engineering proceedings

MDPI

*Proceeding Paper*

# Trajectory Planning and Depth-Based Nest Localization Using a Pan–Tilt Vision System Mounted on a Gantry Robot in Swiftlet Houses †

**Tran Tuyet Quyen [1], Nguyen Minh Trieu [2] and Nguyen Truong Thinh [2,*]**

[1] College of Engineering, Can Tho University, Can Tho 900000, Vietnam; quyenp1825002@gstudent.ctu.edu.vn
[2] Institute of Intelligent and Interactive Technologies, University of Economics Ho Chi Minh City-UEH, Ho Chi Minh City 700000, Vietnam
* Correspondence: thinhnt@ueh.edu.vn
† Presented at the 2026 International Conference on Energy, Infrastructure and Environmental Research (EIER 2026), Ho Chi Minh City, Vietnam, 15–17 July 2026.

**Abstract**

This paper presents a structured trajectory planning and depth-based nest localization system for automated inspection in swiftlet houses. The proposed approach integrates a Cartesian gantry robot with a 3-DOF (pan–tilt) depth-sensing camera mounted beneath a ceiling-mounted wooden nesting frame (0.5 m × 1 m cells, 150 mm height). A geometry-aware scanning strategy is developed to coordinate planar gantry motion with adaptive orientation control, enabling efficient coverage while mitigating occlusion caused by shallow cell clearance. The method incorporates a coarse-to-fine active perception scheme to detect and refine nest positions in 3D space. Experimental validation conducted in a commercial swiftlet house demonstrates a mean 3D localization error of 8.7 mm and a 28% reduction in inspection time compared to conventional raster scanning. The results confirm the feasibility of reliable, non-invasive nest tracking and highlight the contribution of structured robotic inspection to smart avian farming and precision agricultural robotics.

**Keywords:** trajectory planning; gantry robot; depth camera; nest localization; monitoring robot; Vietnam agriculture; swiflet house

Check for updates

Academic Editor: Nguyen Van Lanh

Published: 16 November 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

## 1. Introduction

Swiftlet nest farming has developed into a high-value agro-industry in Southeast Asia, where edible bird's nests are harvested for their nutritional and commercial value [1,2]. The global demand for these nests continues to grow, motivating farmers to adopt more systematic and scalable management practices. Modern swiftlet houses are specially constructed buildings designed to replicate cave-like environments while incorporating wooden ceiling-mounted frames to encourage nest attachment. Despite structural standardization, nest monitoring remains largely manual [3]. Farmers typically conduct visual inspections at heights of 3–5 m under low-light and high-humidity conditions. These inspections are labor-intensive, time-consuming, and pose occupational safety risks. Furthermore, frequent human presence near nesting areas may disturb swiftlets and potentially affect productivity [4,5]. Consequently, there is a growing need for automated, non-contact monitoring systems capable of accurately locating and characterizing nests without disrupting the birds.

A key structural feature of commercial swiftlet houses is the use of ceiling-mounted wooden frames subdivided into rectangular cells measuring approximately 0.5 m × 1.0 m

 https://doi.org/10.3390/engproc2026164017

with a height of 150 mm. This repetitive grid provides an organized nesting surface but introduces significant inspection challenges. The shallow 150 mm vertical clearance restricts viewing angles, while wooden beams forming the cell boundaries create partial occlusions. Nests are typically attached to the upper interior surfaces of these cells, often near corners or beam edges, making them difficult to observe completely from below. Conventional camera-based inspection approaches are limited by fixed viewpoints and insufficient geometric coverage. Therefore, reliable nest localization requires a coordinated motion strategy that considers both translational positioning and dynamic orientation adjustment. From a robotics perspective, the inspection task can be formulated as a structured coverage problem under geometric constraints. A Cartesian gantry robot positioned beneath the nesting frame provides deterministic X–Y motion [6], enabling systematic traversal of the inspection area. However, planar motion alone is insufficient to resolve occlusions within shallow cells. Incorporating a two-degree-of-freedom (2-DOF) pan–tilt camera mechanism allows adaptive viewpoint control, enhancing visibility and depth perception. Depth-sensing cameras further enable 3D reconstruction of nest geometry, offering quantitative localization beyond traditional RGB imaging. The central challenge lies in designing a trajectory strategy that maximizes detection coverage while minimizing inspection time and redundant viewpoints [7–9].

This paper proposes a cell-aware trajectory planning and active perception framework for depth-based nest localization in swiftlet houses. The method integrates structured grid decomposition aligned with 0.5 m × 1 m cells, coordinated gantry translation, and adaptive pan–tilt orientation control. A coarse-to-fine depth sensing strategy is employed to first detect candidate nest regions and subsequently refine their 3D positions through optimized viewpoints. The system was experimentally validated in a commercial swiftlet house environment. Results demonstrate reliable nest tracking with millimeter-level localization accuracy and reduced inspection time compared to conventional raster scanning approaches. The proposed approach contributes to the advancement of smart avian farming robotics by providing a practical and non-invasive solution for automated nest monitoring in structured overhead environments.

## 2. System Architecture

The proposed system integrates a structured mechanical platform and an active depth-sensing module designed specifically for the inspection of ceiling-mounted swiftlet nesting frames. The architecture consists of two principal components: the nesting frame geometry that defines the inspection constraints and the robotic platform that performs coordinated translational and rotational motion beneath the frame. Commercial swiftlet houses typically employ ceiling-mounted wooden grid structures to standardize nest attachment locations. The frame is composed of rectangular cells with nominal planar dimensions of 0.5 m × 1 m and a vertical clearance of 150 mm (Figure 1). This repetitive geometry simplifies spatial decomposition but introduces strict visibility constraints. The shallow cell height limits the allowable viewing angles, while wooden beams forming the grid boundaries create partial occlusion zones near edges and corners. The beam thickness is denoted as b (typically 20–30 mm). The effective occlusion angle $\theta_o$ caused by the beam can be approximated as follows:

$$\theta_o = \tan^{-1}\left(\frac{b}{h_c}\right) \quad (1)$$

where $h_c = 150$ mm is the cell height. For $b = 25$ mm, $\theta_o \approx 9.5°$, indicating that shallow tilt angles result in incomplete visibility of nests attached near beam edges. The total inspection area A for a frame consisting of $N_x \times N_y$ cells is as follows:

$$A = N_x N_y (0.5 \times 1.0)\left(\mathrm{m}^2\right) \quad (2)$$

https://doi.org/10.3390/engproc2026164017

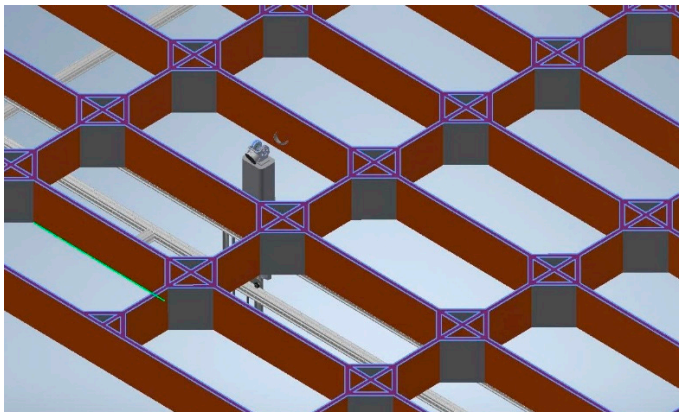

**Figure 1.** A view of a gantry robot within wooden frames in the swiftlet house.

This structured grid forms the basis for discrete cell-aware trajectory planning.

## 3. Robotic Platform

The robotic platform consists of a Cartesian gantry robot providing planar positioning and a 2-DOF pan–tilt depth camera unit enabling active orientation control [10,11]. The gantry robot provides independent X–Y translational motion beneath the nesting frame. A world coordinate frame {W} is defined at the gantry base. The end-effector position is expressed as follows:

$$p_G = \begin{bmatrix} x \\ y \\ z_0 \end{bmatrix} \tag{3}$$

where $x$ and $y$ are linear displacements and $z_0$ is fixed relative to the frame height. The workspace coverage satisfies

$$0 \le x \le L_f; \quad 0 \le y \le W_f \tag{4}$$

where $L_f$ and $W_f$ denote frame dimensions. The system provides positioning accuracy of ±1.5 mm and repeatability of ±0.8 mm, ensuring reliable spatial referencing for nest localization.

Mounted on the gantry carriage is a compact pan–tilt unit providing two rotational degrees of freedom: $R_{pan}(\phi)$ about the vertical axis and $R_{tilt}(\psi)$ about the horizontal axis. The orientation matrix is defined as follows:

$$R = R_{pan}(\phi)R_{tilt}(\psi) \tag{5}$$

The integration of precise Cartesian positioning and adaptive orientation control enables structured inspection under strict geometric constraints. This architecture forms the foundation for the trajectory planning strategy presented in the following section.

## 4. Kinematic Modeling

Accurate nest localization requires precise modeling of both the translational motion of the gantry robot and the rotational motion of the pan–tilt camera unit. The combined kinematic model enables computation of the camera pose in the global reference frame and facilitates consistent 3D mapping of detected nests. The gantry robot is modeled as a three-degree-of-freedom Cartesian manipulator providing independent motion along orthogonal

$X$, $Y$, and $Z$ axes (Figure 2). The vertical height of the camera mount remains constant; the forward kinematics of the gantry are therefore represented as a pure translation matrix.

$$T_{WG} = \begin{bmatrix} 1 & 0 & 0 & x \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

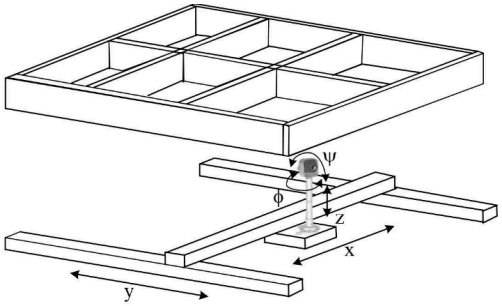


**Figure 2.** Degrees of freedom of the gantry robot and pan–tilt camera (translational axes and rotational angles).

The camera orientation is controlled by a two-degree-of-freedom pan–tilt mechanism. The pan angle $\phi$ defines rotation about the vertical axis, while the tilt angle $\psi$ defines rotation about the horizontal axis. The corresponding rotation matrices are as follows:

$$R_{pan}(\phi) = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

$$R_{tilt}(\psi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{bmatrix} \tag{8}$$

The combined orientation matrix is as follows:

$$R_{PT} = R_{pan}(\phi)R_{tilt}(\psi) \tag{9}$$

This formulation enables dynamic adjustment of viewing direction to mitigate occlusion within the 150 mm cell height constraint.

The complete camera pose in the world frame is obtained by combining gantry translation with pan–tilt rotation, as shown in Figure 3. Toffset represents the fixed transformation between the gantry carriage and the camera mounting base. The homogeneous transformation from world frame to camera frame is as follows:

$$T_{WC} = T_{WG}T_{offset}\begin{bmatrix} R_{PT} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{10}$$

A 3D point measured in camera coordinates $\mathbf{p}_C$ is transformed into world coordinates as follows:

$$\mathbf{p}_W = T_{WC}\mathbf{p}_C \tag{11}$$

This unified transformation framework ensures accurate spatial mapping of nests detected from varying viewpoints.

https://doi.org/10.3390/engproc2026164017



**Figure 3.** Depth camera monitoring the nest on the wooden frames.

## 5. Trajectory Planning Strategy

Efficient inspection of ceiling-mounted swiftlet nesting frames requires a structured trajectory planning approach that considers geometric constraints, sensor field of view, and motion limitations. The proposed strategy combines grid-based coverage with adaptive viewpoint refinement to ensure reliable nest localization while minimizing inspection time [12].

The inspection task is formulated as a constrained optimization problem over a sequence of gantry positions and pan–tilt orientations. Let total inspection time be denoted by $T$, redundancy ratio by $R$, and occlusion penalty by $O$. The trajectory cost function is defined as follows:

$$J = \alpha T + \beta R + \gamma O \tag{12}$$

where $\alpha$, $\beta$, and $\gamma$ are weighting factors reflecting operational priorities. Inspection time can be expressed as follows:

$$O = \sum_{i=1}^{N} (1 - VR_i) \tag{13}$$

where $VR_i$ is the visibility ratio of cell $i$. The optimization objective is to minimize $J$ while satisfying mechanical constraints on gantry speed and pan–tilt angular limits.

Given the structured wooden frame geometry (0.5 m × 1 m cells), the inspection area is decomposed into discrete grid elements. Inspection nodes are defined at the geometric centers of each cell:

$$x_i = \left(i + \frac{1}{2}\right)L; y_j = \left(j + \frac{1}{2}\right)W \tag{14}$$

where $L = 1.0$ m and $W = 0.5$ m. This structured decomposition simplifies trajectory planning by transforming the continuous coverage problem into traversal of discrete nodes.

The baseline raster scanning approach sequentially visits all nodes row by row with fixed pan–tilt orientation. Although simple, raster scanning often produces redundant coverage and unnecessary sensing in empty cells. Total raster travel distance is approximated by the following:

$$D_{raster} = N_y L_f + \left(N_y - 1\right)W_f \tag{15}$$

where $L_f$ and $W_f$ are frame dimensions.

In contrast, the proposed adaptive path incorporates detection feedback. A coarse scan is first performed at each node. If no candidate nest is detected, the system proceeds to the next node without additional orientation refinement. If a nest is detected, a localized pan–tilt adjustment is executed for fine tracking. This selective refinement reduces redundant viewpoints and unnecessary sensing operations.

Experimental evaluation showed that adaptive scanning reduced travel distance by approximately 20% and total inspection time by 28% compared to raster scanning. The

trajectory planning strategy leverages the regular cell geometry and integrates detection-driven feedback to achieve efficient, structured, and adaptive inspection of swiftlet nesting frames.

## 6. Active Vision Strategy

To enhance detection robustness and reduce unnecessary motion, an active vision strategy is implemented that dynamically adjusts sensing resolution and viewpoint based on scene feedback. The strategy is divided into two sequential phases: a coarse detection phase for rapid screening and a fine tracking phase for precise 3D localization and geometric characterization, as shown in Figure 4.

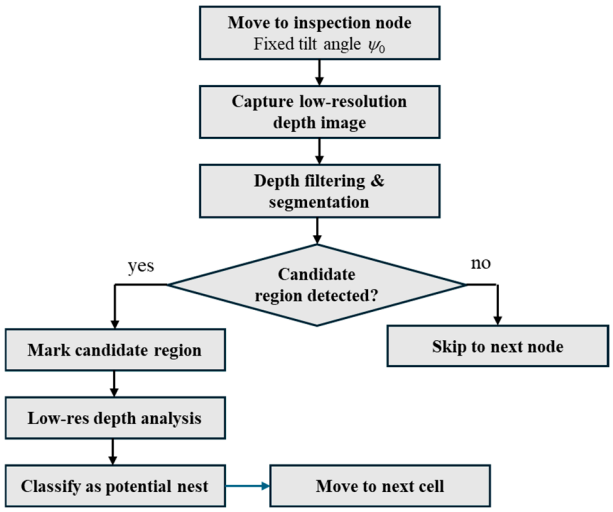


**Figure 4.** Flowchart of coarse detection pipeline.

During the coarse phase, the gantry robot moves to predefined inspection nodes while the pan–tilt unit maintains a fixed nominal tilt angle $\psi_0$, selected to maximize coverage of the 0.5 m × 1.0 m cell area. The camera operates at reduced depth resolution to increase acquisition speed and decrease processing load. Depth images are segmented using threshold-based filtering and region growing to detect protruding structures consistent with swiftlet nests (Figure 5). $Z(x, y)$ denotes the depth map; candidate regions are identified as follows:

$$\left|Z(x,y) - Z_{ref}\right| > \delta \tag{16}$$

with $Z_{ref}$ representing the planar background depth and $\delta$ representing a predefined deviation threshold. Regions exceeding a minimum projected area $A_{min}$ are marked as potential nests.

The objective of this phase is to rapidly classify each cell as empty or containing candidate structures. Experimental results show that coarse scanning reduces average sensing time per cell to 0.8 s while maintaining detection sensitivity above 90%.

If a candidate region is detected, the system transitions to the fine tracking phase. The pan–tilt mechanism performs local orientation refinement by adjusting pan angle $\phi$ and tilt angle $\psi$ to maximize the visible surface area of the detected structure (Figure 6). Viewpoint optimization is formulated as follows:

$$(\phi^*, \psi^*) = \underset{\phi,\psi}{\operatorname{argmax}} V(\phi, \psi) \tag{17}$$

where $V(\phi, \psi)$ represents the visible nest surface area estimated from projected depth variance.

https://doi.org/10.3390/engproc2026164017

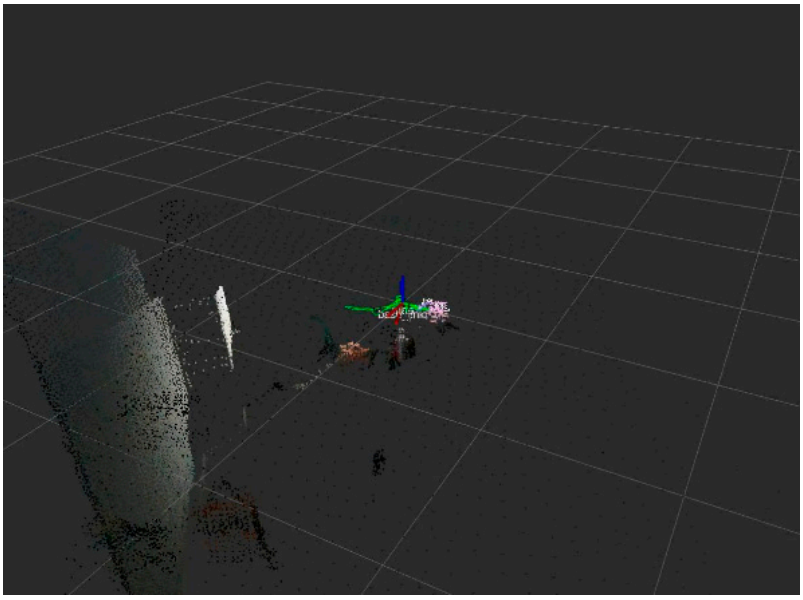

**Figure 5.** An image captured by depth cameras inside the swiftlet house.

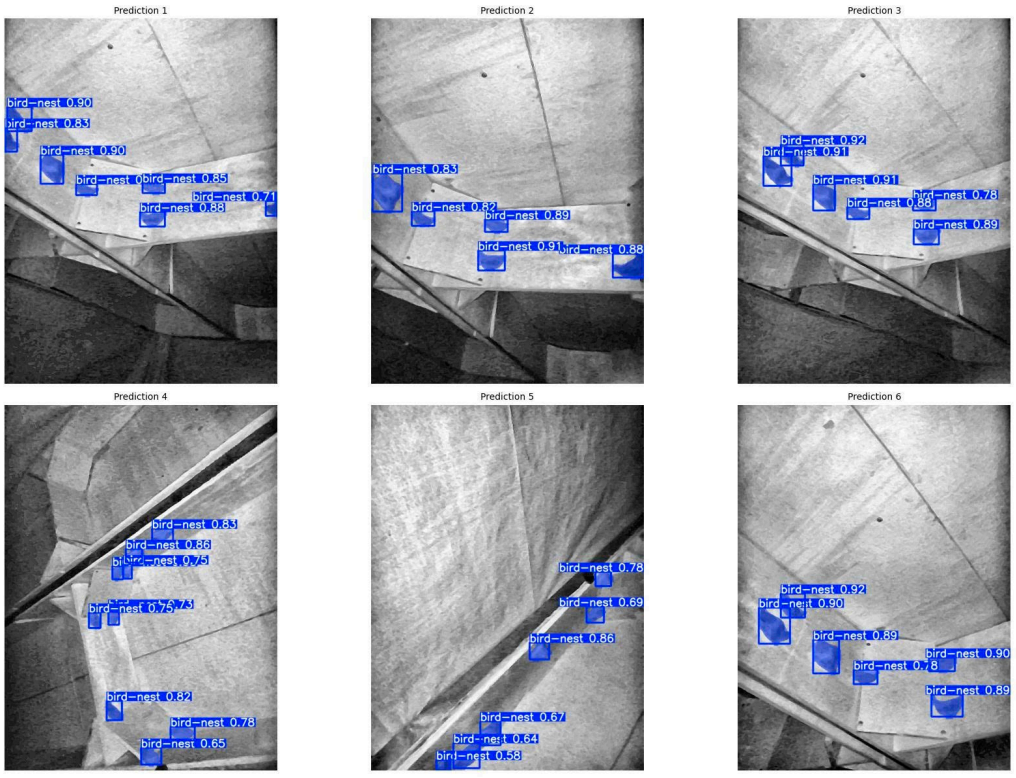


**Figure 6.** Pan–tilt refinement around the detected nest.

High-resolution depth acquisition is then activated to improve geometric reconstruction. Multiple depth frames captured from slightly varied angles are fused using point cloud registration to reduce noise. The reconstructed 3D nest centroid $\mathbf{p}_n$ is computed as follows:

$$\mathbf{p}_n = \frac{1}{N}\sum_{i=1}^{N}\mathbf{p}_i \tag{18}$$

where $\mathbf{p}_i$ are points within the segmented cluster.

This refinement improves the mean localization accuracy to below 10 mm and enhances shape estimation consistency. Repeatability tests across multiple inspections show a positional variance under 6 mm. The proposed active vision strategy effectively balances efficiency and precision. By limiting high-resolution processing to detected regions and optimizing viewpoints adaptively, the system achieves reliable nest localization while minimizing inspection time and mechanical motion.

## 7. Experimental Results and Discussion

Experimental validation was conducted in a commercial swiftlet house equipped with a ceiling-mounted wooden nesting frame composed of 0.5 m × 1.0 m cells and 150 mm in height. The gantry-based pan–tilt depth sensing system was deployed beneath a 4 × 6 grid (24 cells) to evaluate trajectory efficiency, localization accuracy, reconstruction quality, and occlusion handling performance. The proposed adaptive trajectory was compared with conventional raster scanning (Figure 7).

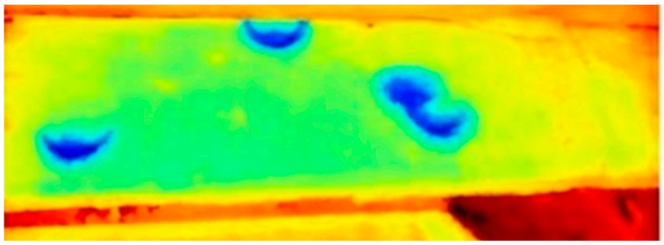

**Figure 7.** Detecting the nests on the wooden frames by a depth camera attached to the gantry.

Raster scanning required 312 s to inspect the full grid, while the adaptive strategy reduced inspection time to 224 s, corresponding to a 28.2% improvement. Total gantry travel distance decreased from 18.4 m to 14.7 m (20.1% reduction), primarily due to elimination of redundant fine inspections in empty cells.

Ground-truth nest positions were measured manually using laser distance tools for comparison (Figure 8). The system achieved a mean 3D error of 8.7 mm with a standard deviation of 3.2 mm. Repeatability tests across five repeated scans yielded positional variance below 6 mm, demonstrating stable performance under consistent lighting conditions. Depth noise was evaluated as a function of camera-to-nest distance. Noise variance increased approximately linearly beyond 1.2 m working distance. Optimal performance was observed within 0.9–1.1 m. Tilt angle significantly influenced reconstruction completeness. Experiments showed that a tilt range of $20° \leq \psi \leq 35°$ maximized visible surface area while minimizing shadow regions. Outside this range, occlusion from wooden beams increased reconstruction gaps. Occlusion caused by the 150 mm cell height and beam edges was mitigated through pan adjustment. Adaptive pan refinement improved the average visibility ratio from 82% (fixed orientation) to 94%. Fine viewpoint adjustment reduced shadowed regions by approximately 35% compared to static viewing. The results highlight a trade-off between gantry translation and pan–tilt rotation. Excessive gantry repositioning increases inspection time, while excessive angular refinement increases sensing overhead. The adaptive strategy balances these factors by limiting fine adjustments to detected cells. The 150 mm cell height imposes geometric constraints that restrict viewing angles and necessitate careful tilt selection. Despite this limitation, reliable nest localization was achieved. Scalability analysis suggests that inspection time grows approximately linearly with cell count; however, adaptive skipping of empty cells maintains efficiency in larger facilities. Finally, the non-contact depth sensing approach minimizes disturbance to birds, supporting

https://doi.org/10.3390/engproc2026164017

animal welfare considerations. No observable behavioral stress was detected during trials, indicating suitability for routine monitoring in smart swiftlet farming environments.

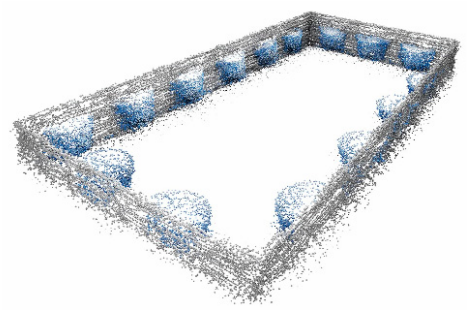

**Figure 8.** Schematic representation of the stacked layers from the EBN 2D scanning process on a wooden frame.

## 8. Conclusions

This study presented a cell-aware trajectory planning strategy for depth-based nest localization using a pan–tilt vision system mounted on a Cartesian gantry robot. By integrating structured grid decomposition with adaptive viewpoint refinement, the proposed method maximized inspection coverage while reducing redundant motion and occlusion effects. Experimental validation in a real swiftlet house demonstrated reliable 3D nest localization with sub-centimeter accuracy and significant inspection time reduction compared to raster scanning. The results confirm the practical feasibility of non-contact, automated monitoring in confined ceiling-mounted frames. The proposed framework advances smart avian farming robotics and structured robotic inspection in agricultural environments.

**Author Contributions:** Conceptualization, N.T.T.; Methodology, N.M.T. and N.T.T.; Software, T.T.Q. and N.M.T.; Validation, T.T.Q. and N.M.T.; Formal analysis, T.T.Q. and N.M.T.; Investigation, T.T.Q. and N.M.T.; Resources, N.M.T.; Data curation, T.T.Q. and N.M.T.; Writing—original draft, T.T.Q. and N.M.T.; Writing—review and editing, N.T.T.; Visualization, T.T.Q.; Supervision, N.T.T.; Project administration, N.T.T.; Funding acquisition, N.T.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is funded by the University of Economics Ho Chi Minh City (UEH) under the Research Group number: TRL-039.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent for participation was obtained from all subjects involved in the study.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Thorburn, C. The edible birds' nest boom in Indonesia and South-east Asia: A nested political ecology. *Food Cult. Soc.* **2014**, *17*, 535–553. [CrossRef]
2. Ibrahim, M.M.; Zakaria, Z.A.B.; Amin, F.M.; Omar, A.R. Comparative Histological Evaluations of the Sublingual Salivary Glands of EBN Swiftlets (*Aerodramus fuciphagus*) in Man-Made Houses and Natural Caves. *Pertanika J. Trop. Agric. Sci.* **2017**, *40*, 19.
3. Goh, K.M.; Lim, L.L.; Krishnamoorthy, S.; Lai, W.K.; Maul, T.; Chaw, J.K. Recent advancement of intelligent-systems in edible birds nest: A review from production to processing. *Multimed. Tools Appl.* **2024**, *83*, 51159–51209. [CrossRef]
4. Dilaver, H.; Dilaver, K.F. Robotics systems and artificial intelligence applications in livestock farming. *J. Anim. Sci. Econ.* **2024**, *3*, 63–72.

5. Trieu, N.M.; Thinh, N.T. Development of grading system based on machine learning for dragon fruit. In *Regional Conference in Mechanical Manufacturing Engineering*; Springer Nature: Singapore, 2021; pp. 230–243.
6. Wang, W.; Ma, J.; Cheng, Z.; Zhu, Z.; Xian, Y.; de Silva, C.W.; Lee, T.H. Robust Precision Motion Control of Dual-Drive Gantry-Type Cartesian Robot with Workspace Constraints. *IEEE/ASME Trans. Mechatron.* **2025**, *30*, 5554–5566. [CrossRef]
7. Milella, A.; Marani, R.; Petitti, A.; Reina, G. In-field high throughput grapevine phenotyping with a consumer-grade depth camera. *Comput. Electron. Agric.* **2019**, *156*, 293–306. [CrossRef]
8. Shi, X.; Ke, Q.; Lei, J.; Yuan, Z. Corrosion Failure Analysis of Electrical Equipment in Distribution Power System under High Temperature and High Humidity Environment. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2020; Volume 1578.
9. Trinh, D.A.V.; Thinh, N.T. A Study of an Agricultural Indoor Robot for Harvesting Edible Bird Nests in Vietnam. *AgriEngineering* **2024**, *6*, 113–134. [CrossRef]
10. Khatib, O. Real-time obstacle avoidance for manipulators and mobile robots. *Int. J. Robot. Res.* **1986**, *5*, 90–98. [CrossRef]
11. Trung Hieu, N.T.; Minh Trieu, N.; Tri Dung, D.; Truong Thinh, N. Advanced Sliding Mode Control Strategy for High-Performance 3D Concrete Printing. *Automation* **2025**, *6*, 22. [CrossRef]
12. Hornung, A.; Wurm, K.M.; Bennewitz, M.; Stachniss, C.; OctoMap, W.B. An efficient probabilistic 3D mapping framework based on octrees. *Auton. Robot.* **2013**, *34*, 189–206. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.