Digital-Twin-Guided Target-Specific Next-Best-View Planning for Building Inspection by UAV Thermography

Hossein Khandandel, Yining Wen, and Kaiwen Chen*

***Abstract*—**Unmanned aerial vehicle (UAV) thermography can localize building-envelope regions for closer inspection, but broad-coverage survey views may leave localized targets distant, oblique, partially visible, or blocked by surrounding geometry. We present a digital-twin-guided, target-specific next-best-view (NBV) method. It fuses registered thermal anomaly predictions from multiple survey views into localized 3D target patches, then evaluates close-range candidates using ray-cast visibility, incidence, ground-sampling distance, and visible patch coverage. Evaluation uses 81 registered UAV survey images and their calibrated poses, yielding 19 target patches from six spatially distinct facade regions and 139 prediction-independent control patches distributed across the reconstructed building envelope. Explicit visibility reasoning increases mean score from 0.551 to 0.792 and produces usable views for 17/19 patches, compared with 12/19 for a matched no-occlusion ablation that ignores visibility during viewpoint selection. The score difference is much smaller on the envelope-control set (0.899 versus 0.880), localizing the benefit to geometrically difficult prediction-derived targets.

I. INTRODUCTION

Unmanned aerial vehicle (UAV) infrared thermography supports building-envelope surveys [1, 2]. Thermal-image quality can be affected by shading, solar radiation, reflections, and limited camera resolution; the present work does not correct these radiometric or environmental effects. A separate geometric problem arises when a broad-coverage trajectory leaves a localized region distant, oblique, only partially visible, or obstructed by surrounding features such as protrusions, setbacks, or vegetation. Such a survey view may be adequate for localization but inadequate for targeted close-range follow-up inspection.

Prior damage-augmented digital-twin studies have registered detected damage in metric building geometry [3]. Gholizadeh Lonbar and Chen registered thermal-anomaly predictions from UAV imagery into a neural radiance field (NeRF)-enhanced 3D building representation and reported subsequent targeted close-range facade acquisition [4]. Song et al. connected facade traversal and defect segmentation to digital-model updates [5], while Liu et al. merged robot observations with pose-matched computer-aided design (CAD) renderings for zero-shot anomaly localization [6]. These systems enable target registration and geometric reconstruction, but do not explicitly use this information to plan target-specific, close-range follow-up viewpoints that account for surrounding occlusions and inspection-view quality.

Previous studies on occlusion-aware view planning have leveraged occlusion information to guide subsequent sensing and identify occlusion-free viewpoints [7, 8]. More recent approaches include receding-horizon next-best-view (NBV) exploration [9], ray-tracing-based planning for large-scale surface reconstruction [10], mapless view selection guided by image-based neural uncertainty [11], neural visibility fields for active mapping [12], and learned policies for damage inspection or revealing occluded targets [13, 14]. Geometry- and building information modeling (BIM)-based inspection planners further integrate visibility with coverage, image scale, ground-sampling distance (GSD), and route sequencing [15, 16]. These studies address individual components of the proposed framework, but do not integrate them to plan target-specific follow-up viewpoints from predicted thermal anomalies.

Given already-registered predictions, the research problem is to construct a localized, multi-view-supported target representation and select a close-range viewpoint that remains observable and satisfies image-quality criteria in the reconstructed surroundings. To address it, this study combines multi-view target construction, target-specific candidate generation, and visibility-gated evaluation of incidence, resolution, and patch coverage before selecting the NBV. The contribution is this application-specific inspection-planning combination, not thermal prediction, NeRF reconstruction, 2D-to-3D registration, generic ray casting or visibility modeling, or generic NBV formulation. Waypoint ordering is applied separately downstream and does not enter the observation-quality score S.

Hossein Khandandel is with the University of Alabama, Tuscaloosa, AL35401 USA (e-mail: hkhandandel@crimson.ua.edu);
Yining Wen is with the University of Alabama, Tuscaloosa, AL35401 USA (e-mail: ywen11@crimson.ua.edu);
Kaiwen Chen, Corresponding Author, is with the University of Alabama, Tuscaloosa, AL35401 USA (phone: 205-348-4682; e-mail: kaiwen.chen@ua.edu)

II. METHODOLOGY

A. Framework and Candidate Viewpoints

Fig. 1 places the proposed planner within the existing inspection pipeline. RGB survey images support NeRF/Instant-NGP reconstruction, while thermal-prediction masks are registered to the reconstructed 3D twin and COLMAP provides calibrated camera parameters. A Python bridge passes registered prediction locations, reconstructed twin geometry used for ray tests, and camera calibration to the planner. The reported contribution spans multi-view target construction, candidate evaluation, and NBV selection. The registered reconstruction is mapped into the gravity-aligned metric twin using the fixed calibration of 16.28m/unit.

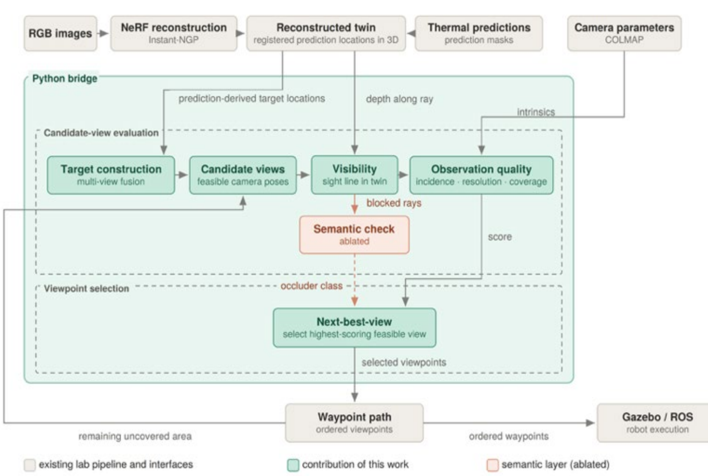

Fig. 1. Framework context for the target-specific NBV planner.

Target patches: Each target patch is a localized reconstructed surface region formed from registered thermal anomaly predictions from multiple survey views. Registered prediction components are ray-cast to the reconstructed surface and fused into 3D regions when supported by at least three views. Target patches are sampled from these fused regions and retain their centroid, outward surface normal, and eight additional directional-extrema points, giving nine representative 3D points for coverage evaluation. Surface normals are estimated hierarchically. Boundary patches use the outward building-footprint normal. For other patches, a local RANSAC plane is first fit to the reconstructed mesh around the patch; if this mesh-based estimate fails, a RANSAC plane fit to the fused 3D patch points is used as a fallback.

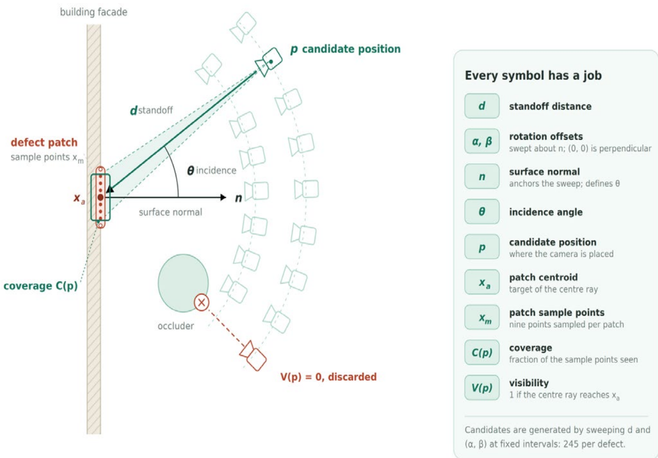

Fig. 2. Candidate-view geometry notation for target-specific NBV evaluation.

In Fig. 2, the labels 'defect patch' and 'per defect' refer to the prediction-derived target patch defined above; the evaluated targets are not independently verified defects.

Candidate Generation: For target patch j on the building facade, let $\mathbf{x}_{a,j}$ be the patch centroid and $\mathbf{n}_j$ its outward unit surface normal. $\mathbf{p}$ is the camera candidate position, with azimuth α, elevation β, and distance d from $\mathbf{x}_{a,j}$. To represent the candidate position in the local patch frame illustrated in Fig. 2, define

$$\mathbf{t}_{1,j} = \frac{\mathbf{z}\times\mathbf{n}_j}{\|\mathbf{z}\times\mathbf{n}_j\|}, \qquad \mathbf{t}_{2,j} = \mathbf{n}_j \times \mathbf{t}_{1,j}, \quad (1)$$

where $\mathbf{z}$ is the vertical axis in the world frame. Therefore, a unit direction from the target toward a candidate camera is

$$\mathbf{u}_j(\alpha,\beta) = \cos\beta\big(\cos\alpha\, \mathbf{n}_j + \sin\alpha\, \mathbf{t}_{1,j}\big) + \sin\beta\, \mathbf{t}_{2,j}, \quad (2)$$

and the candidate position is

$$\mathbf{p}(d,\alpha,\beta) = \mathbf{x}_{a,j} + d\,\mathbf{u}_j(\alpha,\beta). \quad (3)$$

The verified search parameters are $d \in \{3,4,5,6,7\}$ m and $\alpha,\beta \in \{-30°, -20°, -10°, 0°, 10°, 20°, 30°\}$, yielding 245 candidates per target.

Camera Orientation: The camera local-frame reference axes are

$$\begin{aligned} \mathbf{f} &= \frac{\mathbf{x}_{a,j}-\mathbf{p}}{\|\mathbf{x}_{a,j}-\mathbf{p}\|}, \quad \mathbf{r} = \frac{\mathbf{f}\times\mathbf{z}}{\|\mathbf{f}\times\mathbf{z}\|}, \\ \mathbf{u}_c &= \mathbf{r}\times\mathbf{f}, \quad \mathbf{Q} = [\,\mathbf{r}\ \mathbf{u}_c\ -\mathbf{f}\,]. \end{aligned} \quad (4)$$

Here $\mathbf{Q}$ follows the NeRF $(-z)$ forward camera convention; the calibrated OpenCV camera model is used for image projection. Candidate viewpoints are filtered using four feasibility constraints: (1) at least 1 m clearance from the reconstructed mesh, to avoid viewpoints too close to the building surface; (2) inclusion within the plan-view camera-trajectory hull, to restrict candidates to the region spanned by the survey trajectory; (3) a minimum altitude of 3 m; and (4) an upper-height constraint defined by the reconstructed scene extent. These constraints define feasible viewpoint locations only and do not collision-check the connecting flight paths.

B. Visibility-Gated Scoring

Visibility (V): To determine whether a camera-patch ray is blocked, the reconstructed twin is queried along the camera-to-target ray. Let $d_{\text{twin}}(\mathbf{p} \to \mathbf{x}_{a,j})$ denote the first-hit distance along the ray to the patch centroid. Because each candidate is generated at standoff distance d, its expected camera-to-target distance is equivalently $d = \| \mathbf{x}_{a,j} - \mathbf{p} \|$, and visibility is

$$V_j(\mathbf{p}) = \mathbf{1}\big[d_{\text{twin}}(\mathbf{p} \to \mathbf{x}_{a,j}) \ge d - \varepsilon\big], \quad (5)$$

where $\mathbf{1}[\cdot]$ is the indicator function and ε is the metric tolerance for reconstruction error. Thus $V_j = 1$ when the ray reaches the target before any reconstructed blocker and $V_j = 0$ otherwise.

Incidence Angle (A): The angle between the outward surface normal $\mathbf{n}_j$ and the unit target-to-camera direction $\mathbf{u}_j$ is the incidence angle θ_j. Therefore,

$$A_j(\mathbf{p}) = \cos\alpha\cos\beta = \cos\theta_j, \quad (6)$$

so $A_j = 1$ corresponds to a perpendicular on-normal view, while more oblique views receive smaller values.

Resolution at the Surface (R): The calibrated focal lengths are $f_x = 770.0$ px and $f_y = 765.4$ px, giving $f_{\min} = \min(f_x, f_y) = 765.4$ px. The nominal on-axis GSD and task requirement are

$$g_0(d) = \frac{d}{f_{\min}}, \qquad g_{\text{req}} = \frac{L_{\min}}{N_{\min}}, \quad (7)$$

and the resolution score is

$$R_j(\mathbf{p}) = \min\left(1, \frac{g_{\text{req}}}{g_0(d)}\right). \quad (8)$$

Here $L_{\min}$ is the smallest feature size of interest and $N_{\min}$ is its required pixel sampling. Smaller g_0 means finer detail, and $R_j = 1$ once the required GSD is met.

Patch Coverage (C): Let $\mathbf{x}_{j,m}$, $m = 1, \dots, K$, denote the retained patch samples, with $K = 9$ (the centroid and eight directional extrema from the fused 3D patch geometry). For each sample, $I_{j,m}(\mathbf{p}) = 1$ when it lies in front of the camera and projects inside the calibrated image boundary, and $V_{j,m}(\mathbf{p})$ is its individual visibility using the same first-hit ray test. Coverage is

$$C_j(\mathbf{p}) = \frac{1}{K}\sum_{m=1}^{K} I_{j,m}(\mathbf{p}) V_{j,m}(\mathbf{p}). \quad (9)$$

Thus $C_j = 1$ means all nine representative samples are both visible and in frame. Unlike V_j, which tests only the patch centroid, C_j captures partial occlusion and image-boundary clipping.

Visibility-Gated Observation Score: Each feasible candidate is assigned

$$S_j(\mathbf{p}) = V_j(\mathbf{p})\big[w_A A_j(\mathbf{p}) + w_R R_j(\mathbf{p}) + w_C C_j(\mathbf{p})\big],$$
$$w_A + w_R + w_C = 1. \quad (10)$$

The nonnegative weights w_A, w_R, w_C control the relative contribution of incidence, resolution, and patch coverage. Their sum is constrained to one, so the bracketed term is a normalized weighted average. The evaluation uses $w_A = w_R = w_C = 1/3$, so no soft-quality term is favored a priori. Visibility V_j is a hard gate outside the weighted sum: a blocked centroid gives $S_j = 0$ regardless of the other terms. The quantities V, A, R, C, S and the weights are dimensionless; the soft quality terms lie in $[0,1]$.

Best-View Selection: Let $\mathcal{P}_j$ be the candidates that pass the feasibility tests. The target-specific next-best view is

$$\mathbf{p}_j^* = \underset{\mathbf{p}\in\mathcal{P}_j}{\operatorname{argmax}}\ S_j(\mathbf{p}). \quad (11)$$

The argmax returns the feasible camera position with the highest visibility-gated observation score. If no candidate in $\mathcal{P}_j$ has $V_j > 0$, the target is flagged as having no feasible visible view in the search domain. Waypoint ordering is applied only after the target-specific best views are fixed and does not change $\mathbf{p}_j^*$.

III. RESULTS

The building inspection case consists of 81 registered UAV survey images with calibrated camera poses, from which 19 prediction-derived target patches were constructed across six spatially distinct façade regions. To examine whether the proposed planning approach generalizes beyond the thermally predicted regions, 139 prediction-independent envelope-control patches were additionally sampled across the reconstructed building exterior. These control patches represent comparison surface locations rather than defect ground truth and are evaluated within the same digital twin and camera configuration as the prediction-derived targets.

For each target patch, the proposed NBV planner searches for a close-range observation that improves target visibility and observation quality relative to the original survey. Four inspection scenarios are considered for comparison: (1) the as-flown UAV survey, which selects the best available observation among the 81 registered survey poses; (2) a synthetic re-flight survey representing a systematic façade inspection trajectory, with 68 feasible poses generated from 24 azimuths and three flight altitudes; (3) random selection from the feasible candidate viewpoints; and (4) a no-occlusion case that selects viewpoints without considering visibility. All cases are evaluated on the same reconstructed building model and target patches using a common observation-quality formulation incorporating visibility, incidence angle, GSD, and visible patch coverage.

Results are reported primarily using the mean observation-quality score S, with the usable-view rate (V>0) providing a complementary indication of whether each target can be observed from at least one selected viewpoint. The primary case uses a visibility tolerance of 0.05 m and a required GSD of 0.005 m/px. Additional sensitivity cases vary the visibility tolerance (0.01–0.10 m), required GSD (0.003–0.010 m/px), and angular bound (±30°–±40°), and separately exclude the three target patches whose surface normals were estimated using fused-patch RANSAC.

Table I shows that the occlusion-aware method raises mean S from 0.551 to 0.792 on prediction-derived patches, an absolute gain of 0.241 (43.7%), and raises usable views from 12/19 (0.632) to 17/19 (0.895), with five recoveries and no reverse case. The component means explain the mechanism: relative to the no-occlusion ablation, the method accepts lower incidence ($A = 0.816$ versus 0.944) and resolution ($R = 0.771$ versus 0.887) to obtain higher visibility ($V = 0.895$ versus 0.632) and visible patch coverage ($C = 0.789$ versus 0.696). Fig. 5 summarizes this trade-off.

TABLE I Performance comparison of viewpoint selection methods.

Method	Target (n=19)	Control (n=139)
As-flown	0.494 / 0.895	0.329 / 0.748
Re-flight	0.586 / 1.000	0.605 / 0.928
Random	0.497 / 0.647	0.625 / 0.794
No-occl. ablation	0.551 / 0.632	0.880 / 1.000
Occlusion-aware	**0.792 / 0.895**	**0.899 / 1.000**

Note: Entries are mean S / usable $V > 0$; no-view targets remain in the denominator.

The contrast is much smaller on the 139 envelope-control patches: both selectors find usable views for every patch and differ by only 0.019 in mean score (0.899 versus 0.880), compared with 0.241 in score and 0.263 in usable rate on the prediction-derived set. This differential, together with Fig. 3, shows that explicit visibility primarily changes selection where localized targets lie near blocking geometry. For target 12 in region 4, supported by 18 survey views, the best as-flown observation is 53.22 m from the target with a GSD of 69.54 mm/px. The no-occlusion-selected pose is blocked after 3.3 m, whereas the proposed occlusion-aware pose maintains an unobstructed sight line to the target.

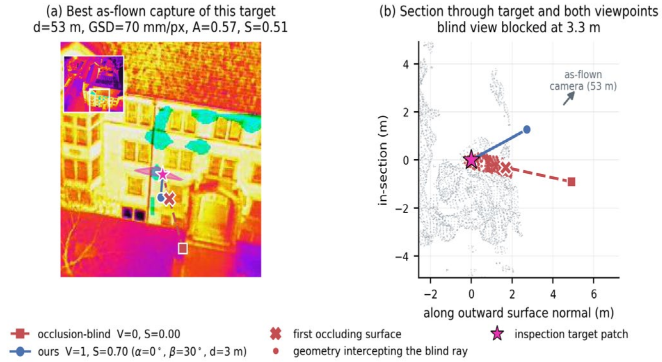

Fig. 3. Occlusion example for target 12 in region 4, comparing the blocked no-occlusion pose with the unobstructed proposed pose.

Fig. 4 further illustrates the Target 12 example using the original UAV RGB imagery and the reconstructed candidate-view geometry. The registered target location and first blocking intersection are projected onto the RGB image, identifying the physical obstruction. The candidate-space evaluation shows the distribution of visible and occluded viewpoints, including the visibility-blind selection and the proposed NBV. Of the 245 nominal candidates, 222 satisfy the position-feasibility constraints and are classified as visible or occluded in Fig. 4(b). Among these feasible candidates, the proposed method selects an unobstructed viewpoint with an observation-quality score of $S = 0.696$.

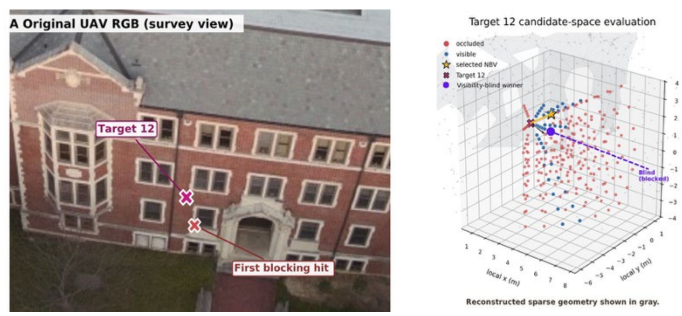

Fig. 4. Target-specific NBV selection for Target 12. (a) Original UAV RGB image. (b) Candidate-space evaluation.

The as-flown and re-flight baselines further show that a visible centroid alone is insufficient: their usable rates are 0.895 and 1.000, but mean scores are only 0.494 and 0.586. Random has a slightly higher usable rate than the ablation (0.647 versus 0.632). This pattern is consistent with visibility-blind optimization favoring close, near-normal poses that can be blocked, while random sampling can occasionally select a clear off-normal pose.

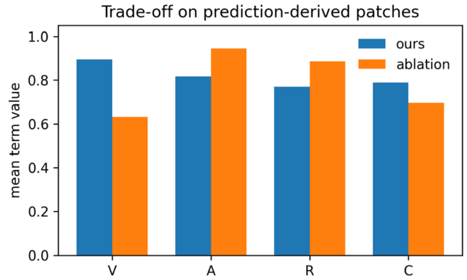

Fig. 5. Component trade-off on prediction-derived patches. Occlusion reasoning increases V and C while accepting lower A and R than the no-occlusion ablation.

IV. DISCUSSION, LIMITATIONS, AND CONCLUSION

The result is not that perpendicular viewing is generally ineffective. On the uniform envelope-control set, the no-occlusion selector already finds a usable view for every patch. The larger prediction-derived difference arises because localized patches can lie near protrusions or setbacks, making a nominally close, near-normal sight line unusable. In this evaluation, the continuous score is more informative than visibility alone because it reveals the trade between nominal incidence and resolution and an observable patch.

Sensitivity checks preserve this interpretation. Varying visibility tolerance from 1–10 cm gives $S = 0.788$–0.796 for the proposed method and 0.547–0.561 for the ablation, with usable rates unchanged. Required GSDs of 3 and 10 mm/px give proposed scores of 0.706 and 0.857. Five of 17 primary views lie on the $\pm 30°$ boundary; widening the cone to $\pm 40°$ raises the proposed score to 0.864 and usable rate to 1.000, indicating that the primary angular domain is binding for some targets; $\pm 30°$ is retained as the prespecified primary search domain. Excluding the three fused-patch RANSAC-normal cases also preserves the comparison ($S = 0.772$, $V = 0.875$ versus $S = 0.542$, $V = 0.625$).

Evidence is limited to one building and six spatially distinct facade regions; 19 patches are spatial samples, not 19 independent defects, and region 4 supplies 10. Targets come from prediction masks rather than independently verified defect ground truth and require support from at least three survey views, so the evaluation is conditioned on regions already observed in the initial survey and does not include fully unseen regions. Consistent with that conditioning, as-flown visibility is 0.895 on prediction-derived patches versus 0.748 on the broader envelope-control set; inference across only six regions remains descriptive. Results also depend on reconstruction, registration, estimated normals, and a finite, bounded candidate grid. This is a static digital-twin planning evaluation and planner replay, not physical closed-loop UAV execution; the waypoint and Gazebo/ROS interfaces shown in Fig. 1 are contextual downstream elements and were not evaluated. Dynamic obstacles, localization uncertainty, collision-checked motion, and online replanning remain outside scope. The evaluation addresses geometric view quality, not independent radiometric accuracy; embedded annotations in some frames prevented clean radiometric validation.

In summary, fusing already-registered predictions across views into localized 3D target patches and evaluating target-specific candidates against surrounding geometry improved mean score from 0.551 to 0.792 and usable views from 12/19 to 17/19. The smaller envelope-control difference supports the narrower conclusion that, in this evaluation, explicit visibility reasoning showed its largest benefit for geometrically difficult prediction-derived targets.

ACKNOWLEDGEMENT—This work was supported by the U.S. NSF Grant (2531557) "Digital Twin and AI-Infused Drones for Energy Retrofitting in Residential Envelopes."

REFERENCES

[1] A. Ficapal and I. Mutis, "Framework for the detection, diagnosis, and evaluation of thermal bridges using infrared thermography and unmanned aerial vehicles," *Buildings*, vol. 9, no. 8, pp. 179, 2019.

[2] T. Rakha, Y. El Masri, K. Chen, E. Panagoulia, and P. De Wilde, "Building envelope anomaly characterization and simulation using drone time-lapse thermography," *Energy and Buildings*, vol. 259, pp. 111754, 2022.

[3] B. G. Pantoja-Rosero, R. Achanta, and K. Beyer, "Damage-augmented digital twins towards the automated inspection of buildings," *Automation in Construction*, vol. 150, pp. 104842, 2023.

[4] A. Gholizadeh Lonbar and K. Chen, "NeRF-enhanced digital twin for building anomaly inspection using unmanned aerial systems (UASs)," *CIB Conferences*, vol. 1, Art. No. 242, 2025.

[5] C. Song, C. Lu, Y. Shi, A. He, J. Lin, and Z. Ma, "Simultaneous digital twin: Chaining climbing-robot, defect segmentation, and model updating for building facade inspection," *Buildings*, vol. 16, no. 3, pp. 646, 2026.

[6] J. Liu, Y. Cao, Y. Chen, C. Li, Y. Du, and H. Zhang, "Towards active real-to-twin inspection: A new paradigm for zero-shot anomaly detection," arXiv preprint arXiv:2605.25407, 2026.

[7] J. Maver and R. Bajcsy, "Occlusions as a guide for planning the next view," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 5, pp. 417–433, 1993.

[8] K. A. Tarabanis, R. Y. Tsai, and A. Kaul, "Computing occlusion-free viewpoints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 3, pp. 279–292, 1996.

[9] A. Bircher, M. Kamel, K. Alexis, H. Oleynikova, and R. Siegwart, "Receding horizon 'next-best-view' planner for 3D exploration," in *2016 IEEE international conference on robotics and automation (ICRA)*, pp. 1462–1468, 2016.

[10] G. Hardouin, J. Moras, F. Morbidi, J. Marzat, and E. M. Mouaddib, "Next-best-view planning for surface reconstruction of large-scale 3D environments with multiple UAVs," in *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 1567–1574, 2020.

[11] L. Jin, X. Chen, J. Rückin, and M. Popović, "NeU-NBV: Next best view planning using uncertainty estimation in image-based neural rendering," in *2023 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 11305–11312, 2023.

[12] S. Xue, J. Dill, P. Mathur, F. Dellaert, P. Tsiotras, and D. Xu, "Neural visibility field for uncertainty-driven active mapping," in *2024 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 18122–18132, 2024.

[13] W. Tang and M. R. Jahanshahi, "Active perception based on deep reinforcement learning for autonomous robotic damage inspection," *Machine Vision and Applications*, vol. 35, pp. 110, 2024.

[14] L. Li and H. Kasaei, "Enhanced view planning for robotic harvesting: Tackling occlusions with imitation learning," in *2025 IEEE international conference on robotics and automation (ICRA)*, pp. 13146–13152, 2025.

[15] S. Gao, L. Bramblett, and N. Bezzo, "Take your best shot: Sampling-based planning for autonomous photography," in *2025 IEEE international conference on robotics and automation (ICRA)*, pp. 14175–14181, 2025.

[16] N. A. Abdulateef, Z. N. Jasim, H. A. Hasan, B. Alsadik, and Y. H. Khalaf, "Autonomous BIM-aware UAV path planning for construction inspection," *Geomatics*, vol. 5, no. 4, pp. 79, 2025.