To, A., Liu, M., Hazeeq Bin Muhammad Hairul, M., Davis, J. G., Lee, J. S. A., Hesse, H. and Nguyen, H. D. (2021) Drone-Based AI and 3D Reconstruction for Digital Twin Augmentation. In: 23rd HCI International Conference, 24-29 Jul 2021, pp. 511-529. (doi:10.1007/978-3-030-77626-8_35).

This is the author's final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.gla.ac.uk/250270/

Drone-based AI and 3D Reconstruction for Digital Twin Augmentation

Alex To[1], Maican Liu[2], Muhammad Hazeeq Bin Muhammad Hairul[2],
Joseph G. Davis[1], Jeannie S.A. Lee[3], Henrik Hesse[4], and Hoang D. Nguyen[5]

[1] School of Computer Science, University of Sydney, Australia
`duto3894@uni.sydney.edu.au,joseph.davis@sydney.edu.au`
[2] Aviation Virtual Pte Ltd, Singapore
`liumaican@gmail.com,hazeeq_hairul@hotmail.com`
[3] Infocomm Technology, Singapore Institute of Technology, Singapore
`Jeannie.Lee@singaporetech.edu.sg`
[4] James Watt School of Engineering, University of Glasgow, Singapore
`Henrik.Hesse@glasgow.ac.uk`
[5] School of Computing Science, University of Glasgow, Singapore
`Harry.Nguyen@glasgow.ac.uk`

Keywords: Digital Twin · 3D Reconstruction · Artificial Intelligence (AI) ·
Information Augmentation · Unmanned Aerial Vehicle (UAV)

Abstract. Digital Twin is an emerging technology at the forefront of
Industry 4.0, with the ultimate goal of combining the physical space and
the virtual space. To date, the Digital Twin concept has been applied
in many engineering fields, providing useful insights in the areas of engi-
neering design, manufacturing, automation, and construction industry.
While the nexus of various technologies opens up new opportunities with
Digital Twin, the technology requires a framework to integrate the dif-
ferent technologies, such as the Building Information Model used in the
Building and Construction industry. In this work, an Information Fusion
framework is proposed to seamlessly fuse heterogeneous components in a
Digital Twin framework from the variety of technologies involved. This
study aims to augment Digital Twin in buildings with the use of AI and
3D reconstruction empowered by unmanned aviation vehicles. We pro-
posed a drone-based Digital Twin augmentation framework with reusable
and customisable components. A proof of concept is also developed, and
extensive evaluation is conducted for 3D reconstruction and applications
of AI for defect detection.

1 Introduction

A Digital Twin is the virtual replication of a physical object. Through mod-
elling and real-time data communication, the Digital Twin simulates the actual
properties and behaviours of its physical counterpart in the physical space, thus,
enable learning, reasoning, and dynamically re-calibrating for improved decision-
making [19, 21]. The tight and seamless integration between the physical and

virtual space in the Digital Twin paradigm makes it one of the most promising enabling technologies for the realization of smart manufacturing and Industry 4.0 [49]

To date, Digital Twin applications have seen success in various industries and domains, including product design, production, prognostic and health management, building and construction, and many others. Recent advances in sensor technologies, big data, cloud computing, social networks, Internet of Things (IoT), Computer-Aided-Design (CAD), 3D modelling, and Artificial Intelligence (AI) allow a massive amount of data to be collected while enabling real-time communication for the realization of the Digital Twin paradigm throughout the complete product's life-cycle [18, 40, 47, 49].

In the Building and Construction industry context, physical objects are buildings and structural components. To generate and capture their virtual counterparts in the virtual space, Building Information Model (BIM) is a common standard that encompasses a large amount of detail on building dimensions and critical components. These components include façade features, dimensions of staircases, slopes of walls, height of railings, etc. The use of BIM provides high-quality preconstruction project visualisation, improved scheduling, and better coordination and issue management. The Digital Twin paradigm in this industry utilises BIM as one of the core technologies to facilitate information management, information sharing, and collaboration among stakeholders in different domains over the building life cycle [1, 10].

In many cases, it is often desirable to obtain 3D models of the physical buildings and landscapes that can be used for enrichment, visualization, and advanced analytics of Digital Twin models [13, 23, 36, 46]. Additionally, other sources of information can be useful for the Digital Twin models, such as contextual information and geographical information systems (GIS). However, there are some limitations with the current BIM technologies that hinder the capability to integrate multiple sources of data. For instance, BIM files are restricted in size, making it difficult to add large artefacts. The BIM format is neither initially designed for the integration of heterogeneous data sources nor capable of capturing real-time updates.

The process to obtain 3D models of the physical buildings and landscapes is also labour-intensive. When 3D reconstruction is done manually using a handheld device, the ability to capture an extensive model of the building is often limited due to physical constraints such as the size of tunnels or large constructions. These issues call for a new and more scalable approach in 3D reconstruction using unmanned aerial vehicles (UAV), optimal scanning methods, and advanced onboard processing algorithms [8, 35, 45, 46].

On the other hand, real-time applications of AI and image analysis of BIM is also underdeveloped. One application of imaging in building maintenance for Digital Twin is defect detection, in which AI algorithms are employed to recognize defect regions such as cracks automatically. To develop such AI models, appropriate training data is required which, however, is often found in 2D format, prompting suitable methods on real-time transformation for AI applicability.

Motivated by the current limitations of building and construction technologies, we aim to innovate BIM for Digital Twin in two broad areas: 1) Develop an Information Fusion framework that extends BIM with a metadata layer to support heterogeneous data integration; 2) Enhance real-time synchronization between the physical space and virtual space in BIM through improved 3D reconstruction methods and real-time scanning.

To this aim, our approach is four-fold: First, we developed a proof of concept Information Fusion framework to facilitate the integration of multiple sources of information to produce useful data representations for BIM applications. It utilises a distributed and fault-tolerant database to store geometry objects (e.g., buildings and structural components) and meta-information (e.g., defects and tagged items) to provide maximal compatibility and highest/raw details; Second, we built a drone-based 3D reconstruction solution for scalable data collection and evaluate major scanning technologies including Light Detection and Ranging (LiDAR) sensor, stereovision, and single-lens camera; Third, we tested our real-time scanning capabilities by performing real-time 2D to 3D mapping from our camera feed at five frames per second. The mapping computation is done on the drone using an onboard miniature computer; Finally, we presented a defect detection use-case as an application of AI in real-time image scanning.

The contributions of our work are as follows. We provided a comprehensive review of Digital Twin technologies in conjunction with AI. We demonstrated an end-to-end proof of concept of the use of BIM for Digital Twin and information fusion. We conducted extensive experiments for the evaluation of 3D reconstruction techniques. Finally, we illustrated the feasibility of AI application in Digital Twin through defect detection with deep learning use-case. Our work provides some insights and theoretical and empirical implications for researchers as well as practitioners in this emerging field.

2   Background

2.1   Digital Twin Technologies and Applications

The concept and model of the Digital Twin were publicly introduced in 2002 by Grieves in his presentation as the conceptual model underlying Product Lifecycle Management [20]. Although the term was not coined at that time, all the Digital Twin's basic elements were described: physical space, virtual space, and the information flow between them. The key enablers of Digital Twin: sensor technologies, cloud computing, Big Data, IoT, and AI have since then experienced growth at an unprecedented rate. Recently, the concept of Digital Twin was formally defined by NASA as a multiphysics, multiscale, probabilistic, ultra-fidelity simulation that enables real-time replication of the state of the physical object in cyberspace based on historical and real-time sensor data.

Tao et al. extended the model and proposed that Digital Twin modelling should involve: physical modelling, virtual modelling, connection modelling, data modelling, and service modelling [49]. From a more structural and technological

viewpoint, Digital Twin consists of sensor and measurement technologies, IoT, Big Data and AI [25, 29].

The applications of Digital Twins span various domains from manufacturing, aerospace to cyber-physical systems, architecture, construction, and engineering.

Digital Twin in Manufacturing Applications of Digital Twin are prominent in smart manufacturing. Due to ever-increasing product requirements and rapidly changing markets, there has been a growing interest in shifting problem identification and solving to early stages of product development lifecycle (also known as "front-loading") [51]. The Digital Twin paradigm fits perfectly because virtual replications of physical products allow early feedback, design changes, quality, and functional testing without entering the production phase.

Tao et al. suggested that a Digital Twin-driven product design process can be divided into conceptual design, detailed design, and virtual verification [48]. Throughout the process, various kinds of data such as customer satisfaction, product sales, 3D model, product functions, and configuration, sensor updates can be integrated to mirror the life of the physical product to its corresponding digital twin. With real-time closed-loop feedback between the physical and the virtual spaces, designers are able to make quick decisions on product design adjustment, quality control and improve the design efficiency by avoiding tedious verification and testing.

During production, simulation of production systems, the convergence of the physical and virtual manufacturing world leads to smart operations in the manufacturing process, including smart interconnection, smart interaction, smart control, and management. For example, Tao et al. proposed a shop-floor paradigm consists of four components physical shop-floor, virtual shop-floor, shop-floor service system driven by shop-floor digital twin data, enabled by IoT, big data, and artificial intelligence [50]. Modeling of machinery, manufacturing steps, and equipment also help in precise process simulation, control, and analysis, eventually leading to improvement of the production process [6]. A similar effort is observed in [36] to evaluate different methods in automated 3D reconstruction in SME factories. The authors explored the use of low-cost stereo vision techniques with Simultaneous Localization and Mapping (SLAM) to generate Digital Twin models of a physical factory floor and machinery.

Digital Twin in Building and Construction Modelling physical buildings and landscapes with Digital Twin brought valuable opportunities to the architecture, construction, and engineering industry, such as improvements in urban planning, city analytics, environmental analysis, building maintenance, defect detection, and collaboration between stakeholders. An important concept in this domain is BIM [26], i.e. a process involving the generation and management of digital representations of physical and functional characteristics of places.

Yan et al. proposed a method for the integration of 3D objects and terrain in BIMs supporting the Digital Twin, which takes the accurate representation of terrain and buildings into consideration [53]. The authors discussed topological

issues that can occur when integrating 3D objects with terrain. The key to solving this issue lies in obtaining the correct Terrain Intersection Curve (TIC) and amending 3D objects and the terrain properly based on it. Models developed by such methods are used for urban planning, city analytics, or environmental analysis.

For preventive maintenance of prestressed concrete bridges, Shim et al. proposed a new generation of the bridge maintenance system by using the Digital Twin concept for reliable decision-making [45]. 3D models of bridges were built to utilise information from the entire lifecycle of a project by continuously exchanging and updating data from stakeholders

Digital Twin also finds application in recording and managing cultural heritage sites. The work by [12] integrated a 3D model into a 3D GIS and bridge the gap between parametric CAD modeling and 3D GIS. The final model benefits from both systems to help document and analyze cultural heritage sites.

From most construction projects, the presence of BIM is prominent due to its wide range of benefits. BIM has received considerable attention from researchers with works aiming to improve or extend its various aspects for e.g social aspect [1], elasticity and scalability [10], sustainability [28], safety [55] and many others.

Digital Twin in Smart Nations Gartner's Top 10 Strategic Technology Report for 2017 predicted that Digital Twin is one of the top ten trending strategic technologies [39]. Digital Twin since 2012 has entered rapid growth stage considering the current momentum with applications in several industries and across variety of domains.

NASA and U.S Air Force adopted Digital Twin to improve production of future generations of vehicles to become lighter while being subjected to high loads and more extreme service conditions. The paradigm shift allowed the organisation to incorporate vehicle health management system, historical data and fleet data to mirror the life of its flying twin, thus, enabled unprecedented levels of safety and reliability [19].

The world's 11th busiest airport, the second largest in the Netherlands, Amsterdam Airport Schiphol built a digital asset twin of the airport based on BIM. Known as the Common Data Environment (CDE), Schiphol's Digital Twin solution integrates data from many sources: BIM data; GIS data; and data collected in real-time on project changes and incidents as well as financial information, documents, and project portfolios. The information fusion capability of Digital Twin presents opportunities to run simulations on potential operational failures throughout the entire complex [3].

Port of Rotterdam built a Digital Twin of the port and used IoT and artificial intelligence to collect and analyse data to improve operations. Digital Twin helps to better predict accurately what the best time is to moor, depart and how much cargo needs to be unloaded. Furthermore, real-time access to information enables better prediction of visibility and water conditions [7].

2.2   Artificial Intelligence in Digital Twin

The rapid adoption of enabling technologies such as IoT, cloud computing, and big data opens up endless opportunities for AI applications in Digital Twin. As a multidisciplinary field, AI encompasses Machine Learning, Data Mining, Computer Vision, Natural Language Processing, Robotics, among many others. AI emerges as a promising core service in Digital Twin to assist humans in decision making by finding patterns, insights in big data, generation of realistic virtual models through advanced computer vision, natural language processing, robotics, etc.

Li et al. proposed a method that uses a concept of dynamic Bayesian networks for Digital Twin to build a health monitoring model for the diagnosis and prognosis of each individual aircraft [31]. For example, in diagnosis by tracking time-dependent variables, the method could calibrate the time-independent variables; in prognosis, the method helps predict crack growth in the physical subject using particle filtering as the Bayesian inference algorithm.

In production, [2] introduced a Digital Twin-driven approach for developing Machine Learning models. The models are trained for vision-based recognition of parts' orientation using the simulation of Digital Twin models, which can help adaptively control the production process. Additionally, the authors also proposed a method to synthesize training datasets and automatic labelling via the simulation tools chain, thus reducing users' involvement during model training.

Chao et al. [14] described an insightful vision of Digital Twin to enable the convergence of AI and Smart City for disaster response and emergency management. In this vision, the authors listed four components in Disaster City Digital Twin, i.e. 1) multi-data sensing for data collection, 2) data integration and analytics, 3) multi-actor game-theoretic decision making, 4) dynamic network analysis, and elaborated the functions that AI can improve within each component.

Another interesting vision of Digital Twin in Model-Based Systems Engineering is described in [33] in which the realization of Digital Twin is progressively divided into four levels 1) Pre-Digital Twin, 2) Digital Twin, 3) Adaptive Digital Twin and 4) Intelligent Digital Twin. In the last two levels: Adaptive Digital Twin and Intelligent Digital Twin, the authors emphasized the tight integration of AI in engineering processes; for example, in level 3, an adaptive user interface can be offered by using supervised machine learning to learn the preferences and priorities of human operators in different contexts, therefore, support real-time planning and decision making during operators, maintenance and support; in level 4, additionally unsupervised machine learning can help discern objects, and patterns in the operational environment and reinforcement learning can learn from continuous data stream from the environment.

Power networks are the backbone of power distribution, playing a central economical and societal role by supplying reliable power to industry, services, and consumers. To improve the efficiency of power networks, researchers in the Energy industry have also been putting initial effort into integrating Digital Twin, and AI for informed decision-making in operation, support, and mainte-

nance [34]. In particular, a virtual replication of the power network is developed. Various time-series measurements from the physical power networks, such as production values, loads, line thermal limits, power flows, etc., are streamed back to the virtual models. Based on the digital models, researchers exploit machine learning algorithms such as reinforcement learning to predict future states of the networks, as well as suggest possible optimal control actions.

2.3 3D Reconstruction

Various 3D scanning technologies are emerging for a range of applications, from outdoor surveying, 3D mapping of cities for digital twins, inspection to autonomous driving. Most of these applications and technologies rely on LiDAR sensors [32, 38, 41, 52]. However, most LiDAR sensors tend to be expensive and heavy, making them less suitable for developing a drone-based surveying solution. Other 3D scanning solutions use a single lens [42, 46] or stereo vision cameras [9, 13, 23, 36] to compute a 3D model of the environment.

Photogrammetry The most common method for 3D reconstruction of outdoor structures is photogrammetry. The 3D representation of complex structures such as buildings, bridges, and even 3D maps of a whole neighbourhood can be generated using a single-lens camera based on the concept of Structure from Motion (SfM) [43].

The steps to create a point cloud or textured mesh is to capture multiple photographs in sequence or randomising order with at least 70% overlapping and at angle part of around 5-10 degrees [42]. This will ensure that the amount of overlap is sufficient for matching photos to have common feature points. Matching the features in different photos allows the SfM algorithm to generate a 3D point cloud [46]. The generated point cloud can be meshed to create a smooth or textured result of the 3D model.

Stereovision Stereovision is a 3D scanning method suitable for smaller or indoor infrastructure projects where higher accuracy is required. The concept uses stereovision cameras (infrared or RGB) to estimate the depth in the field of view of the camera. Stereo Vision uses the disparity between images from multiple cameras to extract depth information [37]. Similar to the binocular vision in humans, when both eyes focus on an object, their optical axes will converge to that point at an angle. The displacement parallel to the eye base (the distance between both eyes) creates a disparity between both images. From the extent of disparity, it is possible to extract the distance of an object and pixel in an image through triangulation [17].

To generate the 3D model from the stereo or depth images, RGB-D cameras require an additional processor to run a process called Simultaneous Localisation and Mapping (SLAM) [5]. As the name suggests, the SLAM concept is able to build a 3D map of the environment in real-time and at the same time estimate

the location and orientation of the camera. SLAM works by scanning the images for key features which can be extracted with Speeded Up Robust Features (SURF) [4] and matched with RAndom SAmple Consensus (RASAC) algorithm between multiple images [16]. These two algorithms work simultaneously, SURF compares two images and extracts matching key points. These key points are then combined with the depth data to allow RANSAC algorithm to determine the 3D transformations between the frames. The transformed key points are optimised into a graph representation resulting in a 3D representation of the environment.

LiDAR Scanning Laser measurements provide another means to obtain depth information of the environment using the concept of time of flight of a light signal reflected at the surrounding. Hence, LiDAR also uses an active approach to obtain depth information similar to RGB-D technology. Still, LiDAR sensors have a much larger range of 100 meters with accuracy in the millimetre range. In recent years, LiDAR sensors have received a lot of attention, mostly due to their extensive use in autonomous driving technologies. This resulted in many available LiDAR sensors, which are affordable and light enough to be installed on drones for aerial scanning of infrastructure projects.

Similar to RGB-D sensors, most LiDAR-based 3D scanning techniques also use a SLAM approach to convert the instantaneous laser-point measurements to 2D or 3D point-cloud representations. GMapping is a common SLAM technique introduced for LiDAR-based mapping, reducing the computation time for the SLAM algorithms [22]. HectorSLAM is the SLAM algorithm used here for the in-house development of a 2D mapping evaluation [27]. It was first developed for Urban Search and Rescue (USAR) scenarios and is suitable for fast learning of occupancy grid maps with low computational requirements. HectorSLAM presents a high update rate simultaneously on a 2D map for lower power platforms and the results yielded were a sufficiently accurate mapping. A more recent SLAM algorithm by Google is called Cartographer [24]. In a comparison study [15], GMapping produced an inaccurate mapping while both the HectorSLAM and Cartographer produced accurate and similar maps.

Many LiDAR-based 3D SLAM frameworks have been proposed specifically for 3D reconstruction and form the foundation for most commercial scanning technologies available. Among the many LiDAR-based 3D SLAM methods, LOAM is a widely used real-time LiDAR odometry estimation and mapping framework that uses a LiDAR sensor and optionally an inertial measurement unit (IMU) [54]. This method achieves real-time performance by separating the SLAM problem into odometry estimation algorithm and mapping optimisation algorithm. The odometry estimation algorithm runs at high frequency with low fidelity, while the mapping optimization algorithm runs at an order of magnitude lower frequency with high accuracy for scan-matching. Since its publication, LOAM has remained at the top rank in the odometry category of various benchmarks. LOAM has since then been commercialized, and its framework is no longer available in the public domain.

The current state-of-the-art 3D SLAM method for LiDAR odometry and mapping is LIO-SAM [44]. It utilizes factor graphs to incorporate multiple measurement factors for odometry estimation and global map optimization. The framework incorporates an IMU to improve the pose estimation and incorporate GPS as an option for additional key factors.

3   Solution Design

The backbone of our solution is an Information Fusion module to extend beyond the current limitations of BIM. The Information Fusion module has an extensive set of APIs, scalable storage, advanced search, and indexing capabilities to fuse multiple data streams, capture different types of BIM artefacts, AI models, and defects while supporting online communication from our drones and management site.

For 3D reconstruction, we present our drone-based setup. The drone has a stereovision camera attached as a cost-effective solution. The main computing unit is a miniature onboard computer responsible for processing the output from the camera feed via USB, and streaming it back to the Information Fusion module in a real-time manner.

To test our defect detection use-case as an application of AI in real-time image scanning, we deployed a deep learning model on the on-board computer. The defects detected from the camera feeds are sent back to the Information Fusion module to fuse with the 3D models and other BIM-related information.

The overall architecture is illustrated in Fig. 1. We also described each component in detail in the following sections.
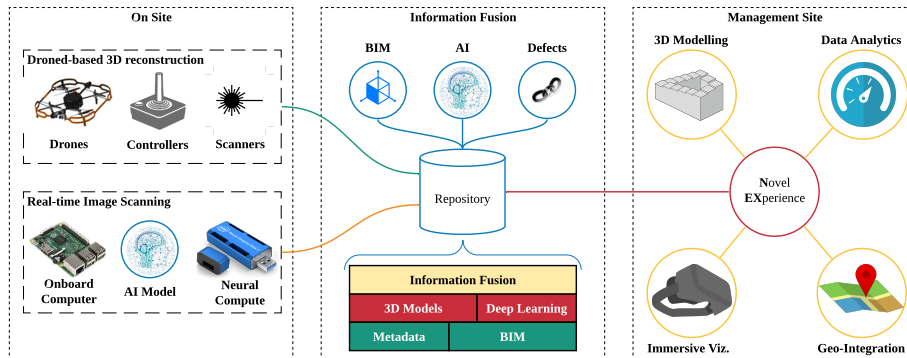

Fig. 1. Overall Architecture

3.1   Information Fusion

To be able to capture heterogeneous data sources including structured, unstructured, images, 3D models, meta-information beyond BIM's capabilities, The Information Fusion module leverages one of the most efficient and well-established NoSQL database systems, Apache Cassandra, originally developed by Facebook, hence, is able to handle a huge amount of data across multiple locations, including on-site and off-site. With an extended database schema, the module offers the ability to store and replicate large BIM files with high data protection and fault-tolerance while also supporting imaging data, defects, and tagged items. We also added an extensive set of API to enable real-time communication from our drone for live streaming of RGB-D images and defect information.

3.2   Drone-based 3D Reconstruction

The stereovision camera used in our drone is an Intel RealSense D435i camera which is more cost-effective compared to a LiDAR sensor. It is an RGB-D camera that produces point-clouds in color instead of black and white. The depth data provides the distance between the camera and the obstacle in its FOV. It has an integrated Inertial Measurement Unit to predict the orientation of the drone and provides a horizontal and vertical FOV of 87 degrees by 58 degrees that allow a 3D map to be generated. Our drone setup is shown in Fig. 2

Fig. 2. Our drone configuration with integration of RPLidar A2 (bottom left) and Intel RealSense D435i (bottom right).

Flight Controller Each drone requires a flight controller to allow the pilot to have precise control over the vehicle and its motors. Even in manual flight, the flight controller translates the throttle command on the radio control to

individual motor commands to stabilize the drone. Flight controllers use several inbuilt sensors to control the vehicle response. In this work, we used a Pixhawk flight controller which allows us to operate the custom drone. The Pixhawk flight controller also supports many additional sensors and companion computers to be integrated.

Onboard Computing We used a companion computer attached to our drone as the main processing unit. In our prototype, a Raspberry Pi4 single-board computer is added to allow additional sensors and features to be integrated. For e.g. it enables features such as obstacle avoidance, automated flight path tracking, or in this work, 3D scanning of the environment. Raspberry Pi4 is utilized in this prototype due to its low cost, high specifications, and large supporting community. The other significant factor for choosing the Raspberry Pi4 is its compatibility with the additional sensors and the Robotic Operating Software (ROS) used. The Intel RealSense D435i camera as mentioned in the previous sections is connected and executed by the Raspberry Pi4 via USB port.
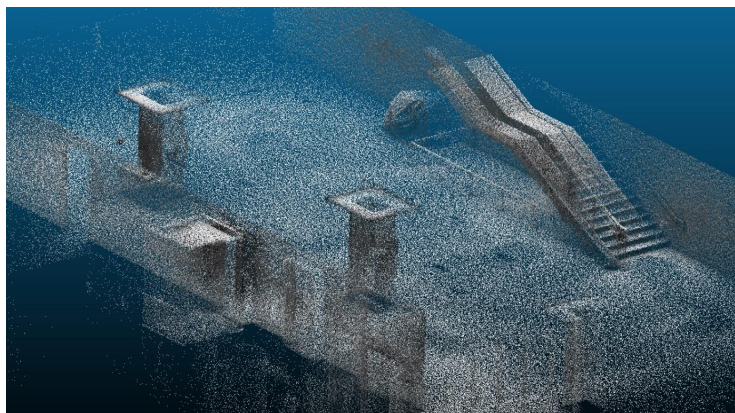

Fig. 3. 3D reconstruction scanning.

Implementation of SLAM Real-Time Appearance Based Mapping (RTABMap) is an open-source SLAM environment [30] with numerous tools to generate maps from RGB-D data. RTABMap has evolved to do online processing, minimal drift odometry, robust localization map exploitation, and multi-session mapping. The approach is based on the SLAM algorithm introduced before and is illustrated in Figure 3, including the different algorithms used to extract the features into the point cloud. Using SLAM as the base to generate point clouds gives the user the flexibility to change parameters or adjust flight paths during the scanning process.

3.3   Real-time Image Scanning

One limitation of Raspberry Pi 4 as compared to conventional computers is the limited processing power. This leads to low frame rates of only 5 frames per second. To provide more processing power to the Raspberry Pi 4 and allow more efficient Real-time Image Scanning, we explored the use of USB accelerators to increase the frame rate. A USB accelerator is a USB stick that contains a Vision Processing Unit aimed at boosting CPU performance. The USB accelerator used in this work is an Intel Neural Compute Stick 2 that is compatible with the Intel Real Sense D435. It also has a toolkit called the OpenVINO toolkit, which allows the companion computer to recognise the NCS2 and make full use of the additional CPU boost. After the implementation of the USB accelerator, the frames rate provided a boost to the CPU of the Raspberry Pi 4 resulting in an average of 12 fps.

3.4   AI for Defect Detection

To further evaluate our Real-time Image Scanning capability, we trained a deep learning model for defect detection using convolutional neural networks. We employed the SDNET 2018, a publicly available dataset, that contains 56,000 images of cracks and non-cracks [11]. The dataset provides various types of cracks, ranging from 0.06mm to 25mm, on different types of surfaces. We trained our classifier engine with multiple backbones, including ResNet18, ResNet50, and VGG; and then the classifiers' performance was evaluated against current baselines. We utilized the best model to classify 2D images, coming from streaming data sources. Our drone (in a simulated environment) captures the Red-Green-Blue (RGB) channels and the depth layer from RGBA images for processing. The drone position and intrinsic camera can be configured to provide the best 3D locationing of the defects for visualizing them in the simulation. The AI defect detection workflow is illustrated in Fig. 4.
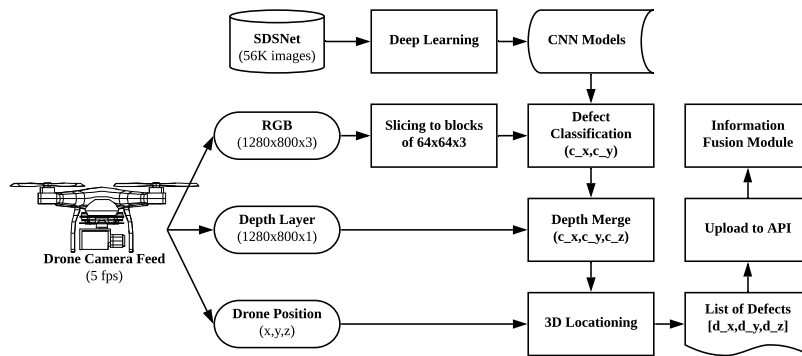

Fig. 4. AI Defect Detection

4   Experiments and Results

We conducted experiments to validate our solution and answer the following questions.

1. How do different scanning technologies perform compared to each other, and how do they perform compared to manual measurements?
2. How does a scanning technology perform when being used as a handheld device vs being used in a drone-based solution? And how are both approaches compared to manual measurements?
3. How do different CNN architectures perform in defect detection?

The detail of the experiments and results are given in the following sections.

4.1   Scanning Performance

We evaluated the performance of three different 3D scanning technologies with the following specific products.

1. Photogrammetry with Pix4D
2. Stereovision with Dot3D/Navisworks
3. 3D LiDAR with geoSLAM/Navisworks.

Table 1. 3D Scanning Technologies

Area of Interest	Actual Dist. (mm)	Stereovision M. (mm)	Rel. %Err	3D LiDAR M. (mm)	Rel. %Err	Photogrammetry M. (mm)	Rel. %Err
1) Ceiling height							
Ceiling height (L1)	3185	3201	0.5%	3189	0.1%		
2) Height of safety barriers							
Balustrade (L2)	1120	1111	-0.8%	1007	-10.1%		
Staircase Railing (L2)	1110	1082	-2.6%	1018	-8.3%		
3) Profile of stairs							
Thread width	1817	1817	0.0%	1845	1.5%		
Riser height	148	149	0.7%	157	6.1%		
4) Dimensions of windows and doors							
Lift door width	1195	1199	0.3%	1204	0.8%		
Toilet door width	1135	1139	0.4%	1116	011.1%		
Corridor width (L2)	2100	2102	0.1%	1964	-6.5%		
5) Gradients of ramp							
Ramp length	3800	3845	1.2%	3851	1.3%	3890	2.4%
Ramp height	295	294	-0.3%	297	0.7%	330	11.9%
Ramp height/length	59/760	64/837	-1.5%	30/389	-0.7%	33/389	9.3%

We manually measured selected areas of interest as well as scan them with the listed products. For photogrammetry, we only included the results for the ramp as the technology is deemed unsuitable for indoor scanning. The results are summarised in Table 1.

The evaluation of the methods showed that stereovision and 3D LiDAR achieve accuracies sufficient for indoor surveying, with stereovision achieving more consistent accuracies. Photogrammetry was found to not be suitable for indoor surveying due to the high inaccuracy of the results.

4.2   Measurement errors

Drone-based Inspection with Stereovision We used the drone-based setup described in Section 3 to compare with manual measurements as well as when being used as a handheld device. The results are given in Table 2.

Table 2. Comparison between handheld and drone-based scanning using the stereovision approach against manual measurements

Area of Interest	Measured Distance	Handheld Distance	Error	% Error	Drone-based Distance	Error	% Error
Room Width	9140 mm	9263 mm	123 mm	1.34%	9260 mm	120 mm	1.31%
Shelf Width	690 mm	688 mm	2 mm	0.29%	691 mm	1 mm	0.14%
Shelf Height	2130 mm	2127 mm	3 mm	0.14%	2127 mm	3 mm	0.14%

Both approaches produce very accurate results with the highest error of 1.3%. In addition, the flight scan results are slightly improved even since the drone only can move around in straight directions (up down, left right, front back) for the scan to be completed. This means that with lesser pitching of the drone the accuracy of results will be improved. This demonstrates that the drone-based concept using stereovision is a feasible approach for automated indoor scanning.

Drone-based Inspection with 2D LiDAR Next, we compared handheld and drone-based scanning using the 2D LiDAR approach against manual measurements. The result is given in Table 3.

Similar to the stereovision approach, the drone-based scan for the 2D LiDAR also shows better accuracy compared to the handheld scanning. Although the difference between the handheld and drone-based readings is small, with the largest being at around 1%, it can be seen that the drone-based scan produces more consistent results, as the drone is more stable than the handheld method.

Comparison of the generated point-clouds from both scanning technologies, stereovision, and LiDAR, shows that using a drone to automate the scanning process has no detrimental effects. In fact, the results demonstrate that drone-based scanning provides a more accurate method compared to the handheld

Table 3. Comparison between handheld and drone-based scanning using the 2D Li-DAR approach against manual measurements

Area of Interest	Measured Distance	Handheld Distance	Error	% Error	Drone-based Distance	Error	% Error
Room Width	9140 mm	9103 mm	37 mm	0.40%	9104 mm	36 mm	0.39%
Shelf Width	690 mm	695 mm	5 mm	0.72%	697 mm	7 mm	1.01%
Door Height	950 mm	945 mm	5 mm	0.53%	949 mm	1 mm	0.11%

approach due to drone stability during flight. Hence, our work demonstrated that it is possible to use 3D scanning technologies integrated on a drone to enable automated indoor surveying.

4.3 Defect Detection Performance

Our drone-based setup scans the surrounding environment and uses the AI model deployed on the on-board computer for inference on the image stream as illustrated in Fig. 5.

Fig. 5. Drone-based defect detection

The detection performance from our three trained models is given in the Table 4.

The ResNet-50 outperformed ResNet-18 by 2% in accuracy; nevertheless, there is a clear improvement of 9% in the recall of crack detection. The results showed that the deeper architecture allowed a better way to recognise cracks

Table 4. Performance Evaluation of CNNs on Crack Detection

	Crack			No Crack			Overall Accuracy
	Precision	Recall	F1-score	Precision	Recall	F1-score	
Resnet18	0.86	0.52	0.65	0.92	0.98	0.95	0.91
Resnet50	0.86	0.61	0.72	0.93	0.98	0.96	0.93
VGG16	0.88	0.59	0.71	0.93	0.99	0.96	0.93

in different forms. VGG-16 has achieved comparable performance with ResNet-50. However, it has a slightly lower performance in terms of F1-score in the Crack category, hence, Resnet-50 is selected as our AI model of choice for defect detection

5 Conclusion

In this paper, we presented a drone-based AI and 3D Reconstruction for Digital Twin augmentation. We illustrated an Information Fusion framework that extends beyond BIM's capabilities to enable the integration of heterogeneous data sources. We developed a proof of concept drone-based 3D reconstruction and real-time image scanning and provided evaluation and comparison results from extensive experiments. Finally, we studied the feasibility of AI applications in real-time image scanning through a defect detection use-case. Our work shows that with Information Fusion, the applicability of BIM can be greatly enhanced because the additional data allows additional applications such as 3D reconstruction to be built on top of BIM. Our empirical experiments also give suggestions to researchers and practitioners that the use of drones, onboard computing, RGB-D cameras, and neural computing unit are viable options for the realisation of large-scale, real-time image scanning and AI in Digital Twin.

6 Acknowledgement

References

1. Adamu, Z.A., Emmitt, S., Soetanto, R.: Social bim: Co-creation with shared situational awareness. Journal of Information Technology in Construction **20**, 230–252 (2015)
2. Alexopoulos, K., Nikolakis, N., Chryssolouris, G.: Digital twin-driven supervised machine learning for the development of artificial intelligence applications in manufacturing. International Journal of Computer Integrated Manufacturing **33**(5), 429–439 (2020)

3. Baumann, J.: Digital Twin Helps Airport Optimize Operations (2019), https://www.esri.com/about/newsroom/arcuser/digital-twin-helps-airport-optimize-operations/
4. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Computer vision and image understanding **110**(3), 346–359 (2008)
5. Beňo, P., Duchoň, F., Tölgyessy, M., Hubinskỳ, P., Kajan, M.: 3d map reconstruction with sensor kinect
6. Botkina, D., Hedlind, M., Olsson, B., Henser, J., Lundholm, T.: Digital twin of a cutting tool. Procedia Cirp **72**, 215–218 (2018)
7. Boyles, R.: How the Port of Rotterdam is using IBM digital twin technology to transform itself from the biggest to the pp. 1–4 (2019)
8. Chan, T., Hesse, H., Ho, S.: Lidar-based 3d slam for indoor mapping. IEEE International Conference on Control, Automation & Robotics (ICCAR) (2021)
9. Choi, S., Zhou, Q.Y., Koltun, V.: Robust reconstruction of indoor scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5556–5565 (2015)
10. Das, M., Cheng, J.C., Shiv Kumar, S.: Bimcloud: a distributed cloud-based social bim framework for project collaboration. In: Computing in Civil and Building Engineering (2014), pp. 41–48 (2014)
11. Dorafshan, S., Thomas, R.J., Maguire, M.: Sdnet2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks. Data in brief **21**, 1664–1668 (2018)
12. Dore, C., Murphy, M.: Integration of historic building information modeling (hbim) and 3d gis for recording and managing cultural heritage sites. In: 2012 18th International conference on virtual systems and multimedia. pp. 369–376. IEEE (2012)
13. Dryanovski, I., Klingensmith, M., Srinivasa, S.S., Xiao, J.: Large-scale, real-time 3d scene reconstruction on a mobile device. Autonomous Robots **41**(6), 1423–1445 (2017)
14. Fan, C., Zhang, C., Yahja, A., Mostafavi, A.: Disaster city digital twin: A vision for integrating artificial and human intelligence for disaster management. International Journal of Information Management **56**, 102049 (2021)
15. Filipenko, M., Afanasyev, I.: Comparison of various slam systems for mobile robot in an indoor environment. In: 2018 International Conference on Intelligent Systems (IS). pp. 400–407. IEEE (2018)
16. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981)
17. Fryer, J.G.: Photogrammetry (2010)
18. Fuller, A., Fan, Z., Day, C., Barlow, C.: Digital twin: Enabling technologies, challenges and open research. IEEE Access **8**, 108952–108971 (2020)
19. Glaessgen, E., Stargel, D.: The digital twin paradigm for future nasa and us air force vehicles. In: 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA. p. 1818 (2012)
20. Grieves, M.: Conceptual ideal for plm. Presentation for the Product Lifecycle Management (PLM) center, University of Michigan (2002)
21. Grieves, M.: Digital twin: manufacturing excellence through virtual factory replication. White paper **1**, 1–7 (2014)
22. Grisetti, G., Stachniss, C., Burgard, W.: Improved techniques for grid mapping with rao-blackwellized particle filters. IEEE transactions on Robotics **23**(1), 34–46 (2007)

23. Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D.: Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments. In: Experimental robotics. pp. 477–491. Springer (2014)
24. Hess, W., Kohler, D., Rapp, H., Andor, D.: Real-time loop closure in 2d lidar slam. In: 2016 IEEE International Conference on Robotics and Automation (ICRA). pp. 1271–1278. IEEE (2016)
25. Kaur, M.J., Mishra, V.P., Maheshwari, P.: The convergence of digital twin, iot, and machine learning: transforming data into action. In: Digital twin technologies and smart cities, pp. 3–17. Springer (2020)
26. Kensek, K.M.: Building information modeling. Routledge (2014)
27. Kohlbrecher, S., Von Stryk, O., Meyer, J., Klingauf, U.: A flexible and scalable slam system with full 3d motion estimation. In: 2011 IEEE international symposium on safety, security, and rescue robotics. pp. 155–160. IEEE (2011)
28. Krygiel, E., Nies, B.: Green BIM: successful sustainable design with building information modeling. John Wiley & Sons (2008)
29. Kusiak, A.: Smart manufacturing must embrace big data. Nature News **544**(7648), 23 (2017)
30. Labbé, M., Michaud, F.: Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. Journal of Field Robotics **36**(2), 416–446 (2019)
31. Li, C., Mahadevan, S., Ling, Y., Wang, L., Choze, S.: A dynamic bayesian network approach for digital twin. In: 19th AIAA Non-Deterministic Approaches Conference. p. 1566 (2017)
32. Liu, X.: Airborne lidar for dem generation: some critical issues. Progress in physical geography **32**(1), 31–49 (2008)
33. Madni, A.M., Madni, C.C., Lucero, S.D.: Leveraging digital twin technology in model-based systems engineering. Systems **7**(1), 7 (2019)
34. Marot, A., Guyon, I., Donnot, B., Dulac-Arnold, G., Panciatici, P., Awad, M., O'Sullivan, A., Kelly, A., Hampel-Arias, Z.: L2rpn: Learning to run a power network in a sustainable world neurips2020 challenge design (2020)
35. Mauriello, M.L., Froehlich, J.E.: Towards automated thermal profiling of buildings at scale using unmanned aerial vehicles and 3d-reconstruction. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. pp. 119–122 (2014)
36. Minos-Stensrud, M., Haakstad, O.H., Sakseid, O., Westby, B., Alcocer, A.: Towards automated 3d reconstruction in sme factories and digital twin model generation. In: 2018 18th International Conference on Control, Automation and Systems (ICCAS). pp. 1777–1781. IEEE (2018)
37. Nair, D.: A guide to stereovision and 3d imaging - tech briefs :: Tech briefs (2012), http://www.techbriefs.com/component/content/article/14925?start=1
38. Nys, G.A., Billen, R., Poux, F.: Automatic 3d buildings compact reconstruction from lidar point clouds. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (XLIII-B2-2020), 473–478 (2020)
39. Panetta, K.: Gartner's Top 10 Strategic Technology Trends for 2017 (2016), http://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/
40. Qi, Q., Tao, F., Hu, T., Anwer, N., Liu, A., Wei, Y., Wang, L., Nee, A.: Enabling technologies and tools for digital twin. Journal of Manufacturing Systems (2019)
41. Sampath, A., Shan, J.: Segmentation and reconstruction of polyhedral building roofs from aerial lidar point clouds. IEEE Transactions on geoscience and remote sensing **48**(3), 1554–1567 (2009)

42. Santagati, C., Inzerillo, L., Di Paola, F.: Image-based modeling techniques for architectural heritage 3d digitalization: limits and potentialities. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **5**(w2), 555–560 (2013)

43. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4104–4113 (2016)

44. Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., Rus, D.: Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. arXiv preprint arXiv:2007.00258 (2020)

45. Shim, C.S., Dang, N.S., Lon, S., Jeon, C.H.: Development of a bridge maintenance system for prestressed concrete bridges using 3d digital twin model. Structure and Infrastructure Engineering **15**(10), 1319–1332 (2019)

46. Spreitzer, G., Tunnicliffe, J., Friedrich, H.: Large wood (lw) 3d accumulation mapping and assessment using structure from motion photogrammetry in the laboratory. Journal of Hydrology **581**, 124430 (2020)

47. Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., Sui, F.: Digital twin-driven product design, manufacturing and service with big data. The International Journal of Advanced Manufacturing Technology **94**(9), 3563–3576 (2018)

48. Tao, F., Sui, F., Liu, A., Qi, Q., Zhang, M., Song, B., Guo, Z., Lu, S.C.Y., Nee, A.: Digital twin-driven product design framework. International Journal of Production Research **57**(12), 3935–3953 (2019)

49. Tao, F., Zhang, H., Liu, A., Nee, A.Y.: Digital twin in industry: State-of-the-art. IEEE Transactions on Industrial Informatics **15**(4), 2405–2415 (2018)

50. Tao, F., Zhang, M.: Digital twin shop-floor: a new shop-floor paradigm towards smart manufacturing. Ieee Access **5**, 20418–20427 (2017)

51. Thomke, S., Fujimoto, T.: The effect of "front-loading" problem-solving on product development performance. Journal of Product Innovation Management: An International Publication of the Product Development & Management Association **17**(2), 128–142 (2000)

52. Wu, B., Yu, B., Wu, Q., Yao, S., Zhao, F., Mao, W., Wu, J.: A graph-based approach for 3d building model reconstruction from airborne lidar point clouds. Remote Sensing **9**(1), 92 (2017)

53. Yan, J., Zlatanova, S., Aleksandrov, M., Diakite, A., Pettit, C.: Integration of 3d objects and terrain for 3d modelling supporting the digital twin. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences **4** (2019)

54. Zhang, J., Singh, S.: Loam: Lidar odometry and mapping in real-time. In: Robotics: Science and Systems. vol. 2 (2014)

55. Zhang, S., Teizer, J., Lee, J.K., Eastman, C.M., Venugopal, M.: Building information modeling (bim) and safety: Automatic safety checking of construction models and schedules. Automation in construction **29**, 183–195 (2013)