Vision-Based Autonomous Localization of Fall Protection Anchor Points on Transmission Towers Using Multi-View Geometric Perception

Chunqing Yang[1], Yu Peng[2], Jian Yu[3], Dongfeng Yu[4], Rui Liu[5], Jiahui Chen[6]
Leshan Power Supply Company, State Grid Sichuan Electric Power Company, Sichuan, 614000, China[1,2,3,4,5]
Electric Power Science Research Institute, State Grid Sichuan Electric Power Company, Sichuan, 610041, China[6]

***Abstract*—This paper presents the first systematic investigation into autonomous UAV-mounted fall protection lanyard (FPL) deployment for high-voltage transmission tower inspections, addressing a critical safety gap in the power industry where falls account for 34% of occupational fatalities. We propose a novel geometry-based solution to overcome three fundamental limitations of existing approaches: the isolated processing of UAV imagery without sensor fusion, unreliable 2D-to-3D spatial correspondence in anchor point detection, and the high annotation costs of supervised learning methods. Our technical contribution establishes a multi-view geometric perception framework that decomposes the FPL anchoring task into ridge line identification and optimal mounting point selection. The method first develops a spacial edge distance perception algorithm specifically for power inspection drones, which computes structural depth through plane-induced homography transformations of temporally matched line features. Subsequently, a mounting position planning algorithm integrates multiview geometric constraints with practical operational requirements including ladder proximity, diagonal steel avoidance, and temporal stability. Experimental validation on real-world power infrastructure data demonstrates superior performance compared to learning-based alternatives, achieving 10.98 MAE in positioning accuracy while maintaining 80ms processing efficiency for real-time operation. The proposed approach eliminates dependency on manual climbing and expert annotations, offering both theoretical advancements in stereo-environment perception for complex structures and immediate field applicability for safer power grid maintenance. This work represents the first formal proposal and comprehensive solution for autonomous FPL deployment in transmission tower inspection scenarios.**

Keywords*—Fall protection lanyard; transmission tower inspection; anchor point localization; multiview geometry; spacial edge distance perception; homography transformation

I. Introduction

Power transmission towers (PTT) are steel lattice structures supporting high-voltage power lines ($110kV$ to ultra-high voltage $\geq$ $1000kV$). These 25-215m tall structures require regular inspection, which still predominantly relies on hazardous manual climbing [1]. In China's power sector (2022), falls accounted for 40% of accidents and 34% of fatalities [2] (Fig. 2), making them the leading occupational hazard. While safety harnesses remain the primary protective measure, their effectiveness is compromised on PTTs that lack integrated fall arrest guide rails (Fig. 1). The absence of such safety infrastructure creates substantial challenges for workers attempting to properly anchor their FPLs to designated mounting points prior to ascending the structure. Therefore, the installation of FPLs urgently requires innovative technical solutions.

Recent years have witnessed the emergence of remotely piloted unmanned aerial vehicles (UAVs) for deploying purpose-engineered fall protection lanyard (FPL) anchoring devices to the apex of high-voltage PTTs. The manual approach inherently has scalability problems due to dependence on scarce specialists while introducing safety risks when operated by novices. Meanwhile, the advances in drone-assisted transmission tower inspection [3], [4], [5], [6], [7], [8], [9] have demonstrated potential for automated FPLs anchoring. To the best of our knowledge, we are the first to formally propose and systematically investigate the problem of autonomous UAV-mounted FPL deployment. While no prior studies have directly addressed this problem, related research on power line facility localization and defect detection has predominantly adopted image-level detector-based methodologies.

However, existing related approaches present three fundamental research gaps that hinder practical deployment: 1) The UAV video imagery are processed in isolation, failing to fully leverage the synergistic information potential between visual data and onboard sensor measurements. 2) Second, while some approaches employ deep learning-based object detection algorithms (e.g. YOLO series [10]) to identify potential anchor points in 2D image space, they fail to establish accurate 3D spatial correspondence, resulting in unreliable detection where visually suitable regions may not correspond to physically actionable locations on the tower structure, particularly when identified points are misaligned with the tower's front face. 3) Third, supervised deep learning requires extensive data annotation by power transmission tower inspection experts, leading to high labeling costs and limited scalability. The fundamental scientific challenge involves developing stereo-environment perception that simultaneously incorporates both the target structure's geometric characteristics and the drone's dynamic operational constraints.

To address these issues, we innovatively proposed an autonomous localization algorithm for FPL mounting points based on multi-view geometry. Drone-assisted FPL deployment involves ascending to operational altitude, horizontal positioning over the tower, and precise anchor point attachment. Our contribution is to guide the drone to automatically find the anchor point during the horizontal positioning stage. In this paper, the original anchoring task is decomposed into two sub-problems in our approach: 1) identifies the ridge line of

the tower's apex that is feasible to attach FPLs, and 2) selects the optimal anchoring point on the ridge line against various requirements. Therefore, we proposed a geometry model to identifies the ridge line of the tower's apex, then to screen the ridge points against specially designed objective functions, which combines geometric perception with practical operational constraints to achieve reliable performance in complex field environments. Experiments revealed that this technical approach enables autonomous drone-based FPL deployment, with potential value for both research and field applications.

Fig. 1. The quadcopter drone operates near the power transmission tower while carrying our specially designed mounting device for fall protection lanyards.

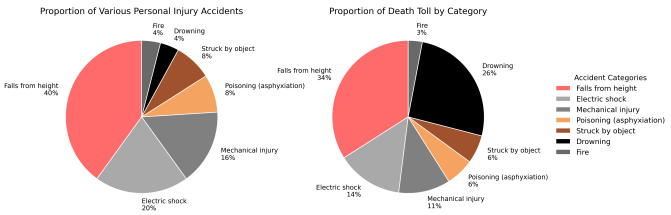


Fig. 2. Component pie-graph of personal injury accidents (left) and death toll (right).

The main contributions of this paper are as follows:

- A scene line segment distance perception algorithm specifically designed for power inspection drones.
- A mounting position planning algorithm for inspection fall protection lanyards.
- Experimental validation on real-world power infrastructure data, demonstrating the effectiveness and advancement of the proposed method.

The rest of this paper is organized as follows. Section II reviews related works. Section III describes the geometry model that links the linear features in images with the their depth in camera's view. Section IV explains the FPL anchoring point localization algorithm. Section V presents the experimental results. Section VI concludes the paper.

II. Related Works

A. Vision-Based UAV Power Line Inspection

Unmanned Aerial Vehicles (UAVs) have become vital for power line inspection [8], [11], offering efficient, high-precision monitoring across diverse environments. Recent developments integrate deep learning models—especially Convolutional Neural Networks (CNNs)—to detect anomalies like insulator fractures, vegetation encroachment, and conductor wear with improved accuracy [12]. These models benefit from large-scale annotated datasets and data augmentation techniques such as color space transformations and edge enhancement, which improve contrast between power line components and background clutter [12], [3]. Federated learning also shows promise for collaborative model training across UAV fleets while preserving data privacy [12], [5].

Despite these advances, spatial perception and geometric reconstruction remain key challenges [13]. Monocular SLAM frameworks like PTI-SLAM enable real-time mapping and localization, achieving trajectory RMSE of 0.1447 m [4]. However, they struggle with textureless surfaces and rapid rotations, leading to misalignment between image features and reconstructed scenes [14], [4], [15], [6]. In contrast, Multi-View Stereo (MVS) methods like DP-MVS offer higher geometric fidelity, achieving RMSE of 3.698 cm [16]. Yet, their computational intensity limits real-time deployment, highlighting a critical research gap: the lack of a unified framework that balances temporal responsiveness with spatial accuracy in complex structural environments. Addressing this gap is essential for reliable UAV-based inspections, particularly for identifying FPL mounting points on transmission towers.

B. Monocular SLAM for Transmission Tower

Monocular SLAM frameworks such as PTI-SLAM offer real-time localization and mapping capabilities for UAV-based power tower inspections [4], [9]. By combining direct and feature-based tracking with semantic filtering, PTI-SLAM achieves trajectory RMSE of 0.1447 m, outperforming GPS alignment methods [4]. Multi-frame fusion and statistical outlier removal further improve point cloud consistency, reducing noise while preserving structural accuracy [4]. However, these systems face key limitations in complex 3D environments.

A major challenge is the poor performance on textureless surfaces—common in metallic tower components—leading to unreliable feature detection and depth estimation errors [15]. Additionally, rapid rotational movements degrade tracking stability, lowering success rates to 6/10 under such conditions [4]. These issues cause misalignment between 2D image features and reconstructed 3D geometry, affecting tasks like defect classification and mounting point identification. Compared to MVS approaches like DP-MVS, which achieve RMSE of 3.698 cm, monocular SLAM lacks the geometric precision needed for detailed structural analysis [16]. Although PTI-SLAM processes frames efficiently at 81 ms per tracking step [4], its metric accuracy remains insufficient for high-fidelity inspection tasks. This highlights a key research gap: the absence of robust multi-view geometric analysis that maintains both real-time performance and spatial coherence in UAV-based power tower inspections.

C. Multi-View Stereo for Transmission Inspections

Multi-view stereo (MVS) methods provide significantly higher geometric accuracy than monocular SLAM, making them suitable for detailed 3D reconstruction of power towers [17]. Techniques like DP-MVS employ PatchMatch-based depth estimation and Delaunay meshing to preserve structural details, achieving RMSE as low as 3.698 cm [16]. Compared to traditional tools such as OpenMVS and COLMAP, DP-MVS

reduces processing time by up to 73%, enhancing efficiency for large-scale modeling [16]. Learning-based approaches like GC MVSNet++ further improve performance by embedding geometric constraints into training, cutting iteration counts by 50% [7], [18].

Despite these gains, MVS methods face challenges in real-time deployment due to high computational demands. They are also sensitive to textureless surfaces where feature correspondence is weak [19]. Progressive Prioritized MVS addresses efficiency concerns by selecting key viewpoints, reducing runtime by 42% while maintaining 98.7% completeness in reconstructed scenes [19]. However, integration into UAV inspection workflows remains limited due to hardware and synchronization requirements. Variable-baseline stereo setups using dual UAVs offer flexibility by adjusting stereo geometry dynamically, improving coverage and reducing occlusion [20]. Yet, they require precise pose estimation [21], which is difficult in GPS-denied or electromagnetically noisy environments. These limitations highlight a key research gap: the need for adaptive MVS frameworks that maintain high geometric fidelity while supporting real-time operation for critical tasks such as FPL mounting point detection.

III. Spacial Edge Perception Model

A. Coordinate system

This study employs a quadrotor UAV system for power transmission tower (PTT) inspection missions. The camera is mounted on the ventral side of the drone via a gimbal with controllable pitch angle. Since we focused on analyzing the relative spatial relationship between the PTT and the UAV, the camera coordinate system is selected as the reference frame for subsequent analysis. To comply with the convention in digital image processing where the y-axis of images points downward, this study defines an right-handed camera coordinate system as illustrated in Fig. 3. The x-axis corresponds to the lateral direction of the UAV, the z-axis aligns with the optical axis toward the observation target. Considering the continuous motion of the UAV in flight, the captured images exhibit dynamic variations across different timestamps. Therefore, the camera coordinate system at time t is adopted as the reference frame fro multi-view geometry analysis.

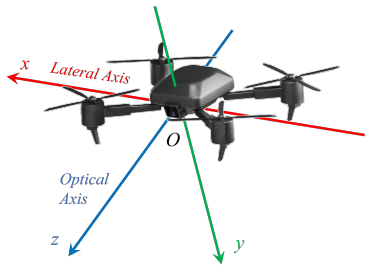


Fig. 3. Coordinate system of the inspection drone's vision.

B. Plane Induced Homography Transformation

Although a PTT appears visually complex, its components primarily consist of planar truss structures. Thus, we first analyze the image flow induced by spatial points on a plane. According to the findings in [22], a moving camera induces a homography transformation in the camera's view of points on a spatial plane. Specifically, let $\mathbf{x}_t \in \mathbb{P}^2$ and $\mathbf{x}_{t+1} \in \mathbb{P}^2$ denote the homogeneous coordinates of a feature point projected at time t and $t+1$, respectively. They satisfy:

$$\mathbf{x}_{t+1} = \mathbf{H}\mathbf{x}_t, \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{3\times3}$ represents the homography matrix. Moreover, we prove that the homography $\mathbf{H}$ satisfies the following proposition:

Proposition III.1. *(Homography): Let the camera coordinate system at time t be the reference frame. If the camera motion from t to $t+1$ is described by a translation vector $\mathbf{c} \in \mathbb{R}^3$ and a rotation matrix $\mathbf{R} \in SO(3)$, and the spatial plane is parameterized by its intercept d and normal vector $\mathbf{n}$ (with the plane equation $\mathbf{n}^T\mathbf{x} + d = 0$), then the homography transformation $\mathbf{H} \in \mathbb{R}^{3\times3}$ induced by this plane in the camera observation can be expressed as:*

$$\mathbf{H} = \mathbf{K}\mathbf{R}\left(\mathbf{I} + \frac{1}{d}\mathbf{c}\mathbf{n}^T\right)\mathbf{K}^{-1}, \tag{2}$$

where $\mathbf{K}$ is the camera's intrinsic parameter matrix, and $\mathbf{I}$ is the identity matrix.

Proof: See Appendix A.

C. Line Depth Approximation From Motion

The Real-Time Kinematics (RTK) system mounted on the drone provides onboard sensors to capture the motion parameters of the camera. Taking the camera frame at time t as the reference, both the translation vector $\mathbf{c} \in \mathbb{R}^3$ and the rotation matrix $\mathbf{R} \in \mathrm{SO}(3)$ at time $t+1$ can be known. Additionally, the camera intrinsic matrix $\mathbf{K}$ can be acquired in advance. Thus, in (2), only the plane parameters (i.e. the intercept d and normal vector $\mathbf{n}$) remain unknown. Since that the camera consistently points toward the monitoring target (i.e. transmission towers), it is reasonable to assume that the target's surfaces are approximately perpendicular to the optical axis, i.e. $\mathbf{n} \approx (0,0,1)^T$. Under this assumption, (2) can be simplified to $\mathbf{H} = \mathbf{K}\mathbf{R}\hat{\mathbf{H}}\mathbf{K}^{-1}$, where

$$\hat{\mathbf{H}} = \mathbf{I} + \frac{1}{d}[\mathbf{0}, \mathbf{0}, \mathbf{c}] \tag{3}$$

Under this assumption, d represents both: (i) the distance from the camera's *optical center* to a quasi-planar surface (approximately parallel to the *principal plane*), and (ii) the orthogonal distance from the surface point to the principal plane. Furthermore, $\mathbf{H}$ is a conjugated transformation of $\mathbf{R}\hat{\mathbf{H}}$, where $\hat{\mathbf{H}}$ is a homology transformation [22]. If the camera's motion, i.e. $\mathbf{R}$ and $\mathbf{c}$ are provided by RTK, it is possible to obtain the d that supplement the target's depth in image.

Although (3) provides the foundation for depth analysis, the visual complexity of PTT scenarios makes point-based analysis unsuitable.

Given that structural edges constitute the most salient features of PTTs (Fig. 6), the depth should be analyzed based on line segments under a homography transformation.

Specifically, if a line segment at time t, denoted as $\mathbf{p} = (\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{P}^2 \times \mathbb{P}^2$, is mapped via homography $\mathbf{H}$ to a segment $\mathbf{q} = (\mathbf{x}'_1, \mathbf{x}'_2) \in \mathbb{P}^2 \times \mathbb{P}^2$ at time $t+1$, then any point $\mathbf{x}$ on $\mathbf{p}$ transformed by $\mathbf{H}$ must satisfy the line equation of $\mathbf{q}$. The line's coefficients are the cross product of the homogeneous endpoints, $\mathbf{x}'_1 \times \mathbf{x}'_2$.

This approach leverages the *duality between points and lines* in parametric space. This geometric constraints enables a line-based solution of d, leading to the following proposition:

Proposition III.2. *Let the camera coordinate system at time t be the reference frame. If a line segment at time t, denoted as $\mathbf{p} = (\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{P}^2 \times \mathbb{P}^2$, is mapped to the segment $\mathbf{q} = (\mathbf{x}'_1, \mathbf{x}'_2)$ at time $t+1$ via the homography transformation $\mathbf{H} \in \mathbb{R}^{3\times3}$ described in Formula (1), then for any point $\mathbf{x} \in \mathbb{P}^2$ on $\mathbf{p}$, its distance d to the principal plane of the camera coordinate system is given by:*

$$d = -\frac{\mathbf{c}^T\mathbf{R}^T\mathbf{K}^T(\mathbf{x}'_1 \times \mathbf{x}'_2)}{\mathbf{x}^T\mathbf{K}^{-T}\mathbf{R}^T\mathbf{K}^T(\mathbf{x}'_1 \times \mathbf{x}'_2)},$$

where, $\mathbf{c} \in \mathbb{R}^3$ and $\mathbf{R} \in SO(3)$ are the camera translation vector and rotation matrix from t to $t+1$, respectively; $\mathbf{K}$ is the camera intrinsic matrix; $\mathbf{x}'_1 \times \mathbf{x}'_2$ denotes the cross product of the homogeneous coordinates of the endpoints, defining the line equation of $\mathbf{q}$.

Proof: Given that $\mathbf{H} = \mathbf{K}\mathbf{R}\hat{\mathbf{H}}\mathbf{K}^{-1}$, for any point $\mathbf{x}$ on line segment $\mathbf{p}$, its corresponding point at time $t+1$, $\mathbf{x}' = \mathbf{K}\mathbf{R}\hat{\mathbf{H}}\mathbf{K}^{-1}\mathbf{x}$, must lie on the line defined by $\mathbf{q} = (\mathbf{x}'_1, \mathbf{x}'_2)$, i.e., $(\mathbf{x}'_1 \times \mathbf{x}'_2)^T\mathbf{K}\mathbf{R}\hat{\mathbf{H}}\mathbf{K}^{-1}\mathbf{x} = 0$. With $\mathbf{R}\mathbf{K}$ known, let $\mathbf{u}' = \mathbf{R}^T\mathbf{K}^T(\mathbf{x}'_1 \times \mathbf{x}'_2)$ and $\mathbf{u} = \mathbf{K}^{-1}\mathbf{x}$, we derive:

$$\mathbf{u}'^T\hat{\mathbf{H}}\mathbf{u} = 0. \tag{4}$$

Substituting (3), the equation expands to:

$$\mathbf{u}'^T\left(\mathbf{I} + \frac{1}{d}[\mathbf{0}, \mathbf{0}, \mathbf{c}]\right)\mathbf{u} = 0. \tag{5}$$

Solving for d from (5) yields:

$$d = -u_3\frac{\mathbf{c}^T\mathbf{u}'}{\mathbf{u}^T\mathbf{u}'}. \tag{6}$$

Since $u_3 = [0, 0, 1][x, y, 1]^T = 1$, the final distance formula is:

$$d(\mathbf{p}, \mathbf{q}) = -\frac{\mathbf{c}^T\mathbf{u}'}{\mathbf{u}^T\mathbf{u}'} = -\frac{\mathbf{c}^T\mathbf{R}^T\mathbf{K}^T(\mathbf{x}'_1 \times \mathbf{x}'_2)}{\mathbf{x}^T\mathbf{K}^{-T}\mathbf{R}^T\mathbf{K}^T(\mathbf{x}'_1 \times \mathbf{x}'_2)} \tag{7}$$

■

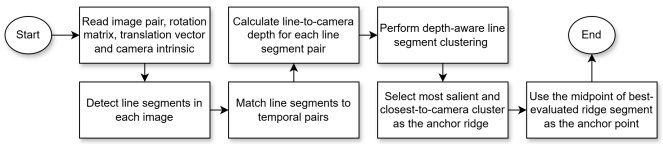


Fig. 4. Workflow of determining anchor ridge that contains fall arrest lanyard.

IV. Planning Safety Rope Anchorage Points

A. Determining Anchor Ridge of Fall Arrest Lanyard

Following the "anchor high, access low" principle for FPL [23], we define the anchor ridge as the load-bearing horizontal edge (the intersection of horizontal and vertical steel structures) at the apex of power transmission towers, which inherently includes appropriate anchor points. In UAV vision, the anchor ridge is the nearest ridge structure relative to the camera. Building upon the spacial line perception model (Section III-C), this subsection presents the detection of the anchor ridge, thus reducing the anchor point positioning problem to screen points along the one-dimensional anchor ridge (discussed in Section IV-B). The core workflow is outlined in Fig. 4, which contains key steps as follows:

1) Line segment detection: First, a line detection algorithm extracts multiple line segments from the current frame at time t, forming a set denoted as $\mathcal{S}^t = \{\mathbf{q}_1, \cdots, \mathbf{q}_m\}$. For all line segment $\mathbf{q} \in \mathcal{S}^t$, their endpoint $\mathbf{q} = (\mathbf{x}'_1, \mathbf{x}'_2)$ are in homogeneous coordinates, $\mathbf{x}'_1, \mathbf{x}'_2 \in \mathbb{P}^2$. Similarly, the prior frame's detection results at $t-1$ are denoted as $\mathcal{S}^{t-1} = \{\mathbf{p}_1, \cdots, \mathbf{p}_n\}$, where $\forall \mathbf{p} \in \mathcal{S}^{t-1}, \mathbf{p} = (\mathbf{x}_1, \mathbf{x}_2)$ and $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{P}^2$. Since lines are salient features in power tower scenes, various line segment detectors (e.g. Line Segment Detector (LSD) [24], Fast Line Detector (FLD) [25], Edge Drawing (ED) [26] or ELSED [27]) are applicable.

2) Temporal line segment matching: Given the two sets of line segments $\mathcal{S}^{t-1}$ and $\mathcal{S}^t$, the corresponding line segments that associate with the same structure on the transmission tower are expected to be paired, yielding a collection of matched line segment tuples.

$$\mathcal{P} = \left\{(\mathbf{p}, \mathbf{q}) | \mathbf{p} \in \mathcal{S}^{t-1}, \mathbf{q} = \arg\max_{\mathbf{q}' \in \mathcal{S}^t} p(\mathbf{p}, \mathbf{q}'), p(\mathbf{p}, \mathbf{q}) > 0\right\}. \tag{8}$$

where, $p(\mathbf{p}, \mathbf{q})$ denotes the probability of line segment $\mathbf{p}$ corresponding to segment $\mathbf{q}$. This can be achieved by line segment's invariant feature extraction and pairwise comparison. To ultimately achieve efficient operation on drone platforms, we adopted the Line Band Descriptor (LBD) [28] to extract stable features from line segments, followed by rapid line matching using KNN-Match algorithm [29]. These algorithms are integrated into the OpenCV library, enabling convenient and efficient implementation.

3) Depth computation and tuple construction: For each line segment pair $(\mathbf{p}, \mathbf{q}) \in \mathcal{P}$, its corresponding edge in the three-dimensional space has a depth d that is the distance from the edge to the camera's principal plane at time t. Eq. (7) is used to calculated the depth d, therefore yielding a set of triplets,

$$\mathcal{R} = \{(\mathbf{p}, d_{\mathbf{p}}) \mid (\mathbf{p}, \mathbf{q}) \in \mathcal{P}, d_{\mathbf{pq}} = d(\mathbf{p}, \mathbf{q})\}. \tag{9}$$

The distance $d(\mathbf{p}, \mathbf{q})$, computed via the previously established spacial edge perception model, reflects the geometric relationship between segment $\mathbf{p}$ and the principal plane.

4) Ridge detection by line segment clustering: Since the anchor ridge represents a prominent geometric feature composed of multiple steel bars, we employ depth-aware agglomerative hierarchical clustering (AHC) [30] to group segmented lines $\mathcal{R}$, generating large-scale linear structures $\mathcal{T} = \{\mathcal{O}_1, \cdots, \mathcal{O}_K\}$ through the aggregation of smaller segments, where $\mathcal{O}_i = \{\mathbf{p}, \cdots\}$ for $i \in \{1, \cdots, K\}$ is the i-th cluster of line segments. The depth-aware is achieved by introducing the depth as the extra feature. Then the anchor ridge is intrinsically retained within the clustered results,

$$\mathcal{T} = \texttt{clustering}_{\text{AHC}}(\mathcal{Q}). \tag{10}$$

To identify the anchor ridge, we select the clustered results with both smaller depth values and greater total segment lengths. Therefore, we have the anchor ridge $\hat{\Omega}$ chosen as

$$\hat{\Omega} = \arg\max_{\Omega \in \mathcal{T}} \sum_{(\mathbf{p}, d_{\mathbf{p}}) \in \Omega} \frac{\|\mathbf{p}\|}{d_{\mathbf{p}}}. \tag{11}$$

The cluster $\hat{\Omega}$ essentially represents the tower's nearest ridge to the camera. If the drone is in proper position, the meaningfulness of $\hat{\Omega}$ if ensured by the geometric stability of the truss structure.

Upon identifying the anchor ridge, the optimal mounting point can be determined through a search along this ridge. The localization algorithm for this process is described in detail in the following section.

B. Locating the Anchor Point for the Safety Lanyard

According to the analysis in Section IV-A, the optimal cluster $\hat{\Omega}$ represents a set of nearest salient edges on the transmission tower related to the drone (*mounting ridge lines*).

This subsection presents the anchor point locating algorithm, which aims to select the midpoint of the most suitable edge $\hat{\mathbf{p}}$ from the cluster $\hat{\Omega}$. To identify the most suitable anchor edge $\hat{\mathbf{p}} \in \hat{\Omega}$, we proposed a multi-objective cost function that evaluates the effectiveness of potential edges,

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p} \in \hat{\Omega}} \left[\psi(\mathbf{p}) + \lambda n(\mathbf{p})) + \omega \phi(\mathbf{p})\right]. \tag{12}$$

where, $\psi(\mathbf{p})$, $n(\mathbf{p}))$ and $\phi(\mathbf{p})$ are functions that evaluate different suitableness of the edge $\mathbf{p}$, and λ and ω are their weight factors.

Specifically, the cost function comprises the following components: 1) *Ladder proximity*: Although multiple suitable mounting points exist along the anchor ridge, we recommend prioritizing positions near the climbing ladder among the available options. Therefore, we model this evaluation using the ψ function. $\psi(\mathbf{p}) = \|\texttt{center}(\mathbf{p}) - \mathbf{x}_{\text{ladder}}\|$ quantifies the horizontal distance between the line segment's midpoint and climbing ladder, prioritizing segments closer to the ladder. Since $\mathbf{p} = (\mathbf{x}_1, \mathbf{x}_2)$, then $\texttt{center}(\mathbf{p}) = (\lfloor \mathbf{x}_1 \rfloor + \lfloor \mathbf{x}_2 \rfloor)/2$ with operator $\lfloor \cdot \rfloor$ representing homogeneous-to-Euclidean coordinate conversion. The $\mathbf{x}_{\text{ladder}} \in \mathbb{R}^2$ is calculated by object detection methods in advance or manually given by the operator. 2) *Vertical frame avoidance*: Considering that the anchor point needs to avoid interference from adjacent vertical steel frames, we define function $n(\mathbf{p})$ to quantitatively evaluate this condition. $n(\mathbf{p}) = |\{\mathbf{q}| \ \mathbf{q} \in \mathcal{R}, \mathbf{q} \notin \hat{\Omega}, e(\mathbf{p}, \mathbf{q}) > 0\}|$ counts adjacent interfering diagonal steel-bars, so as to avoid anchoring in structurally complex areas. λ is a penalty coefficient requiring adjustment. 3) *Length Suitability*: Since steel structures suitable for mounting must possess appropriate lengths (neither too short nor too long), we use the median length of line segments in $\hat{\Omega}$ as the reference length, and then evaluate the deviation between the length of the target segment $\mathbf{p}$ and this reference value. Therefore, $\phi(\mathbf{p}) = (\|\mathbf{p}\| - l_{\text{med}})$, where $l_{\text{med}} = \texttt{median}(\{\|\mathbf{q}\| | \mathbf{q} \in \hat{\Omega}\})^2$ is the median length of line segments in the anchor ridge $\hat{\Omega}$. Finally, the optimal anchor point $\hat{\mathbf{x}}$ is obtained by taking the middle point of $\hat{\mathbf{p}}$, namely, $\hat{\mathbf{x}} = \texttt{center}(\hat{\mathbf{p}})$.

[bt] **Input:** Input parameters $\mathfrak{I}_{t-1}$, $\mathfrak{I}_t$, $\mathbf{R}$, $\mathbf{c}$, $\mathbf{K}$

Output: Result $\hat{\mathbf{p}}$

if $\mathcal{S}_{t-1}$ *not exist* **then**
  $\mathcal{S}_{t-1} \leftarrow \texttt{detect\_line}(\mathfrak{I}_{t-1})$
end

$\mathcal{S}_t \leftarrow \texttt{detect\_line}(\mathfrak{I}_t)$ $\mathcal{P} \leftarrow \texttt{match\_line\_pairs}(\mathcal{S}_{t-1}, \mathcal{S}_t)$; ; // Eq.8

$\mathcal{R} \leftarrow \texttt{calculate\_depth}(\mathcal{P}, \mathbf{R}, \mathbf{c}, \mathbf{K})$; ; // Eq.9

$\mathcal{T} \leftarrow \texttt{line\_clustering}(\mathcal{R})$; ; // Eq.10

$\hat{\Omega} \leftarrow \texttt{select\_cluster}(\mathcal{T})$; ; // Eq.11

$\hat{\mathbf{p}} \leftarrow \texttt{select\_point}(\hat{\Omega})$; ; // Eq.12

return $\hat{\mathbf{p}}$

V. Experimental Results

A. Dataset

Since this study involves specialized experiments in power system applications, our investigation found no relevant publicly available datasets. Therefore, we independently collected experimental data (using devices show in Fig. 5) and manually annotated the optimal mounting positions for fall protection lanyard (FPL), constructing the TOWER dataset. The TOWER dataset originates from real-world power tower inspection tasks, captured by high-resolution cameras mounted on unmanned aerial vehicles (UAVs). During data collection, the UAV's real-time kinematic (RTK) positioning information was synchronously recorded to ensure spatial accuracy. The dataset consists of 2,380 sets of color images, with each set containing two consecutive frames at a resolution of 3840×2160 pixels, providing clear details of the towers and surrounding environments. The corresponding RTK data

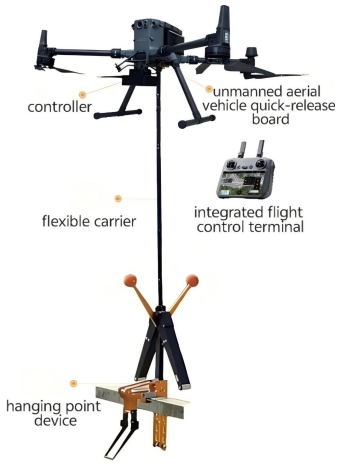


Fig. 5. The power inspection drone and fall protection lanyard equipment used to collect experimental data in this experiment.

Fig. 6. Sample images from the proposed TOWER dataset.

for each frame is stored in JSON format, facilitating image-geolocation alignment. The dataset is divided into training and testing subsets, with the test set containing 620 image pairs and the remaining 5,400 image pairs allocated for training. A key focus of the dataset is the optimal mounting positions for safety ropes. Some images include manual annotations marking the best attachment points, recording both image coordinates and ladder orientation. These annotations support the evaluation of detection algorithms for localization accuracy. Images without viable FPL mounting points are labeled as empty, serving as negative samples to enhance model generalization. The design and annotation methodology of the TOWER dataset provide robust data support for automated power facility inspection and maintenance, while also offering valuable practical insights for computer vision applications in the power industry. Our experimental dataset is planned for future public release.

TABLE I. DATASET DESCRIPTION

Name	Description
Dataset $\mathcal{D} = \{(I_i^0, I_i^1, R_i, t_i, K_i)\}_{i=1}^{N}$	Total $N = 2380$ items
Image Pair $I_0, I_1 \in \{0, \cdots, 255\}^{H \times W \times 3}$	(W,H) = (3840, 2160), 8-bits JPG
Rotation Matrix $\mathbf{R}_i \in \mathrm{SO}(3)$	float type, Json format
Translation Vector $\mathbf{c}_i \in \mathbb{R}^3$	float type, Json format
Camera Intrinsic $\mathbf{K}_i \in \mathbb{R}^{3\times3}$	float type, Json format

B. Hyper-parameter Setting

The experiments were conducted on a PC hardware platform equipped with an i7 processor and 32 GB of RAM to ensure efficient program execution. The software environment utilized Python 3.12, with key dependencies including OpenCV 4.9.0 for line detection algorithms, NumPy 1.26.4 for numerical computation, and Scikit-learn 1.6.1 for clustering algorithm implementation. The results obtained under this configuration effectively reflect the program's performance on typical computing systems, ensuring generalizability and reference value.

The line segment detector in our experiments was the ELSED [27].

The experimental parameters were configured as follows: $T_{\text{ori}} = 0.95$, $T_{\text{len}} = 0.8$, $D_{\text{pos}} = 80$ pixels $\lambda = 5.5$, $\omega = 0.9$.

C. Quantitative Experimental Analysis

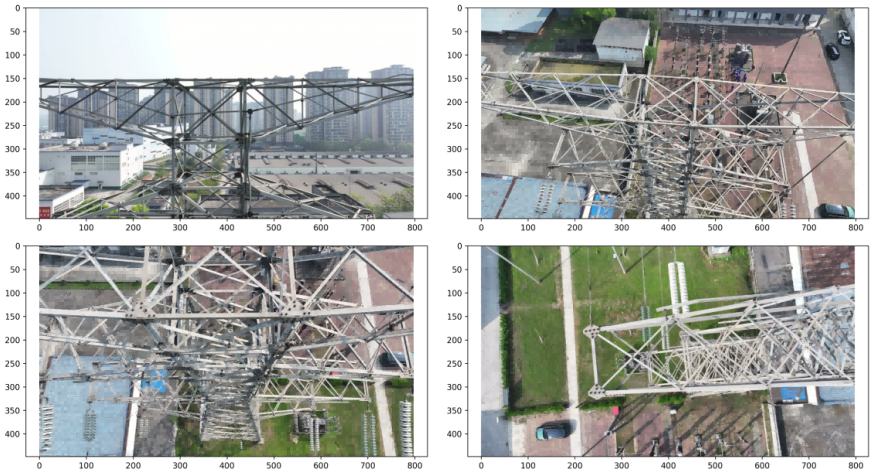
The experimental results demonstrate the effectiveness of our proposed algorithm in accurately identifying anchor ridges and locating optimal mounting points for fall protection lanyards (FPLs), as shown in Fig. 7. The original images are given in Fig. 7a. As illustrated in Fig. 7b, the depth-aware line segment matching successfully isolates structural edges on the transmission tower, with line color represents the depth values, e.g. closer ridges (magenta) clearly distinguished from background clutter (cyan). This enables precise ridge extraction, as shown in Fig. 7c, where the algorithm robustly detects the primary anchor ridges (red lines) and anchor points (black cross) despite complex steel truss interference.

Quantitative evaluation reveals strong alignment between our automated detection and manual annotations (ground-truth red circles). The mean absolute error (MAE) of anchor point positioning is 10.98 pixels, equivalent to sub-optimal positions at typical inspection distances, meeting industrial safety requirements. Our methodology's advantages primarily derive from geometric consistency. The plane-induced homography (Section III-B) maintains spatial coherence in ridge identification, minimizing erroneous detections from diagonal beam structures.

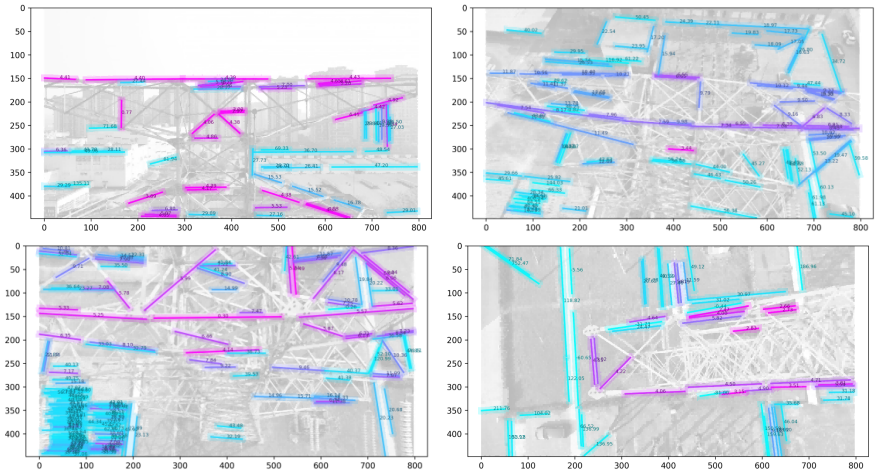
These results validate that our geometry-driven approach outperforms learning-based methods (Table II) in reliability and precision, particularly in preserving spatial relationships critical for FPL deployment safety. The minor deviations from manual markings primarily occur in incorrect line-pair match, highlighting opportunities for future refinement through robust line segment processing.

D. Positioning Accuracy and Efficiency Analysis

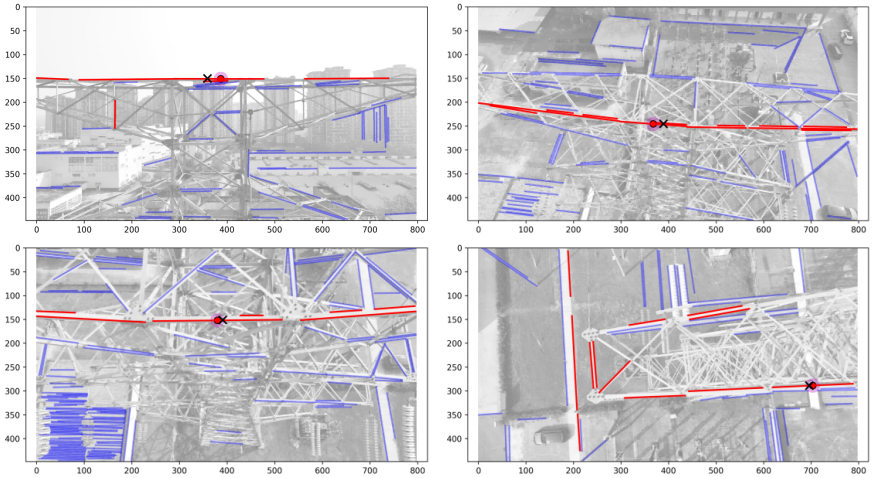
We compared the proposed algorithm with supervised learning-based neural network object detection algorithms.



(a) Original images



(b) Matched line segments with their depth, which is visualized by colormap that cyan indicates far and magenta represents near.



(c) Result of ridge detection (red) and anchoring point localization (cross). The red circle is ground-truth.

Fig. 7. Quantitative analysis of proposed method on various scenarios.

The experimental results are shown in Table I. Vision-based neural models were fine-tuned using our dataset, while the multimodal large model (e.g. DeepSeek V3) was directly invoked through prompt engineering. The analysis of the

experimental results indicates that ResNet30+RE and Our method achieve the best overall performance. Our method demonstrates the strongest robustness with the lowest MAE (10.98) and MedAE (11.02), while also achieving the highest R^2 (0.9725), indicating superior prediction accuracy and model fitting. ResNet30+RE follows closely, with slightly better MSE (265.57) and MAE (11.06) compared to ResNet18+RE, and its R^2 (0.9638) is close to that of Our method, demonstrating strong stability. In contrast, YOLOv5 and DeepSeekV3 exhibit higher MSE (288.72, 420.31), making them more sensitive to outliers, while Template Match is clearly unsuitable due to its excessively large error (MSE=7867.52).

Overall, while ResNet30+RE remains a supervised learning model, our method is unsupervised, eliminating the need for labeled data. This avoids the high costs and potential biases associated with data annotation in supervised learning, making it more advantageous when handling complex or unlabeled datasets.

TABLE II. PERFORMANCE COMPARISON OF DIFFERENT METHODS

Method	MSE	MAE	R^2	MedAE
Template Match	7867.52	86.47	0.6554	90.43
Yolov5	288.72	16.45	0.9107	18.74
ResNet18+RE	276.54	12.12	0.9535	13.16
ResNet30+RE	265.57	11.06	0.9638	12.93
DeepSeekV3	420.31	17.88	0.8989	19.22
Our	268.03	10.98	0.9725	11.02

E. Algorithm Efficiency Experiments

Due to the unique operational environment of UAVs, their environmental perception requires high dynamic responsiveness, making the computational efficiency of UAV algorithms critical. Therefore, we conducted experiments on time efficiency, with comparative results shown in Fig. 8.

The figure illustrates the performance of the proposed algorithm in processing varying numbers of line segments. The x-axis represents different levels of line segment quantities, including the most reliable top 100, 200, and 400 lines (top100, top200, and top400), while the y-axis displays the processing time in milliseconds. Performance at each level is measured using both the mean and standard deviation. As observed, the average processing time increases gradually with the number of line segments. Specifically, the mean processing time for top100 segments is approximately 80 ms, with a small standard deviation, indicating stable performance. These results demonstrate that the proposed algorithm can achieve real-time or near-real-time video processing by controlling the number of line segments to be processed, thereby meeting industrial requirements.

F. Limitations

Through our numerical experiments, we identified several limitations in the current system:

1) Line segment extraction and matching robustness issues: The performance of line detection and temporal matching remains sensitive to environmental variations, such as changes in lighting conditions and complex backgrounds. In extreme cases, this can lead to complete failure in segment extraction or incorrect matching, compromising the system's reliability.

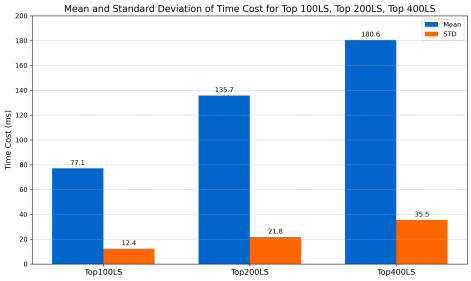


Fig. 8. Performance of the proposed algorithm in processing varying numbers of line segments.

2) Depth estimation accuracy limitations: While the proposed plane-induced homography provides relative depth comparisons between line segments (useful for ridge selection), the absolute depth values exhibit noticeable inaccuracies. This restricts applications requiring precise metric measurements, though it remains effective for ranking structural proximity.

3) Performance degradation under cluttered backgrounds: When inspecting towers with dense interfering line features (e.g. overlapping vegetation or secondary structures), computational latency spiked up to 420% beyond average processing times. This stems from the combinatorial complexity in AHC clustering (Section IV-A) and outlier filtering steps.

These limitations highlight key challenges for real-world deployment, particularly in adverse weather or visually congested environments. Future work should integrate deep learning to stabilize depth estimation and adopt attention mechanisms for robust line feature selection. Addressing these issues will be critical for industrial-grade reliability.

VI. CONCLUSION

This study addresses the safety protection requirements for high-altitude operations on power transmission towers by innovatively proposing an automated fall protection lanyard mounting method based on drone vision, providing an intelligent solution to replace traditional high-risk manual operations. Through multi-view geometric analysis, the system achieves 3D spatial localization of tower edges and optimal mounting point selection, overcoming the limitations of deep learning models in spatial mapping for complex scenarios. The method innovatively integrates multi-dimensional evaluation metrics, including proximity to climbing ladders, avoidance of diagonal steel obstructions, and temporal stability, ensuring that the selected mounting points comply with the "high-hook, low-use" safety principle while maintaining operational convenience. Experimental validation demonstrates that compared to conventional methods, the proposed solution significantly improves positioning accuracy and system intelligence, showing promising potential for pilot applications.

Future research can be deepened in three dimensions: First, expanding algorithm adaptability by optimizing detection models for special structures such as non-standard towers and angled steel members. Second, integrating mechanical verification modules to enable real-time assessment of load-bearing capacity at mounting points. Third, establishing open datasets and standardized testing environments to promote

collaborative technological advancement in the industry. With policy support from State Grid Corporation of China for drone-assisted operations, this technology is expected to become a standard configuration for high-altitude power operations, providing robust safety assurance for smart grid construction. Subsequent research should focus on the integration of drone swarm coordination and digital twin technology to further enhance operational reliability in complex environments.

REFERENCES

[1] W.-T. Chang, C. J. Lin, Y.-H. Lee, and H.-J. Chen, "Development of an observational checklist for falling risk assessment of high-voltage transmission tower construction workers," *International Journal of Industrial Ergonomics*, vol. 68, pp. 73–81, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016981411830132X

[2] E. National Energy Administration, *Compilation of National Electric Power Accidents and Safety Incidents (2022)*. Beijing: China Energy Media Group Co., Ltd., 10 2023, paperback, 16mo.

[3] Y. Zhang, H. Liu, and X. Chen, "Deep learning for aerial power line inspection: A review," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 11, pp. 6789–6801, 2019.

[4] L. Wang, Q. Li, and Y. Zhao, "Pl2dm: A deep learning pipeline for power line defect detection and monitoring," *Remote Sensing*, vol. 12, no. 18, p. 2987, 2020.

[5] V. Solilo, E. Okafor, and A. Nnadi, "Image enhancement strategies for uav-based power line inspection," in *Proceedings of the IEEE International Conference on Unmanned Aircraft Systems*, 2021, pp. 1122–1131.

[6] Z. Liu and M. Zhong, "Federated learning for collaborative uav-based power line inspection," *IEEE Internet of Things Journal*, vol. 9, no. 4, pp. 2890–2902, 2022.

[7] X. Chen, T. Zhang, and Y. Wang, "Pti-slam: Hybrid visual-inertial odometry for uav-based power tower inspection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 175, pp. 213–229, 2021.

[8] D. Lee, T. Kim, and J. Park, "Sparse-view multi-view stereo for uav-based infrastructure inspection," *Sensors*, vol. 22, no. 10, p. 3845, 2022.

[9] R. Zhang, Y. Chen, and D. Li, "Monocular slam performance evaluation for uav-based transmission tower inspection," *Autonomous Robots*, vol. 55, no. 2, pp. 321–339, 2021.

[10] R. Varghese and S. M., "Yolov8: A novel object detection algorithm with enhanced performance and robustness," in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, 2024, pp. 1–6.

[11] A. Garcia, F. Martinez, and R. Lopez, "Learning-based multi-view stereo for aerial inspection tasks," in *Proceedings of the European Conference on Computer Vision*, 2021, pp. 334–349.

[12] S. Fang, J. Zhou, and W. Xu, "Color space transformation and edge enhancement techniques for aerial power line detection," *Sensors*, vol. 20, no. 12, p. 3456, 2020.

[13] W. Tan, F. Huang, and J. Sun, "Progressive prioritized multi-view stereo for efficient uav-based reconstruction," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1890–1897, 2021.

[14] L. Huang, G. Cheng, and X. Lin, "Challenges in textureless region reconstruction for uav vision applications," *Journal of Field Robotics*, vol. 37, no. 6, pp. 987–1004, 2020.

[15] H. Wu, L. Yang, and G. Zhao, "Semantic-aware slam for uav-based power line inspection," *Robotics and Autonomous Systems*, vol. 123, p. 103345, 2020.

[16] Y. Zhou, C. Wu, and M. Li, "Dp-mvs: Detail-preserving multi-view stereo for large-scale infrastructure modeling," *Computer Vision and Image Understanding*, vol. 223, p. 103521, 2022.

[17] Q. Zhao, M. Sun, and Y. Tang, "Outlier rejection in uav vision for power line inspection," in *Proceedings of the IEEE International Symposium on Safety, Security, and Rescue Robotics*, 2021, pp. 123–130.

[18] J. Kim, H. Park, and S. Lee, "Gc-mvsnet++: Geometrically consistent multi-view stereo with enhanced training efficiency," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4321–4335, 2022.

[19] J. Wang, Y. Liu, and H. Xu, "Multi-view stereo reconstruction for complex power infrastructure," *IEEE Access*, vol. 10, pp. 123 456–123 468, 2022.

[20] R. Yang, B. Liu, and K. Zhao, "Variable-baseline stereo vision for uav-based infrastructure inspection," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021, pp. 4567–4574.

[21] X. Li, Y. Zhang, and Z. Wang, "Pose estimation challenges in dual-uav stereo vision systems," *Journal of Intelligent and Robotic Systems*, vol. 99, no. 3, pp. 789–804, 2020.

[22] L. Zeinik-Manor and M. Irani, "Multiview constraints on homographies," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 214–223, Feb. 2002.

[23] S. A. of China (SAC), "Personal fall protection equipment–lanyard," National Standard of China, GB 24543-2009, 2009.

[24] F. Fan, Y. Jiang, K. Han, H. Li, X. Fang, and Y. Fan, "A line segment detector based on subanchor point and line segment combination method," in *2022 37th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 2022, pp. 896–901.

[25] I. C. Baykal and I. C. Yilmaz, "An extremely fast pattern based line detector," in *2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, 2017, pp. 369–376.

[26] C. Akinlar and C. Topal, "Edlines: Real-time line segment detection by edge drawing (ed)," in *2011 18th IEEE International Conference on Image Processing*, 2011, pp. 2837–2840.

[27] I. Suárez, J. M. Buenaposada, and L. Baumela, "ELSED: Enhanced line SEgment drawing," *Pattern Recognition*, vol. 127, p. 108619, Jul. 2022.

[28] L. Zhang and R. Koch, "An efficient and robust line segment matching approach based on lbd descriptor and pairwise geometric consistency," *Journal of Visual Communication and Image Representation*, vol. 24, no. 7, pp. 794–805, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1047320313000874

[29] M.-W. Cao, L. Li, W.-J. Xie, W. Jia, Z.-H. Lv, L.-P. Zheng, and X.-P. Liu, "Parallel k nearest neighbor matching for 3d reconstruction," *IEEE Access*, vol. 7, pp. 55 248–55 260, 2019.

[30] R. Dhivya and N. Shanmugapriya, "An efficient dbscan with enhanced agglomerative clustering algorithm," in *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2023, pp. 1322–1327.

APPENDIX

A. Proof of (2)

Proof: Let $\mathbf{x}_\mathrm{W} \in \mathbb{R}^3$ be an arbitrary point on the a plane $(\mathbf{n}, d)$ in the world frame. Then the $\mathbf{x}_\mathrm{W}$ satisfy,

$$\mathbf{n}^T \mathbf{x}_\mathrm{W} + d = 0 \tag{13}$$

By projecting the $\mathbf{x}_\mathrm{W}$ onto the image planes of camera at t and $t+1$, it satisfy that

$$\mathbf{x} = \mathbf{P}\mathbf{x}_\mathrm{W} \tag{14a}$$
$$\mathbf{x}' = \mathbf{P}\mathbf{R}\left(\mathbf{x}_\mathrm{W} - \mathbf{c}\right) \tag{14b}$$

The (13) can be rewritten as $-\frac{1}{d}\mathbf{n}^T\mathbf{x}_\mathrm{W} = 1$ and substituted into (14b) to obtain,

$$\mathbf{x}' = \mathbf{P}\mathbf{R}\left(\mathbf{x}_\mathrm{W} + \frac{1}{d}\mathbf{c}\mathbf{n}^T\mathbf{x}_\mathrm{W}\right) = \mathbf{P}\mathbf{R}\left(\mathbf{I} + \frac{1}{d}\mathbf{c}\mathbf{n}^T\right)\mathbf{x}_\mathrm{W} \tag{15}$$

Similarly, based on (14a), the $\mathbf{x}_\mathrm{W}$ can be rewrite as $\mathbf{x}_\mathrm{W} = \mathbf{P}^{-1}\mathbf{x}$. By putting this into (15), we have

$$\mathbf{x}' = \mathbf{PR}\left(\mathbf{I} + \frac{1}{d}\mathbf{c}\mathbf{n}^T\right)\mathbf{P}^{-1}\mathbf{x} \tag{16}$$

■