eng.* **2022**, *9*, 837–856. [CrossRef]
10. Wang, X.; Dunston, P.S. Design, Strategies, and Issues towards an Augmented Reality-Based Construction Training Platform. *J. Inf. Technol. Constr.* **2007**, *12*, 363–380.
11. Adascalitei, I.; Baltoi, M. The Influence of Augmented Reality in Construction and Integration into Smart City. *Inform. Econ.* **2018**, *22*, 55–67. [CrossRef]
12. Shin, D.H.; Dunston, P.S. Technology Development Needs for Advancing Augmented Reality-Based Inspection. *Autom. Constr.* **2010**, *19*, 169–182. [CrossRef]
13. Lee, Y.-J.; Kim, J.-Y.; Pham, H.; Park, M.-W. Augmented Reality Framework for Efficient Access to Schedule Information on Construction Sites. *J. KIBIM* **2020**, *10*, 60–69. [CrossRef]
14. Microsoft HoloLens 2—Overview, Features, and Specs | Microsoft HoloLens. Available online: https://www.microsoft.com/en-us/hololens/hardware (accessed on 1 July 2024).
15. Nuernberger, B.; Ofek, E.; Benko, H.; Wilson, A.D. SnapToReality: Aligning Augmented Reality to the Real World. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 1233–1244.
16. Lee, K. Augmented Reality in Education and Training. *TechTrends* **2012**, *56*, 13–21. [CrossRef]
17. Gierwiało, R.; Witkowski, M.; Kosieradzki, M.; Lisik, W.; Groszkowski, Ł.; Sitnik, R. Medical Augmented-Reality Visualizer for Surgical Training and Education in Medicine. *Appl. Sci.* **2019**, *9*, 2732. [CrossRef]
18. Dalager, S.; Majgaard, G. Development of an Educational AR Tool for Visualization of Spatial Figures and Volume Calculation for Vocational Education. In Proceedings of the Virtual, Augmented and Mixed Reality: Applications in Education, Aviation and Industry, Virtual Event, 26 June–1 July 2022; Chen, J.Y.C., Fragomeni, G., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 14–30.
19. Chen, Y.; Wang, Q.; Chen, H.; Song, X.; Tang, H.; Tian, M. An Overview of Augmented Reality Technology. *J. Phys. Conf. Ser.* **2019**, *1237*, 022082. [CrossRef]
20. Paavilainen, J.; Korhonen, H.; Alha, K.; Stenros, J.; Koskinen, E.; Mayra, F. The Pokémon GO Experience: A Location-Based Augmented Reality Mobile Game Goes Mainstream. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 2493–2498.
21. Han, J.-G.; Park, K.-W.; Ban, K.-J.; Kim, E.-K. Cultural Heritage Sites Visualization System Based on Outdoor Augmented Reality. *AASRI Procedia* **2013**, *4*, 64–71. [CrossRef]
22. Malta, A.; Mendes, M.; Farinha, T. Augmented Reality Maintenance Assistant Using YOLOv5. *Appl. Sci.* **2021**, *11*, 4758. [CrossRef]
23. Ajanki, A.; Billinghurst, M.; Gamper, H.; Järvenpää, T.; Kandemir, M.; Kaski, S.; Koskela, M.; Kurimo, M.; Laaksonen, J.; Puolamäki, K.; et al. An Augmented Reality Interface to Contextual Information. *Virtual Real.* **2011**, *15*, 161–173. [CrossRef]
24. Sugiura, A.; Kitama, T.; Toyoura, M.; Mao, X. The Use of Augmented Reality Technology in Medical Specimen Museum Tours. *Anat. Sci. Educ.* **2019**, *12*, 561–571. [CrossRef]
25. Vasilis, S.; Nikos, N.; Kosmas, A. An Augmented Reality Framework for Visualization of Internet of Things Data for Process Supervision in Factory Shop-Floor. *Procedia CIRP* **2022**, *107*, 1162–1167. [CrossRef]
26. Gammeter, S.; Gassmann, A.; Bossard, L.; Quack, T.; Van Gool, L. Server-Side Object Recognition and Client-Side Object Tracking for Mobile Augmented Reality. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition—Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 1–8.
27. Majid, N.A.A.; Mohammed, H.; Sulaiman, R. Students' Perception of Mobile Augmented Reality Applications in Learning Computer Organization. *Procedia-Soc. Behav. Sci.* **2015**, *176*, 111–116. [CrossRef]
28. Sousa, L.; Alves, R.; Rodrigues, J.M.F. Augmented Reality System to Assist Inexperienced Pool Players. *Comp. Vis. Media* **2016**, *2*, 183–193. [CrossRef]

29. Dong, S.; Behzadan, A.H.; Chen, F.; Kamat, V.R. Collaborative Visualization of Engineering Processes Using Tabletop Augmented Reality. *Adv. Eng. Softw.* **2013**, *55*, 45–55. [CrossRef]
30. Garbett, J.; Hartley, T.; Heesom, D. A Multi-User Collaborative BIM-AR System to Support Design and Construction. *Autom. Constr.* **2021**, *122*, 103487. [CrossRef]
31. Li, X.; Yi, W.; Chi, H.-L.; Wang, X.; Chan, A.P.C. A Critical Review of Virtual and Augmented Reality (VR/AR) Applications in Construction Safety. *Autom. Constr.* **2018**, *86*, 150–162. [CrossRef]
32. Kamat, V.R.; Behzadan, A.H. GPS and 3DOF Tracking for Georeferenced Registration of Construction Graphics in Outdoor Augmented Reality. In Proceedings of the Intelligent Computing in Engineering and Architecture, Ascona, Switzerland, 25–30 June 2006; Smith, I.F.C., Ed.; Springer: Berlin, Heidelberg, 2006; pp. 368–375.
33. Zollmann, S.; Hoppe, C.; Kluckner, S.; Poglitsch, C.; Bischof, H.; Reitmayr, G. Augmented Reality for Construction Site Monitoring and Documentation. *Proc. IEEE* **2014**, *102*, 137–154. [CrossRef]
34. De Pace, F.; Kaufmann, H. A Systematic Evaluation of an RTK-GPS Device for Wearable Augmented Reality. *Virtual Real.* **2023**, *27*, 3165–3179. [CrossRef]
35. Job Site Productivity Tools: Digital Twin and Construction-Grade Augmented Reality. Available online: https://www.vgis.io/ (accessed on 23 June 2024).
36. Chai, C.S.; Klufallah, M.; Kuppusamy, S.; Yusof, A.; Lim, C.S. BIM Integration in Augmented Reality Model. *Int. J. Technol.* **2019**, *10*, 1266. [CrossRef]
37. Hübner, P.; Weinmann, M.; Wursthorn, S. Marker-Based Localization of the Microsoft Hololens in Building Models. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII–1*, 195–202. [CrossRef]
38. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
39. Du, Y.; Zhao, Z.; Song, Y.; Zhao, Y.; Su, F.; Gong, T.; Meng, H. StrongSORT: Make DeepSORT Great Again. *IEEE Trans. Multimed.* **2023**, *25*, 8725–8737. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.