3D Reconstruction and Defect Localization in Industrial Environments

Gyum Cho

Delft University of Technology

3D Reconstruction and Defect Localization in Industrial Environments

by

Gyum Cho

Gyum Cho	
Student number:	4840054

Instructor:	Prof. dr. Jan van Gemert
External Supervisor:	Steve Nowee
Project Duration:	Nov, 2025 - July, 2026
Faculty:	Faculty of Electrical Engineering, Mathematics and Computer Science, Delft

Preface

*T*his thesis was written as the final part of the MSc programme in Data Science and Artificial Intelligence at Delft University of Technology. The work was carried out in the field of industrial visual inspection, with a focus on 3D reconstruction and defect localization from endoscopic video of aircraft turbine blades. What began as a reconstruction-focused project gradually developed into a broader study on how reconstructed geometry, defect information, and visualization can be combined to support inspection in challenging industrial settings.

This project was made possible through the support of several individuals and organizations. I would like to thank Prof. dr. Jan van Gemert for his supervision, critical feedback, and continued research guidance throughout the thesis. I would also like to thank AIIR Innovations for providing the industrial context that motivated this work and for sharing valuable data, feedback, and discussions during the different stages of the project. In particular, I am grateful to Steve Nowee for his support, involvement, and constructive discussions throughout the development of the thesis.

Beyond its technical contributions, this thesis has been a valuable learning experience in dealing with uncertainty, iteration, and research decisions in a complex real-world setting. It marked a transition from implementing reconstruction methods to thinking more broadly about how technical results can become useful in practice. I hope that the work presented here provides a meaningful step toward more effective inspection support in industrial environments.

Gyum Cho
Delft, July 2026

Summary

*T*his thesis addresses the problem of defect-aware 3D understanding from industrial endoscopic video of turbine blades. In this inspection setting, the available imagery is constrained by restricted viewpoints, repetitive blade structure, unstable camera motion, reflective metallic surfaces, and limited reliable texture. These properties make conventional reconstruction difficult and also make it challenging to present inspection information in a form that is useful for human interpretation.

The objective of this thesis is therefore not limited to reconstructing turbine geometry alone. Instead, the work focuses on connecting geometric reconstruction to defect localization and visualization, with the broader goal of supporting industrial inspection tasks more effectively. In this sense, the thesis extends reconstruction-focused work toward a pipeline in which reconstructed geometry serves as the basis for spatial defect interpretation and inspection-oriented presentation.

To achieve this, the thesis develops a practical reconstruction workflow for challenging turbine borescope video. A direct application of conventional SfM and COLMAP-based reconstruction was not sufficient, since the turbine scenes often led to unstable camera pose estimation and geometrically weak blade recovery. To obtain a more usable result, the final pipeline combines fixed ROI cropping, frame extraction, segmented sparse reconstruction with refinement, dense reconstruction, and mesh generation. On top of this geometric representation, precomputed image-domain defect annotations are transferred to 3D through the selected sparse reconstruction and projected onto the final mesh surface. This produces a representation in which defect evidence can be inspected not only in the image domain, but also in relation to reconstructed blade geometry.

The final evaluated result shows that this workflow can produce a usable defect-aware 3D representation for a selected turbine inspection sequence. The selected reconstruction case yields a stable sparse model, a dense point cloud, and a Poisson mesh that can be used for downstream defect mapping. The mapping stage produces a defect-colored mesh, sparse 3D defect support points, and timeline-linked outputs that connect spatial defect evidence to the inspection sequence over time. An exploratory user evaluation was then conducted to compare the current 2D image-based view with the 3D geometry-aware representation. The evaluation did not show an overall preference for the 3D view, but it suggested that 3D geometry may support blade-level spatial interpretation.

The main contribution of this thesis lies in connecting three elements that are often treated separately, namely difficult 3D reconstruction, defect localization, and inspection-oriented visualization. Rather than treating reconstruction as an isolated technical goal, the thesis frames reconstruction as one step in a broader inspection-support pipeline. At the same time, the current work remains a proof-of-concept study. The strongest results depend on a selected reconstruction instance, and the pipeline is not yet uniformly reliable across all temporal windows of the sequence. Even with these limitations, the thesis shows that reconstructed geometry can serve as a meaningful intermediate representation for organizing defect evidence, while further work is still needed before this representation can become reliable for practical inspection use.

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
3D	Three-dimensional
A/B	Comparative evaluation between two alternative presentation conditions
AIIR	AIIR Innovations, the industrial partner that provided the inspection context and data
CGAL	Computational Geometry Algorithms Library
COLMAP	Structure-from-Motion and Multi-View Stereo reconstruction software used in the pipeline
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DISK	Learned local feature detector and descriptor used in the reconstruction pipeline
HLOC	Hierarchical Localization framework
JSON	JavaScript Object Notation
KD-tree	K-dimensional tree used for nearest-neighbour on mesh vertices
LightGlue	Learned local feature matching method used together with learned image features
LoFTR	Detector-free local feature matching method based on transformer
MVS	Multi-View Stereo
NeRF	Neural Radiance Fields
Open3D	Open-source library for 3D data processing and visualization
PLY	Polygon file format used for point clouds and meshes
ROI	Region of Interest
SfM	Structure from Motion
SIFT	Scale-Invariant Feature Transform
YOLO	You Only Look Once

Symbols

Symbol	Definition	Unit
d_{mesh}	Diagonal length of the mesh bounding box	[scene unit]
M	Reconstruction model	[-]
$N_{\mathrm{pts}}(M)$	Number of reconstructed 3D points in model M	[-]
$N_{\mathrm{reg}}(M)$	Number of registered images in model M	[-]
p	Sparse 3D defect support point	[scene unit]
r_c	Radius used for defect-color influence on the mesh	[scene unit]
r_t	Radius used for spatial grouping of defect events	[scene unit]
$S(M)$	Lexicographic score used to rank sparse reconstruction	[-]
v	Mesh vertex position	[scene unit]
$w(v,p)$	Influence weight of defect support point p on mesh vertex v	[-]
σ	Standard deviation of the Gaussian weighting function for mesh coloring	[scene unit]

1 Introduction

Aircraft engine inspection is essential for safe operation and continued airworthiness. In practice, many internal turbine components are inspected with a borescope, since this allows the engine interior to be examined without full disassembly. This makes borescope inspection a practical and widely used part of maintenance. At the same time, the process remains difficult. The inspector must interpret damage from image sequences acquired in a confined environment, often under limited illumination and with only partial views of the turbine surface. As a result, uncertainty in visual interpretation can lead to additional inspection effort, delayed maintenance decisions, and increased operational cost [2, 32].

One possible way to reduce this uncertainty is to move from image-only inspection toward a representation that preserves more of the underlying spatial structure. A 3D representation is attractive because it can relate a suspicious observation to blade geometry more directly than an isolated video frame. In principle, this could make defect location easier to interpret and could support a more structured inspection process. However, this is difficult to achieve in turbine borescope footage. The visual conditions are challenging, with restricted viewpoints, reflective metallic surfaces, weak texture, repetitive blade structure, and unstable camera motion. Under these conditions, conventional reconstruction methods often struggle to produce stable camera poses and plausible blade geometry [20, 25].

The work in this thesis starts from this gap between the potential value of 3D inspection support and the practical difficulty of obtaining usable geometry from borescope video. The initial question was whether turbine inspection footage could be transformed into a usable 3D representation. Early work therefore focused mainly on reconstruction itself. In practice, however, reconstruction alone did not provide a sufficiently reliable answer. This shifted the direction of the project. Instead of treating reconstruction as the final objective, the work gradually moved toward a broader pipeline in which reconstruction serves as the geometric basis for defect localization and inspection-oriented visualization.

The central aim of this thesis is therefore to investigate how turbine borescope video can be processed into a defect-aware 3D representation for inspection support. The focus is not only on recovering geometry, but on connecting reconstructed geometry, defect evidence, and visualization in one coherent workflow. In the final pipeline, image-domain defect annotations are linked to a selected 3D reconstruction and transferred onto the reconstructed surface, so that defect information can be interpreted in relation to blade geometry rather than only in relation to individual frames.

This thesis is organized in three main parts. The first part introduces the problem and explains the overall direction of the work. The second part provides the technical background needed to understand the later scientific article, including borescope data constraints, reconstruction from video, dense mesh generation, defect transfer from 2D to 3D, and visualization for inspection support. The third part presents the scientific core of the thesis. It describes the final reconstruction and localization pipeline, evaluates the resulting geometry and defect transfer process, discusses the current limitations, and includes an exploratory user evaluation of the final 2D and 3D inspection representations.

Seen as a whole, the thesis follows a transition from reconstruction as an isolated technical goal toward reconstruction as one component of an inspection-support pipeline. The broader contribution of the work is therefore not only that challenging turbine footage can be reconstructed in 3D under selected conditions, but also that the resulting geometry can be used to organize defect evidence in a more interpretable form for later inspection analysis.

2
Background

2.1. Borescope Inspection and Data Constraints

The scientific article in Part 3 does not treat the input as a generic video sequence. Instead, it considers turbine inspection footage acquired in a constrained borescope setting, where the observed image content is strongly shaped by the inspection environment itself. For that reason, the technical choices in the later chapters cannot be understood without first clarifying what type of data is available, which constraints it imposes, and why these constraints directly affect reconstruction and defect localization.

2.1.1. Visual Characteristics of Turbine Inspection Video

(a) Irregular illumination **(b)** Repetitive blade structure **(c)** Weak local texture

Figure 2.1: Examples of visual constraints in turbine borescope video. Together, these properties make correspondence-based reconstruction more difficult.

A first important observation is that turbine borescope video differs substantially from the imagery that is commonly used in general multi-view reconstruction. As shown in Figure 2.1, the inspection frames are affected by irregular illumination, repetitive blade geometry, and weak local texture. These properties are not minor visual imperfections. They directly determine how much reliable geometric information can be recovered from the video.

In classical multi-view geometry, reconstruction depends on finding image structures that can be recognized again across multiple views. These repeated correspondences are then used to estimate camera motion and scene geometry. In turbine inspection footage, this assumption becomes much weaker. Strong reflections and unstable lighting change the appearance of the same surface across frames. Repetitive blade structures reduce the uniqueness of many local observations. Low-texture metallic regions provide only limited distinctive image content. As a result, the raw inspection video is not only visually constrained, but also geometrically ambiguous from the perspective of feature-based reconstruction.

This difficulty is closely related to the notion of correspondence. Reconstruction pipelines depend on points or patches that can be found again under viewpoint change. Classical local-feature methods such as SIFT were designed to extract repeatable and distinctive image structures for this purpose [19]. Later learned approaches, including SuperPoint and DISK, improved the robustness of point detection and description, while more recent matching methods such as LoFTR and LightGlue further improved correspondence estimation under challenging image conditions [12, 18, 29, 31]. Even so, the underlying problem remains. When the scene provides weak visual distinctiveness, reliable camera pose estimation and plausible 3D recovery cannot be taken for granted. In this thesis, this is one of the main reasons why a more constrained reconstruction pipeline became necessary.

A second important observation is that, under weak appearance, spatial support from masks becomes more valuable. Silhouette-based reconstruction literature formalized this idea through the concept of the visual hull and later through space carving. These formulations show that masks can restrict the set of feasible 3D shapes even when dense appearance-based correspondence is weak. In the present thesis, masks are not used as a complete reconstruction solution on their own. Instead, they provide spatial priors that make later stages of the pipeline more focused and more stable. Blade masks indicate where the relevant turbine structure is located in the frame, while defect masks or defect regions preserve where candidate damage is observed. In this way, the mask system reduces irrelevant image content during reconstruction-oriented preprocessing and preserves localization evidence for the later defect-mapping stage.

2.1.2. Blade Masks, Defect Masks, and Region of Interest Selection

(a) Example of blade support and localized defect region. The blade area is outlined, while the defect candidate is indicated by a small red region.

(b) Example of multiple blade support regions in a raw inspection frame. The masks restrict the relevant blade areas and separate them from dark background, interface elements, and other less informative regions.

Figure 2.2: Examples of blade support regions and localized defect evidence in turbine borescope video. The blade support regions define where geometrically relevant turbine content is located, while the small defect region indicates where candidate damage is observed relative to the blade surface. Such spatial priors are useful both for reconstruction-oriented preprocessing and for later defect transfer to 3D geometry.

A mask can be understood as a spatial support region that identifies which part of the image belongs to the object of interest. In modern computer vision, this idea is closely related to instance segmentation, where the goal is not only to localize an object by a bounding box, but also to estimate its visible extent at the pixel level [14]. In the context of this thesis, the same distinction is useful for turbine inspection video. As illustrated in Figure 2.2, the blade support region indicates where the visible turbine blade is located in the frame, while the defect evidence is a much smaller localized region defined relative to that blade.

This distinction is important because blade support and defect evidence serve different purposes in the pipeline. A blade mask describes the support of the inspected object itself. A defect region describes where candidate damage is visible on that object. The two are therefore not interchangeable. The reconstruction stage uses blade support to restrict the scene region that should contribute geometric information, whereas the localization stage uses the smaller defect region to determine where defect evidence should later be transferred into 3D space.

A related concept is the region of interest, abbreviated as ROI. In general terms, an ROI is the image subregion retained for downstream processing because it contains the information that is relevant to the task. This is especially important in turbine borescope footage, where the raw frame may also contain dark background, endoscope interface elements, strong reflections, and blade-adjacent regions that contribute little to reconstruction while increasing ambiguity. A fixed crop can remove part of this irrelevant content, but a mask-guided ROI can go further by using the spatial distribution of blade support to retain the most informative blade region more consistently.

In this thesis, the role of masks is therefore practical rather than decorative. They do not solve reconstruction on their own, but they provide spatial priors that make later stages of the pipeline more focused and more stable. Blade support regions help concentrate the reconstruction process on the visible turbine structure, while defect regions preserve where candidate damage is observed for the later localization stage. Under the constrained conditions of borescope inspection, this restriction of the problem to the relevant blade region becomes an important step for both reconstruction and defect transfer.

2.1.3. Structure from Motion and Camera Pose Estimation

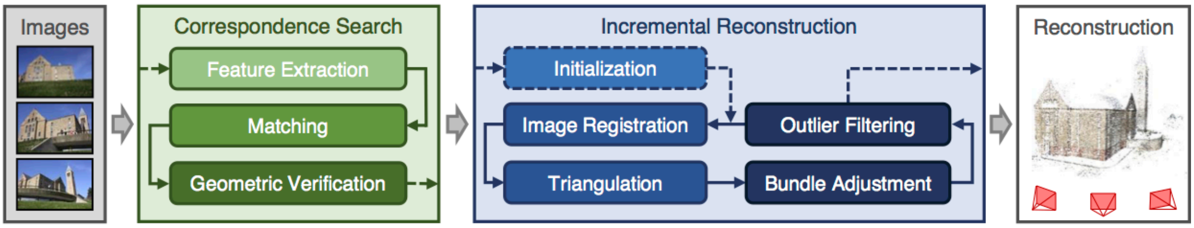

Figure 2.3: Example workflow of a Structure-from-Motion pipeline. Adapted from the Theia Vision Library documentation [30].

Structure from Motion, usually abbreviated as SfM, refers to the recovery of scene structure and camera motion from multiple overlapping images. As illustrated in Figure 2.3, the central idea is that the same physical point is observed in several views. Once these corresponding observations are identified, the viewing geometry can be used to estimate both the camera poses and the 3D location of the point. In practice, SfM produces a sparse reconstruction consisting of registered images, estimated camera poses, and a sparse set of 3D points. This sparse representation is already highly informative, because it establishes which images are geometrically consistent with one another and which scene points are supported by multiple views. In this thesis, this is particularly important because defect localization depends not only on the final mesh, but also on the relationship between registered image observations and sparse 3D point identities [25, 28].

The key quantity in this process is the camera pose. The camera pose specifies where the camera is located in 3D space and how it is oriented with respect to the scene. If the camera poses of multiple views are known, image observations can be triangulated to obtain 3D structure. SfM solves the inverse problem by jointly estimating camera poses and sparse 3D points from image correspondences. This is one of the reasons why SfM is not only a reconstruction method, but also a geometric registration framework. In the present thesis, the recovered camera poses are important because they determine whether image-domain defect evidence can be transferred reliably into 3D space.

A central optimization step in SfM is bundle adjustment. Bundle adjustment jointly refines camera parameters and reconstructed 3D points by minimizing reprojection error over all observations. In practical terms, this means that the camera poses and geometry are not estimated once and then fixed. They are repeatedly refined so that the reconstructed structure becomes more consistent with all image measurements at the same time. In the context of this thesis, this matters beyond reconstruction quality alone. If the recovered pose graph is unstable, then the mapping of defect evidence from image space into 3D also becomes unstable. A stable SfM solution is therefore important for both reconstruction and localization.

For turbine inspection video, this estimation problem is more difficult than in many benchmark reconstruction settings. The reason is not that SfM is conceptually unsuitable, but that its success depends on consistent correspondences, sufficient view overlap, and geometrically informative motion. In the present thesis, the difficulty of obtaining these conditions from turbine borescope video was one of the main reasons why direct reconstruction attempts often remained unstable and why more constrained reconstruction strategies were needed.

2.1.4. Local Features, Matching, and COLMAP-Based Reconstruction

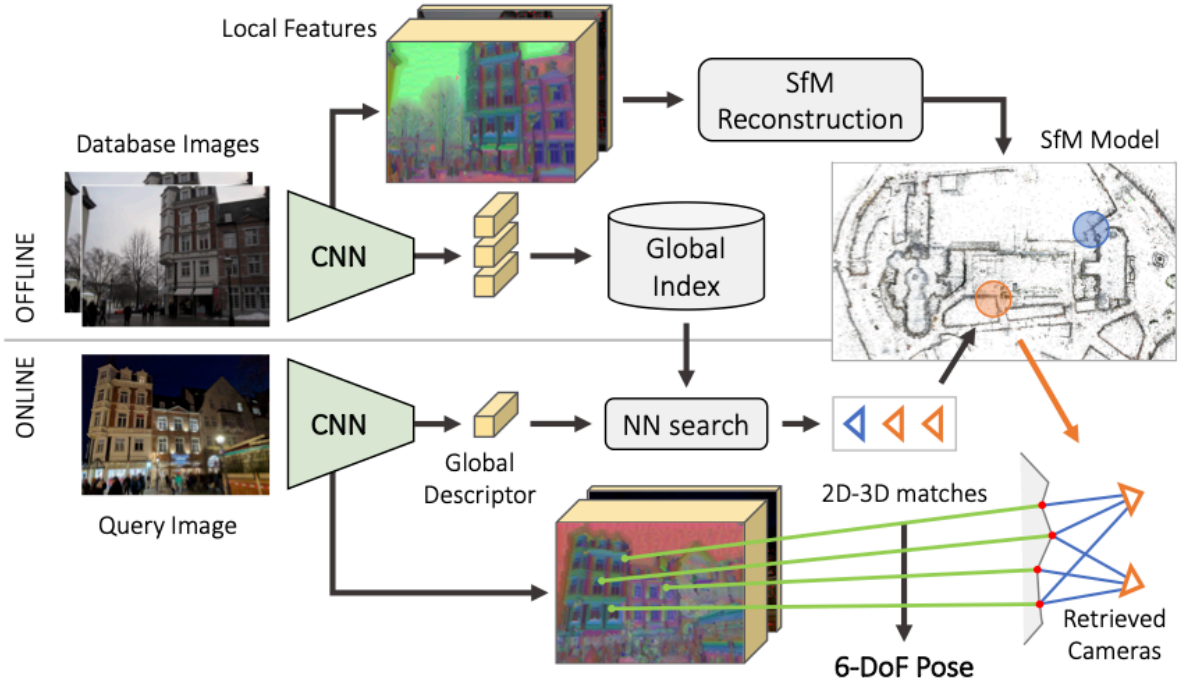

Figure 2.4: Overview of a hierarchical localization workflow. Image retrieval first proposes candidate image relations, and local feature matching then refines correspondences for the most promising pairs. In the present thesis, the same principle is useful because not every pair of turbine inspection frames provides equally informative overlap. Adapted from the official HLOC repository documentation [10].

To estimate structure and motion from images, the pipeline first needs correspondences. A local feature consists of two parts: a detector that selects repeatable image locations, and a descriptor that encodes the local appearance around those locations in a form that can be matched across views. Classical methods such as SIFT were designed to provide scale- and orientation-robust local measurements, while later learned methods such as SuperPoint and DISK replaced hand-crafted stages with learned detectors and descriptors. In broad terms, these methods pursue the same goal. They provide image measurements that remain sufficiently stable across viewpoint and appearance changes so that the later geometric stage can use them [19, 31].

Feature matching is the next step. Once candidate features are extracted from two images, the pipeline must decide which pairs correspond to the same physical scene point. This step is often more difficult than feature extraction itself, because false correspondences can propagate directly into incorrect pose estimates. More recent matching methods therefore incorporate stronger contextual reasoning. In particular, transformer-based approaches such as LoFTR and LightGlue have shown that correspondence estimation can remain more stable under weak texture and visually difficult conditions. This is directly relevant to the present thesis, because turbine borescope video contains exactly the type of repetitive and visually ambiguous content that weakens more conventional matching pipelines [18, 29].

The final reconstruction pipeline in this thesis combines learned local features with a hierarchical reconstruction workflow. As illustrated in Figure 2.4, hierarchical localization uses a coarse-to-fine strategy in which image retrieval first proposes promising image relations and local feature matching then refines correspondences only where needed. This idea is attractive because exhaustive matching across all images quickly becomes expensive and noisy. In the present thesis, the same principle is useful because not every frame pair in a long turbine sequence is equally informative. A hierarchical strategy helps concentrate matching effort on frame pairs that are more likely to share useful overlap [10].

COLMAP is then used as the geometric backbone that turns these correspondences into an actual reconstruction. In practical terms, COLMAP combines feature handling, matching, geometric verification, incremental registration, triangulation, and bundle adjustment in a single reconstruction framework. Its role in the present thesis is twofold. First, it provides the camera poses and sparse 3D points that are required for later dense reconstruction. Second, the sparse point identities later become the geometric bridge between image-domain defect observations and 3D localization. In this sense, the sparse reconstruction is not merely an intermediate visualization. It is the scaffold on which the later defect-mapping stage depends [9, 25].

In the pipeline of this thesis, DISK is used as the local feature representation and LightGlue as the matcher, while COLMAP is used for sparse geometric reconstruction. This choice should be understood as a practical response to the data constraints of turbine inspection video. The goal is not simply to adopt newer components, but to obtain a more reliable correspondence set and a more stable sparse reconstruction under conditions where visual distinctiveness is weak and repeated blade structures reduce local uniqueness.

2.1.5. Segmented Reconstruction and Refinement

A final concept needed for the scientific article is segmented reconstruction. In a long inspection video, not all frames are equally useful for reconstruction. Some windows may contain stronger overlap, more stable motion, or more informative blade visibility than others. A single, fully unconstrained reconstruction over the entire sequence may therefore fail even when a useful local sub-sequence exists. A segmented strategy addresses this by reconstructing overlapping temporal windows separately, comparing their quality, and then refining the most promising region. From a geometric point of view, this can be understood as a robustness strategy. Instead of assuming that the full video sequence forms one stable reconstruction problem, the sequence is searched for sub-problems that are more stable and more internally consistent. This idea is closely aligned with the broader observation that robustness, accuracy, and completeness remain the main challenges in practical SfM systems.

This concept is particularly important for the present thesis, because reconstruction quality was not determined by a single variable alone. A segment with the largest number of registered images did not always produce the most stable dense reconstruction later on. For that reason, segmentation and refinement are not used only to reduce computation. They are used to manage instability in the data itself. This point forms one of the key bridges between the background chapter and the scientific article: once the visual constraints of turbine video are understood, segmented reconstruction becomes a natural response rather than an ad hoc engineering choice.

In summary, the material in this section explains why the thesis eventually adopts a segmented HLOC+COLMAP pipeline. The choice follows directly from the structure of the problem. Camera pose estimation is essential because it underpins later defect localization. Correspondence quality is essential because pose estimation depends on it. Hierarchical matching reduces unnecessary ambiguity. Segmented reconstruction provides robustness when the full video sequence is not equally informative throughout. Together, these ideas form the conceptual basis of the reconstruction stage used later in the thesis.

2.2. Dense Reconstruction and Mesh Generation

The sparse output of SfM is sufficient to recover camera motion and a coarse geometric scaffold, but it is not sufficient for surface-level inspection. A sparse point cloud contains only isolated 3D observations, whereas the later stages of this thesis require a surface representation on which defect evidence can be visualized more directly. For that reason, the sparse reconstruction is followed by dense reconstruction and mesh generation. This section introduces the main concepts behind that process, namely geometric preparation, multi-view stereo, dense fusion, surface reconstruction, and mesh cleanup. These concepts are needed to understand why the scientific article does not stop at sparse reconstruction and why the quality of the dense stage directly affects the usefulness of the final defect visualization.

2.2.1. From Sparse Reconstruction to Dense Surface Recovery

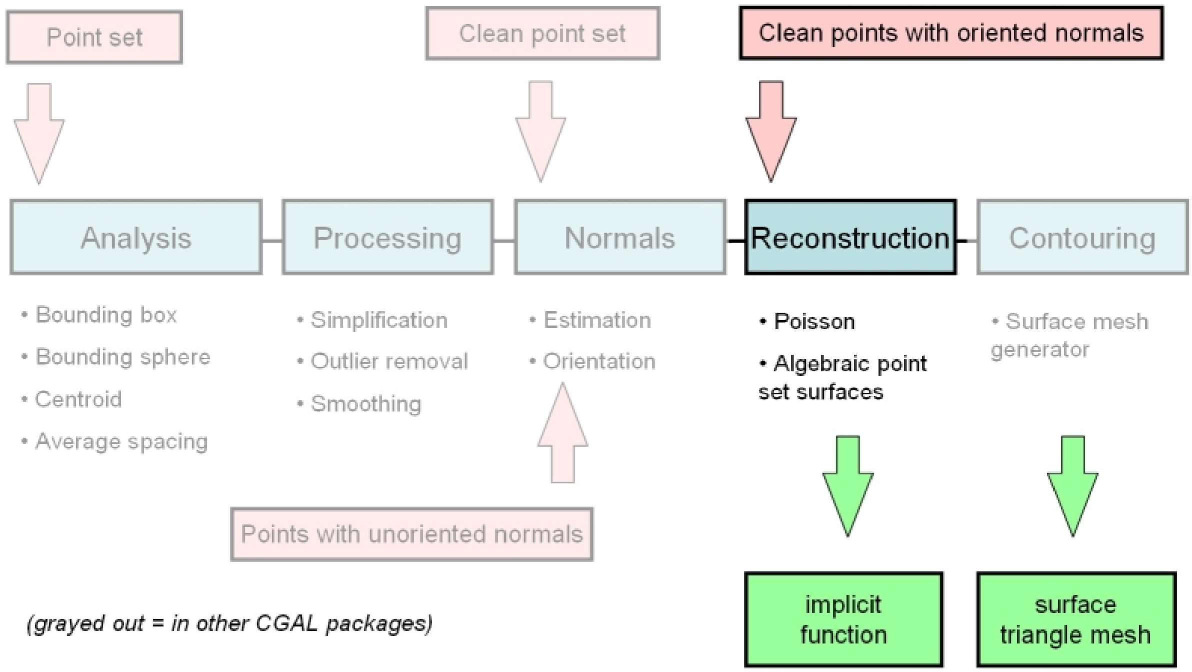

Figure 2.5: Conceptual view of dense surface recovery from geometric input. [6].

Dense reconstruction starts where sparse reconstruction ends. As illustrated in Figure 2.5, the sparse SfM stage provides camera poses and a geometrically consistent but incomplete set of 3D observations. This is already sufficient to recover camera motion and a coarse geometric scaffold, but it is not sufficient for surface-level inspection. A sparse point cloud contains only isolated 3D observations, whereas the later stages of this thesis require a denser surface representation on which defect evidence can be displayed more directly. For this reason, the reconstruction process continues from sparse geometry to dense depth estimation, dense fusion, and finally surface reconstruction [13].

Before dense stereo is applied, the imagery is usually prepared geometrically. In practice, this includes undistortion based on the camera model estimated during the sparse stage. This step matters because dense matching assumes that the geometric relation between neighboring pixels and scene structure is sufficiently regular to support stable depth estimation across views. If lens distortion remains uncorrected, dense matching becomes harder to stabilize. In COLMAP, this transition is reflected explicitly in the dense reconstruction workflow, where image undistortion precedes patch-match stereo, stereo fusion, and meshing [8].

The standard framework for recovering dense geometry from registered views is multi-view stereo, usually abbreviated as MVS. Unlike sparse reconstruction, which relies on a limited number of correspondences, MVS aims to estimate depth for a much larger portion of the visible surface by enforcing consistency across several views with known camera poses. One influential strategy here is PatchMatch stereo. The importance of PatchMatch in this thesis is mainly conceptual. It explains why the dense stage can remain sensitive to weak texture, repeated appearance, visibility changes, and parameter choices such as admissible depth range. Even when the sparse stage has produced a usable pose graph, the dense stage can still fail if local appearance is ambiguous [3].

The output of dense stereo is typically not a final mesh, but a larger collection of depth-supported surface samples. These per-view results must then be fused into a common geometric object. In practical terms, this fusion stage reduces local ambiguity by requiring agreement across multiple views and by combining many local depth estimates into a denser set of 3D samples. In COLMAP, this stage produces a fused point cloud before surface meshing. To obtain a continuous surface, this thesis then relies on Poisson surface reconstruction. The Poisson formulation turns oriented geometric input into a coherent surface through a global reconstruction process, which is why it often produces smoother and more readable surfaces than a direct interpretation of the point cloud alone [15].

In the context of this thesis, this entire progression is important because the final goal is not dense geometry for its own sake. The dense stage is needed because later defect localization and visualization require a surface representation rather than only sparse point support. In other words, dense reconstruction is the step that turns sparse geometric registration into a surface that can later carry defect evidence in a more interpretable form.

2.2.2. Mesh Cleanup and Surface Usability for Inspection

(a) Input mesh with small disconnected components

(b) Largest connected component retained

Figure 2.6: Example of mesh cleanup by connected-component analysis.

Surface reconstruction does not automatically produce a clean final object. In practice, a reconstructed mesh may contain disconnected fragments, small floating components, or geometry that is technically reconstructed but not useful for interpretation. As illustrated in Figure 2.6, this is a common outcome in practical reconstruction pipelines, where the recovered surface may contain several small clusters in addition to the main object. Under such conditions, mesh cleanup can become useful because the reconstructed result is otherwise harder to interpret and less suitable for downstream use [7, 22].

A standard concept used for this purpose is the connected component. Two triangles belong to the same connected component if they are linked through adjacency on the mesh. In practice, small connected components often correspond to reconstruction fragments or noise, while the largest connected component frequently corresponds to the main recovered object. This is why connected-component analysis provides a practical way to isolate the most informative blade-like geometry when the full mesh contains disconnected artifacts [7]. In the context of this thesis, this concept is important because the target of later defect visualization is not an arbitrary fragment, but a coherent surface that can plausibly be interpreted as part of the turbine blade structure.

This point is especially important for visualization. Defect presentation becomes easier when the target surface is coherent and interpretable. A fragmented mesh makes it harder to determine whether a highlighted region belongs to an actual blade surface or to a spurious reconstruction artifact. Mesh cleanup can therefore improve not only appearance, but also interpretability. For this reason, connected-component extraction is considered in this thesis as a useful post-processing concept after dense meshing, even though it is not part of the final defect-mapping workflow.

The importance of mesh quality in this thesis goes beyond geometric completeness alone. The dense mesh is not merely a more detailed version of the sparse point cloud. It is the representation on which defect observations are later projected, colored, grouped, and visualized. If the dense stage fails, the final inspection support becomes weaker even when the sparse reconstruction itself appears acceptable. For this reason, dense reconstruction is not treated here as a purely automatic continuation of SfM. The dense stage introduces its own stability issues, and these directly affect whether the final result becomes useful for downstream defect localization and inspection-oriented presentation.

Taken together, these observations explain why dense reconstruction is necessary for surface-level inspection and why mesh cleanup can be useful when disconnected artifacts reduce interpretability.

Sparse structure alone is not sufficient for surface-level inspection. Dense recovery provides a richer geometric basis, but the resulting mesh may still contain fragmented or weakly interpretable regions. Cleanup should therefore be understood as an optional post-processing tool for improving mesh interpretability, rather than as a mandatory step in the final defect-aware mapping pipeline.

2.3. Defect Localization from 2D to 3D

Recovering geometry alone is not sufficient for inspection support. In the context of this thesis, the geometric model becomes practically useful only once observed defects can be localized on that geometry. For that reason, the reconstruction stage is followed by a localization stage that links image-domain defect evidence to the recovered 3D structure. This stage forms one of the main differences between reconstruction-only work and the present thesis. The aim is not merely to reconstruct turbine geometry, but to connect geometry, defect evidence, and visualization into a single pipeline that is more useful for inspection. The localization stage therefore starts from defect observations in the image domain, then relates these observations to registered views and sparse 3D structure, and finally transfers the resulting information onto a surface representation that can support interpretation more directly.

2.3.1. Image-Domain Defect Evidence and Spatial Priors

The localization stage begins in the image domain. In general terms, localization means identifying where an object or event of interest appears in an image, rather than only deciding whether it is present. In inspection footage, this usually takes the form of a spatially localized region such as a bounding box or a small defect-support region. In the present thesis, the final pipeline starts from precomputed defect annotations in the image domain. These annotations provide the initial evidence that a candidate defect is visible at a particular image location, but they do not yet explain where that evidence lies on the turbine geometry.

This distinction is important. An image-domain defect region is informative, but it is still incomplete from the point of view of inspection support. It identifies where a suspicious observation appears within one frame, yet it does not resolve how that observation relates to the blade surface in 3D or whether similar observations in later frames refer to the same physical region. For this reason, the thesis does not treat image-domain localization as the final objective. Instead, it treats it as the starting point for a later geometric association step.

Under turbine borescope conditions, defect evidence is also difficult to interpret in isolation. The raw frames often contain reflections, dark regions, structural clutter, and repeated blade patterns. In such settings, a localized defect region becomes more meaningful when it is interpreted together with object support information. This is where blade support regions become useful. A blade support region indicates which part of the frame belongs to the visible blade, while the defect region indicates where candidate damage is observed relative to that blade. The two therefore serve different purposes. The blade support region constrains the spatial support of the inspected object, whereas the defect region constrains the support of the damage on that object.

More broadly, this distinction is related to the difference between coarse localization and finer spatial support. In computer vision, object detection usually provides a localized region such as a bounding box, while instance segmentation aims to recover a more precise support of the visible object [14]. The present thesis does not depend on a single segmentation architecture as part of its final pipeline, but the conceptual separation remains useful. Defect evidence narrows down where damage is likely to appear, while blade support restricts where that evidence should later be interpreted geometrically.

Seen in this way, image-domain defect evidence and blade support regions are complementary inputs to the later 2D-to-3D transfer stage. One provides the location of candidate damage in the frame. The other provides the spatial context of the blade surface on which that damage is observed. Together, they define the image-domain information from which later geometric localization begins.

2.3.2. From Registered Views to Surface-Level Defect Localization

Once defect evidence is available in the image domain, the next question is whether that observation can be related to the recovered 3D structure. This is possible only for registered views. A registered view is an image whose camera pose has been successfully estimated during the SfM stage and whose image observations are connected to the sparse reconstruction. This point is fundamental for the present thesis. The 3D localization stage cannot operate on arbitrary video frames. It can only operate on views that are geometrically anchored in the reconstruction. For this reason, the quality of the earlier reconstruction stage matters strongly for localization. An image-domain defect region may be correct as a visual observation, but if the corresponding frame is not registered, that evidence cannot be transferred reliably into 3D.

In a registered view, image observations can be linked to sparse 3D point identities through the point tracks maintained by the SfM system. Conceptually, the idea is straightforward. If a defect region overlaps with image locations that are already associated with sparse 3D points, then that defect observation can inherit a geometric interpretation through those points. In the present thesis, the image-domain defect annotation provides the localized region of interest, while the registered-view structure provides the geometric bridge into 3D. The resulting localization is therefore sparse at first. It does not immediately recover a full defect surface, but it identifies 3D points that are geometrically consistent with the observed defect region. This sparse-point association forms one of the main links between the reconstruction stage and the later visualization stage.

A set of sparse 3D points already provides useful evidence, but it is not yet the most interpretable form for inspection. Sparse points indicate where defect-related observations are geometrically supported, but they still appear as isolated markers. For inspection-oriented interpretation, it is usually easier to understand a defect when the evidence is transferred onto the reconstructed mesh itself. In practical terms, the sparse 3D points act as anchors, while the mesh provides the surface on which defect evidence can be displayed in relation to blade shape. The role of the mesh is therefore not only geometric. It is also interpretive, because it allows localized damage to be shown as part of the reconstructed surface rather than as a detached set of points.

This transition from sparse-point support to surface-level localization is especially important in the present thesis. The goal is not only to show that a defect-related region has geometric support, but also to make that support easier to interpret on the recovered blade surface. In background terms, the underlying idea is simple. Image-domain evidence identifies where the observation begins, sparse 3D association identifies where that evidence is geometrically supported, and mesh-level transfer makes the result easier to read spatially. For this reason, localization in the thesis is not treated as a standalone detection problem. It is treated as a staged transfer of evidence from 2D observations to sparse 3D structure and finally to a surface-level representation.

2.3.3. Repeated Observations and Defect Event Grouping

A final issue is that the same physical defect may appear in multiple frames. If every localized image observation were treated as a separate defect, the result would contain many duplicates and would be difficult to interpret over time. For this reason, localization in inspection video usually needs some form of grouping across observations. One general way to think about this is clustering. Density-based clustering methods such as DBSCAN are well known for grouping observations according to spatial proximity without assuming that the clusters must have a simple shape. In the context of this thesis, the same general principle is useful: repeated defect observations that are close in 3D and close in time can be grouped into a more stable defect event rather than being reported as unrelated detections.

This idea leads directly to timeline-style outputs. A single frame-level localization says where a candidate defect is visible at one moment. A grouped defect event says that related observations likely refer to the same physical damage over part of the video sequence. This distinction is important for inspection-oriented reporting. The goal is not only to count frame-level detections, but to organize repeated observations into a more interpretable record of damage on the reconstructed structure. In the present thesis, this is why the localization stage ultimately produces not only spatial outputs on the mesh, but also structured defect events and timeline outputs that can support later analysis.

2.4. Visualization for Inspection Support

In the present thesis, visualization is not treated as a cosmetic step that is applied after the technical work has already been completed. Instead, visualization is part of the inspection pipeline itself. The reconstructed geometry and the localized defect evidence become practically useful only when they can be presented in a form that supports interpretation. This point is central to the thesis. Reconstruction produces geometric structure, and localization identifies where defect evidence is supported, but inspection support requires one more step: the resulting information must be organized in a way that makes defect location, defect extent, and defect distribution easier to interpret. In this sense, visualization is the stage at which geometric recovery is translated into inspection utility. This role is consistent with the broader literature on information visualization, where visualization is understood not only as display, but as a means of supporting interpretation, exploration, and decision making.

2.4.1. Visualization as Interpretation Support

A useful way to think about visualization in this thesis is to distinguish between visibility and interpretability. A result can be visible without being easy to interpret. For example, defect points or separate defect boxes may technically show where defect evidence exists, yet still require substantial mental effort from the reader to understand how that evidence relates to blade geometry. Inspection-oriented visualization therefore aims to reduce this gap. Rather than simply rendering all available information, it should make the relevant spatial relationships easier to read. In visualization terms, this is closely related to the general principle that a good visual representation should support both overview and detailed inspection, rather than overwhelming the viewer with unstructured information. Shneiderman's well-known visual information-seeking mantra captures this idea in a compact form: overview first, zoom and filter, then details on demand. In the current thesis, that principle motivates the shift from detached markers toward surface-integrated defect presentation and later toward layouts that make comparison across blades easier.

This point also explains why visualization is not independent from the earlier stages of the thesis. A poor reconstruction limits what can be shown reliably, and weak localization limits what can be highlighted meaningfully. Conversely, when a coherent mesh and localized defect evidence are available, visualization can serve as the interface between technical output and practical interpretation. This is why the scientific article in Part 3 does not end with reconstruction metrics alone. The final output of the pipeline must be interpretable by a user who needs to assess blade damage, compare defects, and understand where damage is distributed across the turbine structure.

2.4.2. Surface-Level Visualization versus Marker-Based Visualization

A first important design distinction concerns how defect evidence is drawn on the recovered geometry. One possibility is marker-based visualization, in which defects are shown as separate points, boxes, or auxiliary objects placed in 3D space. This has the advantage of making the detected evidence explicit, but it also separates the evidence from the object surface itself. The viewer then has to infer how the marker relates to the blade surface. In inspection settings, that extra interpretation step can be inconvenient, especially when the geometry is already complex or when multiple defect observations appear close to one another.

A second possibility is surface-level visualization, where the target mesh itself acts as the visualization canvas. In that setting, defect evidence is not shown as a detached object, but as a colored or otherwise modified region on the reconstructed surface. The practical advantage is that the defect is presented directly in relation to the blade geometry. The viewer no longer has to mentally align a floating marker with the object. Instead, the defect becomes part of the surface interpretation. In the context of this thesis, this distinction is particularly important because the project evolved from earlier marker-style outputs toward direct mesh coloring. That evolution should not be understood as a purely aesthetic choice. It is better understood as a shift toward a visualization strategy that reduces the interpretive distance between detected damage and reconstructed geometry. More broadly, this design direction fits well with the literature on visual analysis, where external visual structure is used to reduce cognitive effort and support more direct reasoning from the displayed representation.

2.4.3. Blade-Level and Turbine-Level Organization

Another important distinction concerns the level at which the geometry is shown. A localized defect can be interpreted at the level of a single blade, but inspection often also benefits from understanding how multiple blades relate to one another. A blade-level representation emphasizes local shape and local damage. A turbine-level representation emphasizes distribution, comparison, and structure across the wider object. These two needs are not contradictory. They correspond to different inspection questions.

A local view helps answer where a defect lies on one blade. A more global organization helps answer which blades contain more damage, whether damage is concentrated in certain regions, and how defect observations compare across the reconstructed structure.

This distinction is closely related to overview-versus-detail principles in visualization. Information visualization literature has repeatedly emphasized that large or complex data should be presented in ways that allow both a global overview and local inspection. In the present thesis, the same logic applies to reconstructed turbine geometry. A single blade may be the most suitable level for detailed examination, while a more organized turbine-level representation may be the most suitable level for comparison. This is one of the reasons why the thesis eventually considers not only whether defects can be shown in 3D, but also how the reconstructed blade components can be organized in a way that resembles turbine structure more clearly. The goal of that step is not to claim perfect physical reconstruction of the full turbine, but to provide a more interpretable organizational frame for the defect evidence.

2.4.4. Layout, Comparison, and Inspection Tasks

Once defects are localized on reconstructed blade geometry, the next question is how different visual layouts affect interpretation. This is especially relevant when several blades or several defect-bearing components must be compared. A layout is not a neutral choice. Different layouts support different tasks. A compact arrangement may preserve a turbine-like impression, while a more separated arrangement may improve comparison between components. In visualization research, this broader point is well established: the effectiveness of a visual representation depends on the task it is meant to support. A representation that is suitable for recognition is not necessarily the most suitable representation for comparison, grouping, or explanation. In this thesis, the same reasoning motivates the later comparison between alternative ways of arranging reconstructed blade components for inspection.

This issue also explains why a visualization stage can become a legitimate research contribution rather than a purely presentational add-on. If two visual representations of the same defect evidence lead to different levels of interpretability, then the design of the representation itself becomes relevant to the thesis. The scientific article therefore does not only ask whether reconstruction and localization are possible. It also asks how the resulting information can be presented more effectively. In that sense, the visualization stage extends the thesis from a reconstruction problem into an inspection-support problem.

2.4.5. Visualization as Communication of Technical Results

A final point is that visualization in this thesis has a dual role. It is both an analysis tool and a communication tool. During development, visualization helps determine whether reconstructed geometry is coherent and whether localized defects align with plausible blade regions. At the same time, the final outputs must also communicate the result to others, including supervisors, technical readers, and potential industrial users. In this respect, the visualization stage is also related to the broader literature on visual communication and narrative visualization. Segel and Heer showed that visual representations do not merely present data; they also structure how the viewer understands and follows the underlying message. This idea is directly relevant here. A defect-colored mesh or a turbine-level arrangement does not only show the technical output. It also frames how the output is read.

For the present thesis, this means that the final visualization should make the central message of the work easier to grasp: challenging borescope video can be turned into a geometric representation, defect evidence can be linked to that geometry, and the resulting information can be displayed in a more interpretable way than in isolated image frames alone. This is the reason why visualization appears in the background chapter rather than only at the very end of the scientific article. It is not a decorative afterthought. It is one of the mechanisms through which the technical contribution becomes understandable and useful.

3

Defect-Aware 3D Reconstruction and Localization for Turbine Borescope Inspection

Gyum Cho[1,2]

[1]Delft University of Technology, Delft, The Netherlands.

[2]Computer Vision Lab, Delft University of Technology, Delft, The Netherlands.

Abstract

Borescope inspection is widely used in aircraft engine maintenance because it allows internal turbine components to be examined without full engine disassembly. However, the inspection process remains difficult because image-based observations provide limited geometric context for understanding where a blade defect is located and how it relates to the surrounding surface. Although 3D reconstruction offers a possible route toward more structured inspection support, turbine borescope video is challenging for standard reconstruction pipelines due to restricted viewpoints, weak texture, repetitive blade structures, reflective metallic surfaces, and unstable camera motion.

This thesis presents a defect-aware 3D inspection pipeline for turbine borescope footage. The final method combines fixed region-of-interest cropping, segmented HLOC/COLMAP-based sparse reconstruction, dense mesh generation, and reconstruction-aware transfer of precomputed image-domain defect annotations onto the recovered blade geometry. Rather than treating reconstruction as an isolated end goal, the pipeline uses reconstruction as the geometric basis for defect localization, mesh-level visualization, and inspection-oriented representation.

The final evaluated result shows that a usable defect-aware 3D representation can be produced for a selected turbine inspection sequence. The pipeline yields a sparse reconstruction, dense point cloud, Poisson mesh, defect-colored surface regions, sparse 3D defect support points, and timeline-linked defect outputs. An exploratory user evaluation did not show an overall preference for the 3D view over the current 2D image-based view, but it suggested that 3D geometry may support blade-level spatial interpretation. The current work should therefore be understood as a proof of concept that connects reconstruction, defect evidence, and visualization, while further work is needed to improve robustness, mapping confidence, and practical usability.

Keywords: Borescope inspection, 3D reconstruction, defect localization, turbine blades, HLOC, COLMAP, visualization

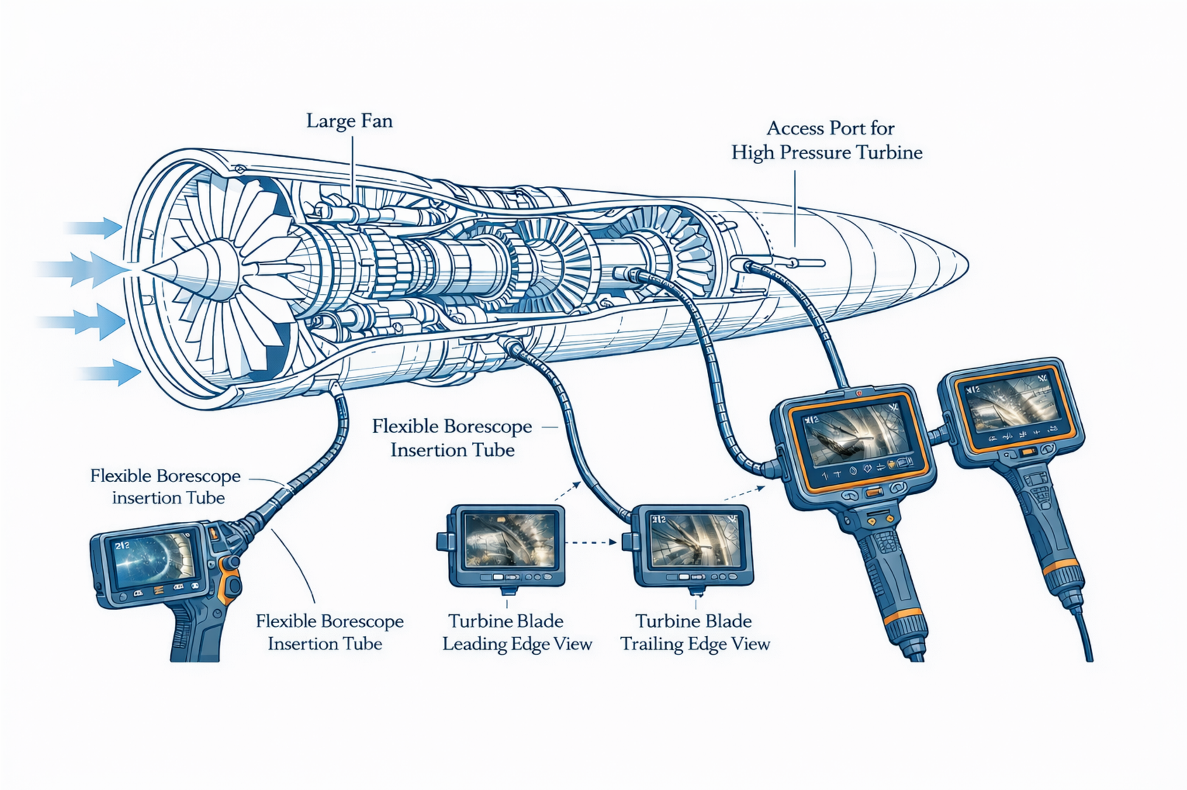

Figure 3.1: Borescope inspection setup for aircraft engine maintenance. A flexible camera is inserted through access ports to observe internal turbine components such as the high-pressure turbine.

3.1. Introduction

Aircraft engine inspection is a necessary part of safe operation and continued airworthiness. In practice, internal turbine components are commonly inspected using a borescope, since this allows the engine interior to be examined without full disassembly [2]. As shown in Figure 3.1, a flexible camera is inserted through access ports to capture internal turbine structures, which makes borescope inspection a practical and widely used method in maintenance workflows.

Despite this advantage, the inspection process remains difficult. The visual information obtained through a borescope is limited to image sequences, where depth, scale, and spatial relationships between structures are not directly observable. Uzun et al. describe borescope inspection as manual, time-consuming, and strongly dependent on operator experience [32]. As a result, the exact location of a defect on the blade surface and its extent relative to the surrounding geometry are not always clear. This introduces uncertainty in interpretation and may lead to additional inspection steps or conservative maintenance decisions.

To address these limitations, both industry and research have explored the use of 3D reconstruction techniques for more structured inspection support [1]. A 3D representation allows defect observations to be related directly to turbine geometry, which makes it easier to reason about location and distribution in a more consistent way. In principle, this can reduce ambiguity compared to purely image-based inspection and provide a more informative basis for later analysis.

However, applying conventional 3D reconstruction methods to turbine borescope video is not straightforward. The inspection setting shown in Figure 3.1 reflects a visually difficult environment. The camera operates under restricted viewpoints and confined geometry, while turbine blades exhibit weak texture, repetitive patterns, and reflective metallic surfaces. These properties make reliable correspondence estimation difficult across frames. In practice, this can lead to unstable feature matching, unreliable camera pose estimation, and incomplete or distorted reconstruction results [25]. Even when reconstruction succeeds partially, the recovered geometry is often not stable enough to support downstream inspection analysis.

The work presented in this thesis starts from this gap between the potential of 3D-based inspection support and the practical difficulty of obtaining usable geometry from borescope video. Early attempts focused mainly on reconstruction itself and explored several reconstruction strategies under different assumptions. These attempts showed that treating reconstruction as an isolated objective does not provide a reliable solution in this setting. Instead, the project gradually moved toward a broader pipeline in which reconstruction serves as the geometric basis for defect localization and inspection-oriented visualization.

The main objective of this thesis is therefore to investigate how turbine borescope video can be transformed into a defect-aware 3D representation for inspection support. To achieve this, a pipeline is developed that combines constrained preprocessing, segmented reconstruction, and geometric transfer of image-domain defect evidence. Rather than aiming for perfect reconstruction, the approach prioritizes obtaining sufficiently stable geometry that can support the mapping of defect observations into 3D space. The resulting representation allows defect information to be visualized directly on the reconstructed surface, which provides a more structured view of the inspection data.

The remainder of this thesis follows the development of this approach. It first describes the progression from earlier reconstruction attempts to the final pipeline design, then explains how defect information is integrated into the 3D representation, and finally evaluates the resulting reconstruction and defect-mapping outputs. The thesis also includes an exploratory user evaluation that compares the current 2D image-based inspection view with the 3D geometry-aware representation developed in this work.

The main contributions of this thesis are threefold. First, it develops a segmented reconstruction workflow for challenging turbine borescope video. Second, it connects precomputed image-domain defect annotations to reconstructed 3D geometry through registered sparse reconstruction. Third, it evaluates the resulting 2D and 3D inspection representations through technical analysis and an exploratory user evaluation.

3.2. Related Work

3.2.1. Borescope Inspection and Blade Defect Analysis

Borescope inspection has already been studied as a practical but difficult setting for turbine blade assessment. In [2], Aust et al. describe gas turbine blade inspection as a process in which defect detection and maintenance decisions remain closely tied to the quality of the available visual evidence. This point is important for the present thesis because it shows that inspection is not only an image-analysis problem. Defect identification and maintenance decision support are closely connected, which means that the value of a technical method depends on whether it improves the interpretation of defect location in a form that is useful for inspection.

A related conclusion is reached in [1], where Aust and Pons compare human operators with advanced inspection technologies in the visual inspection of aero-engine blades. Their study shows that human inspection remains central in practice, but is affected by inconsistency, subjectivity, and limited interpretability. At the same time, the study makes clear that advanced technologies such as image processing and 3D scanning are relevant, but not sufficient on their own. This is one of the main points that shaped the present work. A more advanced representation is useful only when it improves how defect information is interpreted. For this reason, the current thesis does not stop at reconstruction quality alone, but moves toward defect-aware 3D visualization and inspection-oriented representation.

Work on automated defect detection from borescope imagery has also shown that image-domain analysis is feasible. In [27], Shen et al. propose a deep-learning framework for automatic damage detection in aircraft engine borescope inspection, while in [26], Shang et al. develop a deep-learning-based borescope image-processing method for aero-engine blade in-situ damage detection. These studies show that damage can be localized from inspection imagery and that automated support can reduce part of the manual burden. At the same time, their focus remains mainly in the image domain. A suspicious region can be detected in a frame, but its spatial relation to blade geometry remains limited.

This is the point at which the present thesis departs from earlier work. Instead of treating borescope imagery only as a source of 2D defect localization, the current work connects image-domain defect evidence to recovered 3D geometry so that defect location can be interpreted on the blade surface itself. In this sense, the thesis builds on earlier inspection and detection studies, while extending them toward a geometry-aware representation that is more directly aligned with inspection support.

3.2.2. 3D Reconstruction for Turbine Blades and Similar Difficult Scenes

A natural response to the limitations of image-based borescope inspection is to recover the observed structure in 3D. In principle, a 3D representation can place defect observations on blade geometry itself, which makes spatial reasoning more direct than in isolated image frames. However, prior work also shows that this is not a straightforward problem in visually constrained scenes. Reconstruction methods behave differently depending on whether they rely mainly on silhouettes, dense appearance modeling, or explicit multi-view correspondences. This distinction is important for the present thesis, because the turbine setting does not provide equally favorable conditions for all three directions.

One classical direction is silhouette-based reconstruction. Laurentini introduced the visual hull as a shape representation that can be recovered from object silhouettes [17], and Kutulakos and Seitz later extended this line of work through space carving with photo-consistency and visibility constraints [16]. These studies are relevant because they show that masks can provide useful geometric constraints when direct correspondence is weak. This was one of the reasons why silhouette-based ideas were considered in the early stages of the present thesis. At the same time, these methods also make their own limitations clear. They depend strongly on the available viewing range and on the quality of the silhouettes, and they do not naturally provide the stable surface detail needed for inspection-oriented defect interpretation. The lesson taken here is therefore not that silhouettes alone solve the reconstruction problem, but that mask information can still act as a useful geometric constraint in a more practical pipeline.

A second direction is neural scene reconstruction. Mildenhall et al. showed with NeRF that a continuous scene representation and high-quality novel views can be recovered from posed images [21]. This makes radiance-field methods attractive whenever dense 3D structure is desired from image sequences. At the same time, this line of work also highlights an important limitation for the present setting. NeRF relies heavily on reliable camera poses and sufficiently informative view coverage, both of which are difficult to guarantee in turbine borescope footage. Depth-aware variants such as DS-NeRF further reinforce the same point by showing that sparse geometric supervision can improve reconstruction when the available views are limited [11]. For the present thesis, the main lesson is therefore that dense reconstruction in difficult scenes benefits from stronger geometric constraints and more structured inputs, rather than from an unconstrained reconstruction process alone.

A further lesson comes from reconstruction work that is more directly tied to inspection. In [33], Wang and Gan show that 3D reconstruction becomes more useful when treated as part of a broader inspection pipeline rather than as an isolated modeling task. Although their application domain differs from turbine borescope inspection, the underlying point is highly relevant here. The value of reconstruction lies not only in producing geometry, but in supporting later analysis on top of that geometry. This is also the position taken in the present thesis. The goal is not to obtain a visually convincing turbine model in its own right, but to obtain a representation that is stable enough to support defect localization and visualization.

Together, these studies define the broader reconstruction context of the present work. Silhouette-based methods show that masks can constrain geometry, but do not by themselves provide sufficiently rich inspection surfaces. Radiance-field methods show that dense 3D recovery is possible, but require pose and view conditions that are difficult to satisfy in turbine borescope video. Inspection-oriented reconstruction studies further show that geometry becomes more useful once it is linked to downstream analysis. The present thesis builds on these lessons by treating reconstruction as one stage in a larger process that ultimately supports the interpretation of defect location and distribution.

3.2.3. Structure from Motion in Difficult Scenes

Structure from Motion has long been one of the standard approaches for recovering scene geometry and camera motion from image collections. In [28], Snavely et al. show that repeated observations across multiple views can be used to reconstruct sparse 3D structure together with the corresponding camera poses. This line of work establishes the basic promise of feature-based reconstruction. If reliable correspondences can be found across images, geometry can be recovered without requiring an explicit 3D sensor. For the present thesis, this is a natural starting point, since borescope inspection already provides image sequences from which blade geometry might, in principle, be reconstructed.

A more practical and robust formulation of this idea is given by modern incremental SfM pipelines. In [25], Schönberger and Frahm present a reconstruction framework that combines feature extraction, matching, geometric verification, incremental image registration, triangulation, and repeated refinement. This work is especially relevant because it reflects the type of reconstruction backbone that later became standard in practical systems such as COLMAP. The main lesson is that successful reconstruction is not determined by one stage alone. Stable camera poses depend on a sequence of decisions, including the quality of the local features, the reliability of the matches, and the consistency of geometric verification. This point is directly relevant to the present thesis, where unstable pose estimation was one of the main reasons why early reconstruction attempts failed to produce usable blade geometry.

At the same time, classical feature-based reconstruction makes assumptions that become fragile in turbine borescope footage. In [19], Lowe shows that local feature matching depends on distinctive image structures that can be detected and described consistently under changes in viewpoint and scale. This becomes difficult when the scene contains weak texture, repeated patterns, reflective surfaces, and limited viewpoint variation. These are exactly the conditions present in turbine inspection footage. The issue is therefore not that SfM is conceptually unsuitable, but that the visual evidence available to SfM becomes much less stable in this setting. As a result, the challenge shifts from applying SfM as a standard pipeline to making SfM reliable enough for visually ambiguous data.

This point is central to the present thesis. Existing SfM literature shows that feature-based geometry recovery is effective when correspondences are reliable and the available motion provides enough geometric diversity. The turbine setting weakens both assumptions. Blades are visually similar, surface texture is weak, and the borescope viewpoint is constrained by the engine geometry. For this reason, the present work does not use SfM as a black-box reconstruction tool. Instead, SfM is treated as the geometric backbone of the pipeline, while the surrounding stages are adjusted to improve its stability under difficult inspection conditions.

Taken together, these studies explain why SfM remains relevant for the present thesis despite the difficulty of the target setting. Earlier work shows that sparse geometry and camera poses can be recovered effectively from image sequences, and that incremental SfM provides a strong practical foundation for doing so. At the same time, the assumptions behind these pipelines become weaker in turbine borescope footage. The present thesis builds on this observation by keeping SfM as the geometric core of the reconstruction process, while modifying the surrounding stages to make it more reliable for difficult blade imagery.

3.2.4. Learned Features and Matching for Reconstruction

The previous subsection showed that Structure from Motion becomes fragile in turbine borescope footage because the required correspondences are difficult to recover reliably. The next question is therefore not whether reconstruction should still rely on correspondences, but how those correspondences can be made more stable under weak texture, repeated blade structure, and uneven image quality. This is the point at which learned local features and learned matching become relevant.

One of the main weaknesses of classical feature-based reconstruction in difficult scenes is that the quality of the final geometry depends heavily on the quality of the local correspondences. When the image structure is weak or repetitive, hand-crafted features can become less stable, which in turn affects camera pose estimation and sparse reconstruction. In [12], DeTone et al. introduce SuperPoint as a learned interest point detector and descriptor designed for multiple-view geometry tasks. The importance of this line of work for the present thesis is not only that it improves local feature extraction, but that it shows feature extraction itself can be adapted to the reconstruction problem rather than treated as a fixed preprocessing step.

A similar shift appears in learned matching. In [24], Sarlin et al. show that feature matching can be improved by reasoning jointly about candidate correspondences instead of treating them as isolated nearest-neighbor assignments. This is especially relevant in scenes with repeated patterns, where visually similar structures can easily be matched incorrectly. For turbine blade imagery, this point is directly important. Different blades may contain local regions that look similar, which means that matching quality depends not only on local appearance but also on contextual consistency across the image pair.

This direction becomes even more relevant in methods designed specifically for difficult matching conditions. In [31], Tyszkiewicz et al. introduce DISK as a learned local feature framework optimized for producing a high number of correct matches. In [18], Lindenberger et al. present LightGlue as a learned matching method with a stronger focus on efficiency and adaptation to image-pair difficulty. These works are closely related to the final design choices of the present thesis. The move toward DISK and LightGlue was not made simply because they are newer methods. It was made because turbine borescope footage requires stronger matching under weak texture, repeated blade structure, and uneven visual quality across frames. Under these conditions, a learned feature-and-matching stack provides a more suitable basis for sparse reconstruction than a conventional SIFT-based setup.

The way these methods are organized across many images is also important. In [23], Sarlin et al. propose a coarse-to-fine strategy in which global retrieval first narrows the search space and local matching is then applied more selectively. Although that work is framed in visual localization, its underlying logic is highly relevant here. A long image sequence should not be treated as if every frame pair is equally useful. In the present thesis, this idea supports the broader use of structured matching and segmented reconstruction. The lesson taken from this literature is therefore not only that better local features improve reconstruction, but also that matching becomes more reliable when it is applied in a more selective and better organized way.

These studies explain why the present thesis moves beyond a conventional SIFT-based feature-and-matching pipeline and toward a learned matching stack. The previous subsection established that difficult turbine footage weakens the assumptions on which classical SfM depends. The present subsection shows how learned features, learned matching, and more selective pair organization provide one practical response to that weakness. This is the basis on which the later reconstruction pipeline adopts DISK, LightGlue, and structured image-pair selection to stabilize sparse reconstruction before defect evidence is transferred into 3D.

3.2.5. Dense Reconstruction and Mesh Generation

Sparse reconstruction provides camera poses and a set of geometrically consistent 3D points, but this is not yet sufficient for inspection-oriented visualization. A sparse point cloud can indicate the overall structure of the observed scene, yet it does not provide a continuous surface on which defect evidence can be displayed clearly. For this reason, prior work has often extended sparse reconstruction with dense reconstruction and mesh generation.

A standard direction is dense multi-view stereo. Furukawa and Ponce show that dense surface information can be recovered from multiple calibrated views by enforcing photometric and geometric consistency across images [13]. Their work is important in the present context because it makes clear that sparse geometry alone is only the first step. Once camera poses are available, denser surface estimates become possible, which is necessary when the goal is not only to recover structure but also to visualize information on that structure. At the same time, this line of work also shows that dense recovery remains sensitive to the quality of the underlying sparse reconstruction. When camera poses or image correspondences are unstable, the dense stage inherits this instability rather than correcting it.

A more specific dense stereo direction is PatchMatch-based estimation. Bleyer et al. propose PatchMatch Stereo with slanted support windows, showing that dense depth estimation can be improved by propagating and refining depth hypotheses efficiently [4]. This is relevant to the present thesis because it reflects the type of dense reasoning used after sparse reconstruction in practical pipelines. The main lesson taken from this work is that dense reconstruction is not only a matter of producing more points. It also depends on the local reliability of depth hypotheses and on the consistency of image support. For turbine borescope footage, where appearance is weak and reflective, this becomes especially important. Dense reconstruction may still fail even when a sparse reconstruction appears acceptable.

To obtain a usable surface from dense points, a further reconstruction step is required. Kazhdan et al. formulate surface reconstruction from oriented points as a Poisson problem, producing coherent surfaces that are more resilient to noise than local stitching-based approaches [15]. This is directly relevant to the present thesis because the final goal is not only to recover point geometry, but to obtain a mesh that can support defect localization and visualization. A mesh is easier to interpret than a sparse or dense point cloud alone, especially when defect evidence must be interpreted on the blade surface itself. At the same time, Poisson reconstruction also makes a practical point clear. The quality of the final mesh still depends on the quality of the input point set. If dense reconstruction contains instability or fragments, the resulting surface will reflect these weaknesses.

These studies explain the role of dense reconstruction in the present work. Dense multi-view stereo provides a way to move beyond sparse geometry, PatchMatch-based methods show how local depth estimation can be made practical, and Poisson reconstruction provides a route from oriented points to a continuous surface. The present thesis builds on these ideas by treating dense reconstruction and mesh generation not as purely visual post-processing, but as necessary steps toward a 3D representation that can support defect-aware visualization.

3.2.6. Defect Localization from 2D to 3D

A large part of the recent inspection literature has focused on detecting defects directly from images. This is a useful direction, since image-based detection can already reduce part of the manual inspection burden. Bochkovskiy et al. show that modern object detection can localize regions of interest efficiently and with strong accuracy [5]. For the present thesis, this is relevant because turbine defect analysis also starts from image observations. A suspicious region first appears in a frame, and it is natural to localize that region before attempting any geometric reasoning. The main lesson taken from this line of work is that image-domain localization is a practical first step, but not yet a complete inspection solution.

A related point appears in instance-level segmentation. He et al. show that localization can be extended beyond bounding boxes by predicting object masks alongside detection results [14]. This distinction is important for the current work. A bounding box gives a coarse defect location, while a mask provides stronger spatial support for where the relevant region lies in the image. In turbine inspection, this difference matters because the defect itself is meaningful only relative to the blade surface on which it appears. The literature on image-based localization therefore suggests two useful ideas. First, defect evidence can be localized automatically. Second, the spatial support of that evidence can be made more precise when mask information is available.

However, these methods still operate mainly in the image domain. Even when a defect is correctly localized in a frame, its relation to the blade geometry remains limited. This is the main gap that motivates the present thesis. The question is no longer only whether a defect can be detected, but whether the detected evidence can be placed on a reconstructed 3D surface in a way that supports interpretation. Wang and Gan address a closely related idea in building inspection by combining defect detection with 3D reconstruction and visual inspection [33]. Although their application domain is different, the main lesson is directly useful here. Once 2D inspection evidence is connected to 3D scene structure, the result becomes more informative for downstream analysis than image-only inspection alone.

This is the point at which the present thesis departs from earlier image-based defect analysis. Rather than treating detection as the final output, image-domain defect evidence is used as the starting point for geometric association. The defect evidence is first localized in the image, then related to registered views and reconstructed geometry, and finally transferred to the blade surface for visualization. In this sense, the contribution of the present work is not a new detector by itself, but a pipeline that makes image-domain defect evidence more useful by giving it geometric meaning.

As a result, these studies define the background for the localization stage of the present thesis. Image-based methods such as YOLO show that defect regions can be localized efficiently. Segmentation-based methods such as Mask R-CNN show that the spatial support of those regions can be described more precisely. Inspection-oriented reconstruction work further shows that defect information becomes more useful once it is connected to 3D structure. The present thesis builds on these lessons by treating defect localization as a staged process that begins in 2D and becomes more interpretable once it is transferred to reconstructed blade geometry.

3.3. Method

3.3.1. Problem Setting and Pipeline Overview

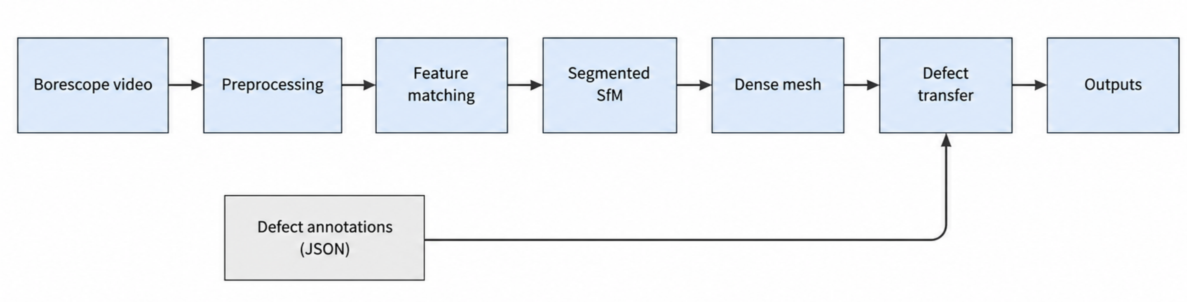

Figure 3.2: Overview of the proposed defect-aware 3D inspection pipeline.

The goal of this work is to transform turbine borescope video into a defect-aware 3D representation that can support inspection. The method takes as input a borescope video of the turbine interior together with time-indexed defect annotations in the image domain. From these inputs, the pipeline produces a reconstructed 3D mesh of the observed turbine structure, a defect-colored version of that mesh, a set of 3D defect support points, and timeline-linked outputs that connect defect observations to the corresponding time intervals.

The method is organized into five main stages. First, the input video is preprocessed to remove the endoscope user interface and to extract a consistent set of reconstruction frames from the target time range. Second, local features and pairwise matches are computed across the extracted frames to establish multi-view correspondences. Third, sparse reconstruction is performed for overlapping temporal segments, after which the most stable sparse model is selected for downstream processing. Fourth, the selected sparse reconstruction is converted into a dense point cloud and then into a 3D surface mesh through Poisson meshing. Fifth, the time-indexed defect annotations are transferred from the image domain to the reconstructed geometry, producing a defect-colored mesh, 3D defect support points, and timeline-linked outputs.

The data flow between these stages is direct. As shown in Figure 3.2, the preprocessing stage produces the frame set used for feature extraction and matching. The matching stage produces the feature and correspondence information needed for segmented sparse reconstruction. The selected sparse model is then passed to the dense reconstruction stage, which produces the Poisson mesh used for downstream analysis. The defect annotation file is provided as a parallel input to the final mapping stage, where image-domain defect observations are associated with the reconstructed geometry and exported in both geometric and timeline-based formats.

This organization reflects the main design choice of the thesis. Reconstruction is not treated as a final objective on its own. Instead, it serves as the geometric basis for defect localization and inspection-oriented visualization. The final output of the pipeline is therefore not only a reconstructed turbine surface, but a representation in which defect evidence can be inspected in relation to the recovered 3D geometry.

3.3.2. Preprocessing and Fixed ROI Cropping

(a) Raw borescope frame **(b)** Fixed ROI overlay **(c)** Cropped reconstruction frame

Figure 3.3: Preprocessing and fixed ROI cropping. The raw borescope frame contains endoscope interface elements and peripheral image regions that do not belong to the turbine surface. The fixed ROI removes these non-structural regions and produces the cropped frame used for the later reconstruction stages.

The reconstruction stage starts from a preprocessing step that converts the input borescope video into a consistent image set for multi-view reconstruction. The purpose of this step is twofold. First, the endoscope user interface and other peripheral image regions are removed, since these regions do not belong to the turbine structure and may introduce unstable image content across frames. Second, the video is reduced to a uniformly sampled frame sequence that covers the target inspection interval while keeping the reconstruction set at a manageable size.

The final preprocessing pipeline operates on the time interval from 2.0 s to 121.0 s of the borescope video. Given the source video frame rate, this corresponds to the frame range from frame 60 to frame 3606. From this interval, 450 reconstruction frames are sampled without allowing duplicate source frames. The frame indices are selected by uniform rounding over the target interval, which yields an approximately uniform temporal spacing of about 0.265 s between consecutive exported frames. This produces a reconstruction set that covers the full target interval without concentrating too strongly on any local temporal region.

Before frame export, each frame is cropped using a fixed region of interest defined in normalized image coordinates as

$$(x_{\min}, y_{\min}, x_{\max}, y_{\max}) = (0.04,\ 0.11,\ 0.86,\ 0.84).$$

For the source resolution of 960×720 pixels, this corresponds to an effective crop of $x = 38 \ldots 824$ and $y = 79 \ldots 603$, resulting in cropped frames of size 787×525 pixels. No resizing is applied after cropping. This means that the exported images preserve the cropped spatial detail of the original footage rather than introducing an additional scaling step.

The cropped frames are written sequentially to the reconstruction image directory as `0000.jpg` to `0449.jpg`. Existing image files in the target output directory are removed before export so that the reconstruction stage always consumes a clean and consistent frame set. No additional filtering is applied at this stage. In particular, the preprocessing does not use blur filtering, exposure filtering, or mask-quality filtering.

This preprocessing design reflects a practical choice made in the thesis. The aim is not to perform aggressive frame selection before reconstruction, but to provide a stable and visually consistent image set that removes obvious non-structural content while preserving as much turbine information as possible. The output of this stage serves directly as the input to feature extraction and matching in the subsequent sparse reconstruction stage.

3.3.3. Segmented Sparse Reconstruction

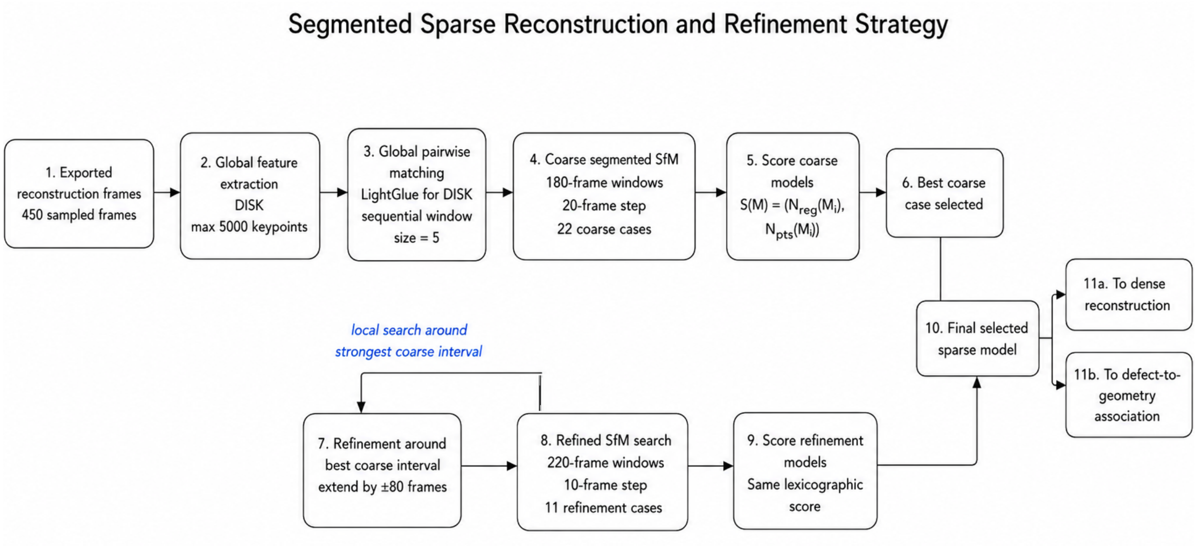

Figure 3.4: Segmented sparse reconstruction and refinement strategy pipeline.

After preprocessing, sparse reconstruction is performed from the exported frame set using a feature-based Structure from Motion pipeline built on HLOC and COLMAP. Local features are extracted with DISK [31], and pairwise matching is performed with LightGlue configured for DISK features [18]. In the final method, feature extraction is carried out with a maximum of 5000 keypoints per image, without grayscale conversion, and with an image resize limit of 1600 pixels. This choice reflects the need to retain sufficient local structure in turbine imagery, where the blade surfaces contain weak texture and repeated patterns.

Image pairs are generated using a sequential sliding-window strategy. For each image, matching is attempted only within a local temporal neighborhood defined by a window size of five frames. A global sequential pair list is created once for the full frame set, and the resulting features and matches are reused across all reconstruction cases. This avoids repeating feature extraction and matching for every segment while keeping the pair graph restricted to temporally adjacent views. In the present setting, this is a practical choice. The borescope video follows a continuous inspection trajectory, so nearby frames are more likely to share useful overlap than distant frame pairs.

To improve robustness, sparse reconstruction is not performed on the full sequence as a single case. Instead, the frame set is divided into overlapping temporal segments. In the final sparse reconstruction pipeline, each segment contains 180 frames and the segment step is 20 frames, which produces strongly overlapping windows. Segments shorter than 30 frames are discarded. This results in 22 coarse reconstruction cases over the target clip. The purpose of this design is to avoid treating the full video as a uniformly stable reconstruction problem. Some temporal intervals contain stronger overlap and more stable visual structure than others, so reconstruction is evaluated locally before the best region is selected for refinement.

Each segment is reconstructed independently with COLMAP using a single shared camera model. The camera model is set to `OPENCV_FISHEYE` and the camera mode is set to `SINGLE`. The mapper is configured with permissive initialization settings to make registration possible under difficult image conditions. These include a minimum model size of three images, an initial minimum number of inliers of 40, and an absolute pose minimum inlier count of 20 with a minimum inlier ratio of 0.10. Geometric verification is kept active so that the relaxed initialization still remains constrained by multi-view consistency.

For each reconstruction case, the sparse model is evaluated using a lexicographic score based on the number of registered images and the number of reconstructed 3D points. Candidate models are ranked by

$$S(M) = \big(N_{\text{reg}}(M),\ N_{\text{pts}}(M)\big),$$

where $N_{\mathrm{reg}}(M)$ is the number of registered images in model M, and $N_{\mathrm{pts}}(M)$is the number of reconstructed 3D points. The model with the highest score is selected. No reprojection-error weighting or custom multi-metric quality score is used in this sparse model selection step. This makes the selection rule simple and reproducible, but it also means that the selected model must be interpreted as the strongest model under this specific ranking criterion rather than as a globally optimal reconstruction.

After the coarse pass, one refinement stage is performed around the best coarse segment. Refinement windows are generated by extending the best coarse interval with a margin of 80 frames on both sides, then scanning this enlarged range with windows of size 220 and a step of 10 frames. Refinement cases shorter than 60 frames are discarded. In the final evaluated run, this produces 11 refinement cases. The best refinement model is then compared against the best coarse model using the same lexicographic score. If the refinement model is strictly better, it replaces the coarse result. Otherwise, the coarse result is retained.

As summarized in Figure 3.4, this segment-and-refine design reflects the main reconstruction strategy of the thesis. Rather than assuming that all parts of the video contribute equally to a single stable model, the method searches for locally reliable reconstruction intervals and forwards only the strongest sparse model to the dense stage. The selected sparse model is then used for dense reconstruction and also serves as the geometric basis for defect-to-geometry association.

3.3.4. Dense Reconstruction and Mesh Generation

(a) Selected sparse reconstruction **(b)** Fused dense point cloud **(c)** Final Poisson mesh

Figure 3.5: Representative outputs of the selected reconstruction case. The sparse model provides registered views and sparse geometric structure, the dense stage produces a fused point cloud, and Poisson meshing converts this geometry into a continuous surface for downstream defect mapping.

The selected sparse reconstruction is next converted into a dense surface representation. This stage is necessary because the sparse SfM model provides only camera poses and a set of reconstructed 3D points, while the final inspection output requires a continuous surface on which defect information can be displayed. As shown in Figure 3.5, the reconstruction therefore progresses from the selected sparse model to a fused dense point cloud and finally to a Poisson mesh. In the final pipeline, dense reconstruction is performed with COLMAP on the sparse model selected from the segmented reconstruction stage, and the resulting mesh is then used in the downstream defect-mapping stage.

The first step of the dense stage is image undistortion. The method uses `colmap image_undistorter` with the original reconstruction images and the selected sparse model as input. This produces an undistorted image set together with a corresponding undistorted sparse model inside the dense workspace. After undistortion, the dense workspace uses a pinhole camera representation. This step is required because the dense stereo stage operates on undistorted imagery rather than directly on the original fisheye-based reconstruction output. Dense stereo is then performed with `colmap patch_match_stereo` using geometric consistency. The explicitly set dense parameters are limited to a maximum image size of 2000 pixels and geometric consistency enabled. Source-view selection follows the workspace configuration generated by COLMAP, which uses automatic source-view selection with up to 20 source images per reference image. No additional non-default PatchMatch parameters are imposed in the final dense reconstruction path. In practice, this means that the dense stage relies mainly on the quality of the selected sparse geometry and on COLMAP's standard dense stereo behavior rather than on extensive manual tuning.

After stereo estimation, the per-image depth results are fused into a dense point cloud using `colmap stereo_fusion`. Geometric fusion is used, producing a dense point cloud stored as `fused.ply`. This fused geometry is then converted into a surface mesh with `colmap poisson_mesher`, which produces the final output mesh `mesh_poisson.ply`. The purpose of this step is to obtain a continuous 3D surface that is easier to inspect and use for defect mapping than a dense point cloud alone.

The dense workspace also contains intermediate artifacts such as undistorted images, the undistorted sparse model, PatchMatch configuration files, depth maps, normal maps, and consistency graphs. These intermediate files are useful for checking the dense reconstruction process, but they are not used directly by the defect-mapping stage. The downstream stage uses the final Poisson mesh generated from the selected reconstruction case.

An important detail of the final method is that no dense-stage retry or fallback logic is used for this selected reconstruction case. In particular, there is no fixed depth-range retry and no automatic alternative dense reconstruction branch. Likewise, connected-component cleanup is not part of the final dense-to-defect workflow. Although a separate utility exists for extracting the largest mesh component, the downstream defect-mapping stage uses the full Poisson mesh generated from the selected dense workspace.

This dense reconstruction design reflects the role of geometry in the thesis. The dense stage is not used to produce a mesh for visualization alone. Its purpose is to generate a surface that is stable enough for later defect transfer and inspection-oriented display. For this reason, the selected sparse reconstruction is converted into a dense point cloud and a Poisson mesh, and this mesh is forwarded directly to the defect-mapping stage.

3.3.5. From Image-Domain Defect Annotations to 3D Geometry

The dense reconstruction stage produces a surface representation of the observed turbine structure, but this surface alone is not yet sufficient for inspection. Defect information is still required before the reconstructed mesh can be used as an inspection-oriented output. In the final pipeline, this information is not generated by online detector inference. Instead, it is provided as precomputed image-domain annotations stored in a JSON file. These annotations represent defect observations over time and serve as the starting point of the 2D-to-3D transfer process.

This design reflects the role of defect information in the present thesis. The goal of the method is not to propose a new detector, but to study how defect evidence that is already available in the image domain can be transferred to reconstructed geometry. For that reason, the annotation file is treated as an external source of localized defect observations. Each annotation is associated with a temporal position in the borescope sequence, so that the image-domain evidence can later be related to registered reconstruction views and, from there, to the final 3D surface.

The mapping stage takes three inputs, namely the image-domain defect annotations, the selected sparse reconstruction, and the final Poisson mesh. In the implementation used in this thesis, the annotation input is stored in `turbine2mask.json`. This file uses timestamp strings as top-level keys, and each timestamp contains defect detections in normalized bounding-box coordinates. The mapping method uses these detection boxes and does not use the polygon masks stored in the same JSON file. The sparse reconstruction is read from the selected SfM model, and the mesh input is the final Poisson surface produced by the dense reconstruction stage.

The first step is temporal association between the annotation stream and the registered reconstruction images. To achieve this, the method reconstructs the sampled image timeline from the video metadata and from the clip settings used during preprocessing. Each registered COLMAP image is matched to a sampled frame by parsing the numeric image filename. For each defect annotation timestamp, the nearest registered image is selected by absolute timestamp difference. This association is accepted only if the time difference is smaller than 0.4 s. The purpose of this step is to ensure that a defect observation is linked only to a reconstruction view that is temporally close to the original image evidence.

This restriction is important because the later geometric transfer is meaningful only when the selected reconstruction view corresponds closely to the annotated inspection moment.

Once the corresponding registered image has been selected, sparse 3D support points are extracted from the SfM model. The normalized defect box is converted to pixel coordinates using the image size of the selected camera view. The method then examines the registered image keypoints and keeps only those keypoints that lie inside the defect box and that have a valid 3D point association. In practice, this means that the defect region is not transferred to 3D directly from the image box itself. Instead, the defect box is used to select the subset of SfM points whose image observations fall inside the annotated defect region. These sparse 3D points form the geometric support of the defect observation. At this stage, the defect evidence has moved from a purely image-domain indication to a geometrically anchored but still sparse 3D representation.

As illustrated in Figure 3.6, the sparse 3D support is then transferred to the mesh. A KD-tree is built over the mesh vertices, and each selected 3D defect point is used to query a local Euclidean neighborhood on the mesh. The coloring radius is defined as

$$r_c = 0.01\, d_{\text{mesh}}, \tag{3.1}$$

where d_{mesh} is the diagonal length of the mesh bounding box. If no mesh vertex falls inside this radius, the nearest single mesh vertex is used instead. This ensures that an isolated support point can still be represented on the surface, although such cases should be interpreted with caution. The role of this step is to convert sparse geometric support into a surface-level representation that can be read more easily during inspection.

(a) Sparse 3D defect support points **(b)** Propagation of defect support to the mesh surface

Figure 3.6: Defect transfer from sparse 3D support to the reconstructed surface. Image-domain defect observations are first associated with sparse 3D support points through the selected SfM model, after which the support is transferred to the Poisson mesh by local mesh-level coloring.

Coloring is applied with a Gaussian decay around each defect support point. For a mesh vertex v and a defect support point p, the influence weight is defined as

$$w(v, p) = \exp\left(-\frac{1}{2}\left(\frac{\|v - p\|}{\sigma}\right)^2\right), \tag{3.2}$$

where $\|v - p\|$ is the Euclidean distance between the mesh vertex and the defect support point, and $\sigma = 0.35\, r_c$. If the mesh does not already contain vertex colors, all vertices are initialized to a neutral gray color. The target defect color is set to red. When multiple defect support points affect the same vertex, the method keeps the maximum influence rather than summing contributions. The final vertex color is then obtained by linearly blending the base gray color with the target defect color according to the strongest local weight. This produces a localized defect-aware surface representation rather than a detached set of 3D markers.

In addition to mesh coloring, the method groups defect observations into events over time. Each accepted detection first produces an event record containing the timestamp and the 3D centroid of the selected defect support. Event grouping is then performed by spatial clustering only. The clustering radius is defined as

$$r_t = 0.02\, d_{\text{mesh}}, \tag{3.3}$$

where d_{mesh} is again the mesh diagonal. Detections are processed in time-sorted order, but no explicit temporal threshold is used during clustering. This means that grouping depends on spatial proximity in 3D rather than on temporal continuity alone. After clustering, the timeline is sorted by start time and cluster identifiers are reassigned sequentially. The result is a set of structured defect events that organize repeated image observations into a more interpretable geometric and temporal record.

The outputs of this stage include a defect-colored Poisson mesh, a point cloud of sparse defect support points, a per-detection event log, and timeline files in both JSON and CSV format. This stage is the central step that turns image-domain defect evidence into a geometry-aware inspection output. The annotation file remains unchanged as the original image input, but the resulting representation makes it possible to inspect the same evidence on the reconstructed turbine surface and to relate it to time-linked defect events.

3.3.6. Visualization Outputs for Inspection Comparison

Figure 3.7: Example of the final 3D inspection representation used for comparison. Defect evidence is displayed directly on the reconstructed blade surfaces as colored regions, which provides a geometry-aware inspection view.

The final stage of the method prepares inspection-oriented visualization outputs from the reconstructed geometry and the mapped defect evidence. The purpose of this stage is not only to display the reconstruction result, but to present defect information in a form that can be compared with a conventional image-based inspection view. In the present thesis, two visualization variants are prepared for this purpose.

The first variant follows the conventional image-domain representation. In this view, defect observations are shown as bounding boxes on the original 2D inspection imagery. Rather than presenting the full video as a continuous sequence, the relevant defect observations are arranged as a gallery of image-based inspection views. This representation preserves the original appearance of the defect evidence and reflects how defect information is commonly interpreted in frame-based inspection.

The second variant uses the reconstructed 3D geometry. As illustrated in Figure 3.7, the mapped defect evidence is shown directly on the turbine surface through mesh coloring. The purpose of this representation is to place defect observations on the reconstructed geometry so that their spatial location can be interpreted relative to the blade surface rather than relative to an isolated image frame. This produces a geometry-aware inspection view in which defect evidence is integrated with the recovered 3D structure.

Both visualization variants are derived from the same underlying inspection sequence and the same defect evidence, but they differ in how that evidence is presented to the user. The 2D variant preserves the original image-domain context, while the 3D variant introduces geometric context through the reconstructed mesh. This distinction is important for the exploratory user evaluation in this thesis, where the two representations are used as the comparison conditions.

At the method level, this stage defines the final inspection outputs of the pipeline. The pipeline does not end at reconstruction or at defect mapping alone. It ends with two comparable visualization forms, namely an image-based gallery with defect bounding boxes and a 3D mesh-based representation with colored defect regions. These outputs provide the basis for evaluating how defect information is interpreted under image-based and geometry-aware inspection views.

3.4. Experiments

3.4.1. Experimental Setup

The experiments are conducted on the final turbine2 pipeline described in the previous section. The input data consist of the borescope video `turbine2.mp4` and the corresponding defect annotation file `turbine2mask.json`. The pipeline uses the clip interval from 2.0 s to 121.0 s of the source video. From this interval, 450 frames are sampled and stored as sequentially numbered reconstruction images. These frames form the image set used for all reconstruction experiments reported in this thesis.

The reconstruction workflow is carried out with HLOC and COLMAP. The sparse stage uses DISK features and LightGlue matching, while the dense stage uses COLMAP image undistortion, dense stereo, stereo fusion, and Poisson meshing. The defect-mapping stage is implemented in Python and uses the selected sparse model, the final Poisson mesh, and the precomputed image-domain annotations. This stage also relies on numerical and geometric processing tools, including `numpy`, `scipy`, `opencv-python`, `plyfile`, and the HLOC binary readers. Final mesh inspection is performed externally through PLY-based visualization, for example in Blender.

All technical reconstruction and mapping results are based on one selected reconstruction case from the segmented and refined sparse reconstruction workflow. The final pipeline uses the reconstruction case `refine_005_0090_0310`, which provides both the sparse model used for defect association and the dense mesh used for downstream visualization. This selected case contains the sparse reconstruction passed to the mapping stage, the dense point cloud `fused.ply`, and the final mesh `mesh_poisson.ply`. For reproducibility, the same selected case is used in both the dense stage and the defect-mapping stage.

The main outputs evaluated in this chapter are the selected sparse reconstruction, the final Poisson mesh, the defect-colored mesh, the sparse defect support points, and the timeline-linked defect outputs. The summary file generated by the final mapping stage reports 220 registered images, 65,087 reconstructed 3D points, 39 matched defect observations, and 9 timeline clusters for the selected reconstruction case. These outputs form the basis of the reconstruction analysis and the defect-mapping analysis presented in the following subsections.

A final reproducibility detail is that the defect-mapping stage does not run online detector inference. Instead, it consumes the pre-existing annotation file `turbine2mask.json`. In addition, the mapping stage assumes the same temporal sampling settings as the reconstruction stage, including the clip range and the total number of sampled frames. The experiments therefore evaluate a fixed technical pipeline in which the reconstruction result and the defect input are both tied to one consistent configuration.

The survey-based comparison between the 2D bounding-box view and the 3D defect-colored visualization is reported separately in the discussion section as an exploratory user evaluation. The present experimental section therefore focuses on the technical outputs of the pipeline, namely reconstruction quality, dense surface generation, defect transfer, and timeline-linked defect mapping.

Table 3.1: Experimental setup and fixed configuration used for the final evaluated pipeline. The table summarizes the input data, preprocessing settings, reconstruction configuration, dense reconstruction stage, and defect-mapping inputs used in the reported experiments.

Category	Setting
Input video	`turbine2.mp4`
Defect annotation input	`turbine2mask.json`
Clip interval	2.0 s to 121.0 s
Sampled reconstruction frames	450 frames
Source frame range	Frame 60 to frame 3606
Fixed ROI	$(x_{\min}, y_{\min}, x_{\max}, y_{\max}) = (0.04,\ 0.11,\ 0.86,\ 0.84)$
Cropped reconstruction frame size	787×525 pixels
Feature extractor	DISK
Maximum keypoints	5000 keypoints per image
Feature matching	LightGlue configured for DISK features
Image-pair strategy	Sequential sliding window with window size 5
Sparse reconstruction backend	HLOC and COLMAP
Camera model	`OPENCV_FISHEYE` with `SINGLE` camera mode
Coarse reconstruction search	180-frame windows with 20-frame step
Number of coarse cases	22 cases
Refinement search	220-frame windows with 10-frame step around the best coarse interval
Number of refinement cases	11 cases
Selected reconstruction case	`refine_005_0090_0310`
Dense reconstruction	COLMAP image undistortion, PatchMatch stereo, stereo fusion, and Poisson meshing
Final mesh used for mapping	`mesh_poisson.ply` from the selected dense workspace
Defect transfer input	Precomputed image-domain bounding-box detections
Temporal matching tolerance	0.4 s
Mesh coloring radius	$r_c = 0.01\, d_{\text{mesh}}$
Timeline clustering radius	$r_t = 0.02\, d_{\text{mesh}}$

Table 3.1 summarizes the fixed configuration used for the reported experiments.

3.4.2. Can the pipeline recover a usable 3D surface from turbine borescope video?

This experiment examines whether the final pipeline can recover a 3D surface that is suitable for downstream defect mapping. The evaluation is based on the selected reconstruction case. In the present thesis, a reconstruction is considered usable if it produces a sufficiently large sparse model, completes the dense stage successfully, yields a continuous Poisson mesh, and is stable enough to serve as the geometric input for the defect-mapping stage.

The selected case produces a sparse reconstruction with 220 registered images and 65,087 reconstructed 3D points. The model uses one shared `OPENCV_FISHEYE` camera and yields a mean reprojection error of 1.1958 pixels and a median reprojection error of 1.1281 pixels. These values indicate that the selected sparse model has a meaningful level of internal geometric consistency, although they do not by themselves guarantee perfect surface accuracy. The same selected case is also stronger than the best coarse segment. The strongest coarse reconstruction, `seg_006_0120_0300`, contains 180 registered images and 54,478 3D points. The selected refinement case therefore improves the coarse result by 40 registered images, which corresponds to about 22.2%, and by 10,609 reconstructed 3D points, which corresponds to about 19.5

The dense stage also completes successfully for the selected case. The fused dense point cloud contains 870,096 vertices, and the final Poisson mesh contains 1,375,199 vertices and 2,711,283 faces.

This is an important result for the present thesis, because sparse geometry alone is not sufficient for surface-based defect display. The dense reconstruction and meshing stages produce a continuous surface representation that can support later annotation and visualization. In the final pipeline, this Poisson mesh is the mesh passed to the downstream defect-mapping stage, which means that the reconstruction is not only visually plausible, but also usable as a geometric input within the technical pipeline.

Table 3.2: Summary of the selected reconstruction and dense meshing result. The selected refined case provides the sparse model used for defect association and the Poisson mesh used for downstream defect visualization.

Quantity	Value
Selected reconstruction case	`refine_005_0090_0310`
Registered images	220
Reconstructed sparse 3D points	65,087
Mean reprojection error	1.1958 px
Median reprojection error	1.1281 px
Dense point cloud vertices	870,096
Poisson mesh vertices	1,375,199
Poisson mesh faces	2,711,283
Final dense point cloud	`fused.ply`
Final Poisson mesh	`mesh_poisson.ply`

Table 3.2 summarizes the main sparse and dense reconstruction outputs of the selected case. The selected case is also clearly stronger than weaker alternatives in the same run. Across the final turbine2 output set, 33 coarse and refined cases are generated. Several of these cases remain substantially weaker. For example, `seg_021_0420_0450` contains only 30 registered images and 9,655 3D points, while `seg_014_0280_0450` contains 41 registered images and 14,593 3D points. In addition, some refinement attempts do not produce a valid sparse model at all. For instance, `refine_010_0140_0360` contains database and log files but no reconstructed `images.bin` or `points3D.bin` model outputs. These weaker cases show that the reconstruction problem is not uniformly easy across the full video, and that a usable surface depends on selecting a locally strong reconstruction interval.

Table 3.3: Comparison between the selected reconstruction case and representative alternative cases. The selected refined case improves on the best coarse result and is substantially stronger than weaker temporal windows in the same reconstruction search.

Case	Type	Reg. images	3D points	Role
`seg_006_0120_0300`	Best coarse	180	54,478	Strongest coarse candidate
`refine_005_0090_0310`	Selected refined	220	65,087	Final selected case
`seg_021_0420_0450`	Weak coarse	30	9,655	Weak alternative
`seg_014_0280_0450`	Weak coarse	41	14,593	Weak alternative
`refine_010_0140_0360`	Failed refinement	–	–	No valid sparse model output

Table 3.3 compares the selected refined case with the strongest coarse case and several weaker alternatives from the same reconstruction search. Together, the results show that the final pipeline can recover a usable 3D surface from the turbine borescope video for the selected reconstruction case. The selected case provides both a substantial sparse model and a large Poisson mesh, and this mesh is later used for defect mapping. For the purposes of this thesis, this answers the reconstruction question positively, but only under the important condition that the pipeline is allowed to select the strongest reconstruction case from the segmented and refined search.

3.4.3. Can image-domain defect evidence be transferred to reconstructed geometry?

This experiment examines whether defect evidence that is available in the image domain can be transferred to the reconstructed 3D geometry. The evaluation is based on the final mapping result built on the selected reconstruction case `refine_005_0090_0310`. In the present thesis, successful transfer means that a defect annotation can be matched to a registered reconstruction view, linked to valid sparse 3D support points, and then displayed on the final Poisson mesh as part of the geometry-aware defect representation.

The image-domain annotation input contains 3,745 timestamp entries. Among these, 134 timestamps contain defect detections, giving a total of 135 detection entries. The same annotation file also contains polygon masks, but the final mapping method uses the bounding-box detections rather than the polygon masks. This means that the experiment evaluates transfer from image-domain defect boxes to 3D geometry, not transfer from dense image masks.

The first requirement for transfer is temporal alignment between the defect annotation and a registered reconstruction image. Under the final pipeline settings, 41 out of the 135 detections can be matched to a registered image within the accepted time tolerance of 0.4 s. The remaining 94 detections are rejected because no sufficiently close registered image is available. This corresponds to a time-match rate of 30.4%. The dominant loss at this stage is therefore temporal mismatch rather than failure of the geometric association itself.

Among the 41 time-matched detections, 39 produce valid sparse 3D support points in the selected SfM model. Only 2 detections remain without valid 3D support after time matching. This gives a 3D-supported transfer rate of 95.1% among the time-matched detections, and an overall 3D-supported transfer rate of 28.9% across all 135 input detections. These numbers show that once temporal alignment is satisfied, the mapping stage usually succeeds in associating the defect evidence with reconstructed 3D structure.

Table 3.4 summarizes the transfer process from the image-domain annotation stream to the final geometry-aware outputs.

Table 3.4: Defect-transfer funnel from image-domain annotations to geometry-aware outputs. Most detections are lost during temporal matching, while almost all time-matched detections obtain valid sparse 3D support. This indicates that the main bottleneck is temporal alignment rather than sparse 3D support extraction after matching.

Transfer stage	Count	Interpretation
Timestamp entries in annotation file	3,745	Total timestamp entries in the input annotation stream
Timestamps with detections	134	Timestamps containing at least one defect detection
Input defect detections	135	Total image-domain bounding-box detections
Time-matched detections	41	Matched to a registered reconstruction image within 0.4 s
Rejected by temporal mismatch	94	No sufficiently close registered reconstruction image available
Detections with valid 3D support	39	Successfully associated with sparse 3D support points
Detections without valid 3D support	2	Time-matched, but no valid sparse 3D support inside the defect box
Unique sparse 3D support points	2,457	Sparse points used for mesh coloring
Timeline clusters	9	Grouped defect events in the final timeline output

The successful defect transfers produce 2,457 unique sparse 3D support points. These points are then used to color the final Poisson mesh, resulting in a defect-aware 3D representation of the turbine surface. In the selected reconstruction case, all 39 successfully transferred detections contribute to the colored mesh. The same stage also produces 39 per-detection records and groups them into 9 timeline clusters. The support size per transferred detection ranges from 10 to 751 sparse points, with a mean of 137.33 points per transferred detection. The resulting timeline spans from 32.7418 s to 76.4534 s in the selected clip.

The failure cases help clarify the main limitation of the current transfer process. Most lost detections are not caused by the absence of 3D support after matching, but by the inability to align the annotation timestamp to a registered reconstruction image. This means that the main bottleneck is the overlap between the annotation timeline and the set of registered reconstruction views. Once a defect box can be assigned to a valid registered image, the subsequent transfer to sparse 3D support usually succeeds.

Taken together, the results show that the final pipeline can transfer image-domain defect evidence to reconstructed geometry for the selected reconstruction case. A total of 39 defect annotations are supported by sparse 3D points and propagated onto the final Poisson mesh, producing both a geometry-aware defect visualization and clustered timeline outputs. For the purposes of this thesis, this answers the transfer question positively, but with an important limitation. The transfer works only for the subset of annotations that can be temporally aligned with registered reconstruction views, so the coverage of the current method is mainly limited by temporal alignment.

3.4.4. Which design choices matter most?

This experiment examines which design choices are most important for obtaining the final usable reconstruction and mapping result. The focus is on four aspects of the pipeline: segmented reconstruction, refinement around the best coarse segment, the fixed preprocessing configuration, and the use of one selected reconstruction case for downstream mapping.

The first important design choice is segmented reconstruction. The final pipeline does not treat the full video as one reconstruction problem. Instead, it evaluates 22 overlapping coarse segments. The quality spread across these segments is large. The number of registered images ranges from 30 to 180, while the number of reconstructed 3D points ranges from 9,655 to 54,478. This spread shows that the reconstruction strength varies strongly across the target clip. Some temporal windows contain enough stable overlap to support a strong sparse model, while others do not. The selected final result therefore depends on searching for a locally strong interval rather than assuming that the full sequence can be reconstructed uniformly.

The second important design choice is refinement. After the coarse pass, the pipeline performs a second reconstruction pass around the strongest coarse interval. This step is not a minor adjustment. The best coarse case, `seg_006_0120_0300`, contains 180 registered images and 54,478 3D points. The best refined case, `refine_005_0090_0310`, increases this to 220 registered images and 65,087 3D points. This corresponds to an increase of 22.2% in registered images and 19.5% in reconstructed 3D points. The final selected case used downstream is therefore not the strongest coarse result, but the strongest refined result. This shows that refinement contributes directly to the quality of the final output.

The fixed preprocessing configuration is also important, although the available evidence is more limited. The final pipeline uses one fixed clip interval, one fixed crop, and one fixed frame-sampling strategy, which together define the 450-frame reconstruction set. The same temporal settings are also used during defect mapping for timestamp alignment. This means that preprocessing is not only a reconstruction step, but also part of the alignment between reconstruction and mapping. At the same time, the present experiment does not include a direct ablation against alternative crop or sampling settings. For this reason, the role of preprocessing can be described as operationally important for consistency, but not as independently validated through a within-run comparison.

A fourth important design choice is the use of one selected reconstruction case for downstream mapping. The mapping stage does not search again over all sparse or dense results. Instead, it uses one fixed sparse model and one fixed Poisson mesh from the selected refined case. This choice matters because the output set contains many weaker reconstruction cases, and only the selected case provides the dense mesh used for the final defect-mapping result. In practice, this means that downstream defect transfer depends on committing to one reconstruction result that is strong enough to support surface-based visualization. This makes the pipeline consistent across dense reconstruction and defect mapping, but it also means that the final result remains dependent on the selected reconstruction case.

Table 3.5 summarizes the evidence for each design choice and how it affects the final reconstruction and mapping result.

Table 3.5: Summary of the main design choices and their observed effect in the final experiment. Segmentation and refinement are directly supported by reconstruction statistics, while preprocessing is mainly supported by its role in maintaining consistency between reconstruction and defect mapping.

Design choice	Evidence in the final experiment	Interpretation
Segmented reconstruction	22 overlapping coarse cases are evaluated. Registered images range from 30 to 180, and sparse 3D points range from 9,655 to 54,478.	Reconstruction quality varies strongly across the clip, so the full sequence should not be treated as one uniformly stable reconstruction problem.
Refinement around best coarse segment	The best coarse case increases from 180 registered images and 54,478 points to 220 registered images and 65,087 points after refinement.	Refinement substantially improves the selected result and directly contributes to the final usable sparse model.
Fixed preprocessing configuration	One fixed clip interval, ROI, and frame-sampling strategy define the 450-frame reconstruction set and are reused for timestamp alignment during defect mapping.	Preprocessing is operationally important for consistency between reconstruction and mapping, although its independent contribution is not isolated by an ablation.
Selected-case consistency	The downstream mapping stage uses one fixed sparse model and one fixed Poisson mesh from `refine_005_0090_0310`.	Using one selected case keeps dense reconstruction and defect mapping consistent, but also makes the final result dependent on the selected reconstruction case.

Taken together, the evidence shows that the most important design choices are segmented reconstruction, refinement, and selected-case consistency. Segmentation makes it possible to search for strong local reconstruction intervals. Refinement improves the best coarse result substantially. Using one selected reconstruction case ensures that the same sparse model and dense mesh are used consistently in the downstream defect-mapping stage. The fixed preprocessing setup is also important for consistency between reconstruction and mapping, although its individual contribution is not isolated by a direct ablation in the final experiment.

3.4.5. When does the pipeline fail or become weak?

The previous experiments show that the final pipeline can recover a usable 3D surface and can transfer a subset of the image-domain defect evidence to reconstructed geometry. This experiment examines where the pipeline remains weak. The goal is not to invalidate the selected successful result, but to identify the main bottlenecks that limit robustness across the evaluated reconstruction cases.

The first weakness appears in sparse reconstruction quality across the segmented search. The final reconstruction search evaluates 33 cases in total, consisting of 22 coarse segments and 11 refinement cases. Among these, one case does not produce a valid sparse model at all. In addition, several cases remain clearly weak. For example, some coarse cases contain only 30 to 50 registered images and fewer than 18,000 reconstructed 3D points. More generally, 7 out of the 33 cases contain fewer than 80 registered images, and 6 out of the 33 cases contain fewer than 25,000 reconstructed 3D points. These results show that reconstruction quality varies strongly across the clip and that not all temporal windows are suitable for downstream geometric analysis.

The second weakness appears at the dense stage. In the final technical pipeline, only the selected reconstruction case is carried forward to dense reconstruction and defect mapping. This means that the experiment demonstrates dense surface generation for one strong selected case, but does not demonstrate that dense geometry can be recovered robustly across all candidate reconstruction windows. In other words, the current workflow depends on selecting one strong local reconstruction interval rather than obtaining dense geometry reliably across many temporal windows. This does not invalidate the selected successful case, but it does show that dense reconstruction remains a limited part of the current proof-of-concept pipeline.

The third and largest weakness appears in the transfer from image-domain defect evidence to reconstructed geometry. The input annotation stream contains 135 defect detections. Among these, 94 detections are rejected because no registered reconstruction image falls within the accepted time tolerance. Only 41 detections remain after temporal matching. Of these, 39 detections produce valid sparse 3D support, while 2 detections remain without valid 3D support inside the annotated bounding box. This means that the dominant loss mode is not failure of the geometric transfer once matching has succeeded, but temporal mismatch between the annotation timestamps and the registered reconstruction views.

A further source of weakness appears in the amount of sparse geometric support available for accepted detections. Even among the successfully transferred defect observations, the number of supporting sparse points varies substantially. The matched support count ranges from 10 to 751 points per transferred detection, and 4 transferred detections contain 20 points or fewer. This does not prevent transfer entirely, but it indicates that some mapped defects are supported by only a small amount of sparse geometry. Such cases are likely to be less reliable for detailed interpretation than detections with stronger spatial support.

Taken together, the experiments show that the main limitation of the current pipeline is not the final 2D-to-3D transfer step itself. Once a defect observation is temporally aligned and valid sparse support exists, transfer to the reconstructed mesh usually succeeds. The main bottlenecks occur earlier in the pipeline: sparse reconstruction quality varies across temporal windows, dense reconstruction is demonstrated only for the selected case, and many defect annotations cannot be temporally matched to registered reconstruction views. These failure modes define the main directions for improvement in future work.

3.5. Discussion

3.5.1. Main Findings

The results show that the proposed pipeline can recover a usable 3D representation from challenging turbine borescope video and can connect a subset of the available defect evidence to that geometry. The main outcome of the thesis is therefore not only that reconstruction is possible in this setting, but that the reconstructed surface can serve as the basis for a defect-aware inspection representation.

The first main finding is that the final pipeline produces a usable geometric result when a strong local reconstruction window is selected. In the selected reconstruction case, the sparse stage provides sufficient image registration and 3D support for the dense stage to complete successfully, and the dense stage produces a Poisson mesh that can be used in later processing. This is an important result for the thesis, because the relevant threshold is not the existence of a sparse model alone. The relevant threshold is whether the pipeline reaches a surface representation that can support defect transfer and inspection-oriented display.

The second main finding is that image-domain defect evidence can be lifted into reconstructed 3D geometry when temporal alignment and sparse support are available. In the selected reconstruction case, the mapping stage produces a defect-colored mesh together with sparse defect support points and timeline-linked defect outputs. This shows that defect observations do not need to remain isolated in 2D image frames. Instead, they can be associated with the reconstructed turbine surface and organized in a form that carries both spatial and temporal information.

The third main finding is that the final usable result depends strongly on the structure of the pipeline itself. The segmented search, the local refinement stage, and the use of one selected reconstruction case are not minor implementation details. These choices determine whether the pipeline reaches a usable surface or remains at a weak or incomplete reconstruction. The experiments therefore show that the final result is not obtained by applying reconstruction once to the full sequence, but by searching for a locally strong interval and carrying that selected case forward to the dense and mapping stages.

At the same time, the results also define the current boundary of the method. The pipeline is effective on the selected reconstruction case, but it is not yet uniformly reliable across the full sequence. Reconstruction quality varies across temporal windows, dense reconstruction is demonstrated only for the selected case, and a large fraction of the defect evidence is lost at the temporal alignment stage before geometric transfer can begin. For this reason, the present thesis should be read as proof-of-concept evidence for an end-to-end defect-aware 3D inspection pipeline, rather than as a claim of full robustness across all candidate windows.

Taken together, the technical experiments support a clear conclusion. The proposed method makes it possible to move from borescope video and image-domain defect observations to a geometry-aware inspection representation on the turbine surface. This is the main technical contribution of the thesis. The current pipeline demonstrates that the approach is feasible for a selected reconstruction case, while also showing that further work is needed before the same level of reliability can be expected across the full sequence.

3.5.2. Exploratory User Evaluation

To complement the technical results, an anonymous user evaluation was conducted. The evaluation compared two inspection representations. Condition A used the current image based inspection representation from AIIR Innovations, where defect evidence is shown in 2D inspection frames. Condition B used the 3D geometry aware representation developed in this thesis, where defect evidence is shown on reconstructed turbine blade geometry.

The comparison was not intended as a final validation of a complete inspection system. The 2D condition represents an existing and more developed inspection view, while the 3D condition represents the current research prototype of this thesis. The purpose was therefore to understand how the current 3D output is perceived by users, where it is weaker than the 2D view, and whether it still provides useful spatial information for inspection.

The survey was distributed anonymously. To reduce the influence of one specific visual example, two questionnaire variants were used. Both variants followed the same structure and compared the same two types of inspection representation, but they used different image sets. In total, eight responses were collected, with four responses for each questionnaire variant.

The survey first asked participants to inspect each condition separately. The questions focused on the number of visible detects, the detects that should receive the highest inspection priority, and the confidence of the participant. After both conditions had been reviewed, the survey asked participants to compare the two conditions directly. The direct comparison focused on four aspects. These were detect counting, identifying the most affected blade, prioritizing detects by apparent size or severity, and preference for practical inspection use. Participants were also asked to briefly explain their preference.

Table 3.6: Summary of the anonymous user evaluation structure. The survey first asked participants to inspect each condition separately, and then asked them to compare the two representations directly.

Survey part	What was measured
Condition specific task	Number of visible detects in each condition
Condition specific task	Highest and second highest priority detect in each condition
Condition specific task	Confidence for each condition on a five point scale
Direct comparison	Easier condition for counting detects
Direct comparison	Easier condition for finding the most affected blade
Direct comparison	Easier condition for prioritizing detects
Direct comparison	Preferred condition for practical inspection use
Open feedback	Short explanation of the participant preference

The direct comparison results are shown in Table 3.7. The result is not uniformly positive for the 3D representation. For identifying the total number of detects, six out of eight respondents preferred the 2D condition. For prioritizing detects by apparent size or severity, five respondents preferred the 2D condition, two preferred the 3D condition, and one reported no clear difference. For practical inspection use, six respondents preferred the 2D condition. These results show that the current 3D representation is not yet clear enough to replace the 2D view for direct inspection tasks.

Table 3.7: Direct comparison results between the 2D and 3D inspection views. The 2D view was preferred for counting, prioritization, and practical use. The 3D view was preferred for identifying the most affected blade, which suggests that the main value of the 3D view is related to spatial interpretation.

Comparison aspect	Condition A 2D	Condition B 3D	No clear difference
Counting detects	6 (75.0%)	2 (25.0%)	0 (0.0%)
Finding the most affected blade	3 (37.5%)	5 (62.5%)	0 (0.0%)
Prioritizing detects	5 (62.5%)	2 (25.0%)	1 (12.5%)
Practical inspection preference	6 (75.0%)	2 (25.0%)	0 (0.0%)

The confidence scores show the same general pattern. As shown in Table 3.8, the average confidence was higher for the 2D condition than for the 3D condition. This means that participants were more certain when using the image based view. This is important because the goal of the 3D representation is not only to show defects in another format, but also to make the defect information easier to interpret.

Table 3.8: Confidence ratings for the two inspection views. Participants reported higher confidence for the 2D condition, which shows that the current 3D presentation still lacks clarity for direct inspection use.

Condition	Mean confidence	Median confidence	Range
Condition A 2D image based view	3.63	4.0	1 to 5
Condition B 3D geometry aware view	2.50	2.0	1 to 5

However, the evaluation also shows that the 3D representation has potential. The only comparison question where the 3D condition was preferred by most respondents was the question about which blade was most affected. Five out of eight respondents preferred the 3D condition for this task. This suggests that the 3D view is currently more useful for understanding spatial organization than for counting or ranking individual defects. In this sense, the 3D representation should not be seen as a direct replacement for the 2D view at this stage. It is more realistic to interpret it as a possible supporting view that adds blade level context to the original image based inspection result.

The qualitative feedback explains why the 3D condition was weaker in several tasks. Several respondents mentioned that the 3D images were too small, low in resolution, or difficult to inspect in the Google Forms layout. Some respondents also wrote that it was not always clear whether the highlighted 3D regions matched the defects in the source images. Other feedback mentioned that the blade orientation was unclear, especially the relation between the blade root and tip. These comments indicate that the weaker result of the 3D condition was strongly related to the current presentation quality.

This is an important limitation of the evaluation. The survey compared a relatively mature 2D inspection representation with a static 3D prototype. The 3D result was shown as a fixed image, although the intended use of a 3D representation would normally require interaction, rotation, zooming, and links back to the original inspection frames. The survey also had a small number of anonymous respondents, and detailed participant background was not collected. For this reason, the result should be interpreted as exploratory feedback rather than as a final conclusion about the value of 3D inspection support.

Overall, the user evaluation shows a clear gap between technical feasibility and practical usability. The technical pipeline can produce a defect aware 3D representation, but the current presentation does not yet make this representation easy enough to use for inspection decisions. At the same time, the stronger result for identifying the most affected blade shows that 3D geometry can still provide useful spatial context. The current evaluation therefore supports a careful conclusion. The 3D representation does not yet outperform the 2D view, but it points to a promising direction for spatial defect interpretation if the level of detail, orientation support, and interaction quality are improved.

3.5.3. What the Results Mean for Inspection

The final result changes how defect evidence can be interpreted during inspection. In a 2D video sequence, a defect indication remains tied to one frame, one viewpoint, and one moment in the inspection pass. This makes interpretation dependent on camera motion and viewing angle, and it can remain unclear whether similar observations across multiple frames refer to the same physical region or to different regions. In the reconstructed 3D representation, the same defect evidence is attached to a surface location on the turbine model. This makes the observation more spatially persistent rather than only frame-local, and gives the inspector a more stable geometric context for interpretation.

A second practical implication is that the final representation combines location and time. The colored mesh shows where suspicious regions are located on the reconstructed surface, while the timeline output shows when these regions were observed during the inspection sequence. This is useful because inspection does not only depend on detecting a suspicious signal. Inspection also depends on being able to relate that signal to the surrounding structure and to the moment at which it appeared. The combined output therefore provides a more structured basis for review than raw frame inspection alone.

At the same time, the present results should be interpreted with the correct scope. The completed technical experiments support the claim that the final pipeline can produce a usable 3D surface for the selected turbine2 case and can transfer part of the image-domain defect evidence to that surface with time-linked outputs. They also support the claim that the representation is inspection-oriented as a technical output. However, the current results do not show that the 3D view is superior to the existing 2D image-based view for all inspection tasks.

The exploratory user evaluation gives a more cautious interpretation. Participants preferred the 3D geometry-aware view for identifying the most affected blade, which suggests that the reconstructed representation can support blade-level spatial interpretation. However, the 2D image-based view was preferred for counting detects, prioritizing detects, and practical inspection use. Confidence ratings were also higher for the 2D view. This means that the current 3D representation should not be interpreted as a replacement for the 2D inspection view. Instead, it is better understood as a supporting representation that can add spatial context, while still requiring a clearer and more interactive interface before it can be practically useful in inspection workflows.

For this reason, the current result is best understood as technical feasibility evidence for geometry-aware inspection support. The thesis shows that defect evidence can be moved from image space to a reconstructed surface and organized with time-linked outputs. The user evaluation suggests that this direction has potential for spatial interpretation, but it also shows that the current static 3D presentation is not yet sufficient to outperform the existing 2D view in practical inspection tasks.

3.5.4. Limitations of the Current Pipeline

The current pipeline demonstrates that defect-aware 3D inspection support is feasible on the selected turbine2 case, but the experiments also make clear that the method remains constrained in several important ways. These limitations do not remove the main contribution of the thesis, but they define the scope within which the current findings should be interpreted.

The first limitation is reconstruction stability across the full sequence. Although the final pipeline produces one strong selected reconstruction case, reconstruction quality varies substantially across the evaluated temporal windows. Some windows yield weak sparse support, and at least one refinement case does not produce a valid sparse model. This means that the current method cannot yet be described as uniformly reliable over the full inspection sequence. Instead, the method should be understood as a pipeline that can recover a usable result when a sufficiently strong local reconstruction interval is present.

The second limitation is the dependence of downstream processing on a single selected dense case. The final Poisson mesh used for defect transfer is obtained from one selected reconstruction window. This is sufficient to demonstrate the end-to-end feasibility of the method, but it also means that dense geometry is not yet shown to be equally reliable across the broader candidate set. In practical terms, the current pipeline still depends on selecting a favorable reconstruction case rather than obtaining equally usable dense surfaces throughout the sequence.

The third limitation is the loss of defect evidence during temporal alignment. The completed experiments show that the largest loss mode occurs before geometric transfer itself. A substantial portion of the image-domain defect detections cannot be matched to a registered reconstruction image within the accepted time tolerance. As a result, the final 3D defect representation remains incomplete relative to the available annotation stream. This limitation is important because it shows that the main bottleneck is not only geometric transfer, but also alignment between the inspection timeline and the registered reconstruction views.

A fourth limitation is the variability of geometric support among accepted defect observations. Even when defect transfer succeeds, the number of supporting sparse 3D points differs substantially across detections. Some mapped regions are supported by many points, while others rely on only a small number of observations. This means that the final defect-colored mesh should not be interpreted as having uniform local confidence across all highlighted regions. The representation is informative, but the geometric support behind individual mapped detections remains heterogeneous.

A fifth limitation concerns the current visualization and user-facing interpretation. The exploratory user evaluation showed that the 3D representation was not preferred over the 2D image-based view for counting detects, prioritizing defects, or practical inspection use. This suggests that the current static 3D output is not yet clear enough to replace the existing 2D inspection view. The 3D representation may support blade-level spatial interpretation, but it still needs better interaction, clearer orientation cues, and stronger links back to the original image evidence before it can become practically useful for inspection workflows.

The current evaluation setup also limits the strength of the conclusions. The results are based on one selected reconstruction case from one turbine inspection sequence. This is appropriate for demonstrating technical feasibility, but it does not yet support broad claims across multiple videos, inspection conditions, or automatic case-selection strategies. In addition, the current pipeline does not perform detector inference online. It uses precomputed image-domain defect annotations as fixed input, which means that the present thesis evaluates geometric transfer and visualization rather than a complete end-to-end detection system.

Taken together, these limitations show that the current thesis should be read as a proof-of-concept study rather than a final operational system. The pipeline demonstrates that 3D reconstruction and geometry-aware defect transfer can be combined for turbine inspection video. At the same time, the experiments and user evaluation show that reconstruction stability, temporal alignment, support consistency, interface clarity, and broader validation remain open challenges.

3.5.5. Future Work

Several directions follow from the current results. A first priority is to reduce the dependence on one selected reconstruction window. The completed experiments show that the quality of the reconstruction varies considerably between temporal segments. The final usable surface is still dependent on a selected reconstruction case. Future work should replace this selected-case dependence with a more explicit quality-based selection strategy. This should include confidence measures for sparse reconstruction, dense reconstruction on multiple strong candidates, and a clearer rule for selecting the final surface used for defect mapping.

A second direction concerns the defect-mapping stage. The current method can transfer defect evidence to 3D when temporal alignment is available and sparse support exists. However, this support is not equally strong for all events. The user evaluation also showed that some participants were not confident that the highlighted 3D regions matched the visible defects. Future work should therefore improve the reliability of the mapping itself. This could be done by using multi-view support, stronger consistency checks across frames, and confidence scores for mapped defect regions.

A third direction is temporal alignment. In the current pipeline, many detections are lost before the geometric transfer stage because they cannot be aligned to registered reconstruction frames within the accepted time threshold. This means that the main bottleneck is not only the 3D mapping method, but also the synchronization between video frames, detection results, and reconstruction images. Future work should therefore use deterministic synchronization between frame extraction, annotation timing, and reconstruction indexing. This would allow more defect evidence to reach the 3D stage.

A fourth direction is the presentation of the 3D result. The user evaluation showed that the current static 3D visualization was not preferred over the 2D view for counting defects, prioritizing defects, or practical inspection use. The feedback suggests that this was partly caused by the way the 3D output was presented. The images were small, the resolution was limited, the blade orientation was not always clear, and the relation between the highlighted mesh regions and the original image evidence was difficult to check. Future work should therefore move from static images to an interactive inspection interface. This interface should support zooming, rotation, blade identifiers, root and tip orientation cues, and direct links between 3D mesh regions and the original 2D frames.

A fifth direction is a larger user evaluation after improving the 3D interface. The current anonymous evaluation was useful for identifying weaknesses, but it remains limited. The number of responses was small, the participant background was not recorded in detail, and response time was not measured as part of the final analysis. Future work should therefore repeat the evaluation with more domain experts and with a more complete interactive prototype. Such a study should measure not only preference, but also counting accuracy, localization accuracy, prioritization accuracy, confidence, and task completion time.

A final direction is system integration. The current result is still a research prototype. It can produce meaningful outputs, but it still depends on selected-case handling and manual organization of intermediate files. Future work should consolidate the pipeline into a reproducible inspection workflow with versioned parameters, automatic quality reporting, and an operator-facing interface. This would move the method closer to a practical inspection support system while keeping the main idea of this thesis intact. The key lesson from the user evaluation is that a technically possible 3D reconstruction is not sufficient on its own. The result also has to be clear, inspectable, and trustworthy for the user.

3.6. Conclusion

This thesis investigated how turbine borescope video can be used to build a defect-aware 3D inspection representation. The starting point of the work was the difficulty of interpreting turbine defects from image sequences alone. Borescope video provides useful visual evidence, but it gives limited spatial context. This makes it difficult to understand where a defect lies on the blade surface and how that defect relates to the surrounding geometry. The goal of this thesis was therefore not only to reconstruct turbine geometry, but to connect reconstruction, defect evidence, and visualization in one inspection-oriented pipeline.

The final pipeline shows that this connection is technically possible for the selected turbine sequence. A direct reconstruction of the full video was not reliable enough, because the footage contains restricted viewpoints, repeated blade structure, reflective surfaces, weak texture, and unstable camera motion. To address this, the final pipeline uses fixed ROI cropping, frame extraction, segmented sparse reconstruction, refinement around a strong reconstruction interval, dense reconstruction, and Poisson mesh generation. In the selected reconstruction case, the pipeline produces a sparse model with 220 registered images and 65,087 reconstructed 3D points. The dense stage then produces a surface representation that can be used for downstream defect mapping. This result shows that usable 3D geometry can be recovered from difficult turbine borescope footage when the reconstruction problem is constrained and a strong local reconstruction interval is selected.

The thesis also shows that image-domain defect evidence can be transferred to reconstructed geometry. The final mapping stage links precomputed defect annotations to registered reconstruction views, associates them with sparse 3D support, and propagates the result onto the final mesh. In the selected reconstruction case, 39 defect annotations are supported by sparse 3D points and organized into 9 timeline clusters. This means that the pipeline does not stop at visual reconstruction. It also creates a connection between observed image evidence, reconstructed blade geometry, and time-linked defect events. This connection is the main technical contribution of the thesis.

The user evaluation gives a more careful view of the practical value of the current output. The evaluation did not show that the current 3D representation is better than the 2D image-based view for all inspection tasks. Participants preferred the 2D view for counting detects, prioritizing detects, and practical inspection use. They also reported higher confidence for the 2D view. This result is important because it shows that technical feasibility is not the same as inspection usability. A 3D representation can be produced, but it still has to be clear enough for users to trust and interpret. At the same time, the 3D view was preferred for identifying the most affected blade. This suggests that the current value of the 3D representation is mainly in blade-level spatial interpretation, rather than in replacing the 2D view.

The current thesis should therefore be understood as a proof-of-concept study rather than as a complete operational inspection system. The strongest result depends on one selected reconstruction case, and the pipeline is not yet uniformly reliable across the full inspection sequence. The defect-mapping stage is also limited by temporal alignment, uneven sparse support, and the use of precomputed image-domain annotations. The user evaluation adds another limitation. The current static 3D presentation does not yet provide enough detail, orientation support, or interaction to make users consistently prefer it over the existing 2D representation.

Even with these limitations, the thesis provides a meaningful step toward geometry-aware inspection support. The work shows that reconstructed geometry can serve as an intermediate representation between image-based defect evidence and human interpretation. It also shows that the main challenge is not only to reconstruct a turbine surface, but to make the reconstructed surface useful for inspection. The final conclusion of this thesis is therefore careful but positive. The current 3D representation does not yet outperform the 2D inspection view as a standalone tool, but it shows potential as a supporting representation for spatial defect interpretation. With stronger reconstruction stability, better temporal alignment, clearer defect mapping, and an interactive inspection interface, the same direction could become more useful for practical turbine inspection workflows.

3.7. AI Statement

Artificial intelligence tools were used during the preparation of this thesis. Canva AI was used for visual post-processing of selected figures, mainly to improve image contrast, background clarity, and readability. ChatGPT was used as an assistance tool for programming support, including code structuring, optimization, debugging, and basic skeleton development.

ChatGPT was also used for writing support. This included grammar checking, paragraph rephrasing, sentence-level rewriting, readability improvement, and academic phrasing.

All research decisions, experimental design, implementation choices, results, interpretations, and final claims were reviewed and controlled by the author.

3.8. Cover Image Credit

Cover image: photograph by Alvaro Pinot, published on Unsplash on January 5, 2018. Used under the Unsplash License and modified for the thesis cover.

References

[1] J. Aust and D. Pons. “Comparative Analysis of Human Operators and Advanced Technologies in the Visual Inspection of Aero Engine Blades”. In: *Applied Sciences* 12.4 (2022). DOI: 10.3390/app12042250, p. 2250.

[2] J. Aust et al. “Automated Defect Detection and Decision-Support in Gas Turbine Blade Inspection”. In: *Aerospace* 8.2 (2021). DOI: 10.3390/aerospace8020030, p. 30.

[3] C. Barnes et al. “PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing”. In: *ACM Transactions on Graphics* 28.3 (2009), 24:1–24:11.

[4] M. Bleyer, C. Rhemann, and C. Rother. “PatchMatch Stereo – Stereo Matching with Slanted Support Windows”. In: *Proceedings of the British Machine Vision Conference*. DOI: 10.5244/C.25.14. 2011, pp. 14.1–14.11.

[5] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao. “YOLOv4: Optimal Speed and Accuracy of Object Detection”. In: *arXiv preprint arXiv:2004.10934* (2020). DOI: 10.48550/arXiv.2004.10934.

[6] CGAL. *Poisson Surface Reconstruction: User Manual*. 2026. URL: `https://doc.cgal.org/latest/Poisson_surface_reconstruction_3/index.html` (visited on 04/14/2026).

[7] CGAL. *Polygon Mesh Processing: Connected Components*. 2026. URL: `https://doc.cgal.org/latest/Polygon_mesh_processing/group__PMP__keep__connected__components__grp.html` (visited on 04/14/2026).

[8] COLMAP. *Command-line Interface*. 2026. URL: `https://colmap.github.io/cli.html` (visited on 04/14/2026).

[9] COLMAP. *Tutorial*. 2026. URL: `https://colmap.github.io/tutorial.html` (visited on 04/14/2026).

[10] CVG. *Hierarchical Localization*. 2026. URL: `https://github.com/cvg/Hierarchical-Localization` (visited on 04/14/2026).

[11] K. Deng et al. “Depth-Supervised NeRF: Fewer Views and Faster Training for Free”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. DOI: 10.1109/CVPR52688.2022.01 2022, pp. 12826–12835.

[12] D. DeTone, T. Malisiewicz, and A. Rabinovich. “SuperPoint: Self-Supervised Interest Point Detection and Description”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. DOI: 10.1109/CVPRW.2018.00060. 2018, pp. 224–236.

[13] Y. Furukawa and J. Ponce. “Accurate, Dense, and Robust Multiview Stereopsis”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.8 (2010). DOI: 10.1109/TPAMI.2009.161, pp. 1362–1376.

[14] K. He et al. “Mask R-CNN”. In: *Proceedings of the IEEE International Conference on Computer Vision*. DOI: 10.1109/ICCV.2017.322. 2017, pp. 2980–2988.

[15] M. Kazhdan, M. Bolitho, and H. Hoppe. “Poisson Surface Reconstruction”. In: *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*. 2006, pp. 61–70.

[16] K. N. Kutulakos and S. M. Seitz. “A Theory of Shape by Space Carving”. In: *International Journal of Computer Vision* 38.3 (2000). DOI: 10.1023/A:1008191222954, pp. 199–218.

[17] Aldo Laurentini. “The Visual Hull Concept for Silhouette-Based Image Understanding”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16.2 (1994), pp. 150–162. DOI: `10.1109/34.273735`.

[18] P. Lindenberger, P.-E. Sarlin, and M. Pollefeys. “LightGlue: Local Feature Matching at Light Speed”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. DOI: 10.1109/ICCV51070.2023.01616. 2023, pp. 17627–17638.

[19] D. G. Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. In: *International Journal of Computer Vision* 60.2 (2004). DOI: 10.1023/B:VISI.0000029664.99615.94, pp. 91–110.

[20] P. Middendorf et al. “Pose Estimation and Damage Characterization of Turbine Blades during Inspection Cycles and Component-Protective Disassembly Processes”. In: *Sensors* 22.14 (2022). DOI: 10.3390/s22145191, p. 5191.

[21] B. Mildenhall et al. “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis”. In: *Computer Vision – ECCV 2020*. DOI: 10.1007/978-3-030-58452-8_24. 2020, pp. 405–421.

[22] Open3D. *Mesh Tutorial*. 2026. URL: `https://www.open3d.org/html/tutorial/Basic/mesh.html` (visited on 04/14/2026).

[23] P.-E. Sarlin et al. “From Coarse to Fine: Robust Hierarchical Localization at Large Scale”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. DOI: 10.1109/CVPR.2019.01300. 2019, pp. 12708–12717.

[24] P.-E. Sarlin et al. “SuperGlue: Learning Feature Matching with Graph Neural Networks”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. DOI: 10.1109/CVPR42600.20 2020, pp. 4938–4947.

[25] J. L. Schönberger and J.-M. Frahm. “Structure-from-Motion Revisited”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. DOI: 10.1109/CVPR.2016.445. 2016, pp. 4104–4113.

[26] H. Shang et al. “Deep learning-based borescope image processing for aero-engine blade in-situ damage detection”. In: *Aerospace Science and Technology* 123 (2022). DOI: 10.1016/j.ast.2022.107473, p. 107473.

[27] Z. Shen et al. “Deep Learning based Framework for Automatic Damage Detection in Aircraft Engine Borescope Inspection”. In: *Proceedings of the 2019 International Conference on Computing, Networking and Communications (ICNC)*. DOI: 10.1109/ICCNC.2019.8685593. 2019, pp. 1005–1010.

[28] N. Snavely, S. M. Seitz, and R. Szeliski. “Photo Tourism: Exploring Photo Collections in 3D”. In: *ACM Transactions on Graphics* 25.3 (2006). DOI: 10.1145/1141911.1141964, pp. 835–846.

[29] J. Sun et al. “LoFTR: Detector-Free Local Feature Matching with Transformers”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 8922–8931.

[30] Theia Vision Library. *Structure from Motion (SfM)*. 2026. URL: `https://www.theia-sfm.org/sfm.html` (visited on 04/14/2026).

[31] M. J. Tyszkiewicz, P. Fua, and E. Trulls. “DISK: Learning Local Features with Policy Gradient”. In: *Advances in Neural Information Processing Systems*. NeurIPS 2020. 2020.

[32] I. Uzun et al. “Damage Detection in Aircraft Engine Borescope Inspection Using Deep Learning”. In: *Neural Computing and Applications* 37 (2025). DOI: 10.1007/s00521-025-11443-8, pp. 20105–20126.

[33] T. Wang and V. J. L. Gan. “Automated joint 3D reconstruction and visual inspection for buildings using computer vision and transfer learning”. In: *Automation in Construction* 149 (2023). DOI: 10.1016/j.autcon.2023.104810, p. 104810.