Automation in Construction 175 (2025) 106153

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Accurate and robust 3D reconstruction of wind turbine blade leading edges from high-resolution images

Jonathan Sterckx [a,*], Michiel Vlaminck [a], Koenraad De Bauw [b], Hiep Luong [a]

[a] Ghent University, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium
[b] ENGIE Laborelec, Rodestraat 125, 1630 Linkebeek, Belgium

ARTICLE INFO

Keywords:
3D reconstruction
Wind turbine blades
Drone-based inspection
Structure from motion
Gaussian splatting

ABSTRACT

Leading edge erosion of wind turbine blades reduces energy production and blade lifetime, a growing issue with larger blades. Effective monitoring is crucial to tracking erosion and controlling maintenance costs. This paper presents an image-based 3D reconstruction method for the leading edges, targeting challenges like textureless surfaces, background motion, and limited image overlap that cause existing methods to fail. Leveraging monocular depth estimation with dense image matching within a sparse reconstruction framework enables superior accuracy and robustness compared to commercial and open-source sparse 3D reconstruction software, achieving a reduction of at least 38% in reprojection error in the conducted experiments. Furthermore, this paper proposes a loss term for Gaussian Splatting – a recent dense reconstruction paradigm – which makes it possible to obtain dense reconstructions without missing patches, in contrast to traditional methods, while maintaining fine surface detail.

1. Introduction

Wind turbines are a critical component of renewable energy infrastructure, contributing significantly to the global effort to reduce reliance on fossil fuels and mitigate climate change [1]. Ensuring the efficiency and longevity of wind turbines is paramount, necessitating regular inspections to detect and address potential issues such as erosion, structural defects, and material fatigue [2,3].

The advent of drone technology has revolutionized wind turbine inspections, enabling non-invasive and highly detailed qualitative imaging. Drones equipped with high-resolution cameras can capture comprehensive images of turbine blades and other components from various angles and elevations, significantly reducing inspection costs and the need for human intervention [4]. This method not only enhances safety by minimizing the risks associated with manual inspections, but also provides high-quality images that allow the automatic detection, localization, and classification of surface defects [5,6].

In recent years, image-based 3D reconstruction has seen notable advancements [7–9], greatly enhancing the potential of drone-based inspections. An accurate 3D reconstruction of the surface of the leading edge holds significant potential for automated defect analysis. One crucial application is the segmentation and quantification of leading edge erosion and other surface defects. Leading edge erosion is an increasingly important issue for wind turbines, as turbine blade sizes keep growing [10]. It can lead to reduced aerodynamic efficiency [11], resulting in lower power output [12–14]. If not detected and addressed early, erosion can cause severe damage, leading to high maintenance costs and operational downtime [3]. Therefore, the ability to accurately detect and quantify leading edge erosion is crucial for maintaining the performance and longevity of wind turbines. While 2D images usually suffice to detect erosion, 3D reconstruction is crucial for quantifying the severity of the damage, providing detailed surface topography, and enabling precise measurements that inform more effective repair strategies and can trigger predictive maintenance actions. The 3D model offers exact dimensions and depth profiling of the eroded areas, allowing for a thorough understanding of the nature and progression of the damage. This quantitative data supports accurate cost estimations and prioritization of repairs. Additionally, 3D reconstructions create a baseline for future inspections, facilitating long-term monitoring and enhancing the accuracy of inspections by minimizing subjectivity.

Traditional 3D reconstruction pipelines often struggle to obtain accurate 3D models from images suffering from limited texture, repetitive structures, illumination changes, oversaturation, and blur [7]. These challenges are commonly encountered during drone-based inspections of wind turbine blades. The uniform color of the blades leads to a lack of texture, making it difficult to find matching keypoints across images, which is a crucial initial step in a 3D reconstruction pipeline. Varying sunlight conditions can cause illumination changes and saturation issues. The motion of the drone platform can result in motion blur, and

* Corresponding author.
E-mail address: jonathan.sterckx@ugent.be (J. Sterckx).

https://doi.org/10.1016/j.autcon.2025.106153
Received 28 August 2024; Received in revised form 19 March 2025; Accepted 24 March 2025

if the images are not perfectly focused, out-of-focus blur will also be present. This blur reduces the level of detail captured in the images and complicates keypoint matching. Therefore, we are motivated to develop a 3D reconstruction method specifically designed to address the challenges of wind turbine blade imaging.

Our method for the robust and highly accurate image-based 3D reconstruction of the leading edge of wind turbine blades consists of two steps.

1. Sparse 3D reconstruction by leveraging accurate, dense, robust pixel matches across images. We present a method to extract high quantities of accurate keypoint tracks from a state-of-the-art dense matcher, RoMa [15], performing well even in challenging image conditions.
2. Dense 3D surface reconstruction with a customized Gaussian Splatting (GS) [16] algorithm. The GS paradigm has recently revolutionized the fields of novel view synthesis and 3D reconstruction, offering a powerful alternative to Multi-View Stereo techniques. The base method can be adapted address various challenges, such as blurry images and specular reflections. We demonstrate the feasibility of GS for reconstructing wind turbine blades and propose a custom loss function.

The main contributions can be summarized as follows.

- **Automatic background masking with monocular depth estimation.**
  Accurate 3D modeling requires a static scene, but turbine images often contain moving backgrounds (e.g., clouds or spinning turbines). We use Depth Anything [17] to generate depth maps that separate the turbine from the background, allowing for reliable foreground-background segmentation. This method is robust to the challenges faced by neural networks trained on limited data, which often struggle to generalize to new turbine images [18].
- **Keypoint track extraction from dense matches.**
  Extracting keypoint tracks from dense matches, which lack explicit keypoint descriptors, requires a novel two-step process. We identify valid image subsets and sample high-confidence tracks within them, balancing execution speed, memory use, and track accuracy.
- **Background-masking loss function for Gaussian Splatting.**
  We introduce a loss term for Gaussian Splatting to exclude background reconstruction, improving efficiency and accuracy, especially with limited image data or semi-transparent backgrounds like the sky.

To evaluate our approach, we quantitatively and qualitatively compare our methods to commercial and open-source software alternatives on various sets of images. The experiments demonstrate that our method can generate an accurate sparse 3D model where the alternatives fail to produce any viable model. Even when other methods successfully generate a 3D model, our approach demonstrates superior accuracy, as indicated by a lower reprojection error. Our Gaussian Splatting-based dense reconstruction achieves full completeness, in contrast to traditional Multi-View Stereo techniques, which often result in numerous missing patches. We also show that this improvement in completeness does not compromise accuracy.

The remainder of this paper is structured as follows: Section 2 reviews related work in general 3D reconstruction, wind turbine blade reconstruction, and keypoint matching. Section 3 outlines our methodology, followed by the results in Section 4 and their discussion in Section 5. Finally, we conclude and propose future work in Section 6.

2. Related work

This section reviews existing research and methods related to 3D reconstruction, focusing on sparse and dense techniques, keypoint matching, and specific applications, such as the reconstruction of wind turbine blades. Various approaches are outlined, including traditional methods and recent advancements like deep learning-based techniques for improving the accuracy and robustness of 3D models.

2.1. Sparse 3D reconstruction

Structure from Motion (SfM) is a photogrammetry technique used to reconstruct a sparse three-dimensional point cloud from a sequence of two-dimensional images. It operates under the principle of simultaneously estimating positions of 3D points in the scene and the extrinsic and intrinsic parameters of the camera(s) capturing the images [19]. The workflow of a traditional, incremental SfM pipeline can be summarized as follows [20] (see Fig. 1).

1. **Feature detection and description:** Keypoints or features are detected in each image using algorithms such as SIFT [21] or ORB [22]. These keypoints are mathematically described based on their local appearance using multidimensional feature vectors.
2. **Feature matching:** Correspondences between keypoints in different images are established by matching their descriptors. This step relies on techniques like nearest neighbor matching to find reliable matches while filtering out outliers with geometric verification.
3. **Incremental reconstruction:** First, an image pair is selected to initialize a two-view 3D reconstruction. The relative camera pose of the initial image pair can be estimated by solving for the essential matrix, based on the keypoint matches, and decomposing it. The estimated poses can then be used to triangulate the matched keypoints. Then, the other images can be registered to the reconstruction, one by one, by solving the Perspective-n-Point [23] problem to estimate their camera pose. The 3D point cloud is extended by triangulating matches from the newly registered images. Techniques such as RANSAC [24] are employed for robust estimation in the presence of outliers such as inaccurate keypoint matches.
4. **Bundle adjustment:** A global optimization process called bundle adjustment refines camera poses and 3D point positions jointly, minimizing the reprojection error and ensuring consistency between observed keypoints in images and their corresponding 3D positions [25].

The most widely used open-source SfM toolbox is COLMAP [20], while Agisoft Metashape [26] remains the leading commercial software for general-purpose 3D reconstruction. Recent advancements to improve SfM accuracy and robustness include approaches like Pixel-Perfect SfM [27] and Detector-Free SfM [7], both of which leverage deep learning to refine keypoint tracking. Furthermore, various methods have been developed to enhance 3D reconstruction quality through post-processing techniques, such as point cloud filtering [28], or by replacing specific modules within the SfM pipeline — for instance, optimizing camera parameters using genetic algorithms [29] or increasing match quantity with particle swarm optimization [30].

2.2. Dense 3D reconstruction

The camera poses and point cloud generated through sparse reconstruction can be leveraged to produce a dense 3D model using dense reconstruction techniques, typically driven by Multi-View Stereo (MVS) methods [31], optionally followed by meshing.

Gaussian Splatting (GS) introduces an innovative approach for explicitly depicting dense 3D scenes, distinct from traditional methods

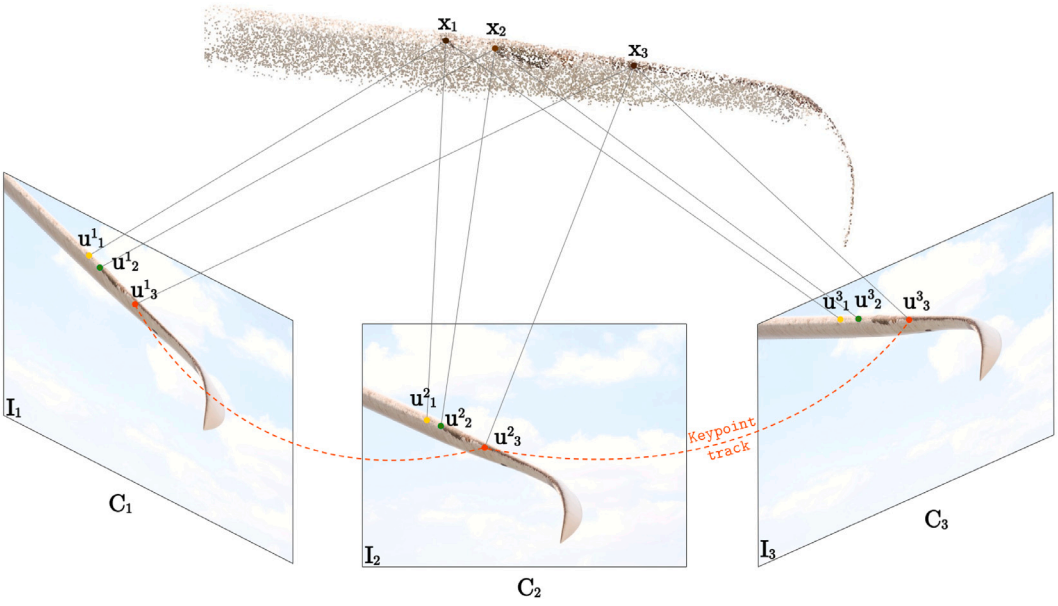

Fig. 1. Structure from Motion jointly estimates a set of 3D scene points $X = \{\mathbf{x}_i \in \mathbb{R}^3 \mid i = 1, \ldots, N_X\}$ and camera poses $C = \{\mathbf{C}_j \in \mathbf{SE}(3) \mid j = 1, \ldots, N_C\}$ from keypoint correspondences across multiple views. Keypoint k in image $\mathbf{I}_l$ is denoted by $\mathbf{u}_k^l$.

like point clouds and meshes. First, the technique initializes a set of 3D Gaussian functions centered on the points obtained during the sparse reconstruction phase. Subsequently, it refines the positions, shapes, and opacities of these Gaussians such that the 3D model resembles the original images when viewed from the perspectives of the estimated camera poses. This iterative process, spanning thousands of iterations, is interleaved by adaptive density control adjustments of the Gaussians. Density control is achieved by splitting and cloning Gaussians, allowing a balance between the model's memory requirements and the accuracy of the reconstruction. Accurate camera pose estimates are crucial for the quality of the dense reconstruction, which is why the accuracy of the preceding sparse reconstruction is of critical importance [16].

The loss function of GS compares each original image with the image generated by rendering the 3D model onto the corresponding viewspace. In other words, the model is trained to produce a reconstruction that looks like the input images when viewed from the corresponding poses, without imposing any geometrical constraints. Lacking any explicit regularization on the 3D geometry, this approach is susceptible to overfitting, particularly in scenarios where images have a low amount of overlap. The rendered images may resemble the input images very well despite inaccuracies in the 3D model such as floating Gaussians (i.e., Gaussians with a high transparency which are not attached to the foreground object) and needle-like artifacts at object surfaces [32].

Researchers have developed various strategies to improve the accuracy of 3D models and reduce overfitting to input images. Mip-Splatting [32] addresses high-frequency artifacts by introducing a 3D smoothing filter that constrains the size of the 3D Gaussians. To produce sharp 3D models from blurry images, Deblur-GS [33] and BAD-Gaussians [34] offer methods specifically designed to handle motion blur. 2DGS [35] replaces volumetric reconstruction with surface reconstruction using 2D Gaussians in 3D space, achieving state-of-the-art reconstruction quality when combined with their novel regularization techniques.

Compared to MVS techniques, Gaussian Splatting offers a more flexible and adaptive approach to 3D reconstruction. MVS relies heavily on finding accurate correspondences between multiple views, which can be challenging in scenarios with repetitive textures or low-contrast surfaces [36]. GS, by refining Gaussian primitives iteratively, can handle such challenges more robustly and produce more consistent models even under difficult imaging conditions. Moreover, GS is generally faster than MVS methods, as it can generate a detailed model in just a few minutes because it represents the scene using a sparse set of 3D Gaussians, which can be efficiently rendered using modern graphics hardware [16]. In contrast, traditional methods rely on complex multi-view image matching and triangulation [37], which are more computationally intensive and slower.

In contrast to NeRF (Neural Radiance Fields), which models scenes implicitly through a neural network [38], Gaussian Splatting uses explicit Gaussian primitives for representation. This explicit nature allows for more precise control over the reconstruction process and facilitates the integration of specific loss terms for improved surface detail and alignment. Additionally, Gaussian Splatting is faster in both rendering and reconstruction, making it a more practical choice for applications requiring detailed surface models, such as wind turbine blade monitoring. The explicit representation also simplifies subsequent analyses, such as surface defect quantification, enhancing its utility in practical inspection and maintenance workflows.

2.3. Keypoint matching

The quality of a dense 3D reconstruction is very sensitive to the accuracy of the estimated camera poses that are provided as input to the dense reconstruction algorithm [37]. Hence, ensuring high quality camera pose estimations is critical for the accuracy of the final 3D model. Since the camera poses are estimated using keypoint matches across images, a good keypoint detector and matcher are critical for an accurate 3D model. Some challenges regarding keypoint detection and matching involve varying light conditions across images, significant viewpoint changes, a lack of texture, and repetitive structures [39]. To tackle these issues, plenty of research has been conducted on improving keypoint detectors and matchers. Recently, traditional methods, typically involving handcrafted keypoint feature descriptors [21,22,40] and brute force or nearest neighbor matching, have been outperformed by deep learning techniques across a wide range of benchmarks [41,42]. Notably, the detector-free matching paradigm has garnered significant attention due to its exceptional effectiveness [7,15,39,43,44]. Whereas traditional detector-based matching consists of first detecting keypoints, mathematically describing these keypoints, and then matching the keypoint descriptors, detector-free matchers skip keypoint detection and description entirely and instead find the most

likely match in the query image for every single pixel in the reference image. These matchers have demonstrated increased robustness to the aforementioned challenges compared to their detector-based counterparts.

2.4. 3D reconstruction of wind turbine blades

Only a few studies have been conducted on assessing the feasibility of commercial photogrammetry techniques such as Agisoft Metashape [26] to obtain 3D reconstructions from images of wind turbine blades. Zhang et al. [45] explored the feasibility of using an unmanned aerial vehicle with a photogrammetry payload to capture images of a wind turbine blade, which were processed using Agisoft Photoscan to create a 3D model for visual inspection. Nielsen et al. [46] also conducted a feasibility study to determine if high-resolution SfM, using Agisoft Metashape, could be used to assess the leading edge roughness of a decommissioned wind turbine blade tip with artificially enhanced erosion. Neither study introduced any new reconstruction techniques. To our knowledge, there have been no prior efforts to create a customized pipeline specifically for the reconstruction of wind turbine blades.

3. Methods

This section outlines the methodologies used for the 3D reconstruction of wind turbine blades. The approach begins with a sparse reconstruction pipeline that incorporates dense feature matching and background masking to improve keypoint tracking and camera pose estimation. This is followed by a dense reconstruction step utilizing Gaussian Splatting for surface modeling. Each stage of the pipeline is designed to enhance robustness and accuracy in challenging imaging conditions, such as specular reflections and low-texture surfaces.

3.1. Considerations for the matcher

Since images of wind turbine blades exhibit most of the challenges faced by keypoint detection and matching, such as lack of texture and the presence of specular reflection, we were motivated to incorporate a state-of-the-art dense matcher, RoMa [15], in the reconstruction pipeline. Aside from the improved robustness to the aforementioned challenges, the large amount of matches per image pair produced by a dense matcher is beneficial for the accuracy of the resulting SfM output, since the errors on the estimated camera poses will vanish as the number of matches increases [47].

Another deciding factor for the quality and robustness of the output of SfM is the length of the keypoint tracks. A keypoint track is defined as a sequence of keypoints across multiple images that correspond to the same 3D point in the scene. Longer keypoint tracks mean that the same 3D points are observed in more images. This redundancy enhances the robustness of estimating the 3D structure and camera parameters. With more observations of the same point, the optimization process in bundle adjustment can better average out noise and errors from individual observations, leading to more accurate results [47].

A final method to improve the quality of the 3D output is to avoid keypoint matches that are located in the background. Since SfM pipelines for dynamic scenes are less effective than pipelines for static scenes [48], masking the background is an important preprocessing step, as it may contain moving clouds or objects. Additionally, even when the background is static, if the images are captured while the turbine is in an idle state, the background may appear to move relative to the blade's reference frame due to its passive rotation. We propose a simple but effective method to obtain the background masks in a fully automated manner, using the powerful monocular depth estimator, Depth Anything [17]. Although less accurate than depth estimation from multi-view inputs, the monocular depth estimator is effective at detecting large depth variations between the foreground and background. Such variations are common in wind turbine blade imaging, where the blade is typically much closer to the camera than the background. Hence, after estimating the depth map of an image and normalizing it to the interval $[0, 1]$, every pixel with a depth larger than a certain threshold can be considered a background pixel. Fig. 2 visualizes our background masking process. Our simple yet effective approach avoids the need for fine-tuning a neural network on images of wind turbines as in [49] or [18]. In turn, we avoid poor results for images that are underrepresented in available training data, such as images with other wind turbines nearby or images of the rotor hub.

3.2. Sparse 3D reconstruction from dense matches

A dense matcher takes a reference image and a query image as input. Its output consists of a warp $\mathbf{W}$ and a certainty map $\mathbf{C}$. Denoting the height and width of the input images by H and W, the warp has dimensions $H \times W \times 2$, where $\mathbf{W}(h, w)$ represents the pixel location (h', w') in the query image which is predicted to correspond to pixel (h, w) in the reference image. The certainty map $\mathbf{C}$ has dimensions $H \times W$, where $\mathbf{C}(h, w)$, ranging between 0 and 1, represents the confidence of the matcher in its predicted match for pixel (h, w) in the reference image.

Extracting keypoint tracks from a dense matcher is not as straightforward as extracting them from sparse matchers. Sparse keypoint detection involves identifying a relatively small set of keypoints in each image and computing descriptors for these keypoints. During sparse matching, similar descriptors are used to establish correspondences between keypoints across different images. Since dense matchers do not rely on keypoint detectors and descriptors, a different method is necessary. Our method for extracting keypoint tracks from dense matchers involves two steps. First, we identify all subsets of images in which valid tracks can be found. Then, within these subsets, we sample a set of tracks with high matching confidences across all images. Special care has to be taken to keep the execution time and memory requirements reasonable, and to avoid drifting or inaccurate keypoint tracks.

Finding image subsets observing a common part of the scene

A dense matcher can detect millions of confident matches in high-resolution images with sufficient overlap. Keeping all these matches requires an infeasible amount of computation time during the incremental reconstruction step of SfM. A naive solution would be to select a fixed amount of M of matches per image pair, where the selection is based on the certainty map of the dense matcher. Keeping in mind that the length of the keypoint tracks plays a major role in the quality of the final 3D reconstruction, this naive selection of the M matches with the highest confidence from each image pair is a suboptimal strategy. There is no guarantee that such a strategy would result in keypoint tracks with more than two elements.

To ensure that each track consists of as many keypoints as possible, we first aim to generate all subsets S of the complete set of images $I = \{\mathbf{I}_l \mid l = 1, \ldots, N_I\}$ for which each subset S contains at least 3 images and consists solely of images observing a common part of the 3D scene, i.e., at least one common 3D point. To find these subsets, we identify all complete subgraphs ("cliques") within the view graph using a well-established, efficient clique-finding algorithm [50]. The view graph is a collection of nodes and edges where each node represents an image and an edge between node i and j signifies that image i and image j contain valid matches. It is either user-defined or automatically determined from exhaustive pairwise matching and geometric verification [20]. A complete view graph – i.e., a set of images in which every pair of images from this set has a certain degree of overlap – is a necessary (but insufficient) condition for the images in this graph to observe a common part of the scene. In other words, the image subsets observing a common part of the scene are contained within the set of possible complete view subgraphs. Fig. 3 exemplifies the identification of complete view subgraphs from a full view graph.

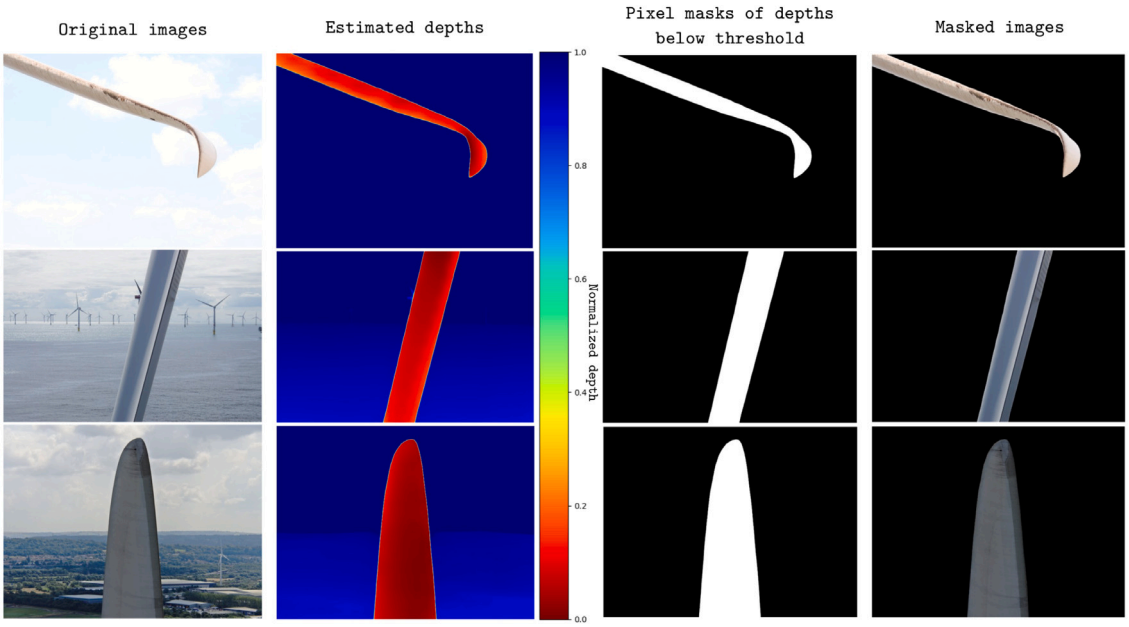

Fig. 2. Output examples of he fully automatic background masker, based on monocular depth maps generated by Depth Anything [17], applied on UAV-captured images. The background/foreground normalized depth threshold is set to be 0.5.

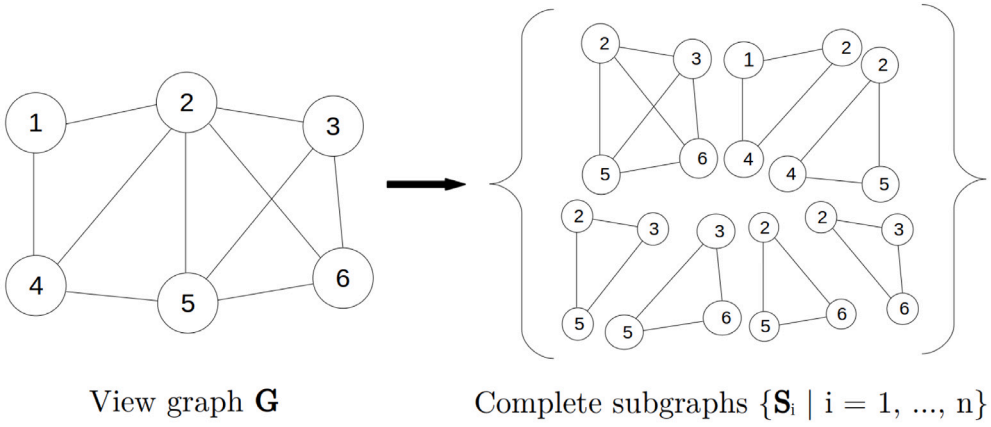

Fig. 3. Identification of complete subgraphs ("cliques") with 3 or more nodes from the view graph. Within each clique, all images are assumed to observe a common part of the 3D scene.

Within each subset S containing the images of a clique, we select a reference image $\mathbf{I}_r$. Next, for each remaining image $\mathbf{I}_q$ in S, we create a binary mask $\mathbf{B}_{r,q}$ with dimensions $H \times W$, where H and W denote the height and width of the images. This mask holds a value of 1 where the certainty map $\mathbf{C}$, resulting from matching $\mathbf{I}_r$ to $\mathbf{I}_q$, exceeds a specified threshold, and 0 otherwise. The final mask, $\mathbf{B}_S$, is the elementwise product of all intermediate masks: $\mathbf{B}_S = \prod_q \mathbf{B}_{r,q}$. From this mask, we randomly sample up to M locations (h, w) where $\mathbf{B}_S(h, w) = 1$. These locations represent keypoints in $\mathbf{I}_r$ for which the matcher found a confident match in each query image of S. For each of these locations, we can extract a keypoint track of length $|S|$ by chaining together the keypoint in $\mathbf{I}_r$ and the keypoints in each $\mathbf{I}_q$ matched to this point. This workflow is generalized and formalized in Algorithm 1. Note that in the rare event where the images in S do not capture a common part of the scene, no confident tracks will be identified. Consequently, this scenario will not affect the quality of the output.

Hierarchical block matching for full-resolution matches

Another challenge posed by extracting keypoint tracks from dense matches is caused by the fact that the matchers process the images by complex neural networks behind the scenes. Processing high-resolution images with neural networks is generally infeasible at full resolution, as processing them on a GPU requires too much VRAM and processing them on a CPU is slow. Hence, images are usually downscaled before passing through a neural network. The output will then also be at a lower resolution than the input, or it will be interpolated to fit the original resolution. To prevent the loss of information associated with downscaling or interpolation, we implemented two-layer hierarchical block matching, allowing us to process the images at full resolution with an intermediate processing step at a lower resolution. For each overlapping image pair, we downscale the images to fit the training resolution of the dense matcher. Then, we consider each image pair again, dividing the reference image of the pair into a grid with cells of equal size. Each cell has dimensions that are equal to or smaller than the training resolution of the dense matcher. For each cell in the reference image, we utilize the results of the coarse matching step to identify the region in the query image that corresponds to this cell. We sample a fixed number of matches originating from the reference cell. Inaccurate matches are filtered out by discarding the outliers of MAGSAC++-based fundamental matrix estimation [51]. The relevant region in the query image is cropped by the bounding box enclosing the matched points in the query image, with an additional 10% margin.

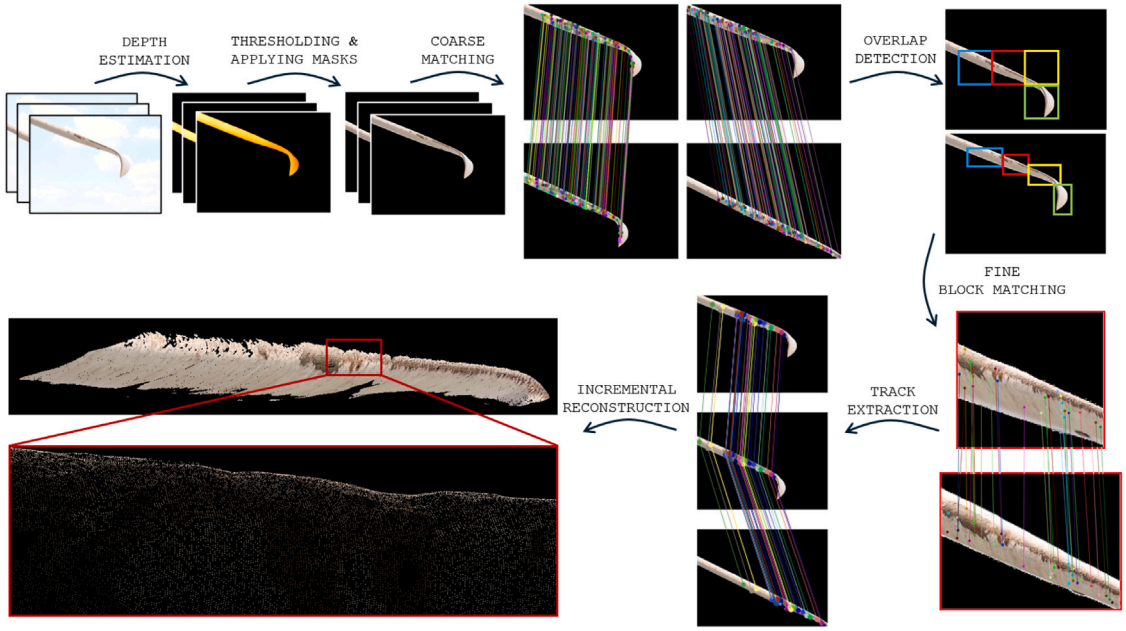

Fig. 4. Overview of sparse reconstruction pipeline.

Algorithm 1 Extraction of keypoint tracks from dense matches and certainty maps

```
I ← LoadImages()
P ← GetImagePairs(I)
W, C ← EmptyLists()
for all p ∈ P do
    w_p, c_p ← DenseMatcher(p)                ▷ Warp & certainty map
    W ← AppendToList(W, w_p)
    C ← AppendToList(C, c_p)
end for
V ← GetViewGraph(I, W)                        ▷ Generate the view graph
S ← GetCliques(V)                             ▷ Find cliques in the view graph
T ← EmptyList()                               ▷ To collect the keypoint tracks
for all s ∈ S do
    r ← SelectReferenceIndex(s)               ▷ Index of the reference image in s
    Q ← GetQueryIndices(s, r)                 ▷ Indices of the query images in s
    W_s ← CollectWarps(W, r, Q)               ▷ The warps from I_r to each I_{q∈Q}
    C_s ← CollectCertaintyMaps(C, r, Q)       ▷ The certainty maps
                                                from I_r to each I_{q∈Q}
    B_s ← BinaryThresholding(W_s, C_s)        ▷ Masks indicating
                                                confident matches from
                                                I_r to each I_{q∈Q}
    t ← GetConfidentTracks(W_s, B_s)          ▷ Sample confident
                                                keypoint tracks
    T ← AppendToList(T, t)
end for
```

The reference grid cells and the query image crops are passed to the dense matcher. Since the dimensions of the cells are equal to or smaller than the training resolution of the matcher, they do not need to be downscaled and no information will be lost. Assuming the query image is captured from approximately the same distance as the reference image, which is typically the case in drone-based inspections, the query image patch will have similar dimensions to the reference cell and will be processed at full resolution.

The hierarchical block matching improves matching accuracy, but this comes at the cost of increased time requirements for the matching step, which scales linearly with the average number of grid cells per image. However, the increase in computation time remains manageable, as masking the backgrounds typically reduces the number of cells in each image to just a few.

The overview of the entire sparse reconstruction pipeline is shown in Fig. 4. For the dense matcher, we opted for RoMa [15], which has set the state-of-the-art in both accuracy and robustness in challenging matching conditions. The keypoint tracks extracted using the method outlined above are then exported to COLMAP [20], where the incremental reconstruction and bundle adjustment steps are performed.

3.3. Dense 3D reconstruction

The capability of Gaussian Splatting to produce accurate outputs despite imperfect input images, combined with its high speed and low memory requirements, makes it an appealing alternative to traditional MVS techniques. In this work, we demonstrate the feasibility of Gaussian Splatting for the dense 3D reconstruction of wind turbine blades by building on top of the 2DGS [35] framework. We chose this framework because we focus on surface reconstruction rather than volumetric reconstruction. For surface reconstruction, selecting 2D primitive elements is more appropriate than 3D elements, as it aligns better with thin surfaces. Additionally, 2D Gaussians enable loss terms that are impractical to compute in 3D Gaussian Splatting, such as the depth distortion loss and the normal alignment loss. The depth distortion loss concentrates the splats by regularizing the intersection depths of camera rays and splats, while the normal alignment loss ensures that all 2D splats are locally aligned with the actual surfaces. These loss terms enhance 3D reconstruction capabilities, with a slight trade-off in novel view synthesis capabilities. This trade-off is acceptable, as our primary interest lies in the 3D model.

When it is desired to only reconstruct the object(s) in the foreground of the input images, background Gaussians can be removed as a post-processing step, e.g. by extracting a 3D bounding box enclosing the relevant object or by space carving using the background masks and camera poses. This approach is suboptimal, as computational resources are wasted when training the model. Reconstructing the background necessitates creating, rendering, optimizing, and densifying Gaussians which will ultimately be discarded. Moreover, by allowing the reconstruction of the background, the model is more likely to be corrupted by floating Gaussians, which are detrimental to the geometric consistency of the 3D model. These artifacts can be caused by, e.g., a low amount

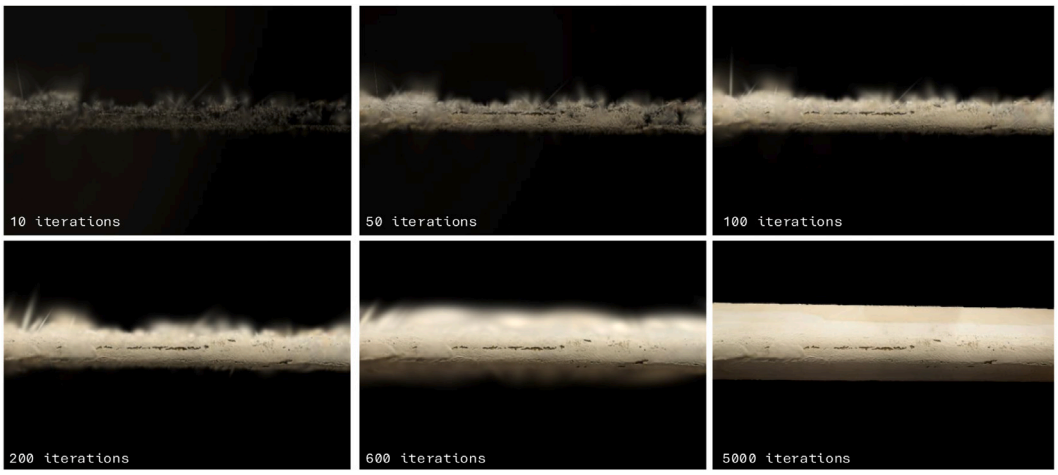

Fig. 5. Training progress of 2DGS with background masking, as seen from a fixed camera pose.

of input images or a semi-transparent background such as the sky. For these reasons, we propose the additional loss term $\mathcal{L}_{\text{mask}}$ which is defined in Eq. (1). This term is added to the other loss terms of 2DGS with a relative weight controlled by the hyperparameter λ. We set λ to 10^{-5} in the experiments.

$$\mathcal{L}_{\text{mask}} = \lambda \sum_{\mathbf{p} \in P_M} [\alpha(\mathbf{p})]^2 \tag{1}$$

A pixel $\mathbf{p}$ which belongs to the set of background pixels P_M contributes to this loss with a value proportional to the square of its rendered opacity $\alpha(\mathbf{p})$. The opacities are calculated during the rendering process in each training iteration of 2DGS and are readily available as auxiliary outputs of the renderer. The background pixels are identified with the binary masks calculated during the preprocessing step of the sparse reconstruction. Fig. 5 visualizes the evolution of the 3D model during training, rendered from a fixed perspective.

4. Results

The following section presents the experimental evaluation of the proposed method, comparing it against alternative techniques for both sparse and dense 3D reconstruction. The experiments assess keypoint matching accuracy, sparse reconstruction quality, and dense model completeness, demonstrating the advantages of the proposed approach across multiple datasets and scenarios.

4.1. Sparse reconstruction

To demonstrate the robustness and accuracy of our RoMa-based matcher, we compare it against two highly regarded keypoint detection and matching methods, each representing a different paradigm in the field. The first baseline employs the traditional SIFT keypoint detector, combined with a 2-nearest neighbor (NN) search and Lowe's ratio test [21], a longstanding method in feature matching known for its reliability and precision. The second baseline leverages the modern deep learning-based methods SuperPoint (SP) [41] for keypoint detection and SuperGlue (SG) [42] for matching. These methods are chosen for their cutting-edge performance and widespread adoption in recent research, providing a robust benchmark for evaluating our approach.

Fig. 6 contains a qualitative comparison of the considered matchers applied on image pairs of the Blade30 dataset [49], which includes drone-based inspections of 30 wind turbine blades under diverse illumination conditions and across various blade types, featuring a wide range of damage and texture variations. We evaluate the matchers on image subsets where the leading edge of the blades is visible and there is sufficient overlap to identify matches.

To quantitatively compare the accuracy of the matchers, we use the symmetric epipolar distance, d_{epi}, as the evaluation metric. This measure is used in computer vision, particularly in stereo vision and SfM, to evaluate the geometric consistency of corresponding points in two images. When dealing with stereo images or multiple views of a scene, the epipolar geometry encapsulates the relationship between the images, defined by the epipolar constraint. The symmetric epipolar distance quantifies the deviation of corresponding points from the (estimated) epipolar lines in both images [47].

Table 1 compares the symmetric epipolar distances, expressed in pixel units, for each of the considered matching methods and for the selected image pairs of the Blade30 dataset [49]. Additionally, the number of inlier matches, M, is reported. Red entries in the table indicate unreliable results, arising from fundamental matrix estimations based on poor matches, as identified through visual inspection of Fig. 6. Bold entries highlight the best results for each experiment.

Fig. 6 illustrates the matching results of each method across these image pairs. For a fair comparison, the input images are processed at full resolution for each method, and outlier matches are all filtered by MAGSAC++ with noise scale $\sigma = 1$. The number of matches in each image pair – visualized by randomly colored lines – is limited to 25. Images are masked using the proposed background masker.

To evaluate our sparse reconstruction pipeline, we compare our approach with Agisoft Metashape [26] and COLMAP [20], which are extensively recognized and influential in the field of 3D reconstruction. These tools represent leading standards in both commercial and open-source software for 3D reconstruction, providing a robust benchmark for our method. We apply the highest-quality preconfigured settings for sparse reconstruction to both methods. Our method samples up to 10^4 keypoint tracks per clique in the view graph, with a minimum match confidence threshold of 0.1.

In Fig. 7, we demonstrate a failure case of the existing sparse reconstruction pipelines. The images originate from blade 15 of the Blade30 [49] dataset. Our method can obtain a sparse reconstruction with sensible camera poses and a large number of points with low reprojection errors. In contrast, COLMAP fails to register any images, and the results produced by Agisoft Metashape are unreliable.

The reprojection error of a 3D point is calculated by projecting this point onto the images containing the keypoints used for its triangulation and measuring the average distance between the reprojection locations and the original keypoint locations. Low reprojection errors indicate accurate camera poses and an accurate sparse point cloud [47].

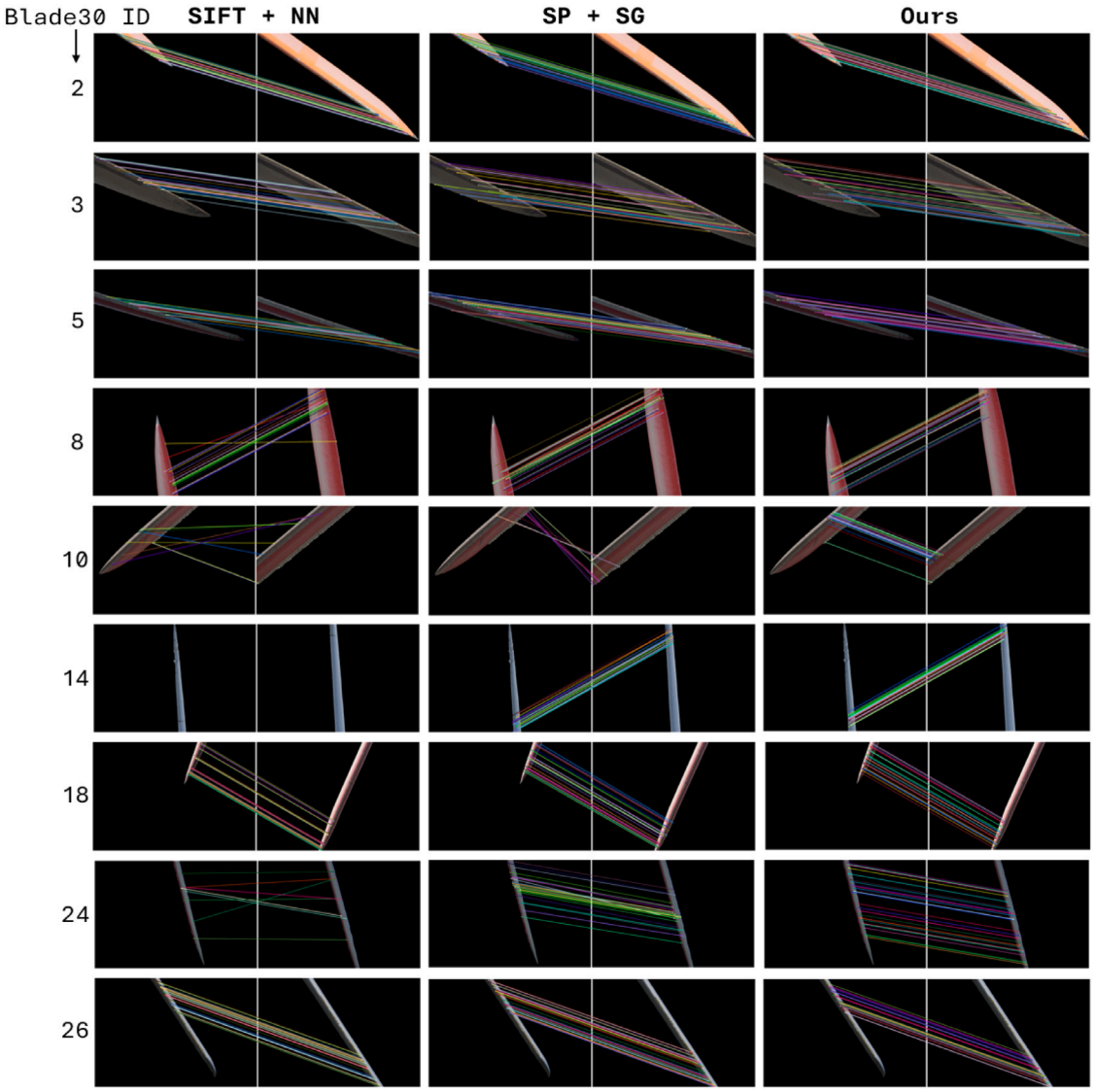

Fig. 6. Qualitative comparison of baseline matchers and the RoMa-based hierarchical block matcher.

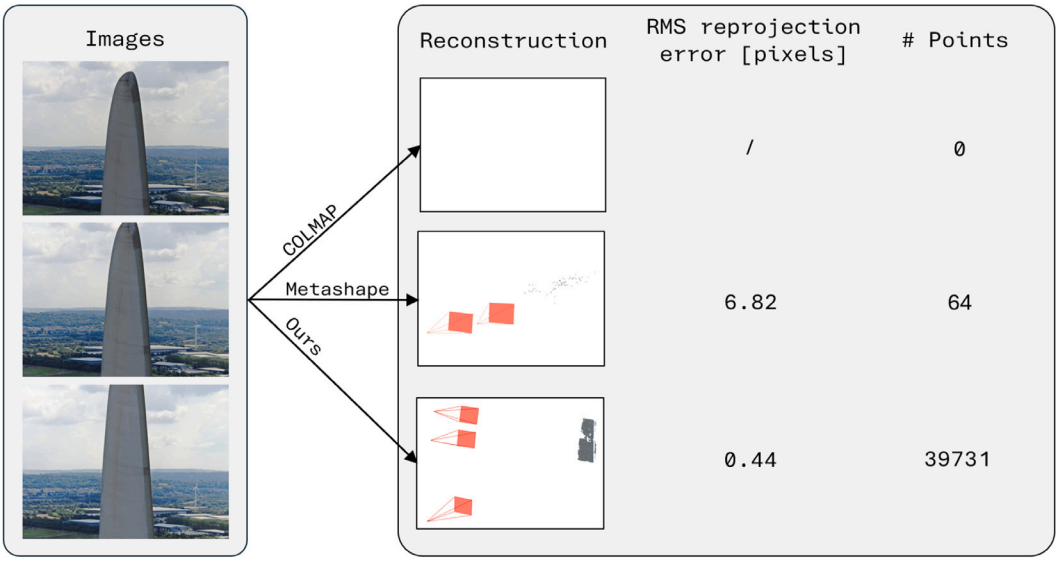

Fig. 7. Failure case in traditional SfM pipelines that the proposed method handles effectively.

Table 1
Comparison of our RoMa-based hierarchical block matcher to baseline matchers.

Blade30 ID	SIFT + NN		SP + SG		Ours	
	d_{epi}	M	d_{epi}	M	d_{epi}	M
2	0.68	329	1.11	138	**0.57**	**8925**
3	0.54	207	1.18	167	**0.33**	**9819**
5	0.49	18	1.21	64	**0.36**	**9487**
8	**0.30**	29	1.15	49	1.05	**6263**
10	0.00	8	0.10	10	**0.53**	**9163**
14	–	0	**0.73**	39	0.87	**8933**
18	0.33	29	1.07	73	**0.28**	**9875**
24	0.31	12	1.23	63	**0.63**	**8574**
26	0.42	381	0.58	104	**0.29**	**8986**

Table 2
Comparison of sparse reconstruction pipeline to COLMAP and Agisoft Metashape.

Dataset ID	COLMAP		Metashape		Ours	
	e_{reproj}	N	e_{reproj}	N	e_{reproj}	N
1	0.68	1578	0.58	2791	**0.42**	**356 867**
2	0.89	3560	1.38	16 802	**0.53**	**115 531**
3	0.88	2132	1.14	12 825	**0.46**	**293507**

The dataset for this experiment is distinct from the one used for comparing the matchers, as the Blade30 dataset lacks sufficient image overlap to construct a 3D model. This dataset comprises three image sets, with samples illustrated in Fig. 8. Sets 1 and 2, provided by ENGIE, include drone inspection images and handheld images of a decommissioned blade, respectively. Set 3, which is publicly available, is detailed in the work by Nielsen et al. [46]. These image sets were chosen to ensure that the baseline methods could still produce valid 3D models. This approach underscores that our method achieves superior accuracy, even where other methods generate viable outputs.

Table 2 reports the root mean squared reprojection errors, e_{reproj}, and the number of points in the point cloud, N, for each method. Bold entries indicate the best results for each experiment. During sparse reconstruction, outlier keypoint tracks are filtered by removing the ones that result in a 3D point with a reprojection error larger than a predetermined threshold. For a fair comparison, we use the same threshold in each experiment.

4.2. Dense reconstruction

We compare the dense reconstruction capabilities of COLMAP, Agisoft Metashape, and our method. Our comparison is limited to a qualitative analysis, as a quantitative comparison would necessitate a ground truth dense 3D model of a wind turbine blade. To the best of our knowledge, no such datasets are publicly available.

To compare the dense reconstruction methods, we acquired a dataset with a high degree of image overlap. This high overlap is crucial for both MVS and GS. Increased image overlap enhances the depth maps of MVS – particularly in textureless or repetitive regions where matching is challenging – due to the benefits of multi-view geometric consistency checks [52]. Furthermore, additional images help to average out noise through depth map fusion. On the other hand, Gaussian Splatting is prone to overfitting when the number of images is insufficient.

Our dataset consists of 23 images of a 50 cm-long section of the leading edge of a decommissioned wind turbine blade. These images were captured using a handheld Nikon D5300 camera from a distance of approximately 1 m and with a focal length of 26 mm. While drone inspections typically occur from a greater distance, using zoom lenses can achieve a similar level of detail, provided that appropriate measures are taken to limit blur during imaging.

Fig. 9 displays dense reconstructions of a section of the leading edge. For a fair comparison of the dense reconstruction methods, each method was provided with the same camera pose estimates – those of our novel sparse reconstruction pipeline – and processed the images at a resolution of 2 megapixels. Other than adjusting the resolution, all parameters remained at the standard settings across all methods. Most notably, our GS-based method exhibits markedly improved completeness over MVS-based approaches. Defining the completeness as the number of foreground pixels with a valid depth value divided by the total number of foreground pixels, the completeness of the depth maps from COLMAP, Metashape, and our method are 65%, 60% and 100%, respectively. This increased completeness can be particularly beneficial for surface defect analysis and mesh extraction. Quantifying the differences in accuracy proves challenging due to the absence of ground truth data. Nevertheless, through qualitative inspection, it is clear that the custom 2DGS can reconstruct intricate surface irregularities with a similar level of detail as the alternatives.

To demonstrate the usefulness of the proposed background loss term, Fig. 10 shows a dense depth map and normal map, with and without masking the background and including the additional loss term. The new loss term is crucial for the 3D geometry, especially in scenarios like drone inspections, where input images typically have limited overlap and do not capture the object from various angles. Despite the significant artifacts in the depth and normal maps in the unmasked case, the foreground rendered by GS remains artifact-free. This is a clear sign of overfitting to the input images and highlights the need for regularization, which is handled by the background loss term. Furthermore, due to its lightweight computation, these improvements do not incur additional runtime costs.

5. Discussion

The results in Table 1 reveal that our approach with RoMa-based matching significantly outperforms the other techniques, achieving the lowest epipolar distance across most images, which reflects higher geometric consistency in the matches. Furthermore, it provides a substantially higher count of valid matches, indicating that our method is more effective in handling the challenges posed by textureless and repetitive blade surfaces and drone-based imaging. The marked improvement in both accuracy and match quantity in our method suggests it is better suited for demanding 3D reconstruction contexts, such as those encountered in wind turbine blade inspection. Moreover, by finding correct matches where other methods fail – which is the case for several of the image sets of the Blade30 dataset – our method will be capable of producing a valid 3D output in more challenging imaging conditions.

Table 2 presents a comparative evaluation of our sparse reconstruction pipeline against COLMAP and Agisoft Metashape, focusing on root mean squared reprojection error and the number of points in the 3D sparse point cloud. Our method consistently achieves the lowest reprojection error and highest point count across all tested datasets, underscoring its superior accuracy and robustness in camera pose estimation and 3D structure recovery. This advantage is largely due to the redundancy afforded by the extensive keypoint tracks extracted from dense matches, which enhances resilience against potential errors in keypoint tracking, as discussed in Section 3.1. By successfully reconstructing sparse models in challenging imaging conditions – where

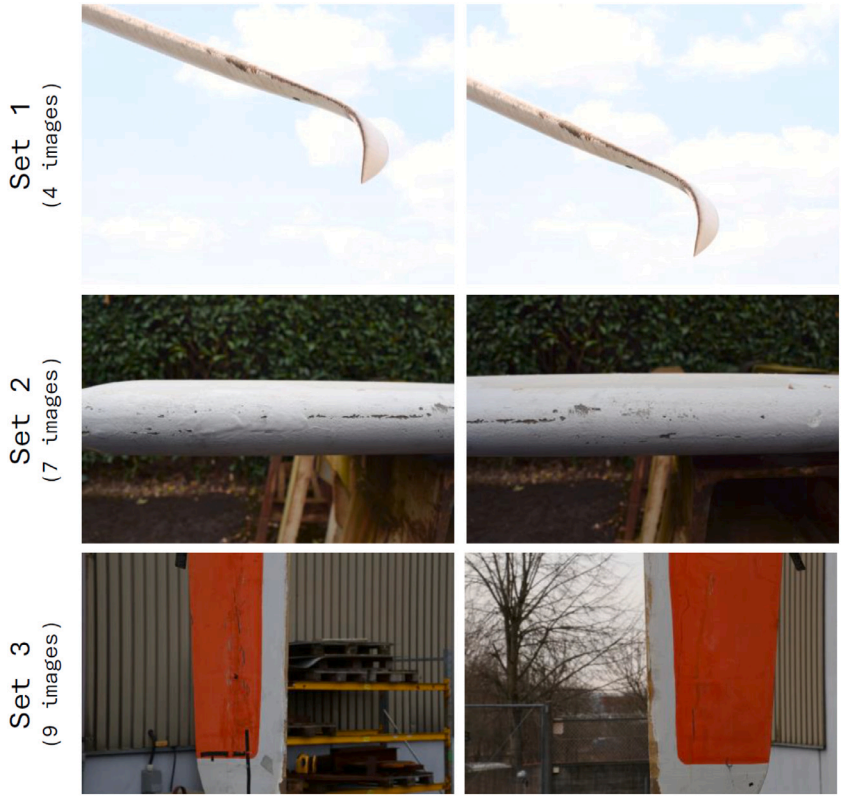

Fig. 8. Samples from the image sets used to compare the sparse reconstruction pipelines in Table 2.

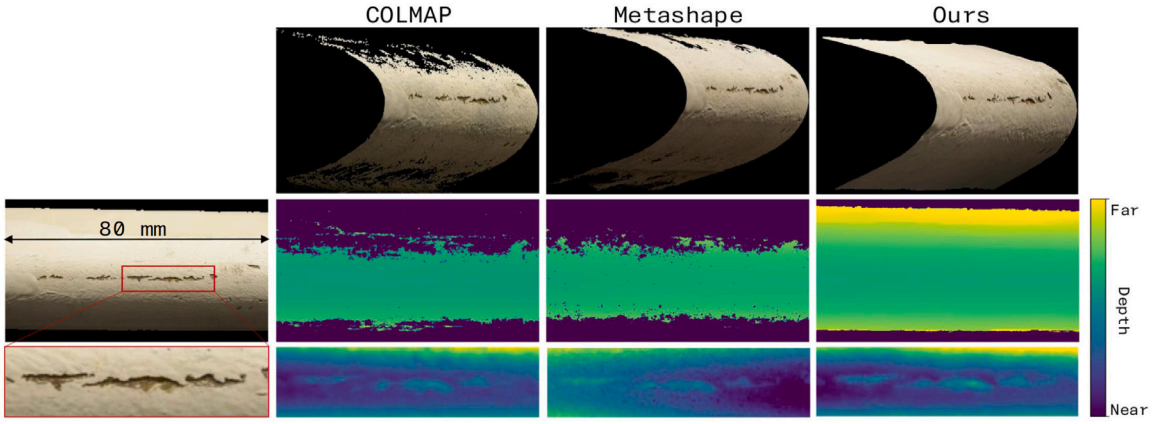

Fig. 9. Comparison of dense point cloud and depth map sections generated by COLMAP, Agisoft Metashape, and 2DGS with background loss.

traditional methods often fail to produce viable models (as shown in Fig. 7) or yield less accurate results – our approach demonstrates enhanced resilience and precision. This combination of reduced re-projection error and a significantly denser point cloud establishes our method as a more reliable solution for high-quality 3D reconstructions in demanding scenarios.

Fig. 7 shows a failure of COLMAP and Metashape to reconstruct a valid sparse 3D model. The images contain very few easily discernible keypoints due to their textureless surface, which presents a challenge for traditional keypoint matching, as implemented in COLMAP and Metashape. Visual inspection of the matches identified by COLMAP and Metashape reveals a low ratio of accurate to inaccurate matches. Both software tools attempt to differentiate correct from incorrect matches using RANSAC. COLMAP is configured to reject image registration when the inlier ratio is too low, resulting in an empty output. In contrast, Metashape generates an output regardless. However, as shown in Fig. 7, this output is unreliable. Note that, even if all outliers were filtered out, the low number if inliers means that the camera pose estimates and 3D point cloud would still be highly sensitive to minor errors or noise in the remaining keypoint matches.

The dense reconstruction was performed using novel Gaussian Splatting techniques. While the original 3D Gaussian Splatting method has been created primarily with novel view synthesis in mind, recent developments have provided creative solutions to regularize the underlying 3D geometry of the model. In this paper, we showcased the feasibility of using 2DGS in combination with a background loss term to create an accurate 3D model from images of a wind turbine blade. The background loss prevents creating Gaussians which are not needed to model the surface of the foreground object. Fig. 10 demonstrates that background masking can be essential to create an accurate 3D model. The enhanced completeness of the dense reconstruction achieved with GS, compared to MVS, is illustrated in Fig. 9. Additionally, despite

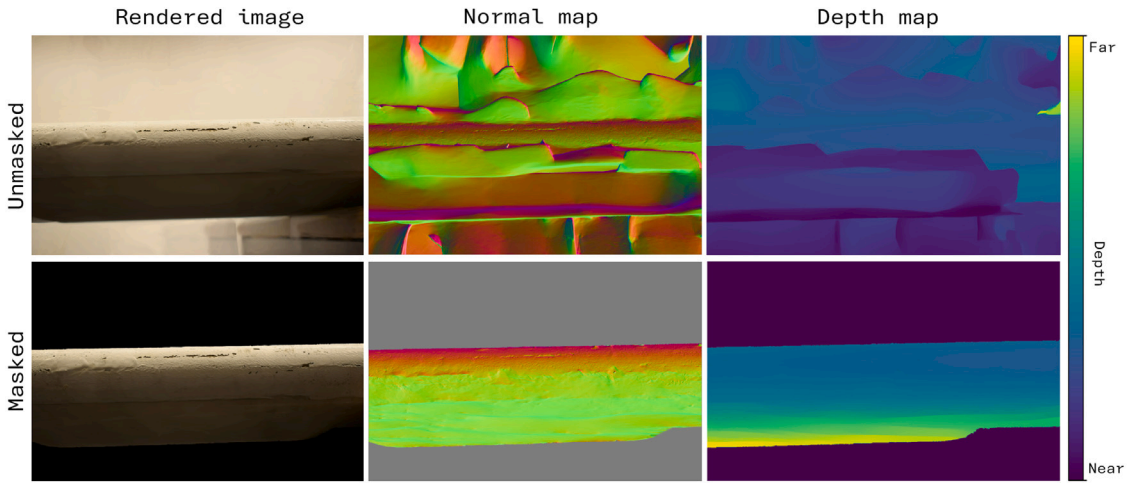

Fig. 10. Effect of background loss on 3D geometry of the Gaussian Splatting model. Surface normals are derived from the depth map and represented using RGB values for visualization.

being a relatively new technology with significant potential for further improvements, GS is already capable of reconstructing very fine surface details.

Having shown how GS can be a viable alternative to MVS techniques for the 3D reconstruction of wind turbine blades, we hope to open up avenues toward exploiting the flexibility of the GS framework by creating or employing customized GS pipelines. Research efforts focused on accurate surface reconstruction and deblurring integrated into GS are expected to converge into a unified framework. As this integration advances, GS is likely to become the preferred method for 3D reconstruction from real drone imagery, which often faces challenges such as out-of-focus blur and motion blur.

Despite its advantages, our method is computationally demanding, especially during the dense matching and keypoint track extraction phases. Furthermore, the current lack of integrated deblurring functionality may restrict the method's effectiveness in high-speed drone inspections, where motion blur could impact reconstruction accuracy.

The quality of a 3D reconstruction is highly dependent on the degree of image overlap. Most public datasets of blade inspections have very limited overlap, rendering them useless for 3D reconstruction. Furthermore, to the best of our knowledge, there are no public datasets with ground truth 3D data of (decommissioned) blades, which is essential for quantitatively validating and refining reconstruction algorithms. The acquisition of datasets with high-overlap drone imagery and datasets with precise ground truth 3D models of blades – e.g., with precision laser scanning – would therefore be of tremendous value to supporting the progression of 3D reconstruction and surface analysis techniques in this field.

6. Conclusion

This paper presented a robust and accurate image-based 3D reconstruction method for the leading edges of wind turbine blades, tackling challenges such as textureless surfaces, background motion, and limited image overlap. By leveraging recent advances in monocular depth estimation, our method enables reliable background-foreground segmentation, making it effective for non-static scenes and overcoming generalization challenges. We enhance traditional photogrammetry techniques by integrating dense image matching within a sparse 3D reconstruction framework, addressing their limitations in producing accurate, viable 3D models. To improve the quality of the dense reconstruction, we introduce a novel loss term for Gaussian Splatting, which reduces floating artifacts, resulting in dense 3D models with accuracy comparable to Multi-View Stereo but with significantly improved completeness. Our technique is highly applicable to the wind energy sector, delivering precise 3D blade models that pave the way for automated erosion quantification and long-term monitoring to enhance maintenance and operational efficiency.

Additional improvements to the dense reconstruction can be achieved by incorporating prior knowledge of blade geometry into the loss function of 2DGS during training to enhance model regularization. Furthermore, implementing adaptive approaches for Gaussian densification can improve reconstruction detail near the blade's leading edge while efficiently managing memory usage. Finally, since drone-based images are typically captured from a few meters away, motion blur and out-of-focus blur may result in significant detail loss, hindering the recovery of a highly detailed 3D model. While our current method does not include a deblurring module, modeling the physical image formation process for motion-blurred images – by jointly learning the parameters of Gaussians and recovering camera motion trajectories during exposure, as proposed in [34] – could effectively address this issue when integrated into the 2DGS framework.

Future work could also focus on automated surface defect analysis based on the dense reconstruction of the leading edge, enabling quantification of erosion and other damage. Comparing the 3D models of the leading edge of a blade across subsequent inspection would enable the study of the progression of leading edge erosion, facilitating improved maintenance strategies. Additionally, incorporating these detailed 3D models into aerodynamic simulations could provide a deeper understanding of how erosion impacts the power output of wind turbines.

CRediT authorship contribution statement

Jonathan Sterckx: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation. **Michiel Vlaminck:** Writing – review & editing, Supervision, Conceptualization. **Koenraad De Bauw:** Supervision, Resources, Conceptualization. **Hiep Luong:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jonathan Sterckx reports financial support was provided by Belgian Federal Government. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors want to thank Mr. Nicolas Quiévy of ENGIE for his valuable support in providing selected images of field blade inspections and ENGIE Laborelec for providing access to blade samples.

Funding: This work was supported by the Energy Transition Fund (ETF) of the Belgian Federal Government (BLEEPID-project; FPS Economy Belgium).

Data availability

Data will be made available upon request, with the exception of image sets provided by ENGIE, which are restricted due to confidentiality agreements.

References

[1] P. Sadorsky, Wind energy for sustainable development: Driving factors and future outlook, J. Clean. Prod. 289 (2021) 125779, http://dx.doi.org/10.1016/j.jclepro.2020.125779.

[2] B.R. Sarker, T.I. Faiz, Minimizing maintenance cost for offshore wind turbines following multi-level opportunistic preventive strategy, Renew. Energy 85 (2016) 104–113, http://dx.doi.org/10.1016/j.renene.2015.06.030.

[3] Z. Ren, A.S. Verma, Y. Li, J.J.E. Teuwen, Z. Jiang, Offshore wind turbine operations and maintenance: A state-of-the-art review, Renew. Sustain. Energy Rev. 144 (2021) 110886, http://dx.doi.org/10.1016/j.rser.2021.110886.

[4] K. Kabbabe Poleo, W.J. Crowther, M. Barnes, Estimating the impact of drone-based inspection on the Levelised Cost of electricity for offshore wind farms, Results Eng. 9 (2021) 100201, http://dx.doi.org/10.1016/j.rineng.2021.100201.

[5] D. Xu, C. Wen, J. Liu, Wind turbine blade surface inspection based on deep learning and UAV-taken images, J. Renew. Sustain. Energy 11 (5) (2019) 053305, http://dx.doi.org/10.1063/1.5113532.

[6] M. Memari, P. Shakya, M. Shekaramiz, A.C. Seibi, M.A. Masoum, Review on the advancements in wind turbine blade inspection: Integrating drone and deep learning technologies for enhanced defect detection, IEEE Access (2024) 33236–33282, http://dx.doi.org/10.1109/ACCESS.2024.3371493.

[7] X. He, J. Sun, Y. Wang, S. Peng, Q. Huang, H. Bao, X. Zhou, Detector-free structure from motion, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 21594–21603, http://dx.doi.org/10.1109/CVPR52733.2024.02040.

[8] B. Fei, J. Xu, R. Zhang, Q. Zhou, W. Yang, Y. He, 3D Gaussian splatting as new era: A survey, IEEE Trans. Vis. Comput. Graphics (2024) 1–20, http://dx.doi.org/10.1109/TVCG.2024.3397828.

[9] S. Wang, V. Leroy, Y. Cabon, B. Chidlovskii, J. Revaud, DUSt3R: Geometric 3D vision made easy, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 20697–20709, http://dx.doi.org/10.1109/CVPR52733.2024.01956.

[10] R. Herring, K. Dyer, F. Martin, C. Ward, The increasing importance of leading edge erosion and a review of existing protection solutions, Renew. Sustain. Energy Rev. 115 (2019) 109382, http://dx.doi.org/10.1016/j.rser.2019.109382.

[11] I.F. Zidane, K.M. Saqr, G. Swadener, X. Ma, M.F. Shehadeh, On the role of surface roughness in the aerodynamic performance and energy conversion of horizontal wind turbine blades: a review, Int. J. Energy Res. 40 (15) (2016) 2054–2077, http://dx.doi.org/10.1002/er.3580.

[12] W. Han, J. Kim, B. Kim, Effects of contamination and erosion at the leading edge of blade tip airfoils on the annual energy production of wind turbines, Renew. Energy 115 (2018) 817–823, http://dx.doi.org/10.1016/j.renene.2017.09.002.

[13] H. Im, B. Kim, Numerical study on the effect of blade surface deterioration by erosion on the performance of a large wind turbine, J. Renew. Sustain. Energy 11 (6) (2019) 063308, http://dx.doi.org/10.1063/1.5115080.

[14] O.S. Özçakmak, D. Bretos, B. Méndez, C. Bak, Determination of annual energy production loss due to erosion on wind turbine blades, J. Phys.: Conf. Ser. 2767 (2) (2024) 022066, http://dx.doi.org/10.1088/1742-6596/2767/2/022066.

[15] J. Edstedt, Q. Sun, G. Bökman, M. Wadenbäck, M. Felsberg, RoMa: Robust dense feature matching, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 19790–19800, http://dx.doi.org/10.1109/CVPR52733.2024.01871.

[16] B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis, 3D Gaussian splatting for real-time radiance field rendering, ACM Trans. Graph. 42 (4) (2023) 1–14, http://dx.doi.org/10.1145/3592433.

[17] L. Yang, B. Kang, Z. Huang, X. Xu, J. Feng, H. Zhao, Depth anything: Unleashing the power of large-scale unlabeled data, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 10371–10381, http://dx.doi.org/10.1109/CVPR52733.2024.00987.

[18] R. Pérez-Gonzalo, A. Espersen, A. Agudo, Robust wind turbine blade segmentation from rgb images in the wild, in: 2023 IEEE International Conference on Image Processing, IEEE, 2023, pp. 1025–1029, http://dx.doi.org/10.1109/ICIP49359.2023.10223165.

[19] O. Özyeşil, V. Voroninski, R. Basri, A. Singer, A survey of structure from motion, Acta Numer. 26 (2017) 305–364, http://dx.doi.org/10.1017/S096249291700006X.

[20] J.L. Schonberger, J.-M. Frahm, Structure-from-motion revisited, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4104–4113, http://dx.doi.org/10.1109/CVPR.2016.445.

[21] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110, http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94.

[22] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: An efficient alternative to SIFT or SURF, in: 2011 International Conference on Computer Vision, 2011, pp. 2564–2571, http://dx.doi.org/10.1109/ICCV.2011.6126544.

[23] S. Li, C. Xu, M. Xie, A robust o(n) solution to the perspective-n-point problem, IEEE Trans. Pattern Anal. Mach. Intell. 34 (7) (2012) 1444–1450, http://dx.doi.org/10.1109/TPAMI.2012.41.

[24] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395, http://dx.doi.org/10.1145/358669.358692.

[25] B. Triggs, P.F. McLauchlan, R.I. Hartley, A.W. Fitzgibbon, Bundle adjustment — A modern synthesis, in: B. Triggs, A. Zisserman, R. Szeliski (Eds.), Vision Algorithms: Theory and Practice, Berlin, Heidelberg, 2000, pp. 298–372, http://dx.doi.org/10.1007/3-540-44480-7_21.

[26] Agisoft metashape, 2024, available online: https://www.agisoft.com.

[27] P. Lindenberger, P.-E. Sarlin, V. Larsson, M. Pollefeys, Pixel-perfect structure-from-motion with featuremetric refinement, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 5987–5997, http://dx.doi.org/10.1109/ICCV48922.2021.00593.

[28] S. El Hazzat, M. Merras, Improvement of 3D reconstruction based on a new 3D point cloud filtering algorithm, Signal, Image Video Process. 17 (5) (2023) 2573–2582, http://dx.doi.org/10.1007/s11760-022-02474-y.

[29] M. Merras, A. Saaidi, N. El Akkad, K. Satori, Multi-view 3D reconstruction and modeling of the unknown 3D scenes using genetic algorithms, Soft Comput. 22 (19) (2018) 6271–6289, http://dx.doi.org/10.1007/s00500-017-2966.

[30] S. El Hazzat, M. Merras, N. El Akkad, A. Saaidi, K. Satori, Enhancement of sparse 3D reconstruction using a modified match propagation based on particle swarm optimization, Multimedia Tools Appl. 78 (11) (2019) 14251–14276, http://dx.doi.org/10.1007/s11042-018-6828-1.

[31] E.K. Stathopoulou, F. Remondino, A survey on conventional and learning-based methods for multi-view stereo, Photogramm. Rec. 38 (183) (2023) 374–407, http://dx.doi.org/10.1111/phor.12456.

[32] Z. Yu, A. Chen, B. Huang, T. Sattler, A. Geiger, Mip-splatting: Alias-free 3D Gaussian splatting, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 19447–19456, http://dx.doi.org/10.1109/CVPR52733.2024.01839.

[33] W. Chen, L. Liu, Deblur-gs: 3d gaussian splatting from camera motion blurred images, Proc. ACM Comput. Graph. Interact. Tech. 7 (1) (2024) 1–15, http://dx.doi.org/10.1145/3651301.

[34] L. Zhao, P. Wang, P. Liu, BAD-Gaussians: Bundle adjusted deblur Gaussian splatting, in: A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, G. Varol (Eds.), European Conference on Computer Vision, Cham, 2025, pp. 233–250, http://dx.doi.org/10.1007/978-3-031-72698-9_14.

[35] B. Huang, Z. Yu, A. Chen, A. Geiger, S. Gao, 2D gaussian splatting for geometrically accurate radiance fields, in: SIGGRAPH 2024 Conference Papers, 2024, pp. 1–11, http://dx.doi.org/10.1145/3641519.3657428.

[36] Y. Wang, Z. Zeng, T. Guan, W. Yang, Z. Chen, W. Liu, L. Xu, Y. Luo, Adaptive patch deformation for textureless-resilient multi-view stereo, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 1621–1630, http://dx.doi.org/10.1109/CVPR52729.2023.00162.

[37] Y. Furukawa, C. Hernández, et al., Multi-view stereo: A tutorial, Found. Trends Comput. Graph. Vis. 9 (1–2) (2015) 1–148, http://dx.doi.org/10.1561/0600000052.

[38] B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi, R. Ng, NeRF: representing scenes as neural radiance fields for view synthesis, Commun. ACM 65 (1) (2021) 99–106, http://dx.doi.org/10.1145/3503250.

[39] J. Sun, Z. Shen, Y. Wang, H. Bao, X. Zhou, 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, http://dx.doi.org/10.1109/CVPR46437.2021.00881.

[40] H. Bay, T. Tuytelaars, L. Van Gool, SURF: Speeded up robust features, in: A. Leonardis, H. Bischof, A. Pinz (Eds.), Proceedings of the European Conference on Computer Vision, Part I 9, Berlin, Heidelberg, 2006, pp. 404–417, http://dx.doi.org/10.1007/11744023_32.

[41] D. DeTone, T. Malisiewicz, A. Rabinovich, SuperPoint: Self-supervised interest point detection and description, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 224–236, http://dx.doi.org/10.1109/CVPRW.2018.00060.

[42] P.-E. Sarlin, D. DeTone, T. Malisiewicz, A. Rabinovich, SuperGlue: Learning feature matching with graph neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 4938–4947, http://dx.doi.org/10.1109/CVPR42600.2020.00499.

[43] J. Yu, J. Chang, J. He, T. Zhang, J. Yu, F. Wu, Adaptive spot-guided transformer for consistent local feature matching, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 21898–21908, http://dx.doi.org/10.1109/CVPR52729.2023.02097.

[44] J. Edstedt, I. Athanasiadis, M. Wadenbäck, M. Felsberg, DKM: Dense kernelized feature matching for geometry estimation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 17765–17775, http://dx.doi.org/10.1109/CVPR52729.2023.01704.

[45] D. Zhang, K. Burnham, L. Mcdonald, C. Macleod, G. Dobie, R. Summan, G. Pierce, Remote inspection of wind turbine blades using UAV with photogrammetry payload, in: 56th Annual British Conference of Non-Destructive Testing, GBR, 2017, url:https://strathprints.strath.ac.uk/63321/1/Zhang_etal_NDT_2017_Remote_inspection_of_wind_turbine_blades_using_UAV_with_photogrammetry_payload.pdf.

[46] M.S. Nielsen, I. Nikolov, E.K. Kruse, J. Garnæs, C.B. Madsen, High-resolution structure-from-motion for quantitative measurement of leading-edge roughness, Energies 13 (15) (2020) 3916, http://dx.doi.org/10.3390/en13153916.

[47] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, ISBN: 978-0-521-54051-3, 2003.

[48] A.P.D. Cin, G. Boracchi, L. Magri, Multi-body depth and camera pose estimation from multiple views, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 17804–17814, http://dx.doi.org/10.1109/ICCV51070.2023.01632.

[49] C. Yang, X. Liu, H. Zhou, Y. Ke, J. See, Towards accurate image stitching for drone-based wind turbine blade inspection, Renew. Energy 203 (2023) 267–279, http://dx.doi.org/10.1016/j.renene.2022.12.063.

[50] C. Bron, J. Kerbosch, Algorithm 457: finding all cliques of an undirected graph, Commun. ACM 16 (9) (1973) 575–577, http://dx.doi.org/10.1145/362342.362367.

[51] D. Barath, J. Noskova, M. Ivashechkin, J. Matas, MAGSAC++, a fast, reliable and accurate robust estimator, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1304–1312, http://dx.doi.org/10.1109/CVPR42600.2020.00138.

[52] J.L. Schönberger, E. Zheng, J.-M. Frahm, M. Pollefeys, Pixelwise view selection for unstructured multi-view stereo, in: European Conference on Computer Vision, 2016, pp. 501–518, http://dx.doi.org/10.1007/978-3-319-46487-9_31.