Automated High-Precision Digital Twin Modeling of Building Façade Defects with GeoBIM-assisted Registration

Jihan Zhang[a], Benyun Zhao[*,a], Guidong Yang[a], Xunkuai Zhou[a, b], Yijun Huang[a], Chuanxiang Gao[a], Xi Chen[*,a] and Ben M. Chen[a]

[a] *Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China*
[b] *School of Electronics and Information Engineering, Tongji University, Jiading, Shanghai, China*

ARTICLE INFO

Keywords:
Unmanned aerial vehicle (UAV)
Digital twin (DT)
Geographic information system (GIS)
Building information modeling (BIM)
Defect registration

ABSTRACT

The utilization of unmanned aerial vehicles (UAVs) for visual inspections has become an emerging solution for large-scale architectural assessments. However, it raises a challenge in effectively managing captured extensive data. Existing research primarily focuses on the integration of such data into pre-existing digital platforms, including image registration within Building Information Model (BIM) or Geographical Information System (GIS). Nonetheless, such approaches necessitate the capture of images from specific points of view and fall short in terms of precise individual defect localization and evaluation. To address these issues, this study proposes an end-to-end solution for digital twin (DT) modelling of the target building for façade defect management. This comprehensive DT encompasses geometric models and image information registered to real-world geographic locations for each defect. A DT-driven registration method is introduced to calculate the real-world geographic coordinate of each defect. Initially, a virtual building and camera model for geographic registration are established using a BIM+GIS approach (named GeoBIM) based on the geographic information associated with the actual building and photographs. Subsequently, virtual images corresponding to real images are generated, and depth calculations are conducted within the virtual domain to infer the distance from the camera to the target defect, thus determining the geographic coordinates of the target. Concurrently incorporating geographic and image features, we propose the dual verification method to merge the duplicated defects on overlapped images. And the relationship between defects and building structures is established through GeoBIM structural retrieval, enriching the semantic information related to defects. This work extends the existing architectural DT efforts by leveraging data generated in the virtual domain to drive real-world operations, effectively mitigating errors associated with defect localization and evaluation for precise maintenance. The proposed solution has been validated on a high-rise building located in Hong Kong, demonstrating the feasibility, effectiveness, and efficiency of the developed approach and its application potential in large-scale built assets.

1. Introduction

Periodic inspection is essential for maintaining the physical and functional conditions of civil infrastructure systems such as bridges, dams, roads, and buildings. For instance, since 2012, the Hong Kong government has initiated the mandatory building inspection scheme to address safety concerns arising from over 50% of private residences that have surpassed a 30-year lifespan [1]. Visual inspection is a common approach aims to identify and locate potential defects caused by infrastructure degradation, such as cracks, spalling, and moisture, to prevent serious safety problems [2, 3]. Traditional visual inspection methods rely on trained engineers for manual identification, characterized by high subjectivity, low accuracy, and low efficiency [4, 5].

The recent trend is to combine robotic technology, such as Unmanned Aerial Vehicles (UAVs), with computer vision technology, such as defect detection algorithms, for automatic data collection and analysis [6, 7, 8]. On the one hand, UAVs equiped with cameras demonstrate significant advantages in terms of safety, cost-effectiveness, and maneuverability [6]. Duque et al. distribute a national survey and find that UAV-enabled infrastructure inspection has been extensively applied and its feasibility has been substantiated [9]. On the other hand, the large amount of data derived from efficient data collection requires rapid automated processing methods, specifically Artificial Intelligence (AI) algorithms. The deep learning-based object detection algorithms have been widely adopted for different structures [10], including tunnels [11], buildings [12], and bridges [13]. The recent novel network models achieve better performance, including Faster R-CNN [14, 15] and the You Only Look Once (YOLO) series, [16], [17], [18], [19].

However, the identified defects on the aerial image data alone offer limited insight for a comprehensive assessment of building. Given the recognized capacity of digital models to encapsulate global attributes, researchers are focusing on defect registration, or image-to-model registration, as a solution to this issue. Zhang et al. proposed that localization of the damage on the building is essential for the building condition assessment [20]. Considering that the target buildings have typically been in operation for over 30 years, there is a high probability that their architectural documents have been lost, and the original design blueprints may not

[*] Corresponding author
✉ byzhao@mae.cuhk.edu.hk (Benyun Zhao); xichen002@cuhk.edu.hk (Xi Chen)
ORCID(s): 0009-0008-7298-2927 (Benyun Zhao); 0000-0003-2168-9057 (Xi Chen)

accurately reflect the as-is condition. So it's necessary to enhance the modeling work in addition to BIM model.

In fact, implementing a modeling method that directly acquires data from the buildings can ensure the fidelity of the model, providing an accurate representation of the current state of the infrastructure, such as Terrestrial Laser Scanning (TLS) [21], or 3D reconstruction method (also known as stereo photogrammetry) [22]. Besides, the integration of Structure from Motion (SfM) and learning-based Multi-View Stereo (MVS) endows stereo photogrammetry with considerable advantages in terms of efficiency and cost [23, 24, 25, 26, 27]. The derived models, which exhibit a high degree of consistency with their physical counterparts and can be used to evaluate their health status for guiding future maintenance, are also referred to as Digital Twins (DTs).

This article introduces a DT-based framework that we developed for automated UAV building façade inspection. The proposed framework registers the detected defects on aerial images to the 3D model (including BIM, stereo photogrammetry models, etc.). This method geo-references the initial 3D model to the Geographic Information System (GIS), constructing a preliminary DT model. Subsequently, a corresponding virtual camera is constructed using UAV IMU data and camera optical parameters, and a depth map corresponding to the original aerial photo is obtained in virtual space. Through corresponding depth alignment, the global coordinates of defects are obtained, and 2D defect images are registered to the 3D model. Besides, the semantic information of building is retrived from BIM which makes the component-level evaluation for precise maintenance. This process is not limited by the format of models or the surface shape of the building (it's applicable to non-flat surfaces). It also doesn't require the UAV to take photos perpendicular to the wall, nor does it need to maintain a constant distance from the wall.

2. Related works

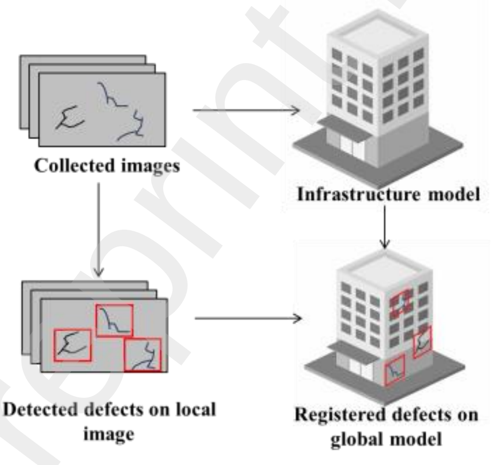

Collected images
Infrastructure model
Detected defects on local image
Registered defects on global model

Figure 1: The detected defects on local image and the corresponding defects registered on model with global reference.

UAV-based visual inspection Visual inspection serves as a valuable tool in detecting preliminary indicators of structural deficiencies in building façade, thereby averting potential severe risks [30]. Moreover, visual inspection imposes minimal equipment requirement, thereby mitigating the costs associated with large-scale evaluations, such as those performed on high-rise buildings and bridges. However, traditional visual inspection requires substantial human resources and faces challenges such as data collection in hard-to-access areas and automated image data processing.

The progression of robotic technology, particularly the advent of UAVs, has significantly facilitated the visual inspection process for large structures. Incorporating advanced sensing technology, lightweight sensors, such as laser scanners, RGB or infrared cameras, can be equipped on UAVs for the capture of video or image data [30]. It amplify the volume of data that can be collected within a brief time span, thereby integrating visual inspection efficiently into routine building maintenance tasks [31]. However, the data collected by UAVs in visual inspection exhibits certain characteristics that presents considerable challenges for processing, including the vast quantity, the high degree of overlap (over 35%), multi-perspective views, and close-range to the surface [28].

The task of conducting precise and efficient defect detection on such image data has surpassed the capabilities of traditional manual identification methods. Computer vision and machine learning technologies have provided solutions for automated defect recognition through their work in object detection. It enables classification and localization of objects of interest in images. Meanwhile, Li et al. adopted Region-based Convolutional Neural Networks (R-CNN) for large-scale surface defect detection in tunnels [11]. Given the increasing scale of scenes and data, detection tasks have higher demands for real-time capability, with more application of YOLO. Despite the achievements of existing research, the detection results are manifested as annotations on 2D images. But it fails to provide a comprehensive assessment of building defects without global information. The envisioned technical approach should consist detection, reconstruction and registration, to match the detected defects with a 3D model, as in Fig. 1.

Digital twin for inspection and defect management DT is supposed to construct the virtual model of the physical parts based on the geometric information, which provide the accurate location and semantic information for the targets [32]. Infrastructure inspection, a critical component of the O&M in building management, has embraced the implementation of DT technology, with numerous successful cases serving as testament to its efficacy. The role of DT in UAV-enabled inspections typically manifests in constructing high-fidelity models of the target facilities to establish a basis for inspection and localization. Initiated as early as 2013, research has already employed TLS-derived point cloud data in the construction of architectural models, consequently setting the foundation for accurate defect localization [33]. The current model construction methods in UAV-enabled inspection include photogrammetry, image-to-BIM projection

Table 1
The comparation between the current and ours registration method

Publication	Application	Method	Associated Technology
Chen et al. [28]	Geo-register 2D GIS spatial model of building façades	Geo-registration	GIS
Tan et al. [29]	Project segemented defect image to BIM based on the UAV GPS localization	Defect registration	BIM
Zhang et al. [20]	Project image to BIM using improved generalised Hough transform	Image-to-BIM registration	BIM
Mohammadi et al. [21]	Integrate BIM with decision support system to provide high fedelity model	Asset management	BIM

moodel, and the simplified Level of Detail (LOD) model. The photogrammetry method captures the topology and geometry of buildings from aerial images, with the advantages of high efficiency, low cost, and ease of integration into automated workflows. However, the derived models lack sufficient structural information, and the accuracy needs to be improved. Chen et al. deployed deep learning method to improve the accuracy of photogrammetry method for DT construction [25]. BIM with semantic information has the potential for further structural analysis and maintenance. Zhang et al. claimed that texturing aerial image onto BIM is preferred than other solutions [20]. But their method only works on flat surface and has limitations on the UAV flight, such as the fixed distance between the UAV and building. The LOD model provides another option as a 3D representation of the asset, which contains simplified geometrical information. Pantoja-Rosero et al. derived the geometric DT from the point cloud and combine them with damage information to form the damage-augmented DT [34]. This facilitates more convenient access by reducing the volume of the model, but it is only suitable for small-scale buildings with simple structure.

Above methods can all construct DTs of the as-is building condition, but most work only focuses on geometric structure, such as the point clouds generated by photogrammetry. The data generated by different methods are incompatible and only viable within their own frameworks. The model built can seldom provide subsequent value for continuous or periodic inspection and management.

Defect registration Defect registration, also known as defect localization, is to locate the defects identified in the images and present the 2D defect information into 3D form [35]. This technology aims to provide fast, comprehensive architectural assessments by correlating image detection results with architectural models [20]. Based on the variations across different model objects and application platforms, it can be divided into three categories: (1) point cloud-based registration; (2) BIM-based registration; (3) Geographic Information System (GIS)-based registration.

Point clouds, serving as models of facilities, are cost-effective and readily accessible, particularly with the development of UAV photogrammetry. They can accurately reflect the as-is building condition, preserving its external visual features and geometric attributes effectively [36]. This makes them a reliable resource for inspection purposes, particularly in the context of aging structures where BIM models are typically absent. Taraben et al. converted the point cloud into voxel representation and visualize the defect with geometric characteristics [37]. Zhang et al. used similar method to construct the finite-element (FE) models for structure damage assessment [38]. Chen et al. constructed the point clouds of the defects and annotate them on the original building point clouds [39]. The inherent limitations of point cloud data impede its interactive capabilities, and the utilization of such data to articulate defect information often faces challenges in delivering substantial and effective insights.

BIM, a widely adopted digital format in the construction industry, possesses the capability to seamlessly integrate both models and inspection data. Musella et al. established the connection between the defects, their associated images, and BIM, but this procedure was manually conducted [40]. Chen et al. proposed an automated method to extract the structure of interest from BIM, improving the accuracy of detection [41]. Automated workflows typically involve coordinate transformation, providing a global reference coordinate system for the registration of images to the BIM [42]. For example, Tan et al. projected the segemented defects to BIM with transformation from the World Geodetic System 1984 (WGS84) to BIM local coordinate [29]. Then, Zhang et al. optimized it with improved generalised Hough transform to ensure the registration accuracy [20]. Based on the preceding analysis, it is evident that the image-to-BIM registration involves several intricate steps, such as complex coordinate transformations. Additionally, the specifications for UAV-captured images and the detected infrastructure surfaces are notably stringent, encompassing factors such as the required distance from the UAV to the wall, the flatness of the surfaces and the inpermissible degree of image overlap.

GIS-based registration processes and analyzes geographical data, enabling the accurate alignment of images or other data with their respective geographical locations. It exhibits superior compatibility with various types of data, encompassing BIM, point clouds, and other model types [43]. Hence, it is extensively utilized in the management of large-scale areas, while BIM is limited to a single building. Chen et al. constructed a 2D GIS spatial model to manage the UAV-collected defect images [28, 44]. Quinci et al. employed GIS for the fusion of multisource and heterogeneous data, aiming to facilitate the assessment and monitoring of infrastructure [45]. This underscores the robust support for dynamic data and the interactive capabilities inherent in GIS. For UAV-enabled inspection, the potential of GIS in managing complex, dynamic, multidimensional data remains to be explored. These works and their methods are listed in the Table. 1

2.1. Research gaps and contribution

The literature review has highlighted the following research gaps:

(1) The existing BIM-based registration methods involve complex coordinate transformation, including the WGS-84 coordinate, region coordinate, projection coordinate, local coordinate, and BIM coordinate [42]. This could compromise the precision of localization, leading to unreliable results.

(2) The current methods for 2D-to-3D registration usually restrict the view points, requiring orthographic and non-overlapping images of the façade. However, in practice, high overlap rate (up to 90%) in data collection is necessary to ensure the completeness of target facilities. So it is required to develop a new method that can automatically and effectively align the unordered inspection photos with 3D models. Besides, the high overlap rate images means that defects are detected multiple times, which is misleading and inaccurate.

(3) 2D registration method ignore the structure of the building and can't deal with the non-flat surface. Also, the semantic information of the defect is ignored, which provides limited insights for structure evaluation.

In an effort to narrow these gaps, we propose a DT-based inspection framework to facilitate UAV-enabled infrastructure inspection. The novelty of this paper is encapsulated in three key aspects:

(1) The UAV-enabled inspection framework is proposed to construct the DT models for buildings, involving model construction, defect detection, and registration.

(2) A GIS-based defect registration method is developed to integrate 2D images with 3D reconstruction models for comprehensive representation. It is the first to solve the problem of defect detection and location in images with irregular POV and overlapped images.

(3) A two-step updating method based on the self-evolution mechanism of DT modeling is proposed. The depth information in virtual model is combined to extend real-world data and correct real-world errors. BIM-based structure retrieval is combined to augment the defect semantic information.

(4) The framework establishes an automated and traceable data flow, offering users an interactive interface to access updated defect models.

3. Methodology

The basic flow chart of this registration method is shown in Fig. 2, which starts with the aerial images collected by the UAV and ends up with a DT of the as-is building condition that integrates defects and a geometric model. The UAV-collected images have additional sensing data in two aspects: optical attributes such as Point of View (POV), image size, etc., and the geographical state of the UAV, including attitude (from Inertial Measurement Unit (IMU)) and position (from Global Positioning System (GPS)).

With the aerial images, the 3D model (either point cloud model or BIM) is constructed by 3D reconstruction algorithm or manual modeling method. Then, the depth texture is generated with the georeferenced model and the corresponding origin aerial image. The detected defects on images are aligned with the distance information on the depth texture to calculate the global coordinate position. After the data fusion, the 2D defects are registered to the 3D model which makes a DT of the target building with as-is conditions.

3.1. Coarse registration

This section introduces and delineates the BIM+GIS approach, which aligns physical objects with virtual counterparts in the virtual space, culminating in the construction of primary DT environment.

While BIM serves as the prevailing model of interest in architectural projects, BIM platforms are typically tailored to the design, construction, and data management of individual buildings. Consequently, their support for regional inspections, particularly those involving unmanned systems and extensive spatio-temporal data, remains notably limited.

Innovatively, we are the first to propose the application of the BIM+GIS approach in the realm of large-scale UAV-based visual inspections, aiming to construct a 4D GeoBIM as the as-is DT representation of architectural defects. The envolved method for primary DT configration entails two fundamental steps: (1) GIS-based georeferencing, encompassing both model and aerial photography POV, and (2) the acquisition of corresponding depth data to facilitate subsequent defect localization.

3.1.1. GeoBIM environment configuration

This sub-section proposes the processes of GeoBIM environment configuration with BIM, GIS and service attributes as in Fig. 3. BIM attribute provides both the geometry model and the semantic metadata of structures, while GIS attribute mainly ensures that elements such as terrain, 3D assets, and world base maps in the environment are consistent with physical scenes. The GeoBIM platform is integrated based on the unreal engine 4.27, which is compatible with GIS and BIM portal by Cesium for unreal API. In addition, the UAVs are modeling with the physical and optical settings.

3.1.2. Pose estimation

The UAV-enabled image data collection involves both optical attributes (image, optical parameters) and pose data (GPS location and camera pose). But the airborne sensors, such as GPS and IMU, are subject to environmental constraints and uncertainties. Therefore, we utilize constraints between images to optimize camera poses. This component of the work integrates an incremental Structure from Motion (SfM) framework, specifically COLMAP, to perform the calibration of multi-view aerial images along with corresponding GPS data as in Fig. 4. Subsequently, the transformation matrices are computed to derive the global geographic information of camera poses.

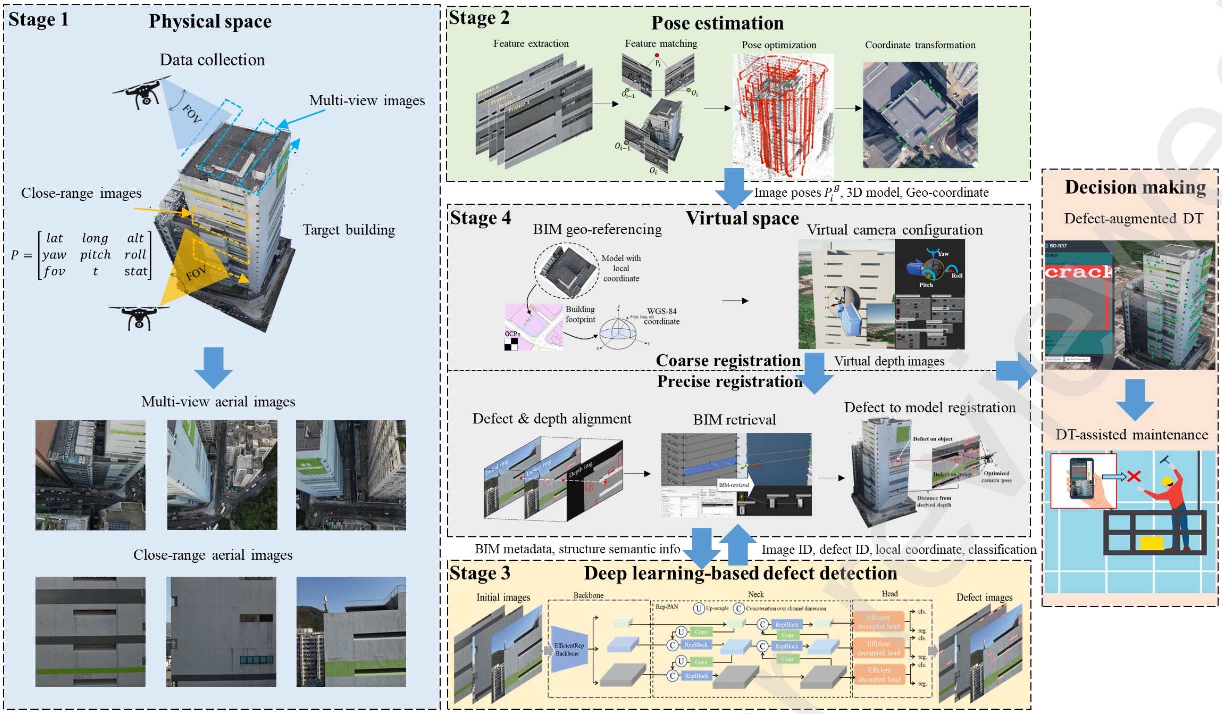


Figure 2: A workflow of autonomous building façade defect detection and management under the DT framework.

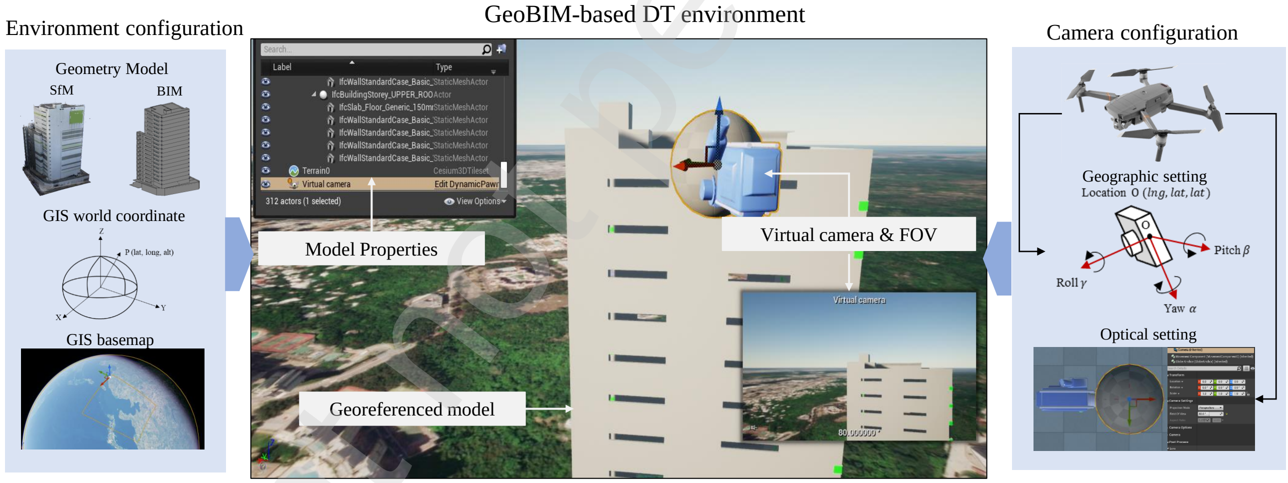


Figure 3: The GeoBIM-based DT environment configuration.

Firstly, for a large batch of consecutive frame images, the ORB algorithm provides fast and robust image feature extraction. Subsequently, spatial matching selects two consecutive frames, i-1 and i, for initialization. The matching of consecutive frames involves fundamental matrix calculation using the Random Sample Consensus (RANSAC) five-point method to eliminate outliers. The resulting image pairs and matching relationships are used for 2D-2D matching based on epipolar geometry, obtaining the relative poses of image pairs. Then, triangulation is employed to generate 3D points, establishing 2D-3D matching relationships, and utilizing the PnP algorithm to solve the camera's poses.

In practice, it is prone to outliers in the matching process, i.e., incorrect matches of feature points. To address this issue, the normalized 8-point algorithm with random sample consensus (RANSAC) is introduced. It comprises the following steps: randomly select eight pairs from all matched keypoints in the image pair, compute the corresponding fundamental matrix F_i (since there is no need to consider the case where the camera centers of matched frames are the same, there is no need to involve the homography matrix H); then, calculate the sampson distance errors for all pairs of matching points with F_i, and if the error is less than a threshold, it is considered an inlier; repeat the above steps until the maximum number of iterations is reached to obtain the optimal match with the maximum number of inliers.

The calculation formula from epipolar constraint for matrix F is as follows:

$$F' = \underset{det F'=0}{\arg\min} \|F - F'\| \\ \begin{cases} p_i^\top F_i p_{i-1} = 0 \\ F_i = K^{-\top} \cdot t^\wedge \cdot R \cdot K^{-1} \end{cases} \tag{1}$$

where, p_i and p_{i-1} are the corresponding points of the physical point P in the image pair; K refers to the intrinsic matrix of camera; $t^\wedge$ is the skew-symmetric of translation matrix from frame i-1 to i; R is the rotation matrix from frame i-1 to i.

The calculation formula for Sampson distance to filer the outliers is as follows:

$$\begin{cases} d(p_{i-1}, p_i) = \dfrac{p_i^\top F_i p_{i-1}}{(F_i p_{i-1})_x^2 + (F_i p_{i-1})_y^2 + (p_i^\top F_i)_x^2 + (p_i^\top F_i)_y^2} \\ d(p_{i-1}, p_i) < \tau \end{cases} \tag{2}$$

where, τ is the maximum error for inliners as an approximate estimate from p_i to $F_i p_{i-1}$. Then number of the corresponding inliers is recorded in this iteration. Given the ratio of inlier is 0.5, the iteration time is set as 1000 to achieve the matching rate of 0.99. The transformation matrix between the image pairs can be decomposed from matrix F. Then the poses of following frames are solved by Perspective-n-Point (PnP) algorithm. The corresponding fomula is as following:

$$[R|t] = \underset{R \in SO(3), t \in \mathbb{R}}{\arg\min} \sum_{i=1}^{n} \min(\left\|p_i - \pi(RP_i^w - t)\right\|^2, d^2) \tag{3}$$

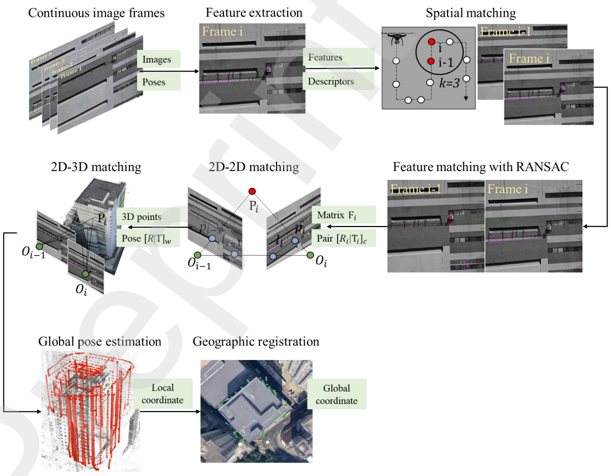


Figure 4: Pose estimation.

where, function π is the projection function to project the 3D point P_i to pixel coordinate and the threshold d is to filter the outliers. Hence, the optimized poses for all the current frames are updated to the camera property in GeoBIM library.

Through the preceding steps, we have filtered and calibrated the poses of each frame; however, these poses are defined within a local coordinate system. To be more precise, the current 3D spatial coordinate system is a Cartesian coordinate system represented by the matrix $[R|t]$. Recognizing that transformation matrix calculations cannot be directly performed in the WGS84 coordinate system, we establish an initial Earth-centered, Earth-fixed (ECEF) coordinate system through a 3D similarity estimation with more than 3 settled frames. Subsequently, we calculate the poses of other frames in the ECEF coordinate system as $P_i^e = \{X_i, Y_i, Z_i\}$. Finally, these poses are transformed into WGS84 coordinates $P_i^g = \{lng_i, lat_i, alt_i\}$, resulting in optimized global pose localization information. The transformation formula from ECEF coordinate to WGS84 coordinate is as follows:

$$\begin{cases} p = \sqrt{X_i^2 + Y_i^2} \\ N = \dfrac{a}{\sqrt{1 - f(2-f)\sin^2(lat_i)}} \\ lng_i = \arctan \dfrac{Y_i}{X_i} \\ alt_i = \dfrac{p}{\cos(lat) - N} \\ lat_i = \arctan \left[\dfrac{z}{p} (1 - e^2 \dfrac{N}{N + alt_i})^{-1} \right] \end{cases} \tag{4}$$

3.1.3. Image registration and depth calculation

The fundamental principle of 3D engine visualization rendering is to transform a scene represented in 3D into a 2D form. The coordinate processing involves various stages such as modeling transformation, viewing transformation, projection transformation, perspective transformation and viewport mapping. Through model and view transformations, the 3D coordinates of objects are converted into 2D coordinates on the screen. Notably, the clipping space, normalized device coordinate space, and screen space are generally inherent, while local space (model space), world space, and camera space are typically user-defined. The entire process is illustrated in the accompanying Fig. 5.

To obtain the distance from the camera to an object, i.e., the depth value, it is necessary to perform a reverse calculation based on the depth texture acquired in screen space. Given that the transformation from local space to eye space (camera space) does not involve scale changes, the depth transformation process can be focused solely on the transition from screen space to view space. As in Fig. 8(e), it's required to calculate the depth in eye coordinate $z_e(i, j)$ from the screen coordinate $z_s(i, j)$, where (i,j) is the pixel coordinate of the depth image.

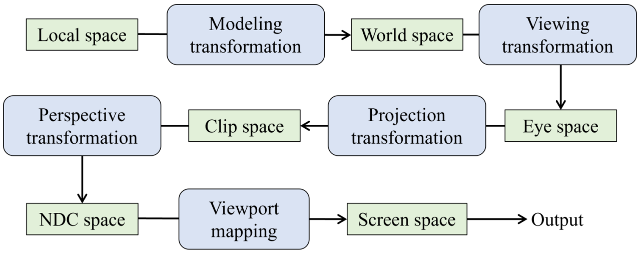


Figure 5: The coordinate transformation pipeline for 3D to 2D rendering.

The depth transformation between the normalized device coordinate (NDC) space $z_n(i,j)$ and the screen space $z_s(i,j)$ is as follows:

$$z_n(i,j) = \frac{2z_s(i,j) - (f_s + n_s)}{f_s - n_s} \tag{5}$$

where f_s is the far plane and n_s is the near plane on screen space with the default value $f_s = 1$ and $n_s = 0$. Accordingly, the depth transformation between the NDC space and the eye space is as follows:

$$z_e(i,j) = \frac{2fn}{(f-n)z_n - (f+n)} \tag{6}$$

As a result, from screen space to view space, the depth value are calculated as the formula:

$$z_e(i,j) = \frac{n}{z_s(i,j)(f-n) + f} \tag{7}$$

3.2. Precise registration

Coarse registration establishes a GeoBIM environment, facilitating the generation of virtual BIM images and physical depth maps that precisely align with POV of physical photographs. For the attainment of global defect registration and semantic annotation, further refinement, known as precise registration, is imperative. This process encompasses the following objectives: (1) the detection and local positioning of defects; (2) the registration of individual defects onto the model; (3) the employment of semantic retrieval for structure alignment.

The essence of this approach is the sophisticated automation of defect detection and registration, achieved through extensive analysis of UAV-captured images. Each detected defect is meticulously cataloged, aligning with a unique identifier within the architectural component, and seamlessly integrated into an cloud-based database system. Remarkably, this method transcends traditional limitations associated with the UAV's POV, the characteristics of the detection subject, and the photographic technique employed. It allows for flexibility in UAV positioning relative to irregular wall surfaces—which need not be perpendicular—and imposes no constraints on the overlap percentage of the defect images, thereby offering an elegant and efficient solution to architectural defect management.

3.2.1. Deep learning-based defect detection

1) **Dataset construction:**

The effectiveness of current learning-based methods in defect detection for large-scale infrastructures is significantly hindered by the lack of a high-quality open-source dataset. To bridge this gap, we present CUBIT-Det, the first high-resolution dataset specifically designed for detecting various defects in extensive infrastructures[46]. This dataset includes 5,527 images captured by unmanned systems, with a remarkable maximum resolution of 8000 × 6000. The dataset's defect images are taken from multiple angles and distances under various lighting conditions, offering a comprehensive array of structural details. This variety ensures the robustness of models in practical inspection scenarios. The dataset covers the three most common types of infrastructure: buildings (65%), pavements (29%), and bridges (6%), focusing on the inspection of three primary defect types: cracks (82%), spalling (12%), and moisture (6%) as shown in Fig. 6.

2) **Real-time detection and localization:**

Based on the self-established dataset, we conduct evaluations on a multitude of state-of-the-art (SOTA) learning-based real-time object detection algorithms to ascertain the optimal solution of the task of defect detection in terms of both speed and accuracy. We train and test 9 SOTA series algorithms (nearly 30 models): YOLOv5 [16], YOLOv6 [17], YOLOv7 [18], YOLOX [19], PP-YOLO [47], PP-YOLOv2 [48], PP-YOLOE [49], PP-YOLOE+ [49] and Faster R-CNN [14].

3) **Evaluation metrics:**

Precision (P), Recall (R), and Average Precision (AP) are the three most commonly used metrics in object detection for infrastructure defect detection. Precision (P) measures the accuracy of detected defects, denoting the ratio of correctly identified defects to all detections made by the model. Recall (R), on the other hand, assesses the rate of missed detections, indicating the proportion of correctly identified defects among all actual defects. Precision and Recall are defined as follows:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

The Average Precision (AP) metric represents the weighted mean of precision scores at each threshold on the precision-recall (PR) curve, using the increase in recall from the previous threshold as the weight. Given the multi-category nature of our detection task, we calculate the AP for each category and then compute the mean Average Precision (mAP) across all categories. The equations for AP and mAP are provided below. Here, AP_i represents the AP for class i, and m is the total number of classes.

$$
\begin{aligned}
AP &= \int_0^1 p(r)dr \\
mAP &= \frac{1}{m}\sum_{i=1}^{m} AP_i
\end{aligned}
\tag{10}
$$

Unlike traditional integral methods used to calculate AP, the computation of AP in MS COCO involves a discretization process: the PR curve is defined as the average of precision values at a set of 101 evenly spaced recall levels [0, 0.01, ..., 1] (from 0 to 1, with the increments of 0.01). The equations for mAP in MS COCO is shown below:

$$
mAP_{COCO} = \frac{1}{m}\sum_{i=1}^{m}\left[\frac{1}{101}\sum_{r\in(0,0.01,...,1)} p * interp(r)\right] \tag{11}
$$

4) **Model selection:**

We visualize the inference time versus $AP_{0.5:0.95}$ of the selected SOTA models in Fig 7. For all series of algorithms, as the model size increases, the inference speed will decrease while the detection capability will improve. However, there is a bottleneck in detection capability, which means that simply enlarging the model to realize the enhancement of detection performance cannot always be effective. From the top-left corner of Fig 7, it becomes clearer that YOLOv6 [17] (red star) networks demonstrate a fabulous trade-off between accuracy and latency on large-scale infrastructure defect detection task.

YOLOv6 [17] divides the input image into a grid, simultaneously predicting bounding boxes and class probabilities for each cell. This model uses anchor boxes—predefined bounding boxes—that are adjusted to fit detected defects like cracks, spalling, or moisture. For each defect, YOLOv6 [17] outputs a bounding box with four parameters: x and y coordinates of the center, and the box's width and height. These parameters, normalized to the image's dimensions, offer a scalable object localization method. Alongside these spatial parameters, the model also outputs a confidence score reflecting the model's certainty in the detection, as well as class probabilities indicating the type of defect detected. The result is a set of bounding boxes, each associated with a defect type and its relative location within the image, providing critical data for subsequent analysis and rectification in architectural maintenance and restoration efforts.

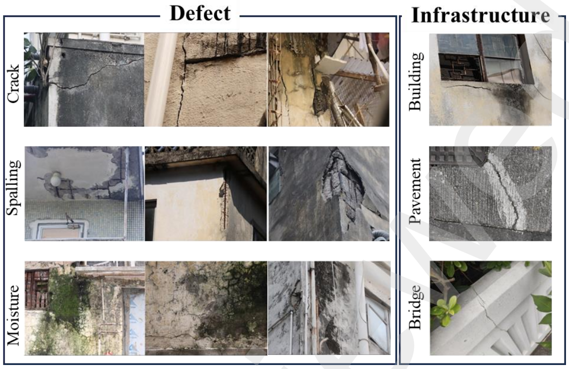


Figure 6: Self-established dataset for infrastructure defect detection. The defect category includes crack, spalling and moisture, and the infrastructure category includes building, pavement and bridge.

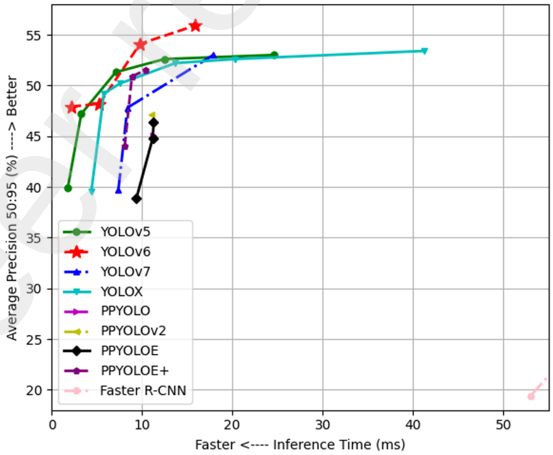


Figure 7: Trade-off performance of different models about inference time versus $AP_{0.5:0.95}$ trained on CUBIT-Det dataset. The further the point is toward the top-left corner, the stronger the detection capability and the shorter the inference time.

3.2.2. Individual defect to model registration

A pivotal step in the integration and management of defect detection outcomes—encompassing defect images, local positioning information, and classification—entails the registration of these results. This process signifies the execution of global positioning and the elimination of redundant defects. Registration ensures that each detected defect is accurately mapped within a comprehensive spatial framework GeoBIM, facilitating precise localization across the entirety of the inspected structure. Moreover, this step is instrumental in streamlining the database by filtering out duplicate entries, thereby enhancing the efficiency of subsequent analyses and interventions. Through the meticulous alignment and consolidation of detection results, the registration phase lays the groundwork for a robust and coherent database of architectural defects, pivotal for effective maintenance and remediation strategies.

We follow the geographic transformation paradigm shown in Fig. 9 to project the detected defects from the image coordinate onto the geo-referenced model. In previous sections, we have already derived the corresponding detected images, GeoBIM images and depth images through GeoBIM environment as in Fig. 8(a)-(d). The defect images contain the origin images with defects marked by red bounding boxes and the local coordinates and defect size on the image plane are recorded individually in data library. For j_{th} defect in the i_{th} image (denoted as defect i_j), we first capture the geographic coordinate of the image center O from the result in the pose estimation section as P_i^g. Next, we translate

O to O'' along the depth axis (vector $\vec{d}(i_j)$) to align it with the facade of the infrastructure. Specifically, the real-world distance corresponding to $\vec{d}(i_j)$ is computed as z_e in previous section. Subsequently, we convert the relative distance between O' and the center of the defect's bounding box i_j into a metric distance (represented by the violet vector $\vec{l}(i_j)$ from O'' to the defect i_j). Finally, we determine the global position of defect i_j by shifting the geographic coordinates of point O'' along the tangential vector $\vec{l}(i_j)$. Following this approach, we automatically determine the global position of each detected defect. The defect category and appearance are also documented to aid in maintenance measures. The corresponding calculation process is detailed in Algorithm 1.

3.2.3. Semantic retrieval for structure matching

Global registration furnishes the geographical coordinates of defects, facilitating the establishment of a one-to-one correspondence between these defects and the geospatially coupled architectural structures within the GeoBIM environment. In this section, we will expound on the methodology for conducting structural retrieval correlated with identified defects through GeoBIM, aiming to construct an assessment of building defects at the structural level. Given that GeoBIM incorporates metadata from BIM and introduces GIS data as well as inspection data from UAVs, as illustrated in Fig. 8. Utilizing collected images along with their fully corresponding GeoBIM-derived images allows for the retrieval of BIM semantic information corresponding to elements present in the images.

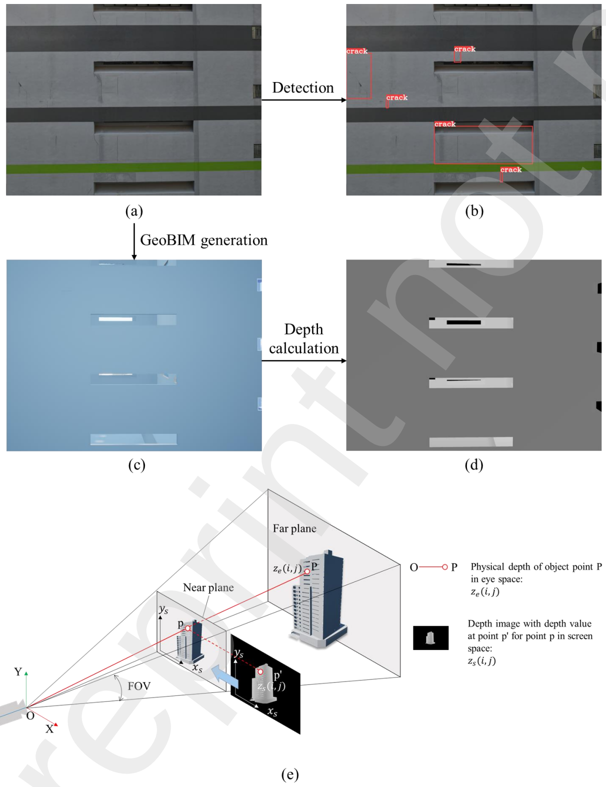


Figure 8: Corresponding images from different sources: (a) original aerial image;(b) detected image;(c) GeoBIM generated image;(d) GeoBIM derived depth image. (e) The transformation process to compute the physical depth $\vec{OP}$ from image to the building surface.

Algorithm 1: Individual defect registration

```
Input: Geographical coordinates of image:
       $P_i^g(lon_i, lat_i, alt_i)$; Distance to the wall of
       defect $d_{i,j}$: $z_e(i,j)$; Projection vector: $\vec{n}$.
Output: Global location of defect $d_{i,j}$:
        $g(i,j)(lng_{i,j}, lat_{i,j}, alt_{i,j})$
Initialize $i$ array of the detected images;
while $t \in 1, 2, ..., i_{max}$ do
    Project from image localization $P_i^g$ to the
      corresponding point on the wall;
    Given the projection vector:
    $\vec{n} = (\alpha_i, \beta_i, \gamma_i)$
    The projection point:
    $P_i^{g*} = P_i^g + z_e(i,0) * \vec{n}$
    for $i \in 1, 2, ..., i_{max}$ do
        The diagonal length of image i:
        $L_D(i) = 2z_e(i,0)\tan\frac{FOV}{2}$
        The tangential vector on the wall from
          image center to defect:
        $\vec{L}(i,j) = L_D(i)\frac{l(i,j)}{l_D(i)}$
        Compute the location of defect $d(i,j)$
        $g(i,j) = P_i^{g*} + \vec{L}(i,j)$;
        Set duplication factor T:
          $T = \sqrt{(0.5/unit_{lng})^2 + (0.5/unit_{lat})^2}$
        for $g(k,t)$: $k \leftarrow 1$ to $i-1$, $t \leftarrow 1$ to $j_{max}-1$
          do
            if $\sqrt{(g(k,t) - g(i,j))^2} > T$ then
                Update the new defect $g(i,j)$;
        $Updatej \leftarrow j+1$
    $Updatei \leftarrow i+1$
Output the projection point
  $g(i,j)(lng_{i,j}, lat_{i,j}, alt_{i,j})$
```

This approach not only enhances the precision of defect assessments but also contributes to the strategic allocation of resources for infrastructure maintenance, underscoring the critical role of GeoBIM in advancing the state-of-the-art in architectural defect management.

As depicted in Fig. 10, while BIM provides essential structural prior knowledge for constructing a building's DT, it does not automatically link to the defects detected within the structure. To bridge this gap, we have developed an automated workflow, which systematically facilitates the integration of BIM with detected defects and other relevant data.

1)BIM registration: This first step involves aligning and registering the BIM data with real-world physical data gathered from the site. This process ensures that the BIM model accurately reflects the current state of the structure, providing a reliable foundation for further analysis. BIM registration serves as the cornerstone for merging virtual and physical data, allowing for precise overlay of detected defects on the BIM model.

2) Knowledge extraction: Upon successful BIM registration, semantic information extraction is executed. This stage involves parsing and interpreting data from the BIM model, such as material properties, structural components, and spatial hierarchies (including specific floor levels and orientations). This extracted knowledge forms the basis of a detailed knowledge base that contextualizes each element of the building.

3) Metadata transference: This step includes transferring and storing metadata that encapsulate both the static data from BIM and dynamic data collected from UAVs, such as flight paths, timestamps, aerial images, and detailed logs of detected defects. This metadata is critical for maintaining a traceable, timestamped record of the building's condition over time, facilitating trend analysis and predictive maintenance.

4) Retrieval mechanism: The final component of the workflow is the development of a sophisticated retrieval mechanism that leverages the consolidated metadata and knowledge base. This mechanism allows users to query the DT for specific information, such as the location and severity of defects, material specifications, or historical data concerning particular structural components. The retrieval system is designed to support complex queries that can combine spatial, temporal, and semantic criteria, offering an intuitive interface for accessing comprehensive building assessments.

Through these interlinked processes, the automated workflow not only enhances the building DT with detailed defect-related information but also transforms it into a dynamic tool for ongoing structural health monitoring and management. This enriched DT provides stakeholders with actionable insights, enabling informed decision-making regarding maintenance and repairs, thus extending the lifecycle and ensuring the safety of the infrastructure.

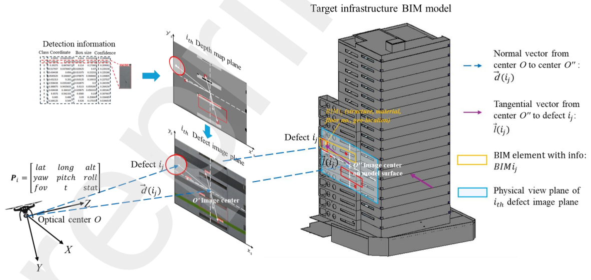

Target infrastructure BIM model
Detection information
Class Coordinate Box size Confidence
i_{th} Depth map plane
i_{th} Defect image plane
Defect i_j
Defect i_j
$\boldsymbol{P}_i = \begin{bmatrix} lat & long & alt \\ yaw & pitch & roll \\ fov & t & stat \end{bmatrix}$
$\vec{d}(i_j)$
$\vec{l}(i_j)$
O' Image center
O'' Image center on model surface
BIMi_j (structure, material, floor no., geo-location)
Optical center O
X
Y
Z
Normal vector from center O to center O'': $\vec{d}(i_j)$
Tangential vector from center O'' to defect i_j: $\vec{l}(i_j)$
BIM element with info: $BIMi_j$
Physical view plane of i_{th} defect image plane

Figure 9: The projection of defects from image to model.

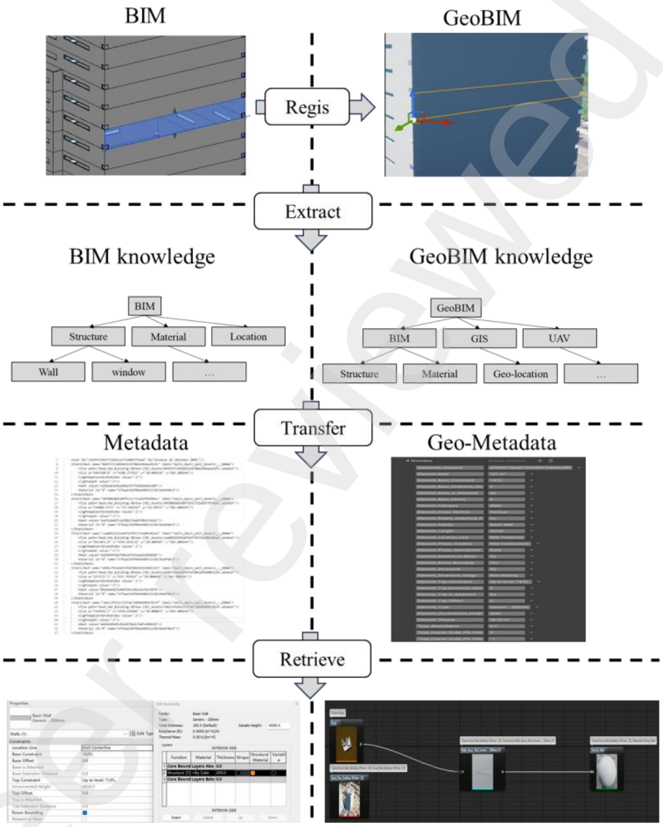


Figure 10: The hierarchy of corresponding BIM and GeoBIM.

4. Implementation

We deploy our proposed inspection framework on various large-scale scenarios to verify its effectiveness and efficiency. Here, we take a large-scale high-rise warehouse (36m×27m×100m) as a representative instance.

To verify the effectiveness and efficiency on real large-scale scenarios, we have deployed the method on a large-scale high-rise warehouse. Fig. 11 illustrates the application of our methodology to a commercial building which rises to a height of 100 meters and spans an area of approximately 27×36 meters, located in the Shatin district of Hong Kong. This 18-story structure was extensively surveyed using three DJI Mavic 2 drones, each equipped with a camera capable of capturing images at a resolution of 8000×6000 pixels. These UAVs were deployed to collect over 1000 aerial photographs to facilitate a detailed analysis of the building's features and conditions. The specific locations from which these images were acquired are detailed in Fig. 11(b).

To process this substantial amount of high-resolution data, the study utilized advanced computing hardware, comprising an Intel(R) Core(TM) i9-10920X CPU and an NVIDIA GeForce RTX 3090Ti GPU. This setup was chosen to ensure robust and efficient handling of the data, enabling precise and timely analysis of the structural integrity of the building.

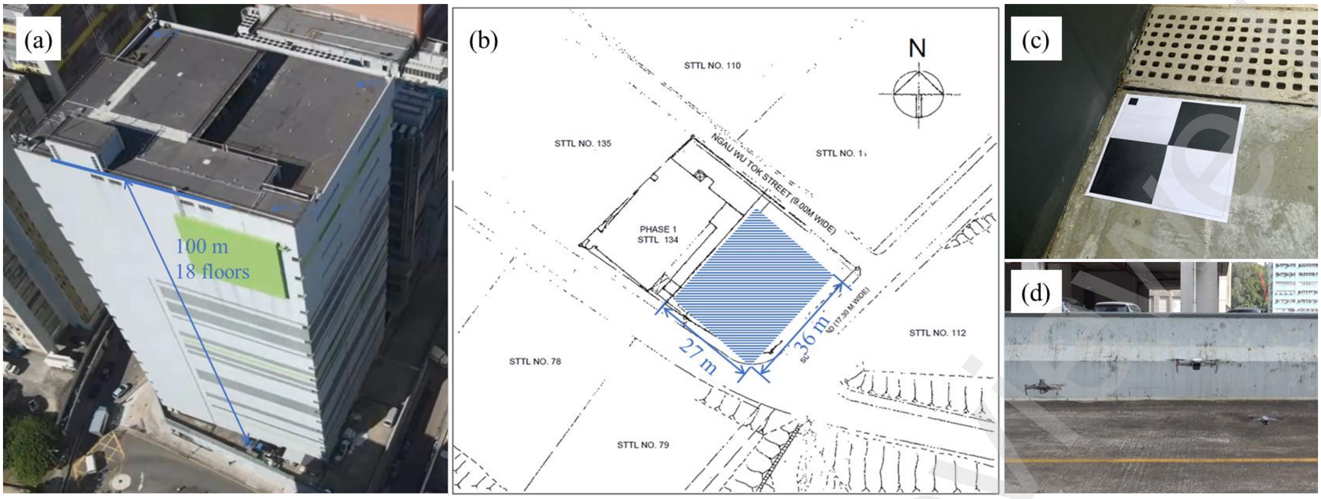


Figure 11: Experiment scene to evaluate the proposed approach: (a) Aerial view of the target building; (b) Footprint of the target building; (c) GCP on the building; (d) Multiple UAVs used in data collection.

4.1. Field experiment results

In the field experiments, the primary focus was on the collection of data via UAVs, with specific attention to ensuring image quality, data completeness, and flight safety. To facilitate this, flight paths were meticulously planned based on the GeoBIM surface model of the structure. Typically, this planning necessitates maintaining the UAVs in a perpendicular orientation to the building's walls while keeping an approximate distance of 10 meters to optimize image capture and data accuracy.

Following these guidelines, the UAVs executed their flights along predetermined trajectories, effectively adhering to the designed flight paths. To enhance the efficiency of data collection and mitigate potential issues such as battery depletion, three drones were deployed simultaneously. This strategy allowed for comprehensive aerial coverage of the target building's surface, achieving complete data acquisition within a span of thirty minutes. This coordinated approach not only maximized the productivity of the data collection phase but also ensured the safety and reliability of the operational process.

4.1.1. Results of individual defect registration

As proposed in Section 3, the registration process involves projecting individual defect images from their original GPS locations to georeferenced 3D model. Specifically, this is achieved by utilizing GeoBIM to generate corresponding coarse registration images and employing depth maps to ascertain physical distances, thereby accurately localizing each defect onto the model.

We collected 1016 aerial images of 8000 × 6000 resolution, covering the exterior surface of the building. These images are then detected and registered to the corresponding SfM model for visualization and evaluation. Fig. 12 illustrates the sequence of intermediate results generated throughout this process. Evaluated by building inspection experts, the detection results achieves more than 80% $\text{mAP}_{0.5}$ accuracy (within 30 FPS), which indicates a high degree of congruence within the dataset and the appropriateness of the chosen model for accurate detection. Fig. 12(a) shows the defect detection outcomes derived from aerial photographs with bounding boxes (red rectangle boxes) on the images. Fig. 12(a) shows the depth maps derived from GeoBIM, which provide crucial spatial information for defect localization. Finally, panel (c) visualizes the registered defects within a WebGIS platform based on Cesium [50], with the defects' central positions marked by green points. From Fig. 12(c), it is evident that the SfM model more accurately represents the as-is condition of the building compared to BIM. The SfM serves as the primary geometrical and visual representation within the DT framework, and the registered defects effectively reflect their corresponding true geographical locations. These results substantiate the efficacy of our methodology. However, errors are inevitable in the data processing phase. To quantify the precision of our approach, we have developed a corresponding validation method.

Considering that the defects have been located in global geographic coordinates, we calculate the defect localization error as the offset between the registered defect positions on GIS-derived images and the centers of the detected defect bounding boxes on original images. To achieve this, we have reconstructed each POV of the camera within the GIS virtual space, which precisely mirrors the actual world settings and incorporates identical geographical and optical features, as shown in Fig. 12(c). Following this setting, we overlay the defect images onto these virtual images to accurately determine the defect localization errors. Here, the gray square masks represent the original defect bounding boxes, and the registered defects are marked with relative green points, as illustrated in Fig. 12 (d). These discrepancies are then converted from pixel measurements to physical units measured in centimeters. The accuracy of our registration method is evaluated by calculating the Mean Absolute Error (MAE),

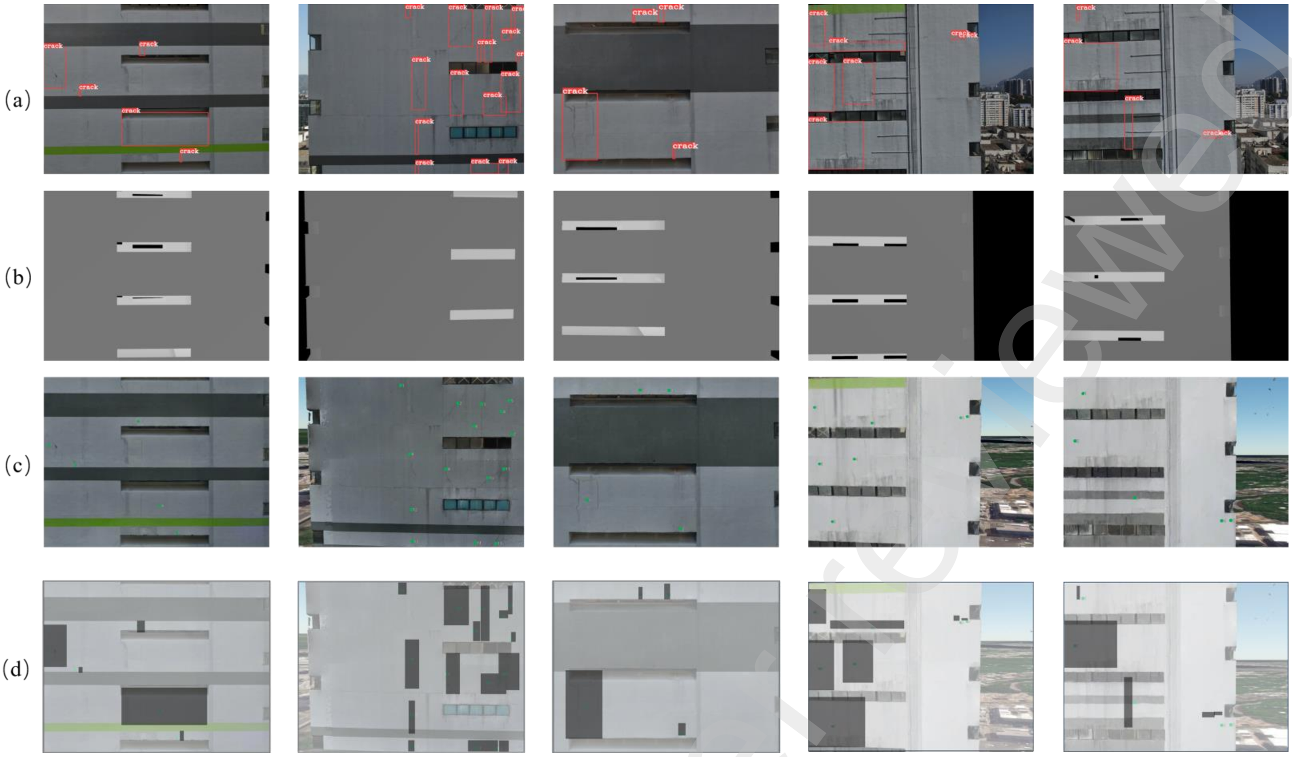


Figure 12: The process of quantitative evaluation of the results of detection defect registration: (a) Detection images; (b) GeoBIM derived depth image; (c) Registration results of individual defects; (d) Mask of defects for registration evaluation.

Table 2
Defect Registration Error for Large-scale Infrastructure (computed over 1016 close-range facade images)

Registration Error (cm)	Mean	MAE	RMSE	IQR
Horizontal	0.490	2.350	4.746	0
Vertical	0.592	1.037	2.385	0
Diagonal	1.360	4.056	7.149	3.747

the Root Mean Square Error (RMSE), and the Interquartile Range (IQR)—the latter being the difference between the first quartile (Q1) and the third quartile (Q3). The results, as tabulated in Table. 2, confirm the centimeter-level accuracy of our approach. These statistical measures provide a robust assessment of our method's precision, ensuring that our registration technique is both reliable and suitable for practical applications in defect detection and localization.

This evaluation methodology is designed to rigorously assess the accuracy of defect localization and the fidelity of the geometric representations in our DT models. By systematically comparing the derived positions and conditions of structural defects against empirical measurements and DT-derived data, we can not only validate the effectiveness of our process but also identify areas for further refinement and enhancement. In fact, both theoretically and in practice, we have demonstrated that this method possesses commendable robustness and scalability, effectively addressing the limitations inherent in existing methodologies. Our approach facilitates the registration of irregular images, the exclusion of non-target areas, the merging of redundant defects, and the verification of model integrity. Details are listed as below:

1) Irregular defect image registration:

In practical applications, the trajectory of UAVs seldom aligns perfectly with the planned flight paths due to factors such as localization errors, planning inaccuracies, and the influence of wind forces. This discrepancy is particularly pronounced in manual flight scenarios where aerial photographs are sometimes captured at skewed angles relative to the building facades. Discarding these images, however, would compromise the integrity of the data collected. Unlike existing methods that require the camera to be perpendicular to flat wall surfaces, our approach demonstrates robust performance even with skewed shooting angles and on arbitrary wall surfaces. As illustrated in Fig. 13, it demonstrates that the defect registration images on the corresponding GIS platform exhibit a favorable match with the original detection images, accurately pinpointing defects at skewed and irregular positions. Our method effectively compensates for these irregularities, ensuring that comprehensive and accurate data are still captured. This adaptability enhances the usability of UAVs in diverse architectural and environmental conditions, thereby broadening the scope of applications for UAV-based surveying and inspection in the field of structural health monitoring.

2) Exclusion of non-target areas:

Due to the reasons mentioned above, aerial photographs captured by UAVs not only encompass the target building but may also include extraneous elements such as adjacent

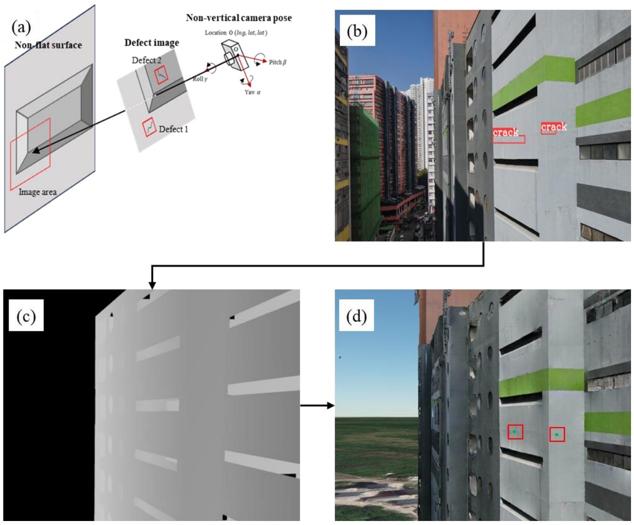


Figure 13: Irregular defect image registration: (a) Original defect image; (b) GeoBIM-derived depth map; (c) Registered defect image (defects are marked as green spots).

structures and trees. Utilizing GeoBIM-derived depth maps, which focus exclusively on the architectural structure itself, aids in the elimination of non-target areas within the images. As depicted in Fig. 14(a), the presence of neighboring buildings can interfere with detection algorithms, leading to erroneous results. Depth maps, as shown in Fig. 14(b), facilitate the direct generation of masked images (Fig. 14(c)), where the target areas are denoted in white (value 1 in image) and non-target areas in black (value 0 in image). Consequently, the final image output (Fig. 14(d)) is purged of non-target areas, thereby also removing incorrect detection outcomes and enhancing the precision of the data. This methodology significantly improves the quality of the analysis by focusing solely on relevant architectural details, thus optimizing the effectiveness of the detection process in urban and complex environments.

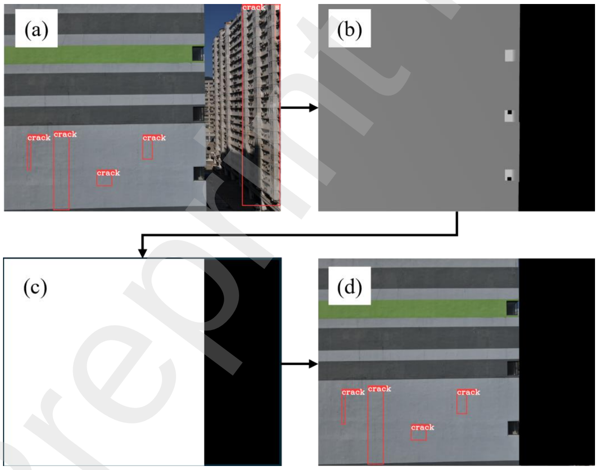


Figure 14: Exclusion of non-target areas: (a) Original defect image with non-target area defect (FP result); (b) GeoBIM-derived depth map; (c) Mask image from depth map; (d) Defect image with target area.

3) Redundant defect merging:

Previous research indicates that UAV-based defect detection tasks often require a high overlap rate—sometimes as much as 90%—which means that the same defect may appear in multiple images, as illustrated in Fig. 15(a). As a result, the final detection output typically includes numerous redundant defect findings, all pointing to the same physical defect. Our method improves upon this by localizing defects to unique global geographic coordinates, enabling the merging of duplicate detection results. This approach ensures the uniqueness of each defect by effectively merging overlapped detections from adjacent images, as demonstrated in Fig. 15(b). By implementing this strategy, we not only streamline the data but also enhance the accuracy of our defect mapping, ensuring that each defect is represented just once in the analysis. This consolidation significantly reduces data clutter and improves the efficiency of subsequent processing and analysis steps, leading to more reliable and actionable insights.

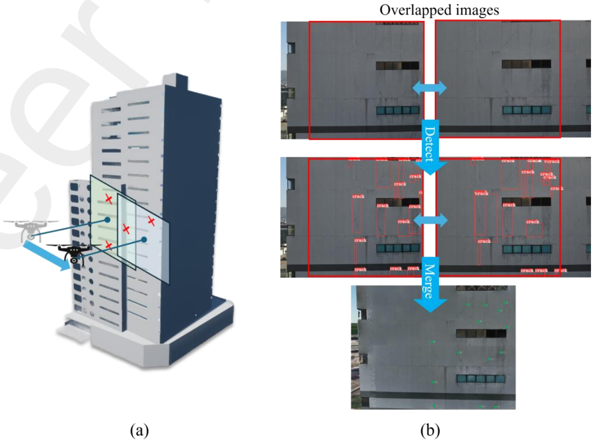


Figure 15: Registration to merge redundant defects on overlapped images: (a) Aerial photography with overlap ratio; (b) Merging of redundant defects (Overlapped area is marked by bounding box).

4) Verification of model integrity:

The results of 3D reconstruction, specifically the 3D models of target buildings, often suffer from issues such as voids and distortions due to insufficient data completeness during collection. Typically, the evaluation of reconstruction methods is conducted on datasets, but such datasets for large-scale architectural scenes are exceedingly rare. Furthermore, generating ground truth for each target building (e.g., through comprehensive laser scanning) is cost-prohibitive and impractical in real-world applications. Therefore, developing effective evaluation methods for assessing the quality of model constructions is a critical need in DT modeling. Our approach offers a feasible quantitative perspective to address this challenge. By comparing the reconstructed models with BIM-derived depth maps from identical POV, we assess the structural integrity of the corresponding constructions. As illustrated in Fig. 16, by comparing depth images from

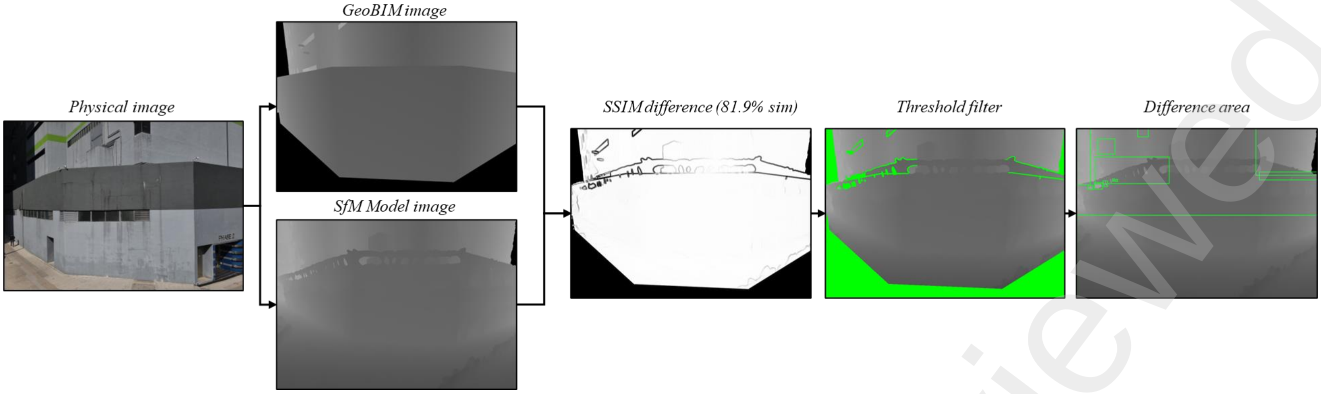


Figure 16: SSIM compare for verification of model integrity.

GeoBIM with corresponding SfM model image, structure defects in the modeling process can be precisely localized. Structural differences between images are quantified using the Structural Similarity Index Measure (SSIM), which is 81.9% for the given sample. The formula of SSIM for image i and j is shown below:

$$SSIM(i,j) = \frac{(2\mu_i\mu_j + C_1) + (2\sigma_{ij} + C_2)}{(\mu_i^2 + \mu_j^2 + C_1)(\sigma_i^2 + \sigma_j^2 + C_2)} \tag{12}$$

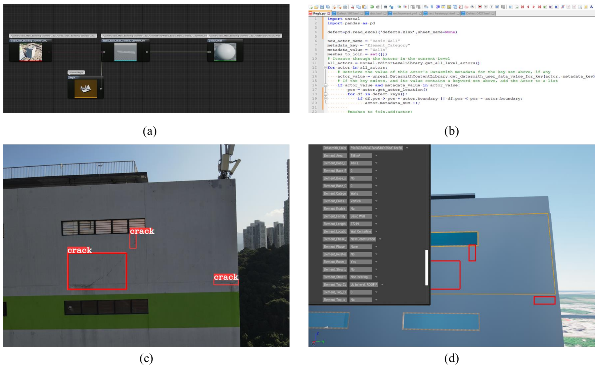
where μ_i and μ_j are the pixel sample mean; σ_{ij} is the covariance of i and j; σ_i^2 and σ_j^2 are respectively the covariance of each image; $c1$ and $c2$ are variables to stabilize the division with weak denominator.

Subsequent differences that exceed a predefined threshold are filtered to delineate predictive bounding boxes, thereby identifying specific defect areas. This comparison allows for the quantitative identification of structural anomalies, such as voids or boundary distortions, at specific locations. Such assessments are instrumental in guiding further data capture and model updates, thereby enhancing the accuracy and utility of the 3D reconstruction process. This method not only improves the fidelity of architectural models but also supports the iterative refinement and updating of DTs, ensuring their applicability and reliability in practical scenarios.

4.1.2. *Results of GeoBIM retrieval*

Using the aforementioned approach, GeoBIM has successfully extracted all structural semantic information from the BIM system and completed geographic registration. As illustrated in Fig. 17(a), detailed information about each architectural element is accessible. To automate the retrieval of structural information corresponding to each defect, we utilized the script in Fig. 17(b) to acquire the geo-position and geometric boundary data of all structures, comparing these with the locations of defects. Fig. (c) presents an image of a specific defect, while (d) shows the structural information corresponding to that defect retrieved via GeoBIM. This method efficiently links each defect with its respective structural location, thereby providing a detailed depiction of the defect distribution within the building structure.

Particularly, since defect detection primarily focuses on the façade of concrete structures, the final data display the distribution of defects across each floor. Each direction represents a different wall surface; for instance, the northwest-facing wall, due to visual obstructions, only includes defect data from the upper floors, with no lower floor defects included. This structure-oriented distribution of defects supports systematic assessments of structural damage in buildings, guiding efficient and targeted maintenance strategies.

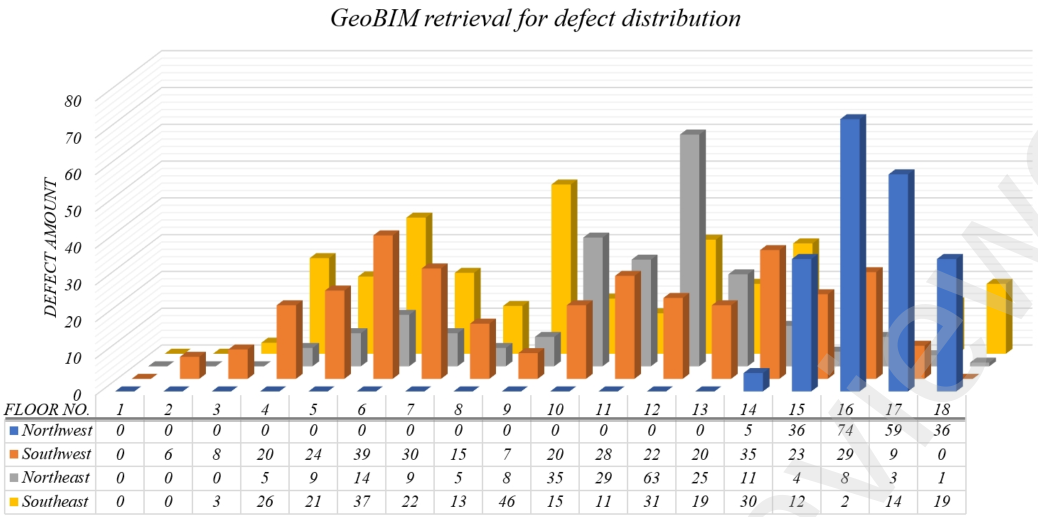


```
import unreal
import pandas as pd

defect=pd.read_excel('defects.xlsx',sheet_name=None)

new_actor_name = "Basic Wall"
metadata_key = "Element_Category"
metadata_value = "Walls"
meshes_to_join = set([])
# Iterate through the Actors in the current Level
all_actors = unreal.EditorLevelLibrary.get_all_level_actors()
for actor in all_actors:
    # Retrieve the value of this Actor's Datasmith metadata for the key set above, if any
    actor_value = unreal.DatasmithContentLibrary.get_datasmith_user_data_value_for_key(actor, metadata_key)
    # If the key exists, and its value contains a keyword set above, add the Actor to a list
    if actor_value and metadata_value in actor_value:
        pos = actor.get_actor_location()
        for df in defect.keys():
            if df.pos > pos + actor.boundary || df.pos < pos - actor.boundary:
                actor.metadata_num ++;
        
        #meshes to join.add(actor)
```

Datasmith_Uniq	59c86354f60437ada5409f85bd14ce80
Element_Area	158 m²
Element_Base_C	18/FL
Element_Base_E	0
Element_Base_is	No
Element_Base_C	0
Element_Catego	Walls
Element_Cross-S	Vertical
Element_Enable_	No
Element_Family	Basic Wall
Element_Length	37219
Element_Locatic	Wall Centerline
Element_Phase_	New Construction
Element_Phase_	None
Element_Relatec	No
Element_Room_I	Yes
Element_Structu	No
Element_Structu	Non-bearing
Element_Top_Co	Up to level: ROOF F
Element_Top_Ex	0
Element_Top_is_	No

Figure 17: Results of GeoBIM retrieval: (a) GeoBIM structure element as scene actor; (b) Scripts for GeoBIM retrieval; (c) Defect images with registered defects; (d) GeoBIM retrieval result for structure element matching.

In the realm of building maintenance, the employment of Building Maintenance Units (BMUs) is critical for executing repair operations. We have advanced this domain by introducing an algorithm that crafts the optimal trajectory for maintenance, derived from the building defect DT model delineated in Fig. 19(a), with the aim of augmenting operational efficiency. The user interface (UI) showcased in Fig. 19(b) is designed to provide unequivocal guidance to both on-site engineers and management staff, facilitating strategy planning and the quantification of benefits, while also paving the way for meticulous oversight.



FLOOR NO.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Northwest	0	0	0	0	0	0	0	0	0	0	0	0	0	5	36	74	59	36
Southwest	0	6	8	20	24	39	30	15	7	20	28	22	20	35	23	29	9	0
Northeast	0	0	0	5	9	14	9	5	8	35	29	63	25	11	4	8	3	1
Southeast	0	0	3	26	21	37	22	13	46	15	11	31	19	30	12	2	14	19

Figure 18: Defect distribution across four facades (indicated by direction) on 18 floors from the GeoBIM retrieval.

Fig. 19(c) exhibits the BMU in action, and in tandem, engineers are mandated to utilize the GPS-enabled mobile apparatus illustrated in Fig 19(d) for the acquisition of field data. This step is imperative for pinpointing forthcoming tasks and overseeing the trajectory of the maintenance project. The method under discussion thus orchestrates and assists on-site maintenance endeavors.

Embarking from UAV data acquisition, proceeding through model construction, to defect identification and cataloging, this methodology culminates in the creation of a high-precision architectural defect DT model. It renders invaluable insights for a comprehensive evaluation and rectification of structural defects. This streamlined and automated sequence of operations not only escalates the efficiency of maintenance tasks but also provides a substantive framework for the sustained supervision of building integrity.

5. Conclusions and future works

5.1. Conclusions

In conclusion, the method introduced in this study demonstrates remarkable scalability and holds substantial practical implications for the automated construction of architectural defect DTs and for guiding real-world maintenance endeavors. By utilizing high-precision, 3D global defect localization through GeoBIM registration and incorporating automated structural adaptation, this approach effectively resolves prevalent issues encountered in existing methods. Such issues include the limited scope of UAV data collection and the challenges in applying conventional techniques to all facets of a building's exterior. Our methodology significantly improves upon these limitations by facilitating precise defect mapping across the entire structure.

This novel end-to-end solution leverages the integration of BIM+GIS, not only to enhance the accuracy of defect localization but also to enable the solution's application on a urban scale for holistic management. By adopting this comprehensive approach, the methodology is capable of executing global control over extensive urban infrastructure, thereby paving the way for smart city management.

The implementation of this method allows for a sophisticated synergy between virtual models and their physical counterparts. This synergy is pivotal in enriching the DT with detailed semantic information, which in turn, refines the maintenance strategies and actions taken on the ground. In essence, it transcends the digital-physical division and aligns the DT paradigm with operational reality.

Validated in the dense urban environment of Hong Kong on a high-rise civil structure, our solution has proven its feasibility, effectiveness, and efficiency. It stands as a testament to the potential of similar large-scale assets, ushering in a new era for architectural maintenance and asset management. By fostering an environment where defects are not merely identified but are contextually understood and addressed, the proposed solution offers a significant leap forward from current practices. It marks a pivotal step towards more resilient and maintainable urban architectural landscapes.

5.2. Future works

Future works stemming from this study could pivot around several key advancements and expansions:

1) Integration with IoT devices: By connecting the digital twin with IoT devices installed on the structure, continuous real-time for sub-surface conditions could be harvested to update the digital twin, providing a dynamic and ever-evolving model that reflects current conditions.

2) Expansion to infrastructure networks: Scaling the approach to include entire infrastructure networks such as bridges, tunnels, and utility systems could provide comprehensive asset management capabilities across urban landscapes.

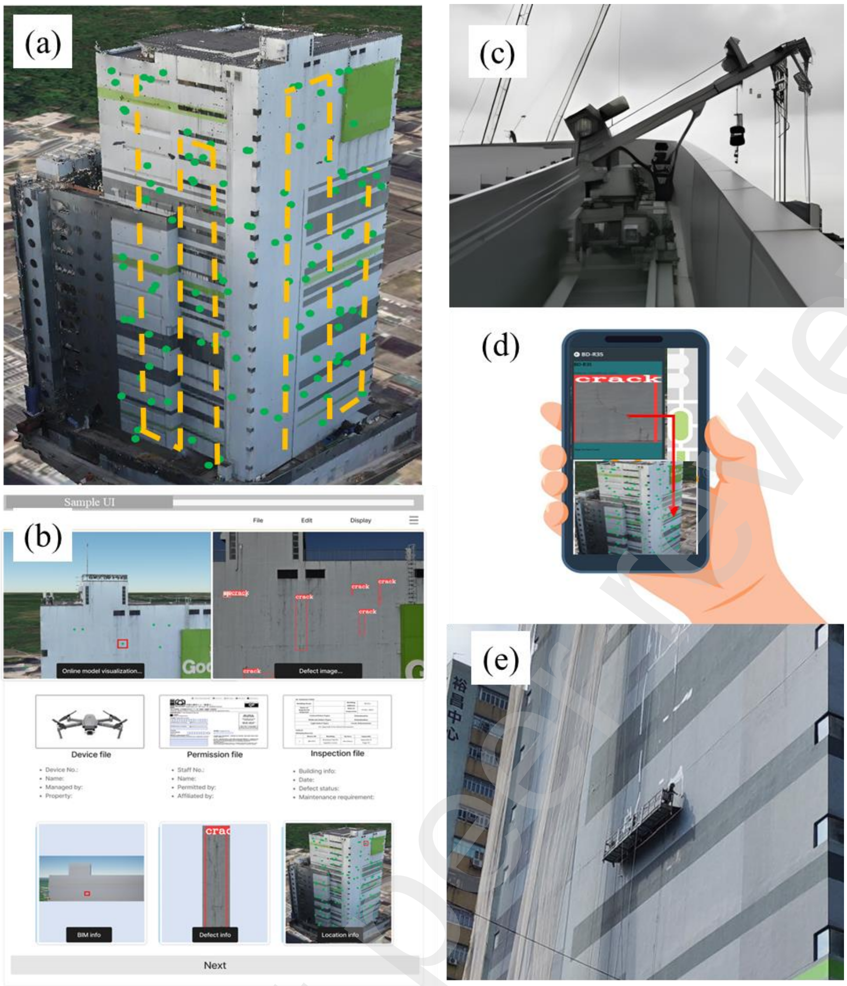


Figure 19: Target building maintenance activities guided by this method: (a) Efficient maintenance path; (b) Data assignment into the web UI; (c) Equipped building maintenance unit; (d) GPS-supported mobile device for defect information access; (e) Onsite maintenance activity.

3) Interoperability with smart city platforms: Ensuring compatibility and integration with emerging smart city platforms would enable the GeoBIM-based method to contribute to broader urban planning and management initiatives.

These potential future works would not only bolster the foundational achievements of this study but also facilitate the adoption of the GeoBIM-based DT model as an industry standard for building maintenance and urban asset management.

CRediT authorship contribution statement

Jihan Zhang: Conceptualization, Investigation, Formal analysis, Writing - Original Draft. **Benyun Zhao:** Investigation, Formal analysis, Writing - Original Draft. **Guidong Yang:** Investigation, Formal analysis, Writing - Original Draft. **Xunkuai Zhou:** Investigation, Formal analysis, Writing - Original Draft. **Yijun Huang:** Investigation, Formal analysis, Writing - Original Draft. **Chuanxiang Gao:** Investigation, Formal analysis, Writing - Original Draft. **Xi Chen:** Conceptualization, Resources, Supervision, Writing - Review & Editing, Project administration. **Ben M. Chen:** Conceptualization, Funding acquisition, Resources, Supervision, Writing - Review & Editing, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in the paper.

Acknowledgements

This work was supported in part by the InnoHK of the Government of the Hong Kong Special Administrative Region via the Hong Kong Centre for Logistics Robotics and in part by the Research Grants Council of Hong Kong SAR (Grant No: 14209020 and Grant No: 14206821).

References

[1] Michele De Filippo, Sasan Asadiabadi, JS Kuang, Dhanada K Mishra, and Harris Sun. Ai-powered inspections of facades in reinforced concrete buildings. 2023.

[2] Billie F Spencer Jr, Vedhus Hoskere, and Yasutaka Narazaki. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*, 5(2):199–222, 2019.

[3] Zhong Wang, Yulun Wu, Vicente A González, Yang Zou, Enrique del Rey Castillo, Mehrdad Arashpour, and Guillermo Cabrera-Guerrero. User-centric immersive virtual reality development framework for data visualization and decision-making in infrastructure remote inspections. *Advanced Engineering Informatics*, 57:102078, 2023.

[4] Judi E See, Colin G Drury, Ann Speed, Allison Williams, and Negar Khalandi. The role of visual inspection in the 21st century. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 61, pages 262–266. SAGE Publications Sage CA: Los Angeles, CA, 2017.

[5] Junjie Chen, Weisheng Lu, Yonglin Fu, and Zhiming Dong. Automated facility inspection using robotics and bim: A knowledge-driven approach. *Advanced Engineering Informatics*, 55:101838, 2023.

[6] Tarek Rakha and Alice Gorodetsky. Review of unmanned aerial system (uas) applications in the built environment: Towards automated building inspection procedures using drones. *Automation in Construction*, 2018.

[7] Sruthy Agnisarman, Snowil Lopes, Kapil Chalil Madathil, Kalyan Piratla, and Anand Gramopadhye. A survey of automation-enabled human-in-the-loop systems for infrastructure visual inspection. *Automation in Construction*, 2019.

[8] Jose Martinez-Carranza and Leticia Oyuki Rojas-Perez. Warehouse inspection with an autonomous micro air vehicle. *Unmanned Systems*, 10(04):329–342, 2022.

[9] Luis Duque, Junwon Seo, and James Wacker. Synthesis of unmanned aerial vehicle applications for infrastructures. *Journal of Performance of Constructed Facilities*, 32(4):04018046, 2018.

[10] Shashi Bhushan Jha and Radu F Babiceanu. Deep cnn-based visual defect detection: Survey of current literature. *Computers in Industry*, 148:103911, 2023.

[11] Dawei Li, Qian Xie, Xiaoxi Gong, Zhenghao Yu, Jinxuan Xu, Yangxing Sun, and Jun Wang. Automatic defect detection of metro tunnel surfaces using a vision-based inspection system. *Advanced Engineering Informatics*, 47:101206, 2021.

[12] Minjuan Zheng, Zhijun Lei, and Kun Zhang. Intelligent detection of building cracks based on deep learning. *Image and Vision Computing*, 103:103987, 2020.

[13] Evan McLaughlin, Nicholas Charron, and Sriram Narasimhan. Automated defect quantification in concrete bridges using robotics and deep learning. *Journal of Computing in Civil Engineering*, 34(5):04020029, 2020.

[14] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 2015.

[15] Yuanbin Wang, Minggao Liu, Pai Zheng, Huayong Yang, and Jun Zou. A smart surface inspection system using faster r-cnn in cloud-edge computing environment. *Advanced Engineering Informatics*, 43:101037, 2020.

[16] G. Jocher. YOLOv5 by Ultralytics, 5 2020.

[17] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, et al. Yolov6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*, 2022.

[18] C. Wang, A. Bochkovskiy, and H. Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022.

[19] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021.

[20] Cheng Zhang, Feng Wang, Yang Zou, Johannes Dimyadi, Brian HW Guo, and Lei Hou. Automated uav image-to-bim registration for building façade inspection using improved generalised hough transform. *Automation in Construction*, 153:104957, 2023.

[21] Masoud Mohammadi, Maria Rashidi, Yang Yu, and Bijan Samali. Integration of tls-derived bridge information modeling (brim) with a decision support system (dss) for digital twinning and asset management of bridge infrastructures. *Computers in Industry*, 147:103881, 2023.

[22] Didit Gunawan Prasetyo Jati. Uav-based photogrammetry data transformation as a building inspection tool: Applicability in mid-high-rise building. *Jurnal Teknik Sipil*, 16(2):113–121, 2021.

[23] Yangming Shi, Jing Du, and Eric Ragan. Review visual attention and spatial memory in building inspection: Toward a cognition-driven information system. *Advanced Engineering Informatics*, 44:101061, 2020.

[24] Haidar Hosamo Hosamo and Mohsen Hosamo Hosamo. Digital twin technology for bridge maintenance using 3d laser scanning: A review. *Advances in Civil Engineering*, 2022, 2022.

[25] Shihong Chen, Gao Fan, and Jun Li. Improving completeness and accuracy of 3d point clouds by using deep learning for applications of digital twins to civil structures. *Advanced Engineering Informatics*, 58:102196, 2023.

[26] Qingxiang Li, Guidong Yang, Chuanxiang Gao, Yijun Huang, Jihan Zhang, Dongyue Huang, Benyun Zhao, Xi Chen, and Ben M Chen. Single drone-based 3d reconstruction approach to improve public engagement in conservation of heritage buildings: A case of hakka tulou. *Journal of Building Engineering*, 87:108954, 2024.

[27] Isabelle Fitkau and Timo Hartmann. An ontology-based approach of automatic compliance checking for structural fire safety requirements. *Advanced Engineering Informatics*, 59:102314, 2024.

[28] Kaiwen Chen, Georg Reichard, Abiola Akanmu, and Xin Xu. Geo-registering UAV-captured close-range images to GIS-based spatial model for building façade inspections. *Automation in Construction*, 122:103503, 2021.

[29] Yi Tan, Geng Li, Ruying Cai, Jun Ma, and Mingzhu Wang. Mapping and modelling defect data from uav captured images to bim for building external wall inspection. *Automation in Construction*, 139:104284, 2022.

[30] João Alencastro, Alba Fuertes, and Pieter de Wilde. The relationship between quality defects and the thermal performance of buildings. *Renewable and Sustainable Energy Reviews*, 81:883–894, 2018.

[31] Ramiro Daniel Ballesteros Ruiz, Alberto Casado Lordsleem Jr, Joaquin Humberto Aquino Rocha, and Javier Irizarry. Unmanned aerial vehicles (uav) as a tool for visual inspection of building facades in aec+ fm industry. *Construction Innovation*, 22(4):1155–1170, 2022.

[32] Cheng Zeng, Timo Hartmann, and Leyuan Ma. Conse: An ontology for visual representation and semantic enrichment of digital images in construction sites. *Advanced Engineering Informatics*, 60:102446, 2024.

[33] Tarvo Mill, Aivars Alt, and Roode Liias. Combined 3d building surveying techniques–terrestrial laser scanning (tls) and total station surveying for bim data management purposes. *Journal of Civil Engineering and Management*, 19(sup1):S23–S32, 2013.

[34] Bryan G Pantoja-Rosero, Radhakrishna Achanta, and Katrin Beyer. Damage-augmented digital twins towards the automated inspection of buildings. *Automation in Construction*, 150:104842, 2023.

[35] Mathias Artus, MSH Alabassy, and Christian Koch. Ifc based framework for generating, modeling and visualizing spalling defect geometries. In *EG-ICE 2021 Workshop on Intelligent Computing in Engineering*, pages 176–186. Universitätsverlag der TU Berlin Berlin, Germany, 2021.

[36] Enrique Valero, Frédéric Bosché, and Alan Forster. Automatic segmentation of 3d point clouds of rubble masonry walls, and its application to building surveying, repair and maintenance. *Automation in Construction*, 96:29–39, 2018.

[37] Jakob Taraben and Guido Morgenthal. Methods for the automated assignment and comparison of building damage geometries. *Advanced Engineering Informatics*, 47:101186, 2021.

[38] Congguang Zhang, Jiangpeng Shu, Yi Shao, and Weijian Zhao. Automated generation of fe models of cracked rc beams based on 3d point

clouds and 2d images. *Journal of Civil Structural Health Monitoring*, pages 1–18, 2021.

[39] Junjie Chen, Weisheng Lu, and Jinfeng Lou. Automatic concrete defect detection and reconstruction by aligning aerial images onto semantic-rich building information model. *Computer-Aided Civil and Infrastructure Engineering*, 38(8):1079–1098, 2023.

[40] Christian Musella, Milena Serra, Costantino Menna, and Domenico Asprone. Building information modeling and artificial intelligence: Advanced technologies for the digitalisation of seismic damage in existing buildings. *Structural Concrete*, 22(5):2761–2774, 2021.

[41] Junjie Chen, Donghai Liu, Shuai Li, and Da Hu. Registering georeferenced photos to a building information model to extract structures of interest. *Advanced Engineering Informatics*, 42:100937, 2019.

[42] Donghai Liu, Junjie Chen, Dongjie Hu, and Zhao Zhang. Dynamic bim-augmented uav safety inspection for water diversion project. *Computers in Industry*, 108:163–177, 2019.

[43] Haishan Xia, Zishuo Liu, Maria Efremochkina, Xiaotong Liu, and Chunxiang Lin. Study on city digital twin technologies for sustainable smart city design: A review and bibliometric analysis of geographic information system and building information modeling integration. *Sustainable Cities and Society*, 84:104009, 2022.

[44] Kaiwen Chen, Georg Reichard, Xin Xu, and Abiola Akanmu. Gis-based information system for automated building façade assessment based on unmanned aerial vehicles and artificial intelligence. *Journal of Architectural Engineering*, 29(4):04023032, 2023.

[45] G Quinci, V Gagliardi, L Pallante, DRJ Manalo, A Napolitano, L Bertolini, L Bianchini Ciampoli, P Meriggi, F D'Amico, and F Paolacci. A novel bridge monitoring system implementing ground-based, structural and remote sensing information into a gis-based catalogue. In *Earth Resources and Environmental Remote Sensing/GIS Applications XIII*, volume 12268, pages 101–111. SPIE, 2022.

[46] Benyun Zhao, Xunkuai Zhou, Guidong Yang, Junjie Wen, Jihan Zhang, Jia Dou, Guang Li, Xi Chen, and Ben M Chen. High-resolution infrastructure defect detection dataset sourced by unmanned systems and validated with deep learning. *Automation in Construction*, 163:105405, 2024.

[47] X. Long, K. Deng, G. Wang, Y. Zhang, Q. Dang, Y. Gao, H. Shen, J. Ren, S. Han, E. Ding, et al. Pp-yolo: An effective and efficient implementation of object detector. *arXiv preprint arXiv:2007.12099*, 2020.

[48] X. Huang, X. Wang, W. Lv, X. Bai, X. Long, K. Deng, Q. Dang, S. Han, Q. Liu, X. Hu, et al. Pp-yolov2: A practical object detector. *arXiv preprint arXiv:2104.10419*, 2021.

[49] S. Xu, X. Wang, W. Lv, Q. Chang, C. Cui, K. Deng, G. Wang, Q. Dang, S. Wei, Y. Du, et al. Pp-yoloe: An evolved version of yolo. *arXiv preprint arXiv:2203.16250*, 2022.

[50] Inc Cesium GS. Cesium, the platform for 3d geospatial. https://www.cesium.com/.