Semi-Automated BIM Generation from Multi-Modal Data for Nuclear Facilities Decommissioning

Léo Sron[1,3], Caroline Chabal[2], Thibaud Durand[1], Baptiste Magnier[3], and Stefan Janaqi[3]

[1] French Alternative Energies and Atomic Energy Commission (CEA), DES, ISEC, DPME, SEIP, LNPA, University of Montpellier, Marcoule, France
{leo.sron,thibaud.durand}@cea.fr

[2] French Alternative Energies and Atomic Energy Commission (CEA), DES, ISEC, DPME, University of Montpellier, Marcoule, France
caroline.chabal@cea.fr

[3] EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, Ales, France
{baptiste.magnier, stefan.janaqi}@mines-ales.fr

Abstract. Building Information Modelling (BIM) is increasingly important for the digital documentation, analysis, and maintenance of complex industrial facilities. In the nuclear sector, and particularly in decommissioning contexts, the creation of accurate BIM models remains a challenging and time-consuming task due to the complexity of the environment, the heterogeneity of available data, and the limited accessibility of many assets. This work proposes an automated methodology for generating BIM-ready 3D representations from multi-modal data, combining point clouds, images, and geometric priors. The approach integrates multi-scale unsupervised decomposition, learning-based semantic and instance segmentation, image-to-3D projection, model retrieval, primitive fitting, parametric instantiation, and surface reconstruction for uncategorised objects. A dedicated data generation pipeline is also introduced to support training, validation, and evaluation of the proposed methods. Preliminary results, obtained in the context of our target field of application, highlight both the strengths and limitations of SAM3D-based 3D generation under varying conditions. The study discusses key practical difficulties, including data sparsity, and incomplete geometry. Overall, this work lays the groundwork for a scalable framework to support semi-automated BIM model creation in demanding industrial environments.

Keywords: Building Information Modelling · Machine Learning · 3D reconstruction

1 Introduction and Motivations

Building Information Modelling (BIM) has emerged as a transformative technology in the management and operation of industrial facilities. By creating detailed digital representations of physical assets, BIM enables integrated design, analysis, maintenance, and lifecycle management. Its adoption offers significant benefits, including improved safety, reduced operational costs, and enhanced compliance with regulatory standards. In the nuclear sector, where facility decommissioning poses unique technical, safety, and regulatory challenges, BIM has the potential to provide critical support for planning, documentation, and risk assessment. Despite its promise, the generation of accurate BIM models in nuclear and other complex industrial environments remains a major challenge. Nuclear facilities are often characterized by highly intricate layouts, with restricted access due to safety or contamination concerns. This results in rare, incomplete, and sparse datasets, compromising open research for this application. Traditional BIM workflows are heavily relying on manual input from domain experts or the availability of descriptive documents, making the modelling process time-consuming, labor-intensive, and prone to human error. In addition, the heterogeneity of available data—including point clouds, photographs, CAD drawings, and schematic diagrams—complicates integration and often requires specialized preprocessing. Recent advances in digital reconstruction and computer vision have opened new possibilities for semi-automated or fully automated BIM model creation. Techniques such as point cloud segmentation, primitive fitting, and model generation have shown potential in industrial applications. However, most existing approaches are limited to specific object types or rely on high-quality, dense data. In nuclear decommissioning contexts, where data is often noisy, incomplete, or unavailable for certain areas, these methods struggle to produce reliable BIM-ready representations. The motivation for this work stems from the need for an automated framework that can leverage multi-modal data to create accurate and semantically meaningful BIM models under challenging conditions. By integrating point clouds, images, and geometric priors, our proposed methodology aims to reduce manual effort while maintaining accuracy and adaptability across a variety of industrial environments. Key components of the proposed methodology include multi-scale unsupervised decomposition, learning-based semantic and instance segmentation, model retrieval, image-to-3D projection, primitive fitting, parametric instantiation, and surface reconstruction for uncategorized objects. To support these methods, we also introduce a dedicated data generation pipeline for training, validation, and evaluation. Because the overall pipeline is a work in progress, preliminary results in nuclear facility contexts focus on the strengths and limitations of the SAM3D model to discuss its potential use in this field of application.

2 Background and Related Work

2.1 Building Information Modelling (BIM)

Building Information Modelling (BIM) is a methodology designed to coordinate the entire lifecycle of a project bridging stakeholders from the architecture, engineering, construction, and operation (AECO) industry. It comprises a digital representation of the physical and functional characteristics of a building or any physical infrastructure and its components. It serves as a shared knowledge

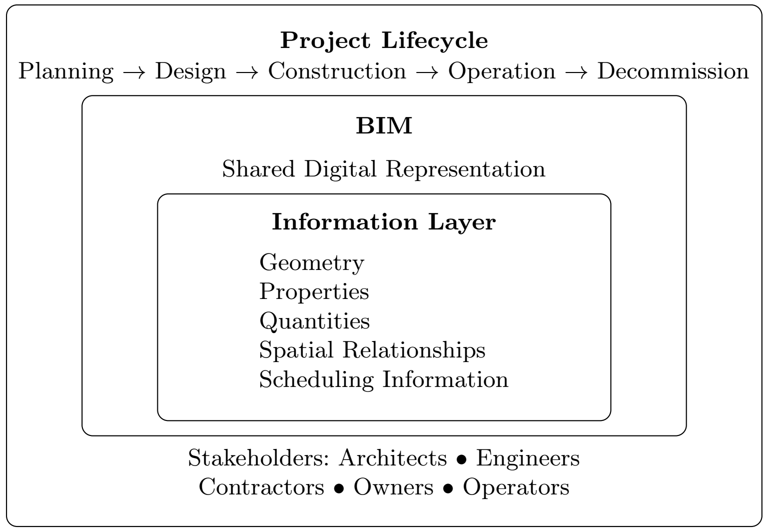


Fig. 1: Role of BIM in managing geometric, semantic, spatial, quantity, and scheduling information across the project lifecycle

resource for information about a facility, assisting users with decision-making throughout the building lifecycle, from planning, conception, to decommissioning (Fig. 1). BIM integrates various data and information about the building project, including the plan, geometry, spatial relationships, geographic information, quantities, and properties of the building components. The methods for creating a BIM model differ depending on the stage of the project in which the BIM ecosystem is introduced. In cases where the BIM is developed as the project progresses, the BIM 3D model is manually created by CAD (Computer Aided Design) engineers, and other information such as water circulation map or electricity circuits are added by field specialists and formally integrated by BIM engineers. For already existing buildings, as-built BIM models are created from scanned data, usually from LiDAR scanners that produce coloured point clouds helping CAD engineers to create actual 3D models. This process, called Scan-to-BIM in the literature, will be the focus of our research.

2.2 Nuclear Industry and decommissioning

BIM is gradually being adopted in the nuclear sector as a practical tool for handling the complexity and volume of information associated with these facilities over their entire lifecycle. Nuclear infrastructures operate under strict safety constraints, involve long service periods, and are subject to rigorous regulatory oversight, which makes reliable data management and coordination essential. In this context, BIM provides a structured digital environment where engineering, operational, and safety-related information can be brought together and maintained consistently [8]. This helps improve communication between stakeholders, supports more accurate design and verification processes, and facilitates maintenance and asset management activities over time. In new nuclear projects, it also contributes to anticipating design issues, improving construction planning, and ensuring alignment with regulatory requirements from the early stages. In the context of nuclear decommissioning, BIM can play an important role in addressing the specific challenges associated with the dismantling of complex and often aging facilities. These projects rely heavily on a detailed understanding of existing structures, despite documentation that is frequently incomplete, outdated, or inconsistent. BIM makes it possible to organize and consolidate various types of data—such as historical plans, material inventories, and radiological information—into a coherent digital model. This supports the preparation and simulation of dismantling operations, helping teams plan interventions more effectively and anticipate potential constraints. It can also contribute to improving safety by enabling better visualization of risk areas and reducing unnecessary exposure for workers. When combined with technologies such as 3D laser scanning or virtual reality [6], BIM further enhances the capabilities of analysis and training. However, its use in decommissioning still requires overcoming several limitations, including data sensitivity, interoperability between tools, and integration of legacy information.

2.3 Automation in BIM Model Creation

Automation in BIM encompasses a broad spectrum of techniques, ranging from rule-based parametric generation to data-driven artificial intelligence approaches. These methods differ in their underlying paradigm, input data, computational requirements, and level of autonomy.

Creating BIM Models from 3D Point Clouds The main source of information when creating a BIM model of an existing facility is the data measured and captured with LiDAR (Light Detection and Ranging) scanners. In addition, existing plans, inventories, and other documentation are used to create different parts of the BIM model. The first step toward generating BIM models from raw point clouds is segmentation: the partitioning of an unstructured point set into groups that are geometrically coherent and/or semantically meaningful. Robust segmentation is critical for downstream steps such as primitive fitting and surface reconstruction, since errors in segmentation propagate to model-fitting and semantic labelling. Geometric methods group points using low-level geometric cues (proximity, normal orientation, curvature, planarity) and model-fitting techniques. Classical tools include region-growing, RANSAC-based primitive detection [10,19](planes, cylinders), Euclidean clustering such as K-means algorithm [15,11], and octree-based partitioning.

Surface Reconstruction Techniques Classic surface reconstruction algorithms can be devided into 2 categories : explicit and implicit methods. Explicit methods like Delaunay triangulation use different strategies to link points to create triangle faces.Implicit methods, like poisson reconstruction, use normals to model isosurfaces that represent the surface shape of the objects. Recent research is now focusing on using machine learning models to generate surfaces with point clouds as input.

3 Proposed methodology

This section describes a multimodal pipeline to convert LiDAR and photographic acquisitions into a semantically rich BIM. Input to the pipeline are *(i)* a colored 3D point cloud produced by co-registered LiDAR scans and *(ii)* panoramic or multi-view photographs whose poses relative to the point cloud are known. The pipeline addresses the problem based on scale and complexity. It begins by extracting easily recognizable geometric structures, followed by clustering the unstructured regions. Subsequently, it applies learned segmentation and image-based segmentation in parallel. Finally, the results are fused to generate fitted 3D models through surface reconstruction and learning-based generation. The overall pipeline is displayed in Figure 2.

3.1 Preprocessing and data structures

Many architectural environments contain dominant geometric structures such as floors, ceilings, and walls that can be robustly identified using geometric properties alone. Detecting these elements early in the pipeline reduces the complexity of subsequent processing stages. To this end, robust model fitting techniques are applied to the point cloud in order to detect large planar surfaces. Algorithms such as RANSAC can be used to identify planar regions corresponding to floors, walls, and other structural components. By computing a direct parameterization, these algorithms simplify the point cloud for later stages, enabling efficient storage of structural elements as mathematical primitives and their direct integration into the reconstructed model. Removing or isolating these easily identifiable structures simplifies the remaining point cloud and allows subsequent analysis to focus on more complex objects present in the scene.

3.2 Multi-scale unsupervised decomposition

Residual points, after removing large structural primitives, are partitioned via density- and feature-aware clustering to produce object candidates. A density-based clustering algorithm like DBSCAN [9] yields an over-segmentation that respects variable point density and occlusions. Additionally, clustering can be performed hierarchically to produce a multi-scale set of proposals with methods like HDBSCAN [14]. When useful, local geometric descriptors (e.g. normals, curvature) and appearance features (like colours) are concatenated with spatial coordinates to form a richer clustering space, improving the separation of proximate but geometrically different objects. Each cluster is annotated with simple geometric statistics (axis-aligned bounding boxes, point count, rotation matrix,

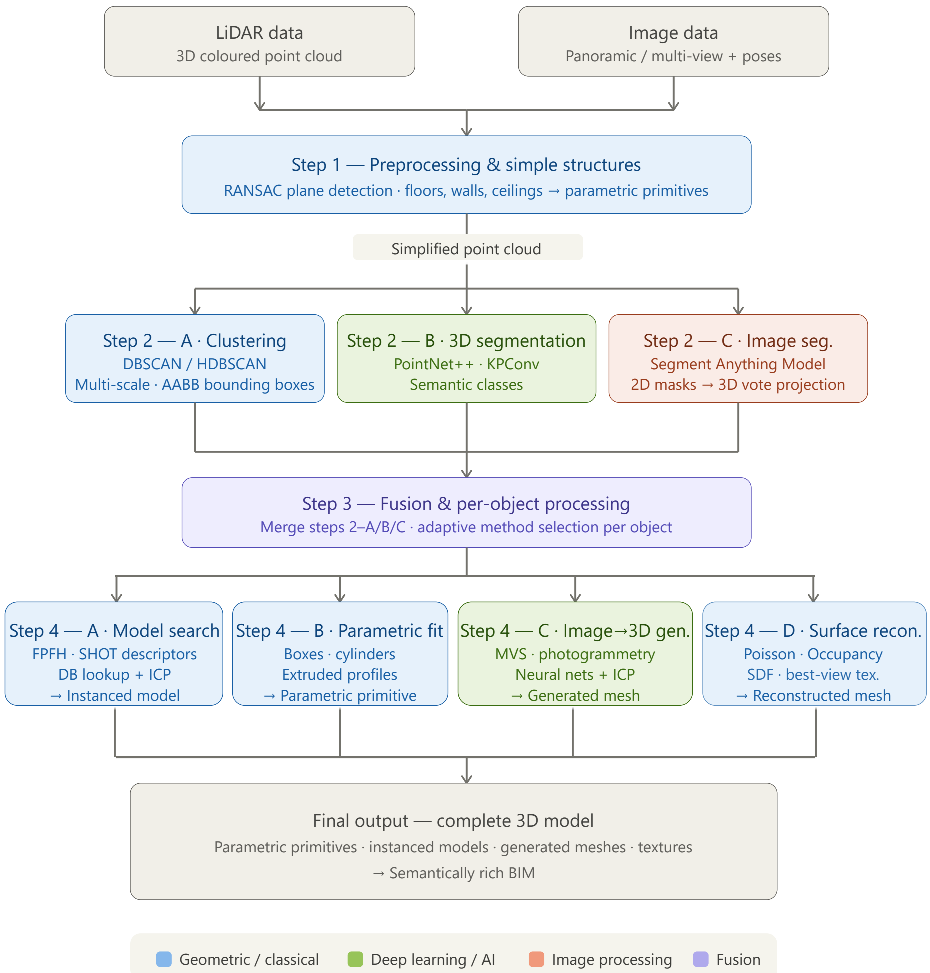


Fig. 2: Overview of the complete pipeline for processing LiDAR data and images. Each stage illustrates the possible techniques employed, from initial data separation to segmentation of both point cloud and images, alignment, and fusion.

translation vector) that support template retrieval, retrieval-based matching, and confidence scoring for following strategies and evaluation.

3.3 Learning-based 3D semantic and instance segmentation

To complement geometry-driven grouping, supervised point-cloud networks predict semantic labels and instance masks directly on the point set. Architectures that capture local and hierarchical geometry—such as point-based networks extending PointNet [16,17], hierarchical point encoders [23] or kernel-based convolutions [20]—have shown strong performance on both indoor and outdoor benchmarks. These models take as input point coordinates plus optionally normals, intensity, and RGB channels, and output per-point class probabilities and instance embeddings.

3.4 Image segmentation and projection to 3D

As part of our multi-modal 3D reconstruction pipeline, we utilize advanced image segmentation techniques on panoramic and multi-view images to extract semantic and instance masks. The Segment Anything family [12,18] and related promptable segmentation systems provide strong zero-shot masks usable in new environments. Each image mask is projected into 3D using known camera poses: pixel rays are intersected with the point cloud and per-point label votes are accumulated across views, storing per-label confidence scores for each 3D point.

Let a 3D point in the world coordinate system be defined as:

$$\mathbf{P}_w = (X_w, Y_w, Z_w)^T. \tag{1}$$

Each camera (LiDAR station) is parameterized by:

- a rotation matrix $R \in \mathbb{R}^{3\times3}$,
- a translation vector $t \in \mathbb{R}^3$.

The transformation from world to camera coordinates is given by:

$$\mathbf{P}_c = (X_c, Y_c, Z_c)^T, \quad \text{with} \quad \mathbf{P}_c = R\mathbf{P}_w + t. \tag{2}$$

The inverse transformation is:

$$\mathbf{P}_w = R^T(\mathbf{P}_c - \mathbf{t}) \tag{3}$$

since $R^T = R^{-1}$ for a rotation matrix.

The rotation matrix R aligns the world coordinate system with the camera frame, while the translation vector t represents the position of the camera in the world. The intrinsic camera matrix K is defined as:

$$K = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{4}$$

where:

- f_x and f_y are the focal lengths in pixels in axes x and y.
- u_0 and v_0 are the coordinates of the principal point of the camera.

We define a virtual image plane in 3D space in the camera coordinate system. A pixel (u, v) is mapped to a 3D point on this plane as:

$$\mathbf{P}_c^{img} = K^{-1}(u, v, 1)^T. \tag{5}$$

This point is transformed into world coordinates:

$$\mathbf{P}_w^{img} = R^T(\mathbf{P}_c^{img} - t). \tag{6}$$

We then define a ray originating from this point toward the direction d_w:

$$r(\lambda) = \mathbf{P}_w^{img} + \lambda d_w, \quad d_w = R^T K^{-1}(u, v, 1)^T, \quad \lambda > 0. \tag{7}$$

Because of the sparsity of a point cloud, the trajectory of the ray projected from the mask could potentially miss a foreground point and hit a point further from the camera position or even never hit a point at all. To tackle this, we search for points within a fixed radius around the ray. Given a set of calibrated cameras, each providing a 2D semantic segmentation mask projected on the 3D point cloud, we assign a class label to each 3D point. Let $M_k(u, v)$ be the segmentation mask of camera k, where each pixel stores a discrete class label in $\{0, \dots, C-1\}$. The observed class for point i in camera k is:

$$c_{i,k} = \begin{cases} M_k(u_{i,k}, v_{i,k}) & \text{if } \mathbf{P}_i \text{ is visible in camera } k, \\ -1 & \text{otherwise.} \end{cases} \tag{8}$$

Each camera provides a noisy observation of the true class of a 3D point. We aggregate these observations using a voting scheme. For each point i and class c, we define the vote count:

$$V_i(c) = \sum_{k=1}^{K} \mathbf{1}(c_{i,k} = c), \tag{9}$$

where $\mathbf{1}(\cdot)$ is the indicator function.

We convert vote counts into class probabilities via normalization:

$$P_i(c) = \begin{cases} \dfrac{V_i(c)}{\sum_{c'=0}^{C-1} V_i(c')} & \text{if } \sum_{c'=0}^{C-1} V_i(c') > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

This yields an empirical estimate of the posterior distribution:

$$P_i(c) \approx P(c \mid \mathbf{P}_i). \tag{11}$$

The predicted class for each point is obtained via Maximum A Posteriori (MAP):

$$\hat{c}_i = \arg\max_c P_i(c). \tag{12}$$

Points with no observations (i.e., $\sum_c V_i(c) = 0$) are assigned as $\hat{c}_i = -1$. This process allows us to tackle the lack of annotated data for semantic and instance segmentation on 3D point clouds of industrial environments by assisting the labelling process.

3.5 Model retrieval, primitive fitting and parametric instantiation

For instances that match known object categories, the pipeline searches a local model database. Retrieval uses geometric descriptors computed from clusters (normalized aspect ratios, PCA signatures, histogram of local geometric features such as FPFH/SHOT, and learned embeddings from a retrieval network). Candidate templates are ranked by descriptor similarity and scaled and/or aligned to the cluster using centroid and plprincipal-axis alignment, followed by point-to-plane iterative closest point (ICP [4,13]) refinement to accommodate partial observations. For objects well-approximated by simple primitives, parametric fitting (such as boxes, cylinders, extruded profiles) is performed with nearest neighbors algorithm. When template fitting residuals exceed thresholds, the instance is flagged for non-template reconstruction.

3.6 3D model generation from image and alignment

When the LiDAR returns are sparse or when fine detail is required, image-conditioned 3D generation is used as a complementary reconstruction path. For segmented image regions, candidate 3D meshes can be generated using single-view or multi-view approaches, including photogrammetry, multi-view stereo (MVS) pipelines, or recent image-conditioned mesh generation methods. Generative models such as SAM3D [7] can generate 3D scenes from 2D single-view image segmentation masks. The aforementioned model demonstrated promising performance on a standard image dataset; therefore, its potential application to the nuclear industry is discussed in Section 3.8. Generated meshes are coarsely placed by matching projected silhouettes obtained with the image segmentation and then refined with ICP against available point returns or with photometric alignment when dense multi-view imagery is present.

3.7 Surface reconstruction for uncategorised objects

For instances without good template matches or reliable image-generation results, explicit surface reconstruction is applied on the high-resolution point set. Classical volumetric or Poisson surface reconstruction techniques produce watertight meshes when coverage is sufficient; when coverage is partial or irregular, neural implicit representations (occupancy/SDF networks) can be fitted to points and optionally supervised with multi-view photometric constraints to improve completeness. Reconstructed meshes receive texture mapping by selecting the best-view images per surface patch and blending using multi-band or seam-aware blending heuristics.

3.8 Preliminary Results

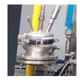
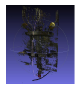
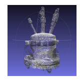
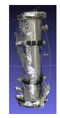
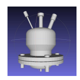
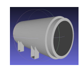
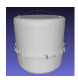
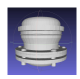
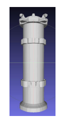
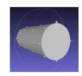
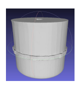
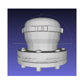
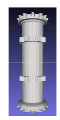
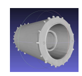
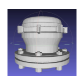
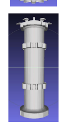
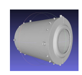
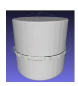
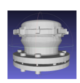
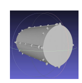
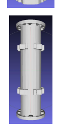
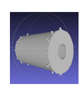
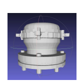
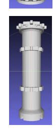
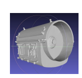
Although the proposed methodology describes a set of techniques considered in this study, the preliminary results presented here focus on the application of the SAM3D model. This choice was motivated by its promising results on more usual field of application. The experiments were conducted on photographs of ELVIS (Fig. 3), a real-life model of a vitrification process [5], composed of components that are likely absent from any public dataset in which SAM3D was trained. The segmentation of the images was done using SAM2 [18].

(a) (b) (c)

Fig. 3: Photos of ELVIS from different point of view: (a) gives a lateral view from behind the calciner, (b) gives a front view, and (c) was taken from the same position as (b) but slightly rotated to the left.

SAM3D has a seed parameter that allows the user to fix the output response of the model for a given input mask and image. For the purpose of evaluating the model, we will show the model response for different random seed values. If our solution were to be used in a production ready application, the seed would need to be fixed as it would not be allowed to let the results of our pipeline at random. Table 1 presents the results for various objects in ELVIS for a variety of seed values compared to the point cloud obtained from a LiDAR scan campaign and the respective 3D mesh models that we use as a reference. Although the 3D reference meshes were created manually by a CAD designer with point cloud and photographs as reference, they lack some visual details. Despite this limitation, the reference meshes provide a reliable structural ground truth for assessing overall shape consistency and geometric plausibility. Across different seed values, SAM3D demonstrates some major variations in local geometric details, while maintaining the global structure relatively stable. This behaviour indicates that the model is primarily constrained by the input mask and image content. The key advantage of SAM3D over conventional approaches lies in its ability to leverage information from the input image to capture fine structural details of the target object. Results on the mist eliminator and the calciner highlight this behaviour the most. However, these finer details seem to lack consistency and precision, particularly visible by the inconsistent number of bolts reconstructed.

Table 1: Visual results of generated 3D models across various seed values.

	solution feed tank	mist eliminator	washing column	calciner
True photo				
Point cloud				
Mesh ref				
Seed 0				
Seed 33				
Seed 42				
Seed 69				
Seed 70				
Seed 99				

3.9 Expected difficulties and limitations

Our proposed methodology is designed to adapt to a variety of industrial environments while producing consistent and reliable 3D models. Despite these efforts, several challenges and uncertainties remain, and the use of SAM3D introduces additional considerations. As a generative model, SAM3D inherently introduces uncertainty in the objects it produces, and its performance is strongly tied to the quality of the input image segmentation. More specifically, over-segmentation caused by texture heterogeneity or occlusion effects can significantly degrade its performance, as it may lead to an excessive number of generated object proposals. This can make the resulting outputs difficult to filter, interpret, and effectively use in downstream tasks without thorough manual inspection.

Fig. 4 illustrates the variation between each seeds for a certain input. This lack of consistency highlights a fundamental reliability issue: the generated models can contain structural details that are inconsistent or physically implausible. In industrial and nuclear contexts, where accurate and trustworthy representations of equipment are essential for planning, maintenance, and safety, such errors can significantly reduce the practical utility of BIM models. While SAM3D can capture fine-grained details, its outputs require careful validation and correction to ensure reliability in high-stakes applications.

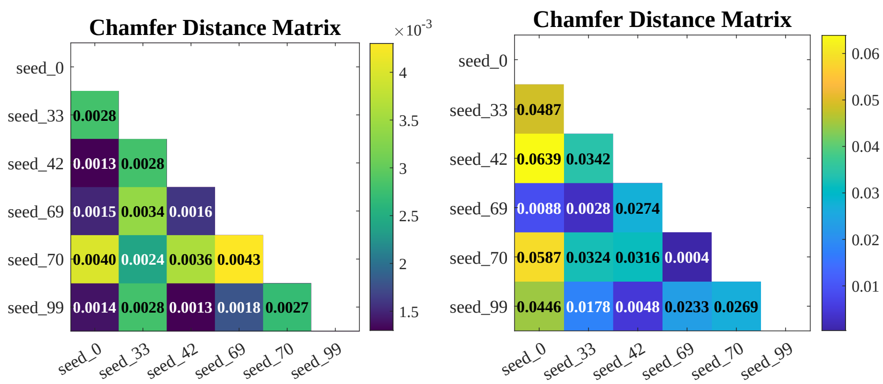


Fig. 4: Pairwise Chamfer distance [3] matrix between seeds for one object generation. The left matrix is associated with the solution feed tank generation; the right matrix is associated with the calciner.

3.10 Data Generation Pipeline

Our pipeline requires a detailed understanding of the environment and data. To prepare for such a task, precise labelled data of our target area of application are mandatory. To this end, we created a pipeline for data generation for experimental use with multimodality in mind. The goal is to reproduce the BIM data generation process applied to varying levels of environment complexity. In that sense, our goal is to create a database combining real world data with simulated data for each experimental environment. Simulated data are acquired using an in-house developed simulator for 3D laser scanning using Unity 3D engine. This simulator enables the possibility of rapidly producing data with different scanning parameters such as the density of points or the introduced noise. The parameters of the LiDAR scanner in the simulator are the horizontal resolution and the vertical resolution (in degrees), noise (in mm), and maximum distance (in m). The resolution of measurement from our LiDAR scanner is measured by the distance between two points at a distance of 10 meters, generally expressed in millimetres (mm). Our simulator also allows us to create panoramic images similar to real laser scanners. The camera simulator can be parametrized by choosing the height and the width of the image, the step in which the panoramic view is divided, and the vertical field of view $\text{fov}_{vertical}$. We can calculate the horizontal field of view $\text{fov}_{horizontal}$ by:

$$\text{fov}_{horizontal} = 2 \arctan\left(\tan\left(\frac{\text{fov}_{vertical}}{2}\right)\frac{width}{height}\right). \tag{13}$$

Given the $\text{fov}_{horizontal}$ and the $\text{fov}_{vertical}$, we can estimate the focal length in pixel f_x and f_y of the virtual camera by:

$$\begin{cases} f_x = \dfrac{width/2}{\tan(\text{fov}_{horizontal}/2)}, \\ f_y = \dfrac{height/2}{\tan(\text{fov}_{vertical}/2)}. \end{cases} \tag{14}$$

By using this connection between real measurements and simulated measurements, we hope to rigorously understand the limit condition in which our working pipeline will be successful and to determine the optimal input data condition for our methodology.

To tackle the lack of annotated data representing industrial and nuclear facilities and equipments, a semi-automatic annotation technique based on image segmentation and mask projection on the point cloud is also proposed. Currently available dataset representing industrial parts are few and lacking (displayed in Table 2). The work from F. Armangeon et al. on IRIS collection of industrial

Table 2: **Dataset comparison**. Empty cells mean the information aren't stated in their respective original paper.

Dataset	CLOI [1]	PSNet5 [22]	Pipework [21]	IRIS-v2 [2]
#Classes	10	5	17	32
#Elements	12,125		4,647	
#Environments	5	4		1
Scanner		TLS (Leica BLK360)	TLS (Leica Geosystems)	360° camera (Xphase Pro X2)
Precision		4mm@10m	12.5mm@10m	
Class examples	Pipe, conduit, ibeam	Pipe, tank, pump	Valve, elbow, tee	

data [2] is similar to our proposed strategy for a semi-automatic point cloud annotation process based on image segmentation, showing its capabilities for the creation of an annotated dataset. These datasets constitute valuable contributions to the study of industrial facilities, yet the data itself suffer from several limitations: for instance, PSNet5 contains a limited number of classes, and the associated point clouds are affected by occlusion and incomplete coverage due to the nature of the scanned environments.

4 Conclusion

This work has presented a semi-automated methodology for generating BIM-ready 3D models from multi-modal data in complex industrial environments, with a particular focus on nuclear facility decommissioning. By integrating point clouds and images, the proposed approach addresses the challenges of data sparsity, incomplete geometries, and heterogeneous information sources. Our preliminary results highlight important information about the possible use of out-of-the-box image-to-3D model, in particular SAM3D, for the generation of domain specific components. The lack of consistency and reliability in the geometry details is a critical drawback. Future work will focus on integrating and evaluating our pipeline on both simulated and real-world data to assess its robustness, scalability, and practical applicability in complex industrial environments.

Acknowledgements The authors would like to thank CEA ISEC for providing access to facilities and working environment.

References

1. Agapaki, E., Glyn-Davies, A., Mandoki, S., Brilakis, I.: Cloi: A shape classification benchmark dataset for industrial facilities. In: ASCE International Conference on Computing in Civil Engineering 2019. pp. 66–73. American Society of Civil Engineers Reston, VA (2019)
2. Armangeon, F., Ehret, T., Meinhardt-Llopis, E., von Gioi, R.G., Thibault, G., Petit, M., Facciolo, G.: An industrial dataset for scene acquisitions and functional schematics alignment. arXiv preprint arXiv:2602.15584 (2026)
3. Barrow, H.G., Tenenbaum, J.M., Bolles, R.C., Wolf, H.C.: Parametric correspondence and chamfer matching: Two new techniques for image matching. Tech. rep. (1977)
4. Besl, P.J., McKay, N.D.: Method for registration of 3-d shapes. In: Sensor fusion IV: control paradigms and data structures. vol. 1611, pp. 586–606. Spie (1992)
5. Chabal, C., Durand, T., Belaazri, H., Delrieu, J., Gauthier, M., Ledoux, A., Père, C.: Developing digital twins to assist in the control of nuclear complex processes. In: Novel & Intelligent Digital Systems Conferences. pp. 611–619. Springer (2024)
6. Chabal, C., Soulabaille, Y.: Virtual reality: a way to prepare and optimize operations in decommissioning projects. In: International Symposium on PREparation for DECommissioning (2016)
7. Chen, X., Chu, F.J., Gleize, P., Liang, K.J., Sax, A., Tang, H., Wang, W., Guo, M., Hardin, T., Li, X., et al.: Sam 3d: 3dfy anything in images. arXiv preprint arXiv:2511.16624 (2025)

8. Depauw, D., Francois, P., Pultier, M., Chabal, C., Perrot, V., Soulabaille, Y., Ardelier, L., Testard, V., et al.: 3d digital simulation tools used to support nuclear facilities decommissioning-benefits for the projects progress and their safety-000033. Tech. rep., Societe Francaise d'Energie Nucleaire-SFEN, 103 rue Reaumur, 75002 Paris ... (2018)
9. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: kdd. vol. 96, pp. 226–231 (1996)
10. 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)
11. Jain, A.K.: Data clustering: 50 years beyond k-means. Pattern recognition letters **31**(8), 651–666 (2010)
12. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
13. Low, K.L.: Linear least-squares optimization for point-to-plane icp surface registration. Chapel Hill, University of North Carolina **4**(10), 1–3 (2004)
14. McInnes, L., Healy, J., Astels, S., et al.: hdbscan: Hierarchical density based clustering. J. Open Source Softw. **2**(11), 205 (2017)
15. McQueen, J.B.: Some methods of classification and analysis of multivariate observations. In: Proc. of 5th Berkeley Symposium on Math. Stat. and Prob. pp. 281–297 (1967)
16. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660 (2017)
17. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems **30** (2017)
18. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: Sam 2: Segment anything in images and videos. In: International Conference on Learning Representations. vol. 2025, pp. 28085–28128 (2025)
19. Schnabel, R., Wahl, R., Klein, R.: Efficient ransac for point-cloud shape detection. In: Computer graphics forum. vol. 26, pp. 214–226. Wiley Online Library (2007)
20. Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: Kpconv: Flexible and deformable convolution for point clouds. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6411–6420 (2019)
21. Yeo, C., Kim, S., Kim, H., Kim, S., Mun, D.: Deep learning applications in an industrial process plant: repository of segmented point clouds for pipework components. JMST Advances **2**(1), 15–24 (2020)
22. Yin, C., Wang, B., Gan, V.J., Wang, M., Cheng, J.C.: Automated semantic segmentation of industrial point clouds using respointnet++. Automation in Construction **130**, 103874 (2021)
23. Zhao, H., Jiang, L., Jia, J., Torr, P.H., Koltun, V.: Point transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 16259–16268 (2021)