



## Article

# Enhancing Digital Twins with Human Movement Data: A Comparative Study of Lidar-Based Tracking Methods

Shashank Karki <sup>1,\*</sup> , Thomas J. Pingel <sup>2</sup>, Timothy D. Baird <sup>1</sup> , Addison Flack <sup>1</sup> and Todd Ogle <sup>3</sup>

<sup>1</sup> Department of Geography, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA; tbaird@vt.edu (T.D.B.); aflack9@vt.edu (A.F.)

<sup>2</sup> Department of Geography, Binghamton University, Binghamton, NY 13902, USA; tpingel@binghamton.edu

<sup>3</sup> University Libraries, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA; jogle@vt.edu

\* Correspondence: shashankkarki@vt.edu; Tel.: +1-540-231-7557

**Abstract:** Digital twins, used to represent dynamic environments, require accurate tracking of human movement to enhance their real-world application. This paper contributes to the field by systematically evaluating and comparing pre-existing tracking methods to identify strengths, weaknesses and practical applications within digital twin frameworks. The purpose of this study is to assess the efficacy of existing human movement tracking techniques for digital twins in real world environments, with the goal of improving spatial analysis and interaction within these virtual modes. We compare three approaches using indoor-mounted lidar sensors: (1) a frame-by-frame method deep learning model with convolutional neural networks (CNNs), (2) custom algorithms developed using OpenCV, and (3) the off-the-shelf lidar perception software package Percept version 1.6.3. Of these, the deep learning method performed best (F1 = 0.88), followed by Percept (F1 = 0.61), and finally the custom algorithms using OpenCV (F1 = 0.58). Each method had particular strengths and weaknesses, with OpenCV-based approaches that use frame comparison vulnerable to signal instability that is manifested as “flickering” in the dataset. Subsequent analysis of the spatial distribution of error revealed that both the custom algorithms and Percept took longer to acquire an identification, resulting in increased error near doorways. Percept software excelled in scenarios involving stationary individuals. These findings highlight the importance of selecting appropriate tracking methods for specific use. Future work will focus on model optimization, alternative data logging techniques, and innovative approaches to mitigate computational challenges, paving the way for more sophisticated and accessible spatial analysis tools. Integrating complementary sensor types and strategies, such as radar, audio levels, indoor positioning systems (IPSS), and wi-fi data, could further improve detection accuracy and validation while maintaining privacy.

**Keywords:** terrestrial lidar; indoor geography; deep learning; dynamic digital twins; computer vision; lidar perception; occupancy and movement



**Citation:** Karki, S.; Pingel, T.J.; Baird, T.D.; Flack, A.; Ogle, T. Enhancing Digital Twins with Human Movement Data: A Comparative Study of Lidar-Based Tracking Methods. *Remote Sens.* **2024**, *16*, 3453. <https://doi.org/10.3390/rs16183453>

Academic Editor: Giuseppe Casula

Received: 5 August 2024

Revised: 10 September 2024

Accepted: 12 September 2024

Published: 18 September 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

The average American spends approximately 87% of their time inside buildings [1,2], yet indoor spaces are much less well mapped compared with outdoor spaces [3]. While satellites and aerial photography have provided ample opportunities to collect data for outdoor spaces over many decades [4–6], tools for examining indoor spaces have been limited [7–9]. Mapping and surveying methods and technology have developed greatly in the last century with the combined efforts of industry, government, and organizations evolving from traditional methods, including chain and compass, to the optical era of theodolites, electronic distance measurement, and aerial photogrammetry [10]. The development of multi- and hyper- spectral imagery in the satellites has allowed the mapping of other properties of the Earth, including temperature [11,12], vegetation [13,14], disaster

monitoring [15,16], and the monitoring of public health [17,18], among other uses [19,20]. Research and development efforts using satellite imagery have grown greatly in recent years, with more derived products like land cover/land use [21], city growth [22], and even temporal analysis of physical phenomena readily available. However, traditional surveying equipment and technologies designed for outdoor spaces are not effective indoors [23]. For example, GPS signals, which have revolutionized outdoor mapping [24,25], are weakened by roofs and walls, leading to inaccurate indoor positioning [26]. Similarly, conventional surveying tools that depend on line-of-sight measurements are ineffective in the geometrically complex and obstructed indoor settings [27].

Recent advancements in sensor technology, particularly lidar and photogrammetry, offer cost-effective means to map indoor spaces with high detail [28,29]. Lidar's precision in generating three-dimensional data makes it ideal for capturing indoor complexities [30], while structure from motion (SfM) photogrammetry provides high-resolution maps and models [9]. Unlike outdoors, where a single satellite can apprehend most of a scene by itself, it is impossible to collect data from a single location for indoor spaces. Indoor spaces are typically more complex than outdoor spaces, with many objects that can obstruct the field of view. This is true at the building scale and within rooms where views are generally obstructed by various objects. Lighting conditions and clutter in indoor spaces make it difficult to differentiate between objects and background elements [31]. As a result, indoor mapping typically has relied on either a mobile mapper or multiple static sensors from different viewpoints. Both approaches are complex and are difficult to combine accurately [32]. Photogrammetry can be particularly difficult indoors, as variable lighting conditions and privacy concerns for people in the space are among the challenges of using image-based data acquisition methods [33]. Both lidar and photogrammetry rely on GPS outdoors to geolocate scans. Mapping indoor spaces with lidar or photogrammetry typically involves carrying a handheld or backpack-mounted 360-degree camera or scanner. However, as GPS cannot easily be used indoors, it is difficult to apply Earth-based coordinates to indoor spaces [34]. To compensate, operators will sometimes scan both the interior and exterior simultaneously or align scans manually based on ancillary geospatial data. The development of indoor positioning systems (IPS) further complements these technologies, enabling real-time data collection within buildings [35,36].

In some ways, indoor maps closely resemble maps of outdoor spaces. For instance, they commonly map areal and linear spaces and boundaries, including doors, floors, roofs, and other structural parts of a building [37]. Common indoor maps include floor plans of buildings and underground railway systems and are commonly used for general navigation or for a specific use, such as to show emergency information like fire exit routes [38]. In outdoor spaces, mapping is generally accomplished by means of global coordinate systems such as latitude and longitude, or more specialized systems such as the universal transverse Mercator are most often used, and these generally rely on a limited set of commonly used reference data. Indoors, features like room numbers and floor numbers are frequently employed as references by which to identify arbitrary positions inside a building area in much the same way that addresses and street names function as a location scheme outdoors. For instance, room 101 may refer to the first room on the first floor's hallway [37]. Methods for defining coordinate reference systems for individual buildings that can be used in tandem with exterior coordinate systems have recently been defined by geographers [39]. Indoor maps have many applications, including indoor navigation, virtual reality, public safety management, and the integration of robotics in daily activities. The use of GIS technology to combine data from indoor and outdoor spaces can have several benefits. The process would involve aligning different coordinate systems used for each type of space and then combining the data into a single, unified spatial data frame. This would allow accurate positioning in all places, allowing for smooth navigation between indoor and outdoor locations [40]. This can be useful for applications such as wayfinding and emergency evacuation, where it is important to provide accurate and reliable navigation both inside and outside of buildings. For the past few decades, GIS for

buildings has often relied on building information models (BIMs), essentially computer-assisted design (CAD) models prepared during the construction of the buildings. Compared with relatively feature-rich conventional outdoor geospatial data, these models tend to be very schematic in nature. They are typically more oriented to building construction and convey information related to the strength and mechanics of the structure [41]. For static modeling of indoor spaces, such as for creating a model of real estate for potential buyers, 360 photography and virtual reality have been widely employed. Matterport is a company that models indoor spaces that can be presented through online platforms and allows potential buyers to view the space without being physically present [42]. The company has also partnered with other organizations to develop applications for disaster response, heritage preservation, and other fields [43]. While BIM models are effective abstractions of indoor spaces, their focus is construction and structural integrity, rather than cartographical functionality. From a mapping perspective, point clouds and meshed models are complex data models, and BIMs are used for navigational and inventory management purposes. Recent developments in mapping indoors include GIS-based solutions, including open standards like City Geographic Markup Language (GML), Indoor Geography Markup Language (IndoorGML), OpenStreetMap (OSM), and the Facilities Information Spatial Data Model (FISDM). These standards map indoor spaces and buildings in different level of detail (LOD) schemes [39]. Esri, a leader in the field of geospatial sciences, is also developing indoor mapping tools such as FISDM and ArcGIS Indoors.

Building on these technological foundations, the concept of *digital twins* has emerged as a transformative paradigm in the field [44]. By integrating data from lidar, photogrammetry, and IPS, digital twins create detailed, representative, and realistic models of buildings. Digital twins are virtual counterparts for physical entities and the data connections between them. They are becoming increasingly popular ways to enhance the performance of physical entities by using computational techniques without affecting the physical entity itself. Digital twins create a mirrored version of physical objects by capturing sufficient information and representing them in the virtual platform simplifying the ability to do mathematical computations [45].

Digital twins can simulate indoor spaces well, with different levels of detail showing different components of a structure [46]. With the integration of GIS into the technology, a digital twin is equipped to depict the geographic dimensions of a physical entity. With the ongoing development of physical–cyber fusion and geospatial technology, environments, entities, and networks are being modeled holistically in digital form, connecting information, systems, and behavior [47].

Our paper contributes to this field by demonstrating how lidar technology can enhance digital twins with real-time human movement tracking in indoor environments. Unlike traditional tracking methods, the use of lidar overcomes challenges such as privacy issues, signal instability and occlusions, which are often present in complex indoor spaces. This novel application of lidar for human movement tracking within digital twin frameworks offers new opportunities for improving the accuracy and responsiveness of virtual representations in real-world settings. Our approach, leveraging lidar's high-precision spatial data, further strengthens the utility of digital twins by ensuring accurate, anonymized tracking of human presence and interactions, contributing to more dynamic and realistic simulations.

These developments support the monitoring of indoor space without affecting ongoing activities. They can be used to simulate the performance of spaces in various scenarios, such as how people's behaviors adjust to different levels of occupancy or changes in environmental settings. Such simulations can help to optimize the use of indoor space and even identify potential issues before they arise in the physical world [48]. This has significantly improved our ability to map and model the physical aspects of indoor environments [49], especially our capability to represent the spatial layout, structural composition, and inter-relationships between objects. These models are not only theoretical but have practical applications across various fields, including real estate [42], emergency response [50,51], and facility operations and maintenance [52,53].

The applications of object detection to indoor spaces are increasingly popular, with smart offices and shared indoor spaces being a major part of modern buildings. The techniques for object detection and tracking can support the many types of spatial analysis. For example, the study of how socio-physical behavior changes based on the layout of space in enclosed areas gives a proper understanding of the association between people and places [54]. Understanding such socio-physical interaction helps planners design and allocate resources to infrastructure before construction, as well as help maintain buildings according to the intensity of use of the resources. By utilizing this, it is possible to optimize the utilization of space and infrastructure, thereby improving the efficiency of the people and resources within [55]. Furthermore, modern technologies such as automated robots can also be deployed for regular household tasks with the idea of where furniture is and how people move.

The theory of space syntax is often used to analyze the spatial layout and organization of built environments, such as cities, buildings, and public spaces. It is based on the idea that the layout and configuration of space can significantly impact how people use and navigate through it [56]. For instance, recognizing majorly occupied spaces, corridors, or hallways would allow the building managers to manage traffic congestion in any setting. For a shared space, this might also mean directing traffic so that people pass through places without difficulties and can still conveniently interact with each other. This would allow planners to plan and lay out the settings of infrastructure components to increase the efficiency of space use, including residential and commercial use cases [57].

Human detection and tracking have been relatively limited to images and videos generated from mostly RGB CCTV cameras. The primary concern with this approach is the privacy concerns of people being captured in the videos and images. Though CCTV cameras are of great use for security concerns in public spaces, for analysis purposes, privacy is a big concern [58]. With lidar, the data are collected as a set of 3D points in space. The lidar streams capture a coarse human shape when people pass the sensor's field of view. This eliminates the risk of people's privacy being compromised with the study. Previous studies have shown that, although RGB-based methods can be highly effective in controlled environments [59,60], privacy issues often hinder their application in public or shared spaces [61]. In contrast, lidar provides an anonymized alternative that preserves privacy without compromising accuracy, making it more suitable for applications in sensitive settings [62].

Lidar also offers additional capabilities, including the ability to work under dynamic lighting conditions [63]. In comparison with structure from motion (SfM) photogrammetry, which often struggles with repetitive textures in indoor spaces and requires complex key-point matching [64,65], lidar streams produce less noisy data and offer higher spatial resolution. The models using SfM require key point matching, which is very complex for a repetitive texture in indoor spaces. Further, the lidar-derived point clouds are inherently spatial compared with images and videos captured, as the data is collected as a collection of  $x$ ,  $y$ , and  $z$  distances relative to the sensor. While prior research has demonstrated the effectiveness of image-based methods in outdoor environments, our work, using lidar, extends this capability to complex indoor environments, offering greater accuracy in positioning and preserving privacy, particularly in shared spaces where human activity may involve sitting or interacting with objects [66,67]. Additionally, though images and videos are easy to analyze objects with, lidar streams tend to give more accurate information about positioning.

Radar, by contrast, operates effectively in adverse weather conditions like fog or rain but provides lower spatial resolution, which limits its accuracy in detecting subtle human movements [68,69]. However, radar's robustness in harsh environments makes it a strong candidate for outdoor applications where lidar's performance might be compromised. Previous studies [62,70,71] comparing radar and lidar for localization and mapping have demonstrated that, while radar is effective in challenging weather conditions, lidar remains the superior choice for environments where precision is important.

The field of object detection has been strongly influenced by advancements in computer vision [72], deep learning and high-performance computing [73]. The ability of computer vision to differentiate people from background objects has enabled computers to assess simple video feeds and extract comprehensive information. Most of these methods are based on detecting the motion of objects using background subtraction and appropriate spatiotemporal filtering techniques. After this, the object's shape, texture, and type of motion can be used to characterize the object as a human being [74]. OpenCV is a comprehensive library of programming functions aimed at real time computer vision. It provides tools and techniques to process images and videos to identify objects, classify human actions, track movements, and monitor moving objects. Deep learning has enhanced object detection capabilities within images or video frames [75]. You Only Look Once (YOLO) is a deep learning model capable of real-time object detection, including prediction of bounding boxes and probabilities [76]. Building upon the YOLO model, the YOLOv5 model is a recent model in object detection, offering improved speed and accuracy [77]. Over the years leading to YOLOv5, each version of YOLO introduced significant improvements. YOLOv2 or YOLO9000, improved the resolution at which the model processes images [78] while YOLOv3 added detection capabilities at different scales, incorporating the use of residual blocks to help the model better generalize a variety of object sizes [79]. Further, YOLOv4 introduced additional features like weighted residual connections, cross-stage partial connections and cross mini-batch normalization. The newer versions, YOLOv6 [80] and YOLOv7 [81], introduced significant enhancements to the detection capabilities. YOLOv6 focuses on optimizing speed and accuracy for industrial applications, achieving impressive frame rates and detection precision across various hardware platforms. YOLOv7, meanwhile, sets a new benchmark for real-time object detectors, surpassing previous models in both speed and accuracy, and demonstrating exceptional performance on the MS COCO dataset without the need for pre-trained weights. While YOLOv6 and YOLOv7 introduce significant enhancements, YOLOv5 is often preferred for its relative stability and efficiency, requiring significantly fewer computational resources.

Convolutional neural networks (CNNs) are vital in deep learning-based object detection models like YOLO and YOLOv5 [76,77]. CNNs are specifically designed to adaptively learn spatial hierarchies of features from input images, making them highly effective for tasks such as object recognition and localization [75,82]. The training of YOLOv5 requires a large annotated dataset that includes diverse objects and scenarios [76]. In recent years, deep learning models like YOLOv4 and YOLOv5 have become standard for object detection across multiple domains, including surveillance and smart cities [83–86]. However, their application in lidar-based human detection remains underexplored, which is an area our research addresses, comparing their performance with other traditional models.

Lidar has also recently been used in real-time streaming and object detection. Companies such as Blickfeld (Munich, Germany), Velodyne (San Jose, CA, USA), Outsight.ai (Paris, France), Seoul Robotics (Seoul, Republic of Korea), and Exwayz (Bidart, France) have been developing specialized lidar sensors and “perception” system software for this purpose. Blickfeld, Seoul Robotics, and Outsight.ai are leading research on lidar technology for the internet of things (IoT), smart cities, and efficient robotics development. Exwayz is leading research in real-time 3D lidar processing, and Velodyne has been a leader in developing lidar sensors. Lidar perception software acts to detect multiple objects through the field of view of multiple sensors [87]. Combined with maps, the detection of movement patterns can facilitate the analysis of how a particular space is being used [88]. This has many applications, including crowd management, emergency response, driver assistance, and transportation safety.

Despite these advancements, there remains a notable gap in capturing an important aspect of indoor spaces—human behavior—though new geographic approaches are emerging [89]. Digital twins are mostly static representations that do not currently represent human activities and dynamics within built spaces. Earlier works utilizing lidar for indoor space mapping have focused primarily on static structures rather than dynamic human

movement [90,91]. Our study expands upon this by integrating deep learning models to track and analyze human behavior in real time, bridging a gap in current lidar-based mapping techniques. These dynamic digital twins offer a more holistic understanding of how buildings serve their users and can enhance simulation and predictive analytics without disrupting the physical environment. By simulating various scenarios and assessing the impacts of design and operational changes, these models may help inform decisions related to space utilization, user experience, and operational efficiency.

To build towards a more dynamic digital twin, this project uses a network of fixed-mounted lidar sensors within the mixed-use Building X at University X. The choice of lidar technology over camera-based methods was deliberate, prioritizing the anonymity of individuals within the space using lidar's ability to generate precise 3D models of indoor environments without capturing identifiable information, thus offering a balance between detailed spatial data and privacy considerations [92]. Designed to encourage interaction and collaboration among its occupants, the study space within Building X aligns with the project's broader objectives. The research is part of a larger project, titled the Building Ecology Project, which aims to investigate various methodologies, including sensor-based and qualitative and quantitative social approaches, to understand how patterns of use, movement, and place evolve over time and respond to various disturbances and shifts in resource distribution. Additional portions of the project include interviews, questionnaires, and other methods to understand how this newly constructed building becomes a place imbued with meaning. Ultimately, the project aims to understand human–building interactions from a geographic perspective, enhancing our ability to design spaces that adapt to and fulfill the complex needs of their users.

This study involves the development and evaluation of lidar-based algorithms to track occupancy and movement in this shared indoor space. Two methods use computer vision as their basis [93]—deep learning [73,76,82] and OpenCV [94]—and utilize a rasterized version of the lidar data. A third method uses Blickfeld's lidar perception software, Percept version 1.6.3., to detect moving objects in real time within the lidar data stream. To assess these methods, we use traditional metrics, such as F1 score, recall and precision. This comparative analysis aims to highlight the capabilities and limitations of current technologies in capturing the intricacies of human activity in indoor spaces.

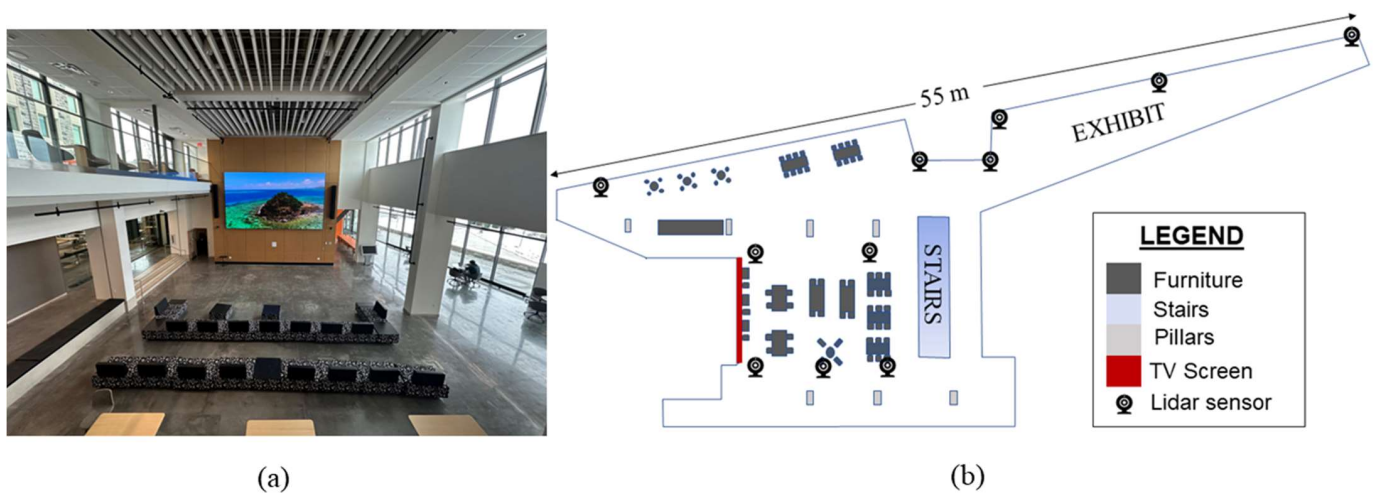
## 2. Materials and Methods

To better understand aggregate occupancy and movement patterns of people in public spaces while maintaining privacy, the study evaluated three different lidar-based object detection methods for people moving through indoor spaces: (1) Blickfeld's lidar perception software Percept version 1.6.3., using a full 3D-point cloud; (2) deep learning; and (3) OpenCV-based approaches on a projected 2D raster of the lidar data. These approaches each have different strengths and challenges. Because Percept uses the entire 3D point cloud it has additional information that can be applied to segmentation and detection; however, it is also more computationally demanding and more affected by noise in the data. Another drawback is that it can only be run on streaming real time data and cannot be operated on data that are recorded. The deep learning and OpenCV approaches were developed to meet these drawbacks while maintaining a high degree of accuracy.

The selection of the three object detection methods was driven by a combination of concerns, as follows: data compatibility, computational efficiency, and detection accuracy. Initial explorations included various methods, including NetCDF analysis using space-time cubes, deep learning with ArcGIS version 3.2.1., and several computer vision techniques, such as mean shift algorithms. However, these methods exhibited significant drawbacks. The space-time cube method proved unfeasible due to the data's high volume (~177 billion pixels per day), and processing capabilities. Deep learning approaches in ArcGIS struggled with the high resolution of the lidar data, leading to suboptimal performance. Similarly, mean shift algorithms using OpenCV did not yield satisfactory results in detecting people with requirements for noise filtering and preprocessing. Other

image processing techniques, such as subtracting median images, also fell short in accuracy and were computationally intensive. The chosen methods—YOLOv5, OpenCV and Percept—were selected because they provided a balanced solution to these issues. Specifically, they offered reliable detection of people in the lidar streams while managing computational resources effectively.

The study was conducted in the first-floor common areas of University X's newly constructed Building X, a 225,000 ft<sup>2</sup>, mixed-use building. Opened in August 2021, Building X was designed to promote interaction and collaboration among approximately 600 undergraduate residents. In addition to its function as a residence hall, Building X was designed as a dynamic venue for academic and extracurricular activities, and contains several types of space, including classrooms, studios, galleries, lounges, and a makerspace. This study specifically targets the main lobby (i.e., community assembly) and its attached hallways and exhibit space, areas characterized by considerable student activity, collaborative work, and social interaction (Figure 1). This is an ideal location for examining space use patterns and interactions among people from diverse backgrounds that use the shared space to work, move through, and interact during different hours of the day, days of the week, and across different weeks of the year.



**Figure 1.** Overview of Building X's public space: (a) arrangement of the community assembly and the general layout. (b) Schematic diagram of the study area.

Eleven Blickfeld Cube 1 (Blickfeld, Munich, Germany) lidar sensors were used to image the space. The Blickfeld Cube is a small (60 mm × 82 mm × 50 mm, 275 g), forward-facing, solid-state sensor with a configurable interface to adjust the parameters including field of view, scan resolution, and network time synchronization among other settings. The Cubes used in this study were configured to a field of view (FOV) of 72 × 30 degrees, a resolution of 230 scan lines, and a refresh rate of 2.4 Hz. The range precision, serving as a metric for accuracy, was <2 cm (bias-free RMS, 10 m, 50% reflectivity target), ensuring high measurement consistency under these conditions [95]. Additionally, a comparison of point density across 10 point clouds yielded a mean density of 578 points per square meter (median = 647), further confirming the high resolution and granularity of the data captured by the system. Point density varied across the point cloud based on the overlap of the sensors at any given point and had a 5–95th percentile range of 120 to 1846 points per square meter.

Reported data include position, velocity, shape, and trajectory information. Data from the Cube is sent via an ethernet port via TCP/IP to a computer that collects the data. The sensors' placement and orientation were aligned to a pre-collected lidar model of Building X's interior physical space (collected using a GeoSLAM Zeb Horizon lidar scanner, GeoSLAM, Nottingham, United Kingdom) to facilitate alignment with a building coordinate reference system (BCRS), as described by Chen and Clarke (2019) [39]. Construction of

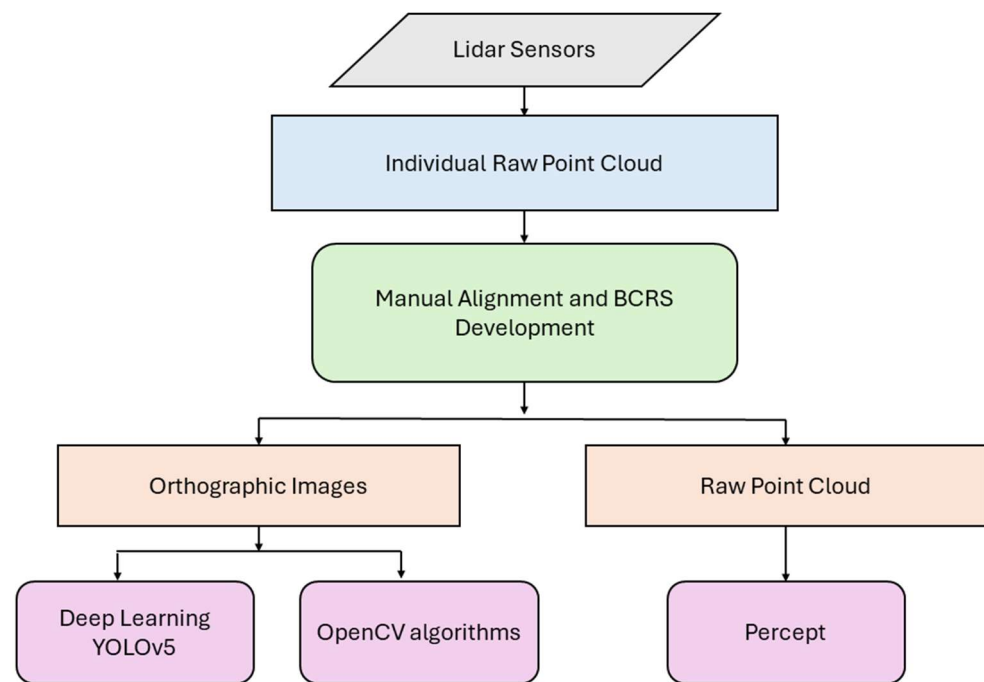
a BCRS involves defining a building-specific origin point and axes, and a transformation equation from that datum to a geographic coordinate reference system. The 3D point clouds from the 11 Blickfeld Cube sensors were manually aligned using CloudCompare software version 2.12, based on a pre-recorded GeoSLAM scan of the same space. Manual alignment was necessary as the iterative closest point (ICP) algorithm did not yield satisfactory results due to the complex indoor environment. The manual alignment process was critical for determining the exact positions of the sensors and establishing the building coordinate reference system (BCRS). Ultimately, the study's 11 sensors were then fused into a single point cloud with each sensor's unique numeric ID embedded under the 'pointsourceID' field.

Each of the 11 lidar sensors in Building X was continuously monitored by a computer connected with a private network. Custom logging software utilized the Blickfeld scanner Python library to retrieve point cloud data from all sensors once per second. Each sensor directly incorporates its position and orientation and transmits its coordinates in the BCRS developed for the study. The fused data were then cropped to contain only the study area. A second filter removes all floor and ceiling points so that only objects, walls and furniture are visible. Two 2D orthographic images are then created by projecting the same 3D point cloud data as PNGs with a 5 cm resolution, capturing both maximum elevation and maximum intensity (scaled 0–255). These projections use an orthographic view, maintaining the same sensors for both 3D point cloud capture and 2D image creation. A full 3D point cloud frame is recorded in LAZ (a compressed version of the LAS file format, which stands for LIDAR aerial survey) format every 10 s. These logs have been generated for Building X since mid-August 2022. The layout of the study area posed significant challenges for sensor deployment. The study area includes numerous pillars, objects, and multi-level floors, which affected the placement of the sensors. The variable ceiling height also imposed restrictions in the lower ceilings, like in the north hallway, meaning that sensors had to be placed with a relatively horizontal view instead of a more top-down view as was the case in the main community assembly space. The deployment required achieving a balance between sensor coverage overlap and minimizing occlusions caused by structural elements.

Daily logs are then aggregated in three ways. First, all logs are combined into hourly network common data form (NetCDF) files. NetCDFs are multidimensional datasets—in this case a stack of rasters taken once per second—and with the same two dimensions of maximum elevation and intensity. Second, each hour's data are aggregated for both dimensions to provide an hourly summary of activity. This measures how many non-zero pixels were active during that hour. Third, each day's images are aggregated into an animation using FFmpeg software version 4.4.2. An example of a portion of a day's animation is available as Video S1 in the Supplementary Materials.

Three methods of object detection were used: deep learning, OpenCV-based methods, and Percept software. Object detection is often applied to RGB/color images from a camera (or frames from video) using shape, size, color and texture information. Object tracking uses video or successive image frames to identify trajectories of detected objects over time. In this study, the lidar data frames were rendered as images, but had different properties than conventional RGB images that are more typically used. The point cloud output from the sensors were used by Percept software whereas the orthographic images rendered from these point clouds were used as the input for the deep learning and OpenCV based methods (see Figure 2). These were single-band (0–255) grayscale images that represented intensity and maximum elevation (as a Digital Surface Model or DSM). Points detected by the lidar would flicker within a few centimeters, as is typical with low-power laser scans of this type. However, when rendered as raster images, this would likewise manifest as a characteristic flickering in the raster along the edges of objects that made frame-to-frame comparisons noisy. These characteristics made the OpenCV-based object detection comparatively difficult.





**Figure 2.** Schematic flow chart showing data fusion and outputs for each of the detection methods.

The use of different methods in this study highlights a key comparison between simple, resource-efficient algorithms and proprietary software. The deep learning and OpenCV-based methods, despite being open source, demonstrated comparable performance when applied to orthographic images rendered from lidar point clouds. This presents an important contrast, as these algorithms require fewer computational resources and are more cost effective compared with proprietary software like Percept. Percept’s direct use of point clouds, combined with its resource-intensive proprietary algorithms, offer advantages in terms of ease of integration and computational power. The study aims to illustrate that open-source algorithms, while less resource intensive, can still deliver comparable results, emphasizing the balance between simplicity, cost efficiency, and performance.

To evaluate object detection in this study, we implemented three distinct methods: YOLOv5, OpenCV-based approaches, and Percept software. Below, we describe the setup and performance of each method in detail, starting with YOLOv5.

### 2.1. YOLOv5

The YOLOv5 model was chosen for our deep learning-based object detection, due to its proven balance between speed and accuracy [96]. The model was trained on an 80% sample of the manually labeled set and validated against the same 20% as all other approaches. The training process was conducted using a custom configuration (YAML) file, describing the location of the images, number of classes and class names. The training was conducted over 100 epochs with a batch size of 16 and an image resolution of  $640 \times 640$  pixels. The training was undertaken for each of the different data modalities, including the intensity returns, DSM returns, and an approach that combined intensity and DSM into a two-band tiff.

### 2.2. OpenCV’s Background Subtraction and Frame Differencing

Two OpenCV approaches were employed to detect moving objects: background subtraction [97] and frame differencing [98]. The background subtraction method is a commonly used method to detect objects in computer vision and works by separating the moving parts of a scene from the stationary part, or background. It is used to find foreground objects by isolating them and comparing them with a frame where no objects are present. For the background subtraction algorithm, the function `cv2.createBackgroundSubtract-`

orMOG2() was used with a history length of 500 frames, meaning a moving window of 500 frames was used to model the background. A variance threshold of 16 was used that dictated the algorithm's sensitivity to changes in the scene. This setup was chosen to strike a balance between accurately detecting genuine motion and minimizing artifacts such as edge flickering and 'ghosting' effects caused by minor spatial and temporal misalignments.

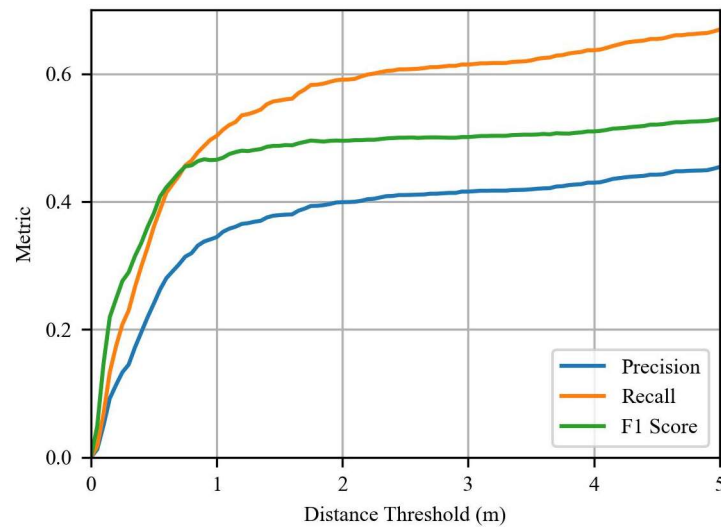
Frame differencing is a technique to check differences between two video frames. It is mostly used for removing effects of lighting in videos or tracking swiftly moving objects within the video. It is computationally less intensive compared with the background subtraction method. Here, a threshold of 25 intensity levels (intensity recorded as 0–255), measuring the intensity difference between pixels in consecutive frames, was applied to distinguish actual motion from background noise, including filtering out lidar noise. For both methods, morphological opening with a  $3 \times 3$  square kernel was used affecting a 1-pixel distance in all directions to clean the image and contour detection to identify and track moving objects. For contour detection, a value of 15 pixels squared was used in both the methods. By setting this threshold value for contour, the algorithm effectively ignores minor variations and focuses on substantial changes—those with a size larger than the threshold, significantly enhancing the algorithm's precision. These threshold values were iteratively optimized after running the model with our dataset.

### 2.3. Blickfeld Percept Software

In contrast to the image-based methods described above, Percept uses a mean shift algorithm to detect and track moving objects in the 3D lidar stream. Mean shift algorithm is a non-parametric, iterative algorithm that identifies clusters by identifying the highest density of data points. It effectively identifies the underlying structure of data by locating the maxima of a density function, making it widely used for clustering and image segmentation tasks [99]. Several tuning parameters were available in the software to improve detection. These parameters included the minimum number of frames per initialization, exponential decay, minimum weight threshold, minimum number of neighbor points, minimum points for a cluster and linear acceleration noise. Over the course of a month, these parameters were iteratively tuned to improve detection and reduce noise, with minimum number of frames for initialization set at 10, exponential decay at 0.005, minimum weight threshold at 0.17, minimum points for a cluster at 15, average radius of objects at 0.33 m, minimum number of neighbor points at 3, and neighbor radius at 0.5 m. After implementing and tuning the three object detection methods—YOLOv5, OpenCV, and Percept—the next step was to assess their accuracy in detecting and tracking objects within the study space.

### 2.4. Accuracy Assessment

The reference dataset for accuracy assessment was generated by manually labeling 3600 frames using Microsoft's visual object tagging tool (VOTT) version 2.2.0. The frames were randomly selected from periods of peak activity, with 80% used for training the deep learning model and 20% for validation. Accuracy was assessed using the Hungarian algorithm to match detected points with the reference set, with a threshold distance of 2 m. This threshold was selected after a sensitivity analysis of values up to 5 m (Figure 3). Two meters was sufficient to overcome the misalignment between points that occurred because of the sub-second temporal misalignment between Percept data and the raster files, while still being small enough to exclude clear misses (e.g., nearby false positives). During the assessment, detections outside the study space were observed. These detections were primarily due to the reflected signals from the building's glass walls, creating false positives. These detections were clipped and removed from the accuracy assessment. Traditional metrics like F1 score, precision, and recall were used. A video of each of these detection methods is included in Videos S2–S4 of the Supplementary Materials.



**Figure 3.** Distribution of variation in accuracy metrics for different distance thresholds. A notable increase in all three metrics as the threshold distance extended up to 2 m, beyond which the improvements began to saturate.

### 3. Results

This section presents the results of object detection in the study space. Each method's performance is analyzed based on object detection accuracy and efficiency, with quantitative evaluations provided for F1 score, precision, and recall. These metrics highlight the strengths and limitations of each technique, which are detailed in the following subsections.

#### 3.1. Object Detection

This section presents the comparative analysis of each of the object detection techniques: deep learning approaches on intensity, maximum elevation (i.e., DSM), and combined rasters; OpenCV background subtraction and frame differencing on intensity rasters; and Blickfeld's Percept lidar tracking software. Each method's efficacy is evaluated based on its ability to accurately detect objects of interest within the frames. The analysis is illustrated by comparing manually labeled points (shown in blue) against the algorithm-detected points (shown in green). Quantitative scores for F1 ratio, recall and precision provide performance metrics. The descriptive result for each method is as shown in Table 1.

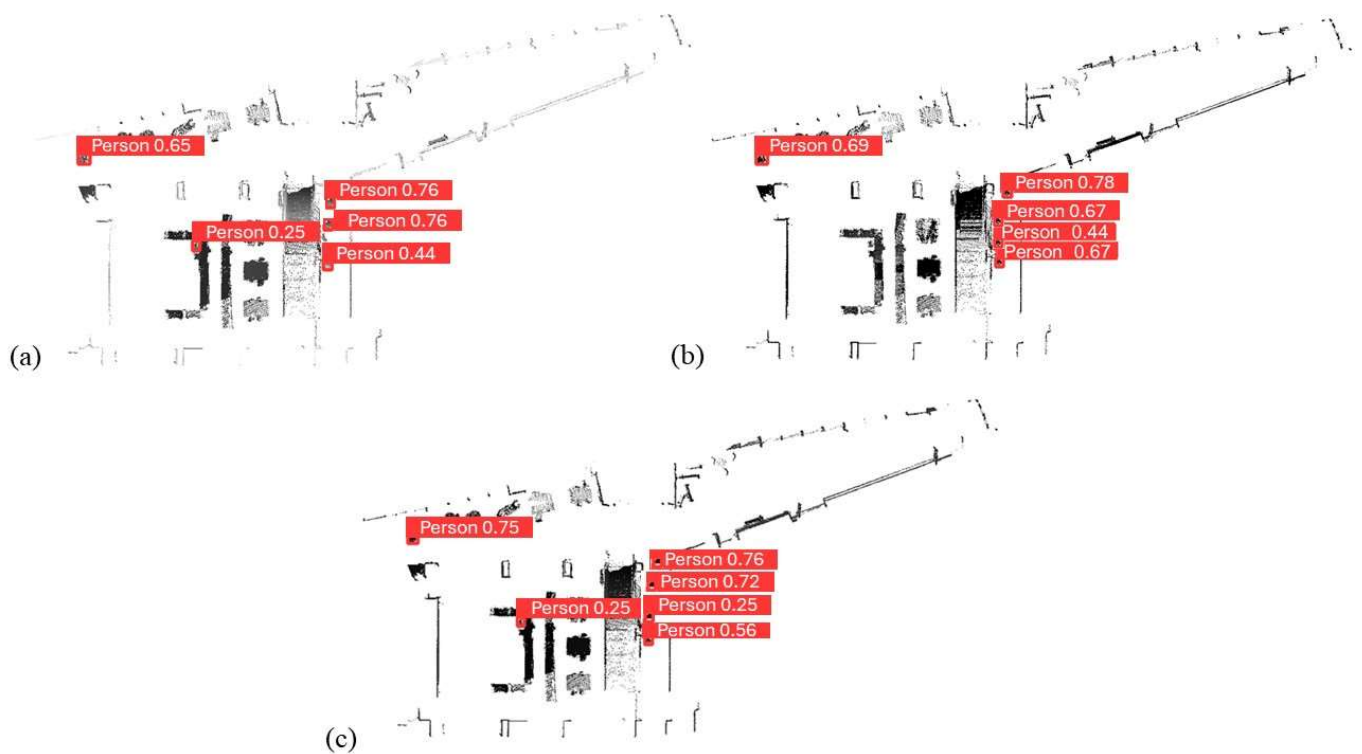
**Table 1.** Performance summary of deep learning methods, OpenCV's background subtraction and frame differencing models and Percept in the detection of human movement within the study area. Numbers in brackets for Percept discount for false positives that were in fact true detections of people sitting on furniture.

Model	Deep Learning			OpenCV		Blickfeld's Percept
	DSM	Intensity and DSM Combined	Intensity	Background Subtraction	Frame Differencing	
F1	0.879	0.875	0.869	0.575	0.438	0.606
Precision	0.843	0.826	0.827	0.858	0.491	0.507
Recall	0.919	0.930	0.916	0.433	0.395	0.754
Sum (TP)	1061 (91.9%)	1074 (93.0%)	1058 (91.6%)	500 (43.3%)	456 (39.5%)	699 (60.5%)
Sum (FP)	197 (17.1%)	226 (19.6%)	221 (19.1%)	83 (7.2%)	472 (40.9%)	680 (58.9%)
Sum (FN)	94 (8.1%)	81 (7.0%)	97 (8.4%)	655 (56.7%)	699 (60.5%)	228 (19.7%)

In the following sections, we present the results of each object detection method—deep learning, OpenCV-based approaches, and Blickfeld's Percept software.

### 3.1.1. Deep Learning

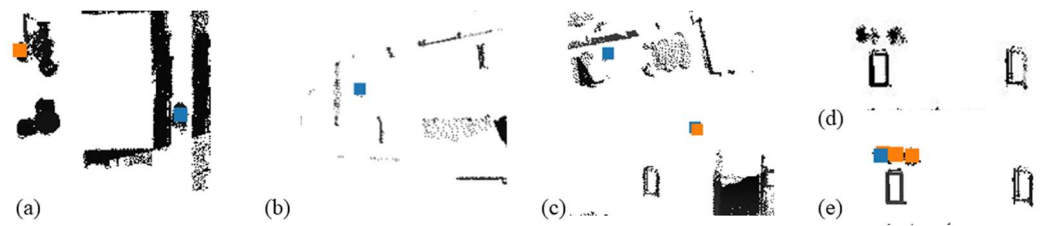
The YOLOv5 deep learning model was applied to three different raster surfaces derived from the lidar: (1) the maximum elevation returns in a grid cell (i.e., a digital surface model or DSM), (2) the maximum intensity within a grid cell, and (3) a two-band image constructed from the other two. These are illustrated in Figure 4a–c, respectively.



**Figure 4.** Sample frames showing detections and confidence levels from the YOLOv5 models trained on (a) maximum elevation (DSM) returns, (b) intensity returns, and (c) their combination as dual bands within the same image. Most detections are the same but minor differences are evident between different methods.

Overall, the deep learning approaches performed similarly well in all three metrics. The maximum F1 score was 0.879 and was similarly high for precision (0.843) and recall (0.930). The lowest false positive (FP) rate of the three was 17.1% and the lowest false negative (FN) rate was 7.0% (all results summarized in Table 1). Of the three methods, the DSM approach had the best F1 score (0.879), largely due to its better performance on precision (0.843). The two-band intensity + DSM approach performed slightly worse on F1 and precision (0.875, 0.826), but slightly better on recall (0.930). Intensity performed lowest on all three metrics (0.869, 0.827, 0.916 respectively), but all methods performed well and within a few percentage points of each other.

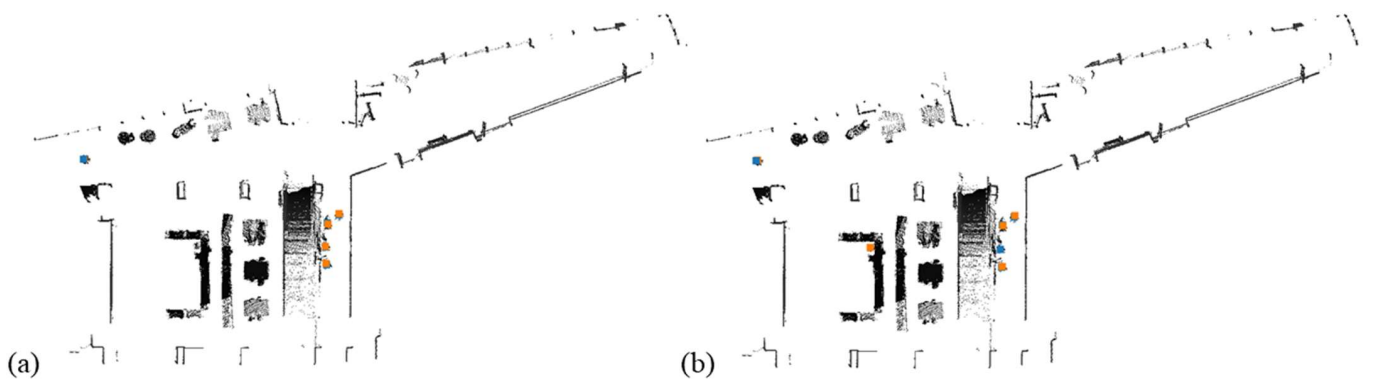
Deep learning was able to accurately detect people moving in the study space but was unable to detect people sitting on or standing near furniture, even when specifically trained to. Errors included missing detections, multiple detections for a single person, and false negatives in areas with sparse lidar coverage. Error scenarios are shown in Figure 5.



**Figure 5.** Error scenarios: (a) a missed detection due to atypical size and shape of point cloud cluster, (b) a missed detection due to inadequate coverage, (c) a missed detection due to an individual blending into the furniture as they linger, (d) a sample point cloud cluster where a single individual's lidar signature is disassociated, and (e) additional false positives from a single individual due to misalignment of sensors. The manually labeled points are shown in blue and detections shown in orange.

### 3.1.2. OpenCV

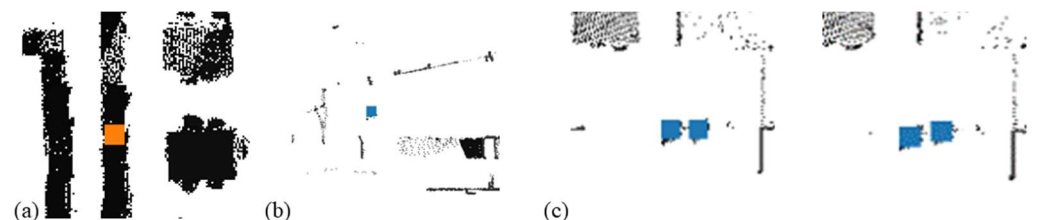
The OpenCV framework implemented two computer vision methods for object detection on the intensity rasters: (1) background subtraction and (2) frame differencing. These are illustrated in Figure 6a,b.



**Figure 6.** Sample frames showing detections from the OpenCV models using (a) background subtraction and (b) frame differencing.

Background subtraction achieved a high precision rate of 0.858 but struggled with recall (0.433), resulting in an F1 score of 0.575. Frame differencing had a precision rate of 0.491 and a recall rate of 0.395, resulting in an F1 score of 0.438.

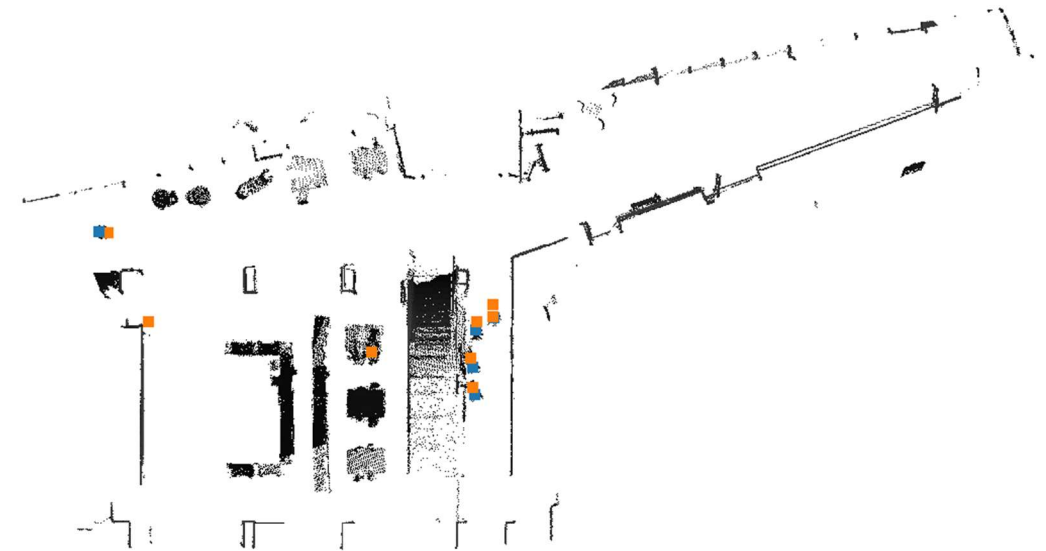
The OpenCV models had significant error scenarios, including detections due to edge flickering, low sensor coverage, and stationary individuals. Each model took approximately one hour to process a day's data. Error scenarios are demonstrated in Figure 7.



**Figure 7.** Error scenarios: (a) False positive due to raster edge flickering effects, (b) a missed detection due to lower density lidar coverage, (c) missed detections where individuals linger in the same space for some time and are not detected by the algorithm. The manually labeled points are shown in blue and detections shown in orange.

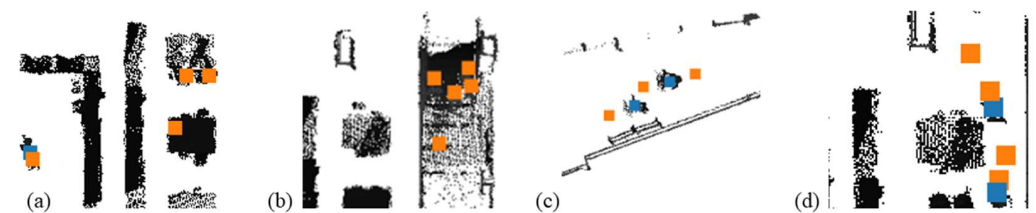
### 3.1.3. Blickfeld's Percept

Blickfeld's Percept lidar perception software records instances of movement within the lidar stream in real time. The results are shown in Figure 8.



**Figure 8.** Sample frame showing detections by Blickfeld's Percept, with manually labeled points shown in blue and detections shown in orange.

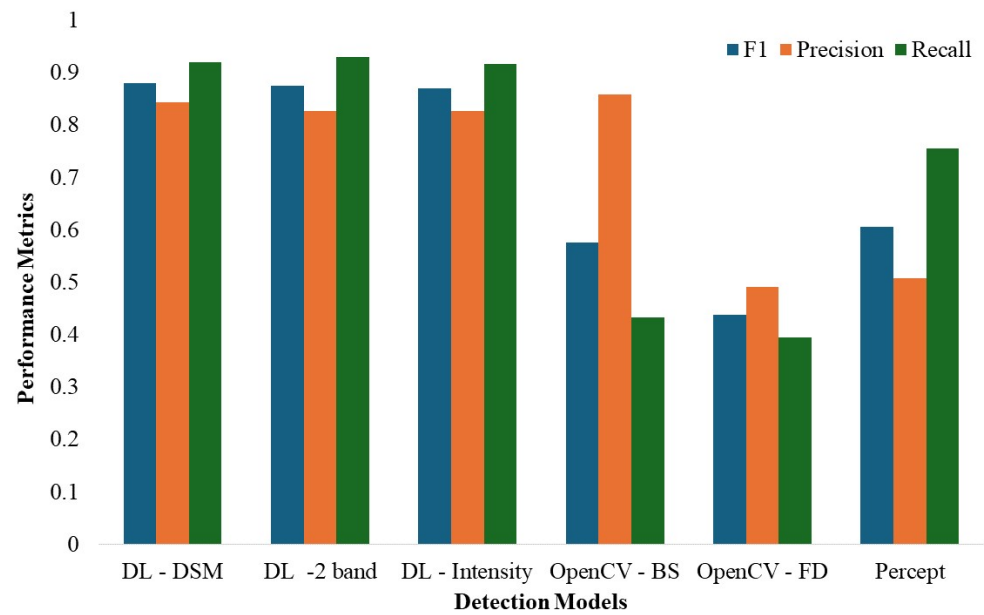
Percept achieved an F1 score of 0.606 and a recall rate of 0.754, indicating reasonable skill at recognizing movements. However, the precision rate was low (0.507) due to a high number of false positives. Detailed metrics are shown in Table 1. Percept was better at detecting people near furniture and on staircases but produced false positives due to temporal misalignment and multiple detections for single objects. Error scenarios are shown in Figure 9.



**Figure 9.** Error scenarios: (a) false positives as people in the furniture move, (b) false positives on the staircase, (c) a misalignment error between detections and ground truth points attributed to time differences between raster logs and Percept data, and (d) false positives due to multiple signatures of the same object as a result of time misalignments between sensors in Percept. The manually labeled points are shown in blue and detections shown in orange.

### 3.2. Comparison of Detection Techniques

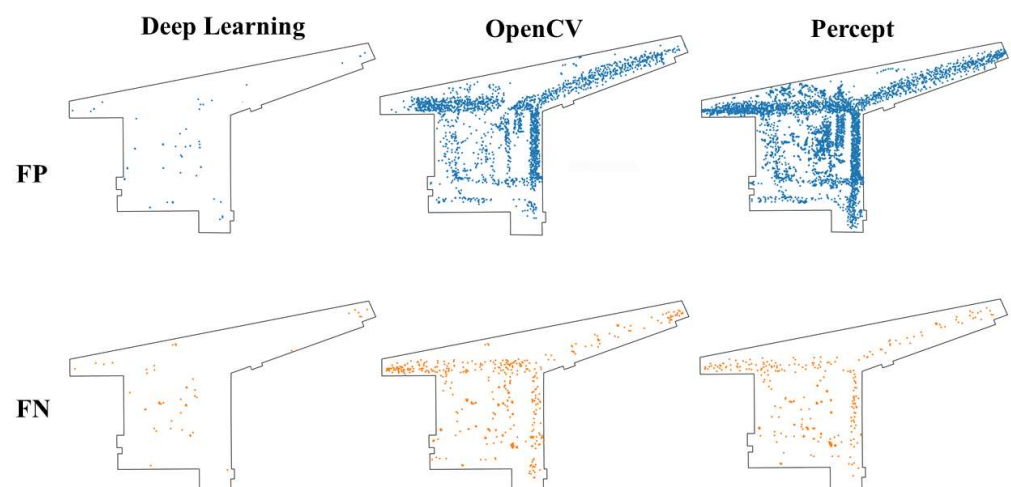
Overall, the deep learning approaches performed similarly well in all three metrics. The maximum F1 score was 0.879 and was similarly high for precision (0.843) and recall (0.930) (Figure 10). The lowest false positive (FP) rate of the three was 17.1% and the lowest false negative (FN) rate was 7.0% (all results summarized in Table 1 below). Of the three methods, the DSM approach had the best F1 score (0.879), largely due to its better performance on precision (0.843). The two-band intensity + DSM approach performed slightly worse on F1 and precision (0.875, 0.826), but slightly better on recall (0.930). Intensity performed lowest on all three metrics (0.869, 0.827, 0.916 respectively), but all methods performed well and within a few percentage points of each other.



**Figure 10.** Chart showing performance summary of detection methods including deep learning approaches for DSM, intensity and dual band; OpenCV's background subtraction and frame differencing models; and Percept.

### 3.3. Spatial Distribution of False Positives and False Negatives

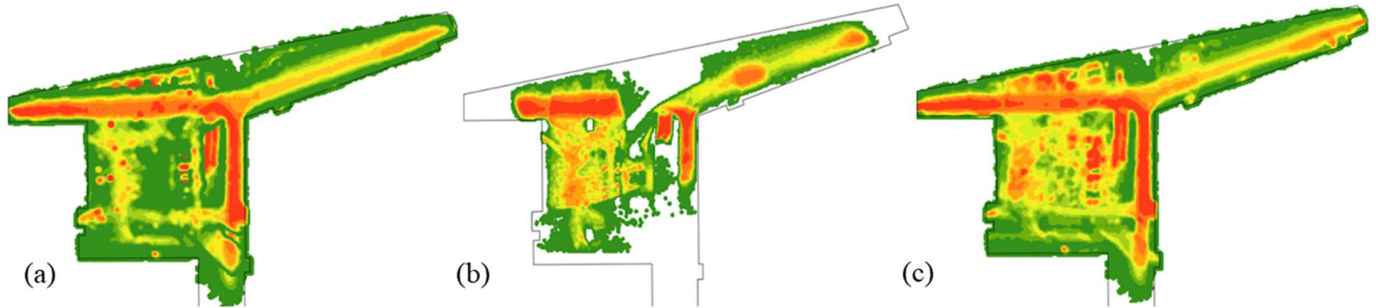
Each of the methods had its own unique pattern of error, specifically of false positives and false negatives (see Figure 11). The deep learning approach showed a much smaller number of FP and FN relative to the other two methods. Generally, patterns of both types of error were proportional to occupancy; however, some specific local differences were apparent. Here, both types of error were less likely to be in movement corridors and more likely to be near entrances and furniture. OpenCV had a large cluster of FPs around the movement corridors, but notably few near the northwest entryway and center vestibule near the foot of the staircase, the latter likely because of particularly high sensor coverage there. OpenCV FNs were clustered near entryways. Percept produced a comparatively high FP rate throughout the space due to sensitivity to temporal misalignment and genuine detections of people near furniture that were not included in the validation samples. Most FNs from Percept were in movement corridors.



**Figure 11.** Spatial distribution of detection errors across the following models: deep learning using DSM, OpenCV background subtraction and Percept, with false positives (FP) in blue and false negatives (FN) in orange. Deep learning has fewer FP and FN than OpenCV and Percept.

### 3.4. Aggregate Movement Analysis and Temporal Patterns

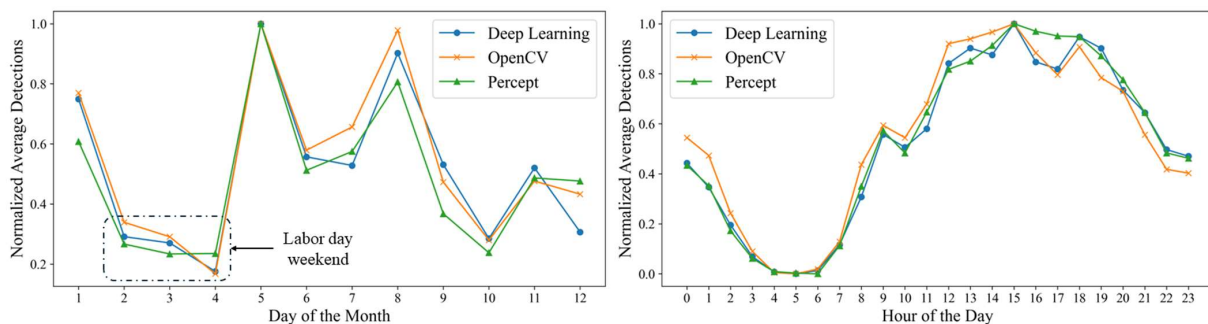
Each method (deep learning using DSM, OpenCV using background subtraction, and Percept) was applied to the lidar data for the month of September 2023 to analyze spatio-temporal patterns of occupancy. A kernel density estimation technique was used to identify and visualize the areas of high and low movement and occupancy (Figure 12).



**Figure 12.** Kernel density estimation (KDE) map of movement patterns detected by the detection models. (a) Deep learning, (b) OpenCV, and (c) Percept.

Across the detection models, the spatial patterns of use were similar. Each highlighted the northwest entryway leading towards campus, the staircase to the second floor, and the hallway, leading to the elevator as the main hotspots of activity. Notable differences between models, however, were observed. The deep learning and Percept models show the northeast corridor leading to the Blacksburg downtown area as a lesser hotspot. The deep learning model additionally showed corridors of orthogonal movement in the study area. OpenCV underrepresented the activity near entrances and the periphery compared with the other two models. This is because the OpenCV model was less sensitive to people entering the space and only detected them after they had been present for a few consecutive frames. Alternatively, Percept demonstrated a higher sensitivity to detections involving furniture, and showed more hotspots in the locations where furniture was located. This indicates that, while all three methods show similar locations for high activity within the space, Percept was better at detecting static presence or occupancy within the space.

To visualize the temporal trends in different models, the detections were aggregated by taking a daily average and normalizing them using a maximum–minimum normalization scheme. This was undertaken for a continuous two-week period in September and for hours of each day (see Figure 13). The normalized data show similar trends across all detection methods, with only minor differences. For example, Percept provides noticeably lower estimates for September 12 and 13 and notably higher estimates at 4 and 5 pm than the other two methods. The Labor Day (September 4) weekend (i.e., days 2–4) is clearly visible as the least active of the first 14 days in September.



**Figure 13.** Normalized aggregate average detections for each model (deep learning in blue, OpenCV in orange and Percept in green) within the study area for the month of September 2023. Overall trends between methods are very similar, with only some minor differences between them.



#### 4. Discussion

Each of the three approaches was successful in tracking human movement through the study space using both rasterized and full 3D lidar data streams. The deep learning method was the most successful, followed by Percept, and then the OpenCV approach. Notably, the false positive and false negative rates were considerably higher for OpenCV and Percept. Percept was moderately worse than deep learning at tracking overall movement in space but was better at tracking individuals on furniture. The logging and analysis of orthographic lidar images provided a clear benefit over using the Percept software alone. These images (1) served as a useful background layer to interpret Percept detections, (2) allowed for improved post-analysis, and (3) required less computing power than 3D analysis. They also helped to illustrate that the lidar representations preserve anonymity. Although the raster images themselves did not track furniture in the sense of  $(x, y, \text{time})$  coordinates, the furniture itself was represented in the data and could, in theory, be recovered later. However, full 3D tracking proved especially important on places like large staircases with usable activity space underneath that are not as easily represented in 2.5D.

Deep learning has been applied in point clouds for point segmentation in many domains in both 3D [100,101] and 2D analyses [102]. Given that deep learning models use each frame individually to identify people, our initial assumption was that this approach would be less effective than OpenCV, which allows for ready comparison between frames. The YOLOv5 deep learning model relies on unique textures and intensity variations to identify people in the study space. Moreover, through extensive training on a large dataset, YOLOv5 learned the specific shapes and structural characteristics of human movement within the space. Unlike sliding window and region proposal techniques, YOLOv5 considers the entire image during both training and testing, thereby capturing contextual information that enhances detection accuracy. YOLOv5's ability to learn generalizable representations of objects across diverse scenarios efficiently enhances its ability to detect people in the projected lidar frames [76]. The deep learning model showed similar performance for all three lidar derivatives (i.e., DSM, intensity, and the dual band combination of the two), with the DSM having the best overall performance.

While the OpenCV model was generally the lowest performing of the three approaches, its challenges are evident and opportunities for improvement remain. Its poor performance was partly due to the persistent edge flickering noise in the orthographic images that occurs between consecutive frames during the conversion of the 3D lidar data into the 2D orthographic images. Low power lidar units often display inconsistency in their scans [103,104]. When rasterized, the small changes of points can fall just within neighboring pixels, causing them to activate; as a result, the resulting images of consecutive frames appear to flicker, or jitter, affecting all the objects within the space. This noise was persistent in our data throughout the study space. OpenCV models were more affected by the flickering noise because they tracked movement by analyzing the difference between consecutive frames, whereas deep learning only looks at a single frame at a time. Morphological opening and the detection size parameter were adjusted to mitigate the flickering; these were partially successful but did not improve OpenCV's performance to the level of the deep learning or Percept methods. More aggressive use of opening resulted in the loss of lidar point clusters in low-coverage areas. The grid size for the rasterized lidar data was set to 5 cm; attempts to use larger grid sizes were similarly unsuccessful at mitigating the flickering effect, because the edge pixels were still activated by the unstable lidar point displacements near their edges. This represents a fundamental limitation of the raster data structure.

Percept detected people moving and pausing, whereas deep learning only detected movement well. However, while Percept does track stationary people, it loses track of them if they remain stationary for a brief period of time (~15 s). This limits the ability to continuously track individuals in scenarios where stationary behavior is common (e.g., students using the furniture for studying). Reconnecting the two (or more) separate unique IDs generated for a predominantly stationary individual can be done but requires additional

post-processing of data. Additionally, Percept's effectiveness relies on careful parameter tuning. With default settings, we found Percept produced a high number of false positives, often repeatedly in specific locations. Of the three methods, Percept was most prone to ghosting errors caused by time misalignments between sensors. These could not be eliminated even when the precise time protocol was used. If the temporal misalignments that caused the ghosting errors could be effectively addressed, Percept's performance in tracking both moving and stationary individuals could be commensurate with that of deep learning models. Percept's voluminous data requirements and resource-intensive post-analysis could be improved with the implementation of a more efficient method to log data in XYZT format, where T is the date and time of detection.

The detection of people on furniture was a challenge for all of the methods. Deep learning struggled with detecting stationary people and people on furniture, even when models were specifically trained to detect seated people and bare furniture. The inclusion of multiple classes in the training process further reduced the ability of the model to detect people moving in the space, which was of primary interest. OpenCV was better at detecting seated people, but the contour size filter designed to differentiate between noise and actual detections often filtered out detections on the furniture.

Percept's false positives included detections of people sitting on furniture—a good result, but one that negatively impacted its precision and F1 scores when only movement was considered, as it was for this study. Percept was better able to detect people on furniture as movement included more volumetric surface area (rather than just a perimeter). A subsequent analysis of Percept's detections suggests that only about 40 percent of its false positives were due to people sitting on the furniture. Adjusting for these would improve Percept's precision from 0.507 to 0.631, and F1 from 0.606 to 0.690, still considerably lower scores than the deep learning approaches. As each model had strengths and weaknesses, a hybrid detection model that combined these approaches, harnessing deep learning's frame-based detections, OpenCV's simplicity and Percept's capability to detect motion in occupancy and movement, could be superior to any of the individual approaches. As an example, deep learning could be applied to the difference between frames, rather than frames themselves, thus integrating OpenCV into its operation.

Sensor coverage plays a crucial role in determining the accuracy and reliability of object detection methods. Sensor placement required access to wired ethernet (power was provided over ethernet lines), but also needed to be sufficiently unobtrusive to maintain the overall atmosphere of the newly constructed building. The field of view of the lidar sensors was also limited. Most areas were covered by multiple sensors except for the long hallways near the doors. In these areas especially, there were missed detections with all methods, due to the reduced coverage. A combination of multiple sensors could increase sensor coverage and improve overall detection performance. One option for enhancing coverage is integrating lidar with radar, as their complementary capabilities allow for seamless operation in both indoor and outdoor settings.

## 5. Conclusions

Recent advancements in modeling indoor spaces have primarily focused on static maps and models, such as building information models (BIM) and other CAD-derived digital 3D models. However, these models do not adequately address the dynamic aspects of indoor environments, particularly human movement. This study examined how lidar-based tracking algorithms can capture the spatial dynamics of indoor spaces while maintaining the privacy of the people using them. The utilization of 2D lidar data, in the form of projected frames, presents a practical solution for efficient data storage and analysis. However, in indoor spaces with complex vertical structures or multiple floors, 2D data may not fully capture the spatial relationships.

Our assessment of object detection models revealed that deep learning outperformed OpenCV and Percept, in terms of accuracy. However, each approach had its strengths and weaknesses. Deep learning performed well operating only on individual frames but had

trouble tracking people when they sat on furniture, even when specifically trained to do so. OpenCV-based frame differencing and background subtraction performed worse than expected due to the edge flickering noise present as a result of the rasterization of the lidar data but could be developed further. Percept was particularly good at tracking people on staircases and when they sat on furniture but was the most resource intensive.

Future work should focus on model optimization (a combination of two or more of these models), explore alternative data logging techniques, and investigate innovative approaches to mitigate the computational challenges associated with 3D data analysis. Upon addressing these challenges, the accuracy, efficiency, and applicability of these technologies across various settings will be enhanced, paving the way for more sophisticated and accessible spatial analysis tools.

Additionally, the detection models could benefit from integrating complementary sensor types and strategies configured to protect privacy. These methods could improve detection accuracy and validation by providing additional context. One approach involves using microphones to capture audio levels. This can detect areas of significant activity, which can be cross-referenced with lidar detections to confirm actual human presence. Additionally, implementing indoor positioning systems (IPS) on individuals can provide precise location data to validate and enhance the accuracy of lidar-based detections, resolving ambiguities such as distinguishing between closely spaced individuals. An additional validation mechanism could involve the installation of sensors on entry and exit points, which can monitor the number of people entering and leaving the space. This ensures that the number of detected individuals aligns with entries and exits, preventing unexplained appearances in the lidar data. Moreover, cross-validating lidar detections with wi-fi usage data within the building could provide another layer of verification. Wi-fi data indicate the presence and movement of individuals based on their device connections, offering a non-invasive method by which to confirm occupancy patterns.

By integrating these complementary methods, future studies can enhance the robustness and accuracy of tracking systems in indoor environments, leading to more sophisticated and accessible spatial analysis tools while maintaining privacy.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs16183453/s1>, Video S1: Composite lidar visualization; Video S2: OpenCV background subtraction; Video S3: Deep Learning; Video S4: Percept.

**Author Contributions:** Conceptualization, S.K., T.J.P., T.D.B., T.O. and A.F.; methodology, S.K., T.J.P., T.D.B., T.O. and A.F.; software, S.K., T.J.P. and A.F.; validation, S.K. and T.J.P.; formal analysis, S.K. and T.J.P.; investigation, S.K., T.J.P., T.D.B., T.O. and A.F.; data curation, S.K., T.J.P. and A.F.; writing—original draft preparation, S.K.; writing—review and editing, S.K., T.J.P., T.D.B. and T.O.; visualization, S.K. and T.J.P.; supervision T.J.P. and T.D.B.; project administration, T.J.P. and T.D.B.; funding acquisition, T.J.P. and T.D.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the US National Science Foundation, grant number BCS-2149229.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Day Biehler, D.; Simon, G.L. The Great Indoors: Research Frontiers on Indoor Environments as Active Political-Ecological Spaces. *Prog. Hum. Geogr.* **2011**, *35*, 172–192. [[CrossRef](#)]
2. Klepeis, N.E.; Nelson, W.C.; Ott, W.R.; Robinson, J.P.; Tsang, A.M.; Switzer, P.; Behar, J.V.; Hern, S.C.; Engelmann, W.H. The National Human Activity Pattern Survey (NHAPS): A Resource for Assessing Exposure to Environmental Pollutants. *J. Expo Sci. Environ. Epidemiol.* **2001**, *11*, 231–252. [[CrossRef](#)] [[PubMed](#)]

3. Odonohue, D. Everything You Need To Know About Indoor Navigation And Mapping—April 14, 2024. 2022. Available online: <https://mapscaping.com/indoor-navigation-and-mapping/> (accessed on 14 April 2024).
4. Sabins, F.F., Jr.; Ellis, J.M. *Remote Sensing: Principles, Interpretation, and Applications*, 4th ed.; Waveland Press: Long Grove, IL, USA, 2020; ISBN 978-1-4786-4506-1.
5. Otero, R.; Lagüela, S.; Garrido, I.; Arias, P. Mobile Indoor Mapping Technologies: A Review. *Autom. Constr.* **2020**, *120*, 103399. [[CrossRef](#)]
6. Schowengerdt, R.A. *Remote Sensing: Models and Methods for Image Processing*; Elsevier: Amsterdam, The Netherlands, 2006; ISBN 978-0-08-048058-9.
7. Basiri, A.; Lohan, E.S.; Moore, T.; Winstanley, A.; Peltola, P.; Hill, C.; Amirian, P.; Figueiredo e Silva, P. Indoor Location Based Services Challenges, Requirements and Usability of Current Solutions. *Comput. Sci. Rev.* **2017**, *24*, 1–12. [[CrossRef](#)]
8. Dwiyasa, F.; Lim, M.-H. A Survey of Problems and Approaches in Wireless-Based Indoor Positioning. In Proceedings of the 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Alcalá de Henares, Spain, 4–7 October 2016; pp. 1–7.
9. Zhou, B.; Ma, W.; Li, Q.; El-Sheimy, N.; Mao, Q.; Li, Y.; Gu, F.; Huang, L.; Zhu, J. Crowdsourcing-Based Indoor Mapping Using Smartphones: A Survey. *ISPRS J. Photogramm. Remote Sens.* **2021**, *177*, 131–146. [[CrossRef](#)]
10. Wolf, P.R. Surveying and Mapping: History, Current Status, and Future Projections. *J. Surv. Eng.* **2002**, *128*, 79–107. [[CrossRef](#)]
11. Jiménez-Muñoz, J.C.; Sobrino, J.A. A Generalized Single-Channel Method for Retrieving Land Surface Temperature from Remote Sensing Data. *J. Geophys. Res. Atmos.* **2003**, *108*, 4688. [[CrossRef](#)]
12. Tomlinson, C.J.; Chapman, L.; Thornes, J.E.; Baker, C. Remote Sensing Land Surface Temperature for Meteorology and Climatology: A Review. *Meteorol. Appl.* **2011**, *18*, 296–306. [[CrossRef](#)]
13. Kogan, F.N. Remote Sensing of Weather Impacts on Vegetation in Non-Homogeneous Areas. *Int. J. Remote Sens.* **1990**, *11*, 1405–1419. [[CrossRef](#)]
14. Xie, Y.; Sha, Z.; Yu, M. Remote Sensing Imagery in Vegetation Mapping: A Review. *J. Plant Ecol.* **2008**, *1*, 9–23. [[CrossRef](#)]
15. Haque, U.; Hashizume, M.; Kolivras, K.N.; Overgaard, H.J.; Das, B.; Yamamoto, T. Reduced Death Rates from Cyclones in Bangladesh: What More Needs to Be Done? *Bull. World Health Organ.* **2012**, *90*, 150–156. [[CrossRef](#)] [[PubMed](#)]
16. Liu, H.; Weng, Q. Enhancing Temporal Resolution of Satellite Imagery for Public Health Studies: A Case Study of West Nile Virus Outbreak in Los Angeles in 2007. *Remote Sens. Environ.* **2012**, *117*, 57–71. [[CrossRef](#)]
17. Brooker, S.; Michael, E. The Potential of Geographical Information Systems and Remote Sensing in the Epidemiology and Control of Human Helminth Infections. In *Advances in Parasitology*; Remote Sensing and Geographical Information Systems in Epidemiology; Academic Press: Cambridge, MA, USA, 2000; Volume 47, pp. 245–288.
18. Hay, S.I. An Overview of Remote Sensing and Geodesy for Epidemiology and Public Health Application. In *Advances in Parasitology*; Remote Sensing and Geographical Information Systems in Epidemiology; Academic Press: Cambridge, MA, USA, 2000; Volume 47, pp. 1–35.
19. Harris, R. *Satellite Remote Sensing. An Introduction*; Routledge and Kegan Paul: London, UK, 1987.
20. Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. Available online: <https://www.hindawi.com/journals/js/2017/1353691/> (accessed on 29 November 2022).
21. Talukdar, S.; Singha, P.; Mahato, S.; Shahfahad, P.; Pal, S.; Liou, Y.-A.; Rahman, A. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sens.* **2020**, *12*, 1135. [[CrossRef](#)]
22. Alqurashi, A.F.; Kumar, L.; Sinha, P. Urban Land Cover Change Modelling Using Time-Series Satellite Images: A Case Study of Urban Growth in Five Cities of Saudi Arabia. *Remote Sens.* **2016**, *8*, 838. [[CrossRef](#)]
23. Tesoriero, R.; Tebar, R.; Gallud, J.A.; Lozano, M.D.; Penichet, V.M.R. Improving Location Awareness in Indoor Spaces Using RFID Technology. *Expert Syst. Appl.* **2010**, *37*, 894–898. [[CrossRef](#)]
24. Cihlar, J. Land Cover Mapping of Large Areas from Satellites: Status and Research Priorities. *Int. J. Remote Sens.* **2000**, *21*, 1093–1114. [[CrossRef](#)]
25. Cracknell, A.P. The Development of Remote Sensing in the Last 40 Years. *Int. J. Remote Sens.* **2018**, *39*, 8387–8427. [[CrossRef](#)]
26. Peterson, B.; Bruckner, D.; Heye, S. Measuring GPS Signals Indoors. In Proceedings of the Institute of Navigation ION GPS-97, Kansas City, MI, USA, 16–19 September 1997; pp. 615–624.
27. Koyuncu, H.; Yang, S.-H. A Survey of Indoor Positioning and Object Locating Systems. *Int. J. Comput. Sci. Netw. Secur. (IJCSNS)* **2010**, *10*, 121–128.
28. Fonstad, M.A.; Dietrich, J.T.; Courville, B.C.; Jensen, J.L.; Carbonneau, P.E. Topographic Structure from Motion: A New Development in Photogrammetric Measurement. *Earth Surf. Process. Landf.* **2013**, *38*, 421–430. [[CrossRef](#)]
29. Liu, X. Airborne LiDAR for DEM Generation: Some Critical Issues. *Prog. Phys. Geogr. Earth Environ.* **2008**, *32*, 31–49. [[CrossRef](#)]
30. Chen, J. Grid Referencing of Buildings. In Proceedings of the Adjunct Proceedings of the 14th International Conference on Location Based Services, Zurich, Switzerland, 15–17 January 2018; pp. 38–43.
31. Pintore, G.; Mura, C.; Ganovelli, F.; Fuentes-Perez, L.; Pajarola, R.; Gobbetti, E. State-of-the-Art in Automatic 3D Reconstruction of Structured Indoor Environments. *Comput. Graph. Forum* **2020**, *39*, 667–699. [[CrossRef](#)]
32. Kim, T.-H.; Park, T.-H. Placement Optimization of Multiple Lidar Sensors for Autonomous Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 2139–2145. [[CrossRef](#)]

33. Zlatanova, S.; Sithole, G.; Nakagawa, M.; Zhu, Q. Problems In Indoor Mapping and Modelling. In Proceedings of the ISPRS Acquisition and Modelling of Indoor and Enclosed Environments 2013, Cape Town, South Arfica, 11–13 December 2013; Volume XL-4-W4. pp. 63–68.
34. Kjærgaard, M.B.; Blunck, H.; Godsk, T.; Toftkjær, T.; Christensen, D.L.; Grønbaek, K. Indoor Positioning Using GPS Revisited. In *Pervasive Computing*; Floréen, P., Krüger, A., Spasojevic, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 38–56.
35. Ijaz, F.; Yang, H.; Ahmad, A.; Lee, C. Indoor Positioning: A Review of Indoor Ultrasonic Positioning Systems. In Proceedings of the Advanced Communication Technology (ICACT), 2013 15th International Conference, PyeongChang, Republic of Korea, 27–30 January 2013; p. 1150, ISBN 978-1-4673-3148-7.
36. Mautz, R. Overview of Current Indoor Positioning Systems. *Geod. Ir Kartogr.* **2009**, *35*, 18–22. [[CrossRef](#)]
37. Li, K.-J. Indoor Space: A New Notion of Space. In *Web and Wireless Geographic Information System*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–3.
38. Montello, D.R. You Are Where? The Function and Frustration of You-Are-Here (YAH) Maps: Spatial Cognition & Computation; Volume 10, pp. 2–3. Available online: <https://www.tandfonline.com/doi/abs/10.1080/13875860903585323> (accessed on 16 May 2024).
39. Chen, J.; Clarke, K. Indoor Cartography. *Cartogr. Geogr. Inf. Sci.* **2019**, *47*, 1–15. [[CrossRef](#)]
40. Giudice, N.A.; Walton, L.A.; Worboys, M. The Informatics of Indoor and Outdoor Space: A Research Agenda. In Proceedings of the Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness, San Jose, CA, USA, 2 November 2010; Association for Computing Machinery: New York, NY, USA, 2010; pp. 47–53.
41. Newell, R.G.; Sancha, T.L. The Difference between CAD and GIS. *Comput.-Aided Des.* **1990**, *22*, 131–135. [[CrossRef](#)]
42. Sulaiman, M.Z.; Aziz, M.N.A.; Bakar, M.H.A.; Halili, N.A.; Azuddin, M.A. *Matterport: Virtual Tour as A New Marketing Approach in Real Estate Business During Pandemic COVID-19*; Atlantis Press: Amsterdam, The Netherlands, 2020; pp. 221–226.
43. Shults, R.; Levin, E.; Habibi, R.; Shenoy, S.; Honcheruk, O.; Hart, T.; An, Z. Capability of Matterport 3d Camera for Industrial Archaeology Sites Inventory. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *42*, 1059–1064. [[CrossRef](#)]
44. Batty, M. Digital Twins. *Environ. Plan. B Urban Anal. City Sci.* **2018**, *45*, 817–820. [[CrossRef](#)]
45. Jones, D.; Snider, C.; Nassehi, A.; Yon, J.; Hicks, B. Characterising the Digital Twin: A Systematic Literature Review. *CIRP J. Manuf. Sci. Technol.* **2020**, *29*, 36–52. [[CrossRef](#)]
46. Blair, G.S. Digital Twins of the Natural Environment. *Patterns* **2021**, *2*, 100359. [[CrossRef](#)]
47. Wang, P.; Yang, L.T.; Li, J.; Chen, J.; Hu, S. Data Fusion in Cyber-Physical-Social Systems: State-of-the-Art and Perspectives. *Inf. Fusion* **2019**, *51*, 42–57. [[CrossRef](#)]
48. Schluse, M.; Rossmann, J. From Simulation to Experimentable Digital Twins: Simulation-Based Development and Operation of Complex Technical Systems. In Proceedings of the 2016 IEEE International Symposium on Systems Engineering (ISSE), Edinburgh, UK, 3–5 October 2016; pp. 1–6.
49. El-Sheimy, N.; Li, Y. Indoor Navigation: State of the Art and Future Trends. *Satell. Navig.* **2021**, *2*, 7. [[CrossRef](#)]
50. Purohit, A.; Sun, Z.; Mokaya, F.; Zhang, P. SensorFly: Controlled-Mobile Sensing Platform for Indoor Emergency Response Applications. In Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks, Chicago, IL, USA, 12–14 April 2011; pp. 223–234.
51. Tashakkori, H.; Rajabifard, A.; Kalantari, M. A New 3D Indoor/Outdoor Spatial Model for Indoor Emergency Response Facilitation. *Buuld. Environ.* **2015**, *89*, 170–182. [[CrossRef](#)]
52. Gunduz, M.; Isikdag, U.; Basaraner, M. A Review of Recent Research in Indoor Modelling & Mapping. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B4*, 289–294. [[CrossRef](#)]
53. Wei, Y.; Akinci, B. A Vision and Learning-Based Indoor Localization and Semantic Mapping Framework for Facility Operations and Management. *Autom. Constr.* **2019**, *107*, 102915. [[CrossRef](#)]
54. Jens, K.; Gregg, J.S. How Design Shapes Space Choice Behaviors in Public Urban and Shared Indoor Spaces—A Review. *Sustain. Cities Soc.* **2021**, *65*, 102592. [[CrossRef](#)]
55. Zimring, C.; Joseph, A.; Nicoll, G.L.; Tsepas, S. Influences of Building Design and Site Design on Physical Activity: Research and Intervention Opportunities. *Am. J. Prev. Med.* **2005**, *28*, 186–193. [[CrossRef](#)] [[PubMed](#)]
56. Hillier, B.; Leaman, A.; Stansall, P.; Bedford, M. Space Syntax. Available online: [https://journals.sagepub.com/doi/abs/10.1068/b030147?casa\\_token=uXzG9WNvYzgAAAAA:ERkqLR5WTKPhvr6x7eJdFTkX9kpy-\\_yLZ5qbaReN\\_oNI\\_ak2juuD9OshMTg8VycVWj5xc\\_JLbsOK](https://journals.sagepub.com/doi/abs/10.1068/b030147?casa_token=uXzG9WNvYzgAAAAA:ERkqLR5WTKPhvr6x7eJdFTkX9kpy-_yLZ5qbaReN_oNI_ak2juuD9OshMTg8VycVWj5xc_JLbsOK) (accessed on 15 December 2022).
57. Petrovska, N.; Stevanovic, A. Traffic Congestion Analysis Visualisation Tool. In Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spain, 15–18 September 2015; pp. 1489–1494.
58. Birnhack, M.; Perry-Hazan, L. School Surveillance in Context: High School Students’ Perspectives on CCTV, Privacy, and Security. *Youth Soc.* **2020**, *52*, 1312–1330. [[CrossRef](#)]
59. Munaro, M.; Lewis, C.; Chambers, D.; Hvass, P.; Menegatti, E. RGB-D Human Detection and Tracking for Industrial Environments. In *Intelligent Autonomous Systems 13*; Menegatti, E., Michael, N., Berns, K., Yamaguchi, H., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 1655–1668.
60. Edelman, G.; Bijhold, J. Tracking People and Cars Using 3D Modeling and CCTV. *Forensic Sci. Int.* **2010**, *202*, 26–35. [[CrossRef](#)]
61. Video Surveillance and Public Space: Surveillance Society vs. Security State | SpringerLink. Available online: [https://link.springer.com/chapter/10.1007/978-3-031-11756-5\\_14](https://link.springer.com/chapter/10.1007/978-3-031-11756-5_14) (accessed on 8 September 2024).

62. A Detailed Comparison of LiDAR, Radar and Camera Technology. Available online: <https://insights.outsight.ai/how-does-lidar-compares-to-cameras-and-radars/> (accessed on 8 September 2024).
63. Günter, A.; Böker, S.; König, M.; Hoffmann, M. Privacy-Preserving People Detection Enabled by Solid State LiDAR. In Proceedings of the 2020 16th International Conference on Intelligent Environments (IE), Madrid, Spain, 20–23 July 2020; pp. 1–4.
64. Nielsen, M.S.; Nikolov, I.; Kruse, E.K.; Garnæs, J.; Madsen, C.B. Quantifying the Influence of Surface Texture and Shape on Structure from Motion 3D Reconstructions. *Sensors* **2023**, *23*, 178. [[CrossRef](#)]
65. Kang, Z.; Yang, J.; Yang, Z.; Cheng, S. A Review of Techniques for 3D Reconstruction of Indoor Environments. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 330. [[CrossRef](#)]
66. Wu, D.; Liang, Z.; Chen, G. Deep Learning for LiDAR-Only and LiDAR-Fusion 3D Perception: A Survey. *Intell. Robot.* **2022**, *2*, 105–129. [[CrossRef](#)]
67. Zhang, J.; Singh, S. LOAM: Lidar Odometry and Mapping in Real-Time. In Proceedings of the Robotics: Science and Systems X, Berkeley, CA, USA, 12–16 July 2014.
68. Fritsche, P.; Kueppers, S.; Briese, G.; Wagner, B. Fusing LiDAR and Radar Data to Perform SLAM in Harsh Environments. In *Informatics in Control, Automation and Robotics: 13th International Conference, ICINCO 2016, Lisbon, Portugal, 29–31 July 2016*; Madani, K., Peaucelle, D., Gusikhin, O., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 175–189; ISBN 978-3-319-55011-4.
69. Bilik, I. Comparative Analysis of Radar and Lidar Technologies for Automotive Applications. *IEEE Intell. Transp. Syst. Mag.* **2023**, *15*, 244–269. [[CrossRef](#)]
70. Mielle, M.; Magnusson, M.; Lilienthal, A.J. A Comparative Analysis of Radar and Lidar Sensing for Localization and Mapping. In Proceedings of the 2019 European Conference on Mobile Robots (ECMR), Prague, Czech Republic, 4–6 September 2019; pp. 1–6.
71. Antonarakis, A.S.; Saatchi, S.S.; Chazdon, R.L.; Moorcroft, P.R. Using Lidar and Radar Measurements to Constrain Predictions of Forest Ecosystem Structure and Function. *Ecol. Appl.* **2011**, *21*, 1120–1137. [[CrossRef](#)]
72. Zou, Z.; Shi, Z.; Guo, Y.; Ye, J. Object Detection in 20 Years: A Survey. *arXiv* **2019**, arXiv:1905.05055. [[CrossRef](#)]
73. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 1–13. [[CrossRef](#)]
74. Paul, M.; Haque, S.M.E.; Chakraborty, S. Human Detection in Surveillance Videos and Its Applications—A Review. *EURASIP J. Adv. Signal Process.* **2013**, *2013*, 176. [[CrossRef](#)]
75. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, Canada, 7–12 December 2015; Curran Associates, Inc.: Red Hook, NY, USA, 2015; Volume 28.
76. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2016**, arXiv:1506.02640.
77. Wang, J.; Chen, Y.; Dong, Z.; Gao, M. Improved YOLOv5 Network for Real-Time Multi-Scale Traffic Sign Detection. *Neural Comput. Applic* **2023**, *35*, 7853–7865. [[CrossRef](#)]
78. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
79. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
80. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976.
81. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *arXiv* **2022**, arXiv:2207.02696.
82. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
83. Molchanov, V.V.; Vishnyakov, B.V.; Vizilter, Y.V.; Vishnyakova, O.V.; Knyaz, V.A. Pedestrian Detection in Video Surveillance Using Fully Convolutional YOLO Neural Network. In Proceedings of the Automated Visual Inspection and Machine Vision II, SPIE, Munich, Germany, 29 June 2017; Volume 10334, pp. 193–199.
84. Garg, R.; Singh, S. Intelligent Video Surveillance Based on YOLO: A Comparative Study. In Proceedings of the 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3), Mumbai, India, 3–4 December 2021; pp. 1–6.
85. Nguyen, H.H.; Ta, T.N.; Nguyen, N.C.; Bui, V.T.; Pham, H.M.; Nguyen, D.M. YOLO Based Real-Time Human Detection for Smart Video Surveillance at the Edge. In Proceedings of the 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE), Phu Quoc Island, Vietnam, 13–15 January 2021; pp. 439–444.
86. Kannadaguli, P. YOLO v4 Based Human Detection System Using Aerial Thermal Imaging for UAV Based Surveillance Applications. In Proceedings of the 2020 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 8–9 November 2020; pp. 1213–1219.
87. Sualeh, M.; Kim, G.-W. Dynamic Multi-LiDAR Based Multiple Object Detection and Tracking. *Sensors* **2019**, *19*, 1474. [[CrossRef](#)] [[PubMed](#)]
88. BenAbdelkader, C.; Cutler, R.; Davis, L. Motion-Based Recognition of People in EigenGait Space. In Proceedings of the Fifth IEEE International Conference on Automatic Face Gesture Recognition, Washington, DC, USA, 20–21 May 2002; pp. 267–272.
89. Villarreal, M.; Baird, T.D.; Tarazaga, P.A.; Kniola, D.J.; Pingel, T.J.; Sarlo, R. Shared Space and Resource Use within a Building Environment: An Indoor Geography. *Geogr. J.* **2024**, e12604. [[CrossRef](#)]

90. Chan, T.H.; Hesse, H.; Ho, S.G. LiDAR-Based 3D SLAM for Indoor Mapping. In Proceedings of the 2021 7th International Conference on Control, Automation and Robotics (ICCAR), Singapore, 23–26 April 2021; pp. 285–289.
91. Zhou, L.; Koppel, D.; Kaess, M. LiDAR SLAM With Plane Adjustment for Indoor Environment. *IEEE Robot. Autom. Lett.* **2021**, *6*, 7073–7080. [[CrossRef](#)]
92. Sharif, M.H. Laser-Based Algorithms Meeting Privacy in Surveillance: A Survey. *IEEE Access* **2021**, *9*, 92394–92419. [[CrossRef](#)]
93. Szeliski, R. *Computer Vision: Algorithms and Applications*; Springer Nature: Cham, Switzerland, 2022; ISBN 978-3-030-34372-9.
94. Culjak, I.; Abram, D.; Pribanic, T.; Dzapo, H.; Cifrek, M. A Brief Introduction to OpenCV. In Proceedings of the 2012 Proceedings of the 35th International Convention MIPRO, Opatija, Croatia, 21–25 May 2012; pp. 1725–1730.
95. Schulte-Tiggens, J.; Förster, M.; Nikolovski, G.; Reke, M.; Ferrein, A.; Kaszner, D.; Matheis, D.; Walter, T. Benchmarking of Various LiDAR Sensors for Use in Self-Driving Vehicles in Real-World Environments. *Sensors* **2022**, *22*, 7146. [[CrossRef](#)]
96. Terven, J.; Cordova-Esparza, D. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Make* **2023**, *5*, 1680–1716. [[CrossRef](#)]
97. KaewTraKulPong, P.; Bowden, R. An Improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection. In *Video-Based Surveillance Systems: Computer Vision and Distributed Processing*; Remagnino, P., Jones, G.A., Paragios, N., Regazzoni, C.S., Eds.; Springer: Boston, MA, USA, 2002; pp. 135–144; ISBN 978-1-4615-0913-4.
98. Bradski, G. The OpenCV Library. *Dr. Dobb's J.* **2000**, *25*, 120–125.
99. Comaniciu, D.; Meer, P. Mean Shift: A Robust Approach toward Feature Space Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [[CrossRef](#)]
100. Jhaldiyal, A.; Chaudhary, N. Semantic Segmentation of 3D LiDAR Data Using Deep Learning: A Review of Projection-Based Methods. *Appl. Intell.* **2022**, *53*, 6844–6855. [[CrossRef](#)]
101. Zamanakos, G.; Tsochatzidis, L.; Pratikakis, I.; Amanatiadis, A. A Comprehensive Survey of LIDAR-Based 3D Object Detection Methods. *Comput. Graph.* **2021**, *99*, 153–181. [[CrossRef](#)]
102. Chen, G.; Wang, F.; Qu, S.; Chen, K.; Yu, J.; Liu, X.; Xiong, L.; Knoll, A. Pseudo-Image and Sparse Points: Vehicle Detection With 2D LiDAR Revisited by Deep Learning-Based Methods. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 7699–7711. [[CrossRef](#)]
103. Elaksher, A.; Ali, T.; Alharthy, A. A Quantitative Assessment of LIDAR Data Accuracy | EndNote Click. *Remote Sens.* **2022**, *15*, 442. [[CrossRef](#)]
104. Glennie, C.L.; Hartzell, P.J. Accuracy assessment and calibration of low-cost autonomous lidar sensors. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *XLIII-B1-2020*, 371–376. [[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.