Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/NA10.20250319
Published: June 17, 2025

AI-Powered Retrofitting for Smart Reverse Engineering: An Approach of NeRF-Based Multi-object Recognition in Complex Ambience with Noise

Chenxi Tao
Department of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA, USA
ctao40@gatech.edu

Dongin Shin, Youngouk Kim
Korea Electronics Technology Institute
Seongnam-si, Gyeonggi-do, Korea

Hong-in Won
Korea Institute of Industrial Technology
Cheonan-si, Chungcheongnam-do, Korea

Seung-Kyum Choi and Roger Jiao
Department of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA, USA

Abstract

In the context of smart reverse engineering and robotic warehousing, efficient and accurate 3D model generation remains a persistent challenge, especially in cluttered and noisy environments. This paper presents an AI-powered retrofitting framework that reconstructs 3D digital twins from monocular video using Neural Radiance Fields (NeRF), offering a cost-effective alternative to traditional hardware-intensive 3D scanning. The proposed pipeline integrates NeRF-based volumetric reconstruction with HSV-based background filtering, DBSCAN clustering, and KD-tree-based spatial recovery to recognize multiple objects from complex, real-world scenes. Demonstrated through an automotive case study, this approach enables component-level recognition and digital modeling without the need for physical tags or high-end scanners. The resulting point clouds are compact, pose-consistent, and directly applicable to downstream tasks such as CAD alignment and robotic manipulation. By addressing key limitations of conventional reverse engineering methods, this work offers a robust and scalable solution for 3D modeling under real-world industrial constraints.

Keywords

Smart Reverse Engineering, AI-Powered Retrofitting, Neural Radiance Fields (NeRF), 3D Object Subtraction, Multi-Object Recognition.

1. Introduction

Reverse engineering plays a critical role in industrial design and manufacturing, particularly in sectors such as automotive engineering, where digital models must frequently be recreated from existing physical components. Conventional methods typically rely on structured light systems, laser scanning, or LiDAR to capture object geometries. While these approaches provide high accuracy, they are often cost-prohibitive, labor-intensive, and susceptible to performance degradation in cluttered or noisy environments.

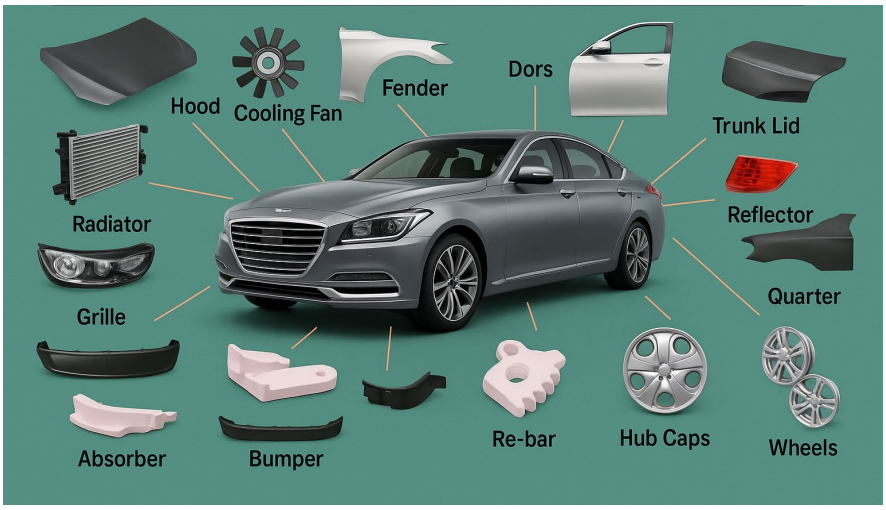

Figure 1. Illustration of missing CAD files for automotive components

Figure 1 illustrates a representative scenario in which CAD files for essential automotive components are unavailable. This poses significant challenges for tasks such as repair, inspection, and robotic manipulation. In such contexts, reverse engineering becomes indispensable. However, the limitations of traditional techniques, including their dependency on controlled capture conditions and expensive equipment, restrict their applicability in real-world, dynamic settings.

Retrofitting, in this context, refers to the enhancement of existing physical systems through digital augmentation, allowing them to be incorporated into modern data-driven workflows without requiring complete redesign or replacement. As the industrial sector increasingly seeks to extend the functionality of legacy assets, retrofitting offers a practical and scalable solution. Recent advancements in artificial intelligence and neural rendering have enabled a new class of retrofitting techniques that bypass the need for specialized hardware. By utilizing visual data from monocular video, AI-powered retrofitting methods can reconstruct digital representations of physical objects, thereby facilitating tasks such as geometric inspection, robotic planning, and CAD integration.

This paper introduces an AI-powered retrofitting framework for smart reverse engineering that reconstructs 3D digital twins from monocular video input. The core of the system is based on Neural Radiance Fields (NeRF), which synthesizes high-fidelity volumetric scenes from sparse 2D images. To address the challenges associated with complex, real-world environments, the framework integrates a post-processing pipeline comprising HSV-based background filtering, density-based spatial clustering (DBSCAN), and spatial recovery using KD-tree proximity analysis. This pipeline yields compact and pose-consistent 3D point clouds that are immediately usable for CAD alignment, digital archiving, or robotic manipulation.

The proposed approach is validated through an automotive case study, in which component-level digital models are reconstructed using only handheld video captured under minimal constraints. By eliminating the need for high-end scanning equipment and manual post-processing, this method offers a scalable and cost-effective alternative for industrial 3D modeling and highlights a transformative direction for reverse engineering practices under real-world operational conditions

2. Related Work

2.1 Retrofitting and Smart Reverse Engineering

Traditional retrofitting in reverse engineering often involves 3D scanning and CAD conversion using structured light or LiDAR. While accurate, these methods are typically costly, static, and impractical in cluttered or unstructured environments. AI-powered retrofitting offers a shift by leveraging monocular video and neural rendering to reconstruct object geometry without specialized hardware. Recent studies (Šlapak et al., 2023) have demonstrated NeRF-based retrofitting in industrial robotics, while MDPI Sensors (2023) compare NeRF against photogrammetry to highlight its cost-effectiveness and adaptability. Such methods align with broader trends in data-driven reverse engineering, particularly for automotive and manufacturing domains.

2.2 3D Reconstruction Techniques

Neural Radiance Fields (NeRF), introduced by Mildenhall et al. (2020), model scenes as continuous volumetric functions by mapping 3D coordinates and camera directions to color and opacity. This has enabled photorealistic view synthesis and dense 3D reconstruction from sparse image input. Extensions such as NeRF++ (Chen et al. 2021), Mip-NeRF (Barron et al. 2021), and Point-NeRF (Wang et al. 2021) address various limitations including unbounded scenes, anti-aliasing, and geometry extraction. SA3D (Jumpat et al., 2023) and Open-NeRF (Zhang et al.2023) further integrate subtraction and open vocabulary decomposition. These advancements have enhanced NeRF's suitability for real-world applications including robotics and digital twin creation.

2.3 3D Scene Background Subtraction

Robust subtraction is essential for isolating object-level representations from reconstructed scenes. Classical methods like DBSCAN (Ester et al. 1996) and HSV filtering remain lightweight and effective in real-time or resource-constrained scenarios. Toolkits like the Point Cloud Library (Rusu and Cousins 2011) provide efficient implementations of these traditional clustering and filtering algorithms, making them suitable for integration with NeRF post-processing. Meanwhile, PointNet++ (Qi et al.2017) and fusion-aware convolution networks (Zhang et al. 2020) have demonstrated powerful deep learning approaches for point cloud subtraction. For NeRF-generated scenes, post-hoc subtraction using clustering and color thresholding remains a practical solution, especially when computational cost and generalizability are priorities. Our work builds on this hybrid subtraction paradigm by combining traditional techniques with NeRF-rendered 3D models. Recent methods such as InsegNeRF (Benaim et al., 2022) have explored direct instance subtraction within NeRFs, though they typically require annotated multi-view supervision and high computational overhead.

3. System Architecture for AI-Powered Retrofitting and Subtraction

This section presents the architecture and internal logic of the proposed 3D object subtraction system. We first introduce a high-level pipeline that outlines the key steps from video acquisition to object isolation. This is followed by a structured functional analysis using the IDEF0 methodology to model each subprocess in terms of inputs, outputs, controls, and mechanisms.

3.1 Retrofitting Pipeline Overview for 3D Modeling and Subtraction

The pipeline begins with a simple monocular video captured under consistent lighting and a green-screen backdrop, serving as the raw input for NeRF-based modeling and subtraction, as shown in Figure 2.

The pipeline begins with a short video captured under controlled lighting conditions, typically with a green-screen background to facilitate later filtering. The video is converted into individual frames using a clipping tool (e.g., FFMPEG). These frames are processed using COLMAP, which extracts visual features and estimates camera poses — key inputs for NeRF model training.

The process includes the following stages:

1. Frame Extraction: The video is parsed into individual frames using FFMPEG, capturing diverse angles of the object.
2. Structure-from-Motion: COLMAP analyzes these frames to extract keypoints and reconstruct camera poses for each image.
3. NeRF Modeling: These inputs are used to train a NeRF model, which synthesizes a volumetric 3D scene of the object embedded in its environment.

4. Subtraction Pipeline:
    - Color Filtering: HSV thresholding removes background based on green-screen hue ranges.
    - Clustering: DBSCAN extracts dense object clusters from the filtered scene.
    - Point Recovery: A KD-tree mechanism reintroduces valid object points removed due to color similarity.

The final output is a clean, pose-consistent 3D point cloud representing only the object, suitable for validation or downstream inference.

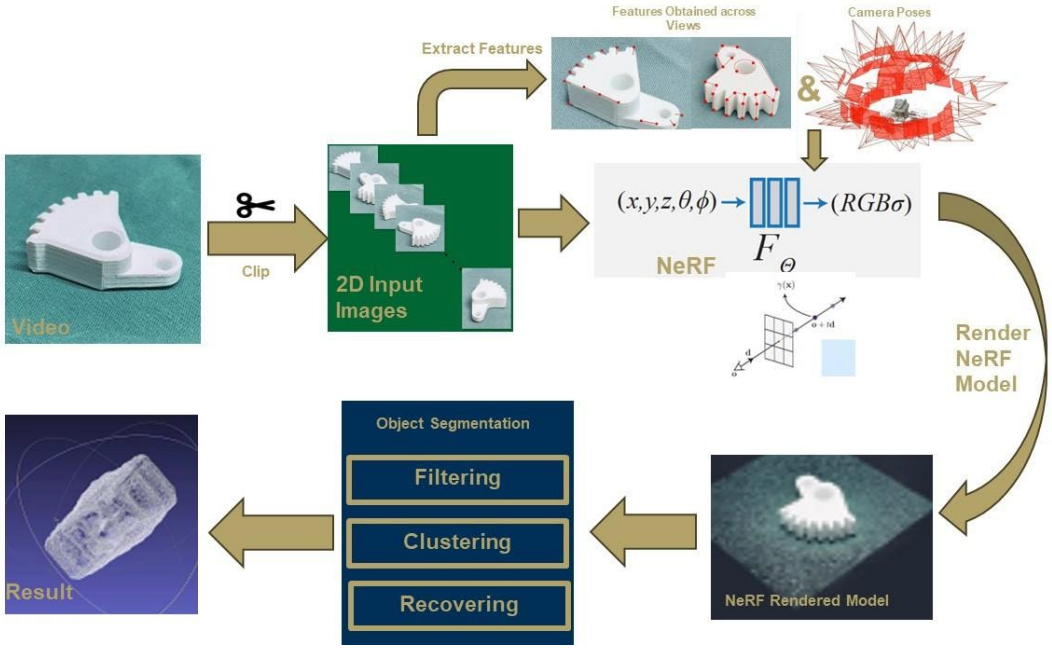

Figure 2. Overview of the AI-powered pipeline from monocular video to 3D object point cloud

3.2 Functional Decomposition of the Retrofitting Framework

To provide a structured view of the system's internal operations, we adopt the IDEF0 modeling technique, breaking down the core function — Generate 3D Digital Twin — into two main subprocesses, as illustrated in Figure 3:

- A1: Reconstruct 3D Object via NeRF-Based Digital Twin Modeling
  Converts raw video into a volumetric scene using NeRF. Tools such as FFMPEG (for frame splitting) and COLMAP (for pose estimation) feed into this modeling process, governed by controls like capture settings and training parameters.
- A2: Recognize Object from 3D Scene
  Operates on the NeRF-generated point cloud to isolate the object. The subtraction chain combines traditional filtering (HSV), density-based clustering (DBSCAN), and spatial consistency (KD-tree recovery) to produce clean object-level geometry.

This decomposition illustrates a modular, scalable framework that facilitates automation, enhances accuracy, and reduces dependency on manual annotation or high-end capture devices.

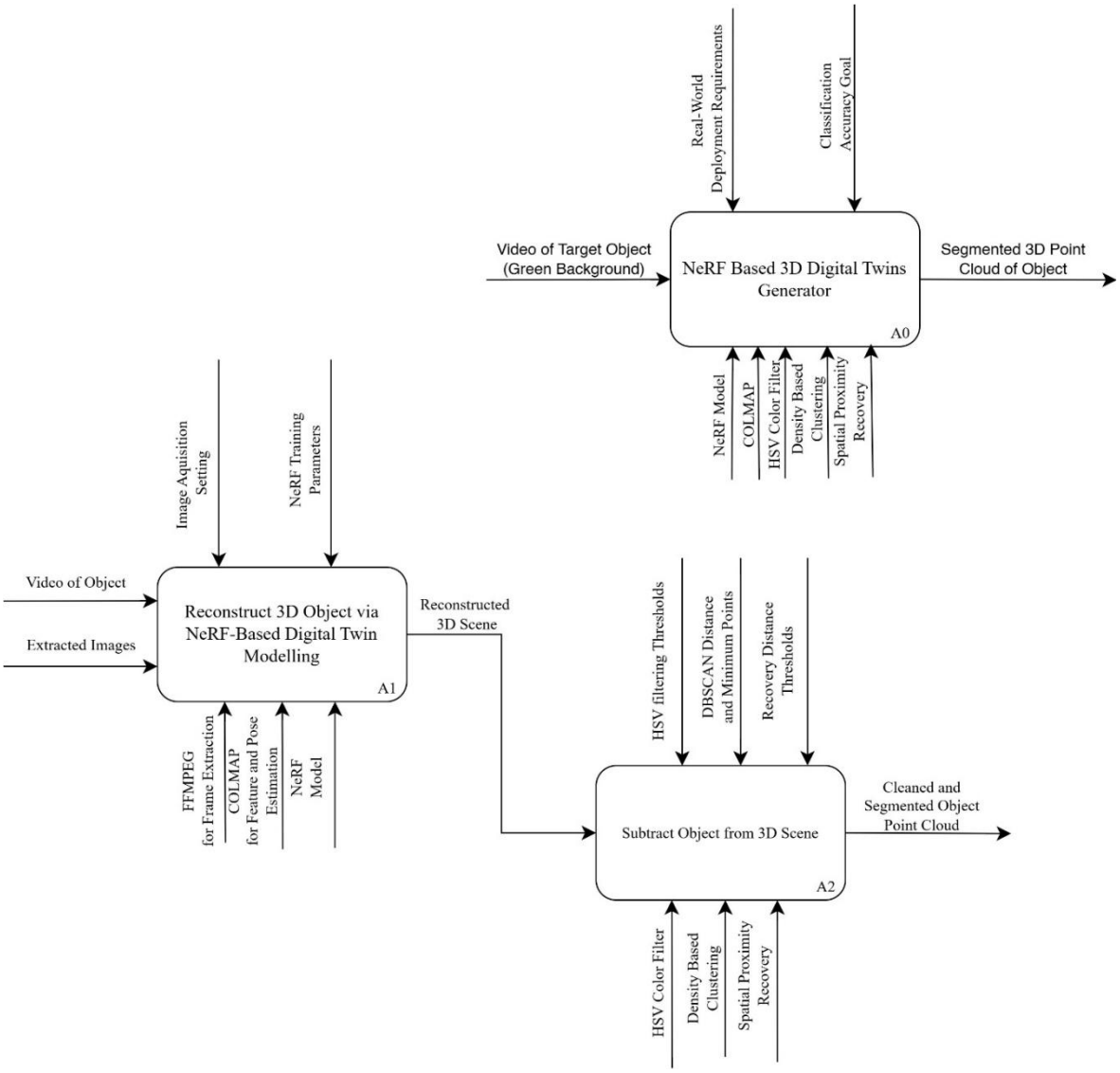

Figure 3. IDEF0 functional breakdown of the proposed system into NeRF-based modeling and subtraction stages

4. Implementation of Smart Retrofitting Pipeline for Multi-Object 3D Modeling

This section outlines the technical implementation of the proposed NeRF-based pipeline for generating 3D digital twins from monocular video. The process is divided into two main stages: (1) 3D object reconstruction using Neural Radiance Fields (NeRF), and (2) subtraction of the object from the reconstructed scene using color filtering, clustering, and spatial recovery techniques. Each stage contains subtasks designed to ensure fidelity, robustness, and minimal manual intervention.

4.1 3D Subtraction and Reconstruction via NeRF

4.1.1 Image Acquisition and Preprocessing

The process begins with recording a short video of the target object using a mobile device under controlled lighting conditions and against a green background. This background simplifies subtraction in later stages. Using FFMPEG, the video is sampled into approximately 120 image frames that capture a range of viewpoints around the object as shown in Figure 4.

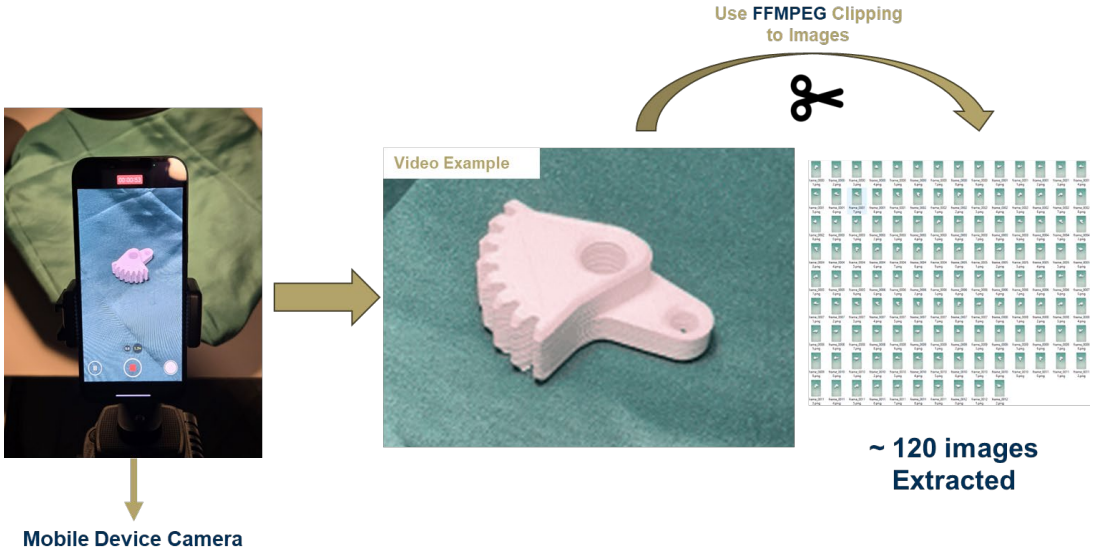

Figure 4. Example frames extracted from a green-background video using FFMPEG

4.1.2 Feature Extraction and Camera Pose Estimation

The extracted images are processed using COLMAP, an open-source structure-from-motion (SfM) and multi-view stereo (MVS) system. COLMAP detects key visual features (e.g., corners, textures) and estimates both intrinsic and extrinsic camera parameters. The output includes a sparse 3D point cloud and camera pose information, which are formatted into a transforms.json file compatible with NeRF training as shown in Figure 5.

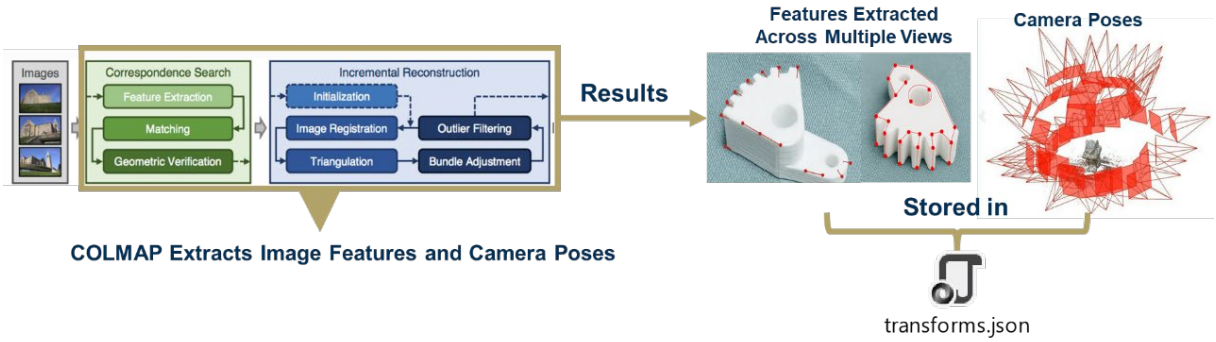

Figure 5. Camera pose estimation and sparse point cloud reconstruction using COLMAP

4.1.3 3D Scene Reconstruction via NeRF

Using camera parameters and images, a NeRF model is trained to synthesize a continuous volumetric scene. NeRF maps each 3D point and viewing direction (x, y, z, θ, ϕ) to a corresponding RGB color and volume density, producing novel views through volumetric rendering. This allows full 3D reconstruction from limited 2D inputs. The result is a dense, high-fidelity 3D scene that includes both object and background elements, as shown in Figure 6.

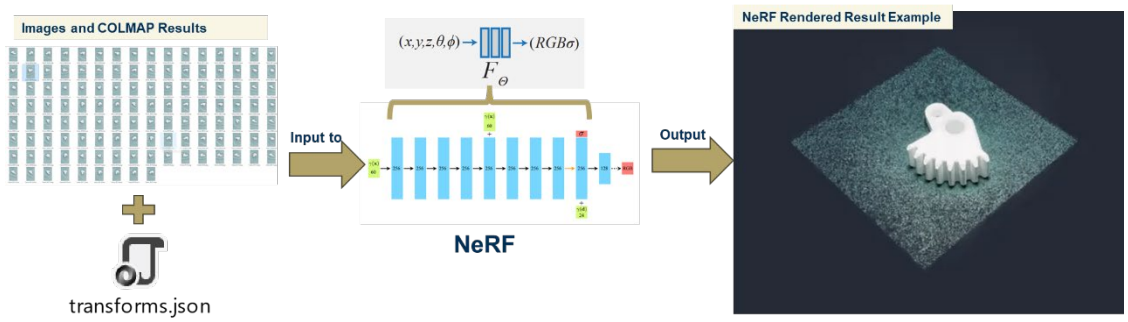

Figure 6. Photorealistic 3D scene reconstructed by NeRF using multi-view images and COLMAP-derived camera poses

4.2 3D Object Reconstruction via NeRF

After 3D reconstruction, subtraction is applied to isolate the object of interest. This stage involves three subtasks: background filtering via HSV color thresholding, noise removal via DBSCAN clustering, and recovery of mistakenly removed points via spatial proximity.

4.2.1 Green Background Subtraction via HSV-Based Color subtracting

The first stage of recognition leverages the HSV (Hue-Saturation-Value) color space to robustly remove the green background used during video capture. Unlike RGB, HSV separates chromatic content (hue) from luminance (value), making it more resilient to changes in lighting intensity or shadowing. This makes it ideal for filtering out uniform backgrounds that may have slight illumination variation across frames.

To determine the optimal HSV thresholds for background removal, we performed a histogram analysis over all NeRF-rendered points, as shown in Figure 7. This histogram visualizes the distribution of Hue, Saturation, and Value across the point cloud, enabling informed threshold selection.

- Hue Channel: Peaks in the histogram near the 66–101 range indicate the dominant green tones introduced by the backdrop. These values correspond to the green hue in HSV space and form the basis for hue thresholding.
- Saturation Channel: The saturation values of background points fall consistently within the 24–67 range, showing a medium to high saturation typical of artificial green screens.
- Value Channel: The value distribution of green background pixels tends to concentrate between 141–206, capturing the brightness range of the lit screen under controlled lighting.

By applying these empirically derived thresholds:

- Hue $\in$ [66, 101]
- Saturation $\in$ [24, 67]
- Value $\in$ [141, 206]

we effectively remove the majority of green-labeled points. This filtering step drastically simplifies the scene by eliminating irrelevant background data, retaining primarily the object and its immediate spatial context.
Figure 7 not only illustrates the statistical backing for these chosen thresholds but also justifies their generalizability across different video samples captured with similar green-screen setups. This data-driven calibration improves robustness across sessions and minimizes manual tuning.

Following this HSV filtering, the resulting point cloud — shown in Figure 8 — contains significantly reduced background noise, preparing it for DBSCAN-based geometric subtraction.

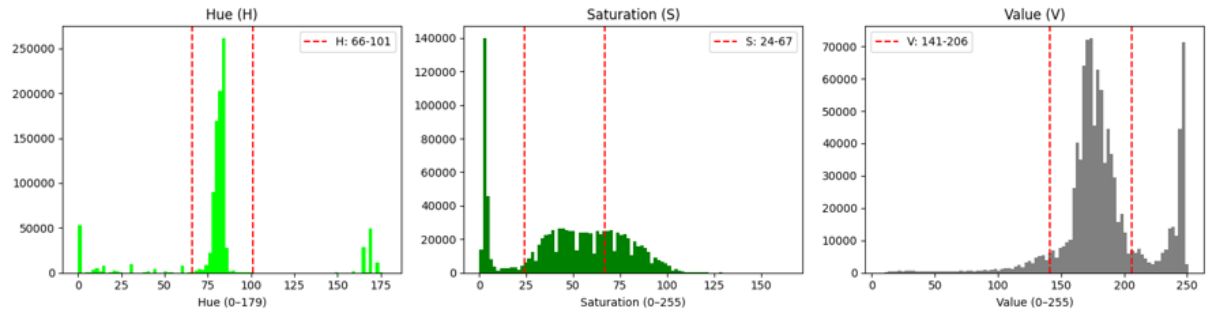

Figure 7. HSV histogram analysis used to determine background removal thresholds

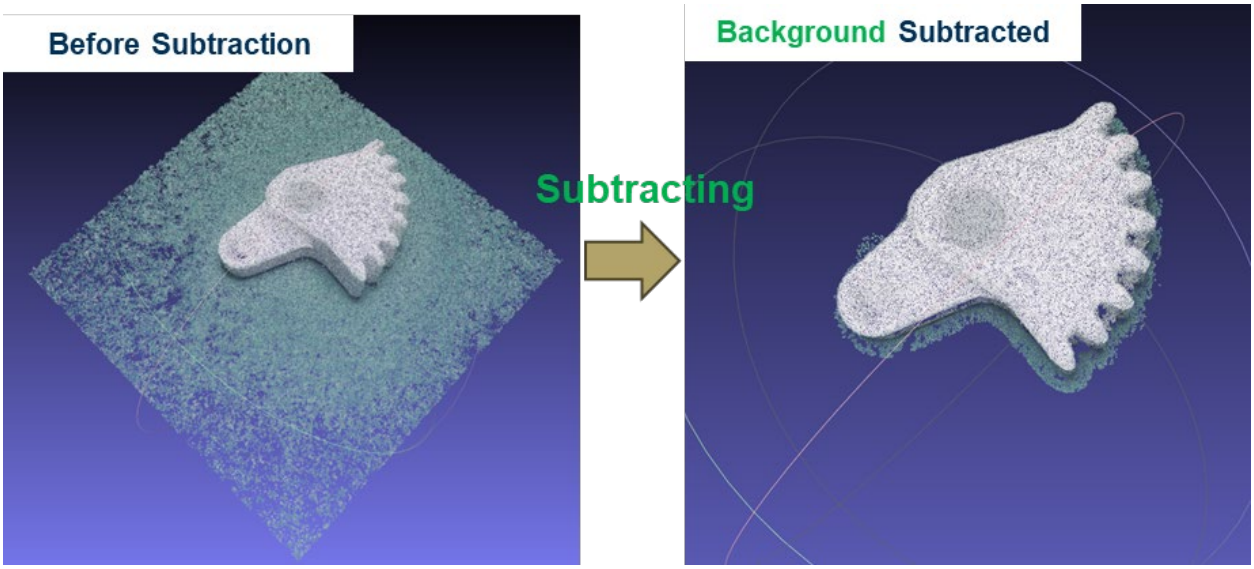

Figure 8. Resulting point cloud after applying HSV-based green background filtering

4.2.2 Noise Removal via DBSCAN Clustering

After the initial background removal using HSV filtering, some non-object points may still persist in the point cloud. These include floating noise, disconnected background fragments, and regions with low point density. To further isolate the object, we apply **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**—a widely used unsupervised algorithm for identifying high-density clusters in spatial data.

DBSCAN offers key advantages in this context:

- It does not require prior knowledge of the number of clusters.
- It is effective in detecting arbitrary-shaped clusters.
- It automatically labels sparse regions as outliers, making it ideal for cleaning noisy NeRF renderings.

The algorithm works by grouping points that have at least MinPts neighbors within a radius ε. For our application:

- ε = 0.02 m is chosen based on the typical point spacing in NeRF outputs at object scale.
- MinPts = 10 is selected to ensure that only sufficiently dense regions are retained.

As shown in Figure 9, applying DBSCAN to the HSV-filtered point cloud results in several clusters. Among these, the largest connected cluster is assumed to correspond to the main object. Smaller clusters and outliers, often located farther away or loosely connected, are considered noise and discarded.

Figure 9 visualizes the intermediate result:

- Points in the retained cluster are marked in color (or kept visible).
- Noise and fragments are labeled as outliers and suppressed.

In Figure 10, we show the cleaned and retained object geometry after the application of DBSCAN:

- Only the densest and most spatially coherent region remains.
- The structure of the object is preserved with minimal internal loss.
- Unwanted residuals have been fully removed.

This stage is critical for ensuring that subsequent processing steps — such as KD-tree recovery or CAD alignment — operate on a clean, accurate representation of the object. The use of DBSCAN, with tuned spatial parameters, thus forms a robust mid-level filter in the smart retrofitting pipeline.

.

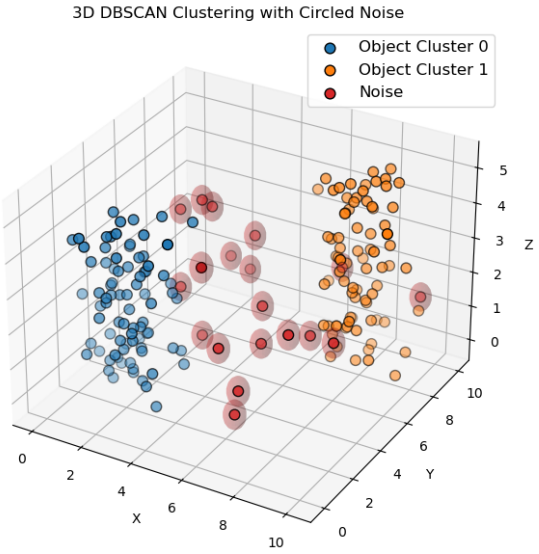

Figure 9. DBSCAN clustering applied to the filtered point cloud. The algorithm identifies dense regions likely to belong to the object while discarding noise and isolated fragments

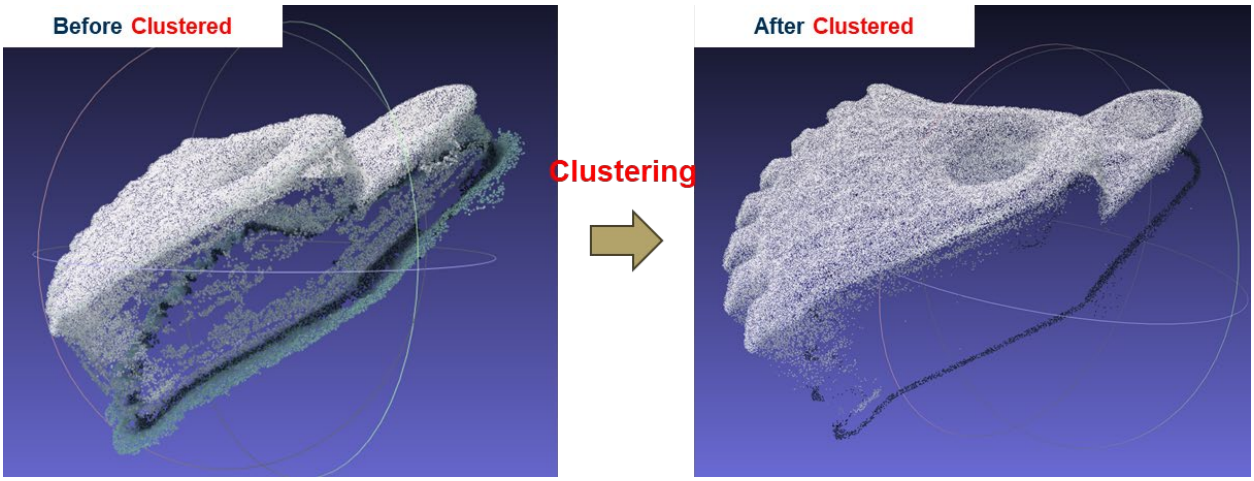

Figure 10. Final object cluster retained after DBSCAN

4.2.3 Recovery of Valid Points via Spatial Proximity

Some valid object points—especially those with color similarities to the background—may have been incorrectly removed during HSV filtering. To recover them, a KD-tree is constructed from the DBSCAN-cleaned point cloud. Points from the original NeRF cloud are queried against this KD-tree; if their nearest neighbor lies within a specified threshold τ (e.g., 0.02 m), they are reintroduced into the final result. This step improves subtraction completeness. An example result is shown in Figure 11.

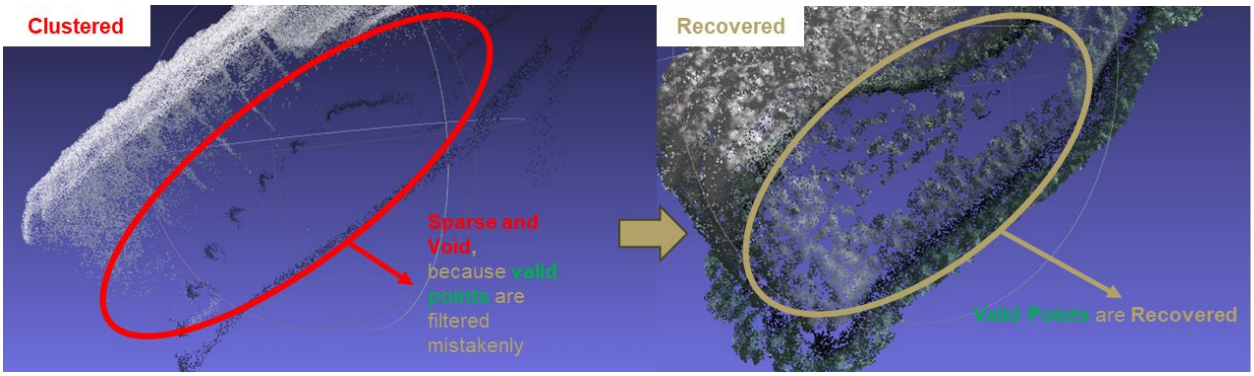

Figure 11. Recovery of valid object points using spatial proximity

5. Conclusion

This paper introduces an AI-powered retrofitting framework for smart reverse engineering, targeting practical deployment in dynamic, resource-constrained settings such as warehouses or automotive assembly lines. By leveraging Neural Radiance Fields (NeRF) for dense scene reconstruction and integrating it with HSV-based filtering, DBSCAN clustering, and spatial recovery, the system generates clean, pose-consistent 3D point clouds suitable for CAD alignment and robotic tasks.

Through a car industry use case, we demonstrate how monocular video input can be transformed into digital twins of components without expensive scanning hardware or manual intervention. The pipeline supports multi-object recognition, handles ambient noise, and facilitates rapid deployment in industrial workflows.
Future work will explore extending this retrofitting paradigm to full-scene digital modeling, real-time inference, and integration with robot learning systems. Additionally, comparative evaluation with traditional reverse engineering pipelines will further highlight the cost-performance trade-offs of this AI-powered approach.

Acknowledgements

This research was funded by the Ministry of Trade, Industry and Energy (MOTIE), Korea, through the “Global Industrial Technology Cooperation Center (GITCC) Program,” supervised by the Korea Institute for Advancement of Technology (KIAT) (Task No. P24680172) and by the "Manufacturing Innovation R&D Program", supervised by the Korea Institute of Industrial Technology (Task No. EH250001).

References

Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P. and Debevec, P., Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5855–5864, 2021.

Benaim, S., Xu, D., Chen, T. and Freeman, W. T., InsegNeRF: Instance segmentation using neural radiance fields, *arXiv preprint arXiv:2203.12341*, 2022.

Chen, A., Xu, Z., Zhao, F., Su, H. and Yu, T., NeRF++: Analyzing and improving neural radiance fields, *arXiv preprint arXiv:2010.07492*, 2021.

Ester, M., Kriegel, H.-P., Sander, J. and Xu, X., A density-based algorithm for discovering clusters in large spatial databases with noise, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 226–231, Portland, Oregon, USA, August 2–4, 1996.

Jumpat, P., Cen, Y., Zhang, Y. and Wang, Y., SA3D: Segment anything in 3D with NeRFs, *arXiv preprint arXiv:2310.12262*, 2023.

Krizhevsky, A., Sutskever, I. and Hinton, G. E., ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 25, pp. 1097–1105, 2012.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R. and Ng, R., NeRF: Representing scenes as neural radiance fields for view synthesis, *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 405–421, Glasgow, UK, August 23–28, 2020.

Pavoni, G., et al., Comparative assessment of neural radiance fields and photogrammetry in digital heritage: Impact of varying image conditions on 3D reconstruction, *Sensors*, vol. 23, no. 10, 2995, 2023.

Qi, C. R., Yi, L., Su, H. and Guibas, L. J., PointNet++: Deep hierarchical feature learning on point sets in a metric space, *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

Rusu, R. B. and Cousins, S., 3D is here: Point Cloud Library (PCL), *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–4, Shanghai, China, May 9–13, 2011.

Šlapak, R., et al., Neural radiance fields in the industrial and robotics domain: Applications, research opportunities, and use cases, *arXiv preprint arXiv:2308.07118*, 2023.

Wang, Y., Wang, S., Wang, Q., Chen, D. and Yu, G., Point-NeRF: Point-based neural radiance fields, *arXiv preprint arXiv:2201.08845*, 2021.

Zhang, H., Li, F. and Ahuja, N., Open-NeRF: Towards open vocabulary NeRF decomposition, *arXiv preprint arXiv:2310.16383*, 2023.

Zhang, Y., Zhao, H., Liu, S., Lin, S. and Jia, J., Fusion-aware point convolution for 3D scene understanding, *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 15814–15825, 2020.