The Study of Object 6D Pose Estimation for High-Density Aerosol Environment

Woojin Son [a], Geonhwa Son [b], Taejoo Kim [c], Yukyung Choi [d,e]*
[a]Dept. of Artificial Intelligence, Sejong Univ, Seoul 05006, Republic of Korea
[b]Dept. of Artificial Intelligence and Robotics, Sejong Univ, Seoul 05006, Republic of Korea
[c]Dept. of Convergence Engineering for Intelligent Drone, Sejong Univ, Seoul 05006, Republic of Korea
[d]Sejong University, Seoul 05006, Republic of Korea
[e]Dept. of Artificial Intelligence and Robotics Institute (AIRI), Seoul 05006, Republic of Korea
[*]Corresponding author: ykchoi@rcv.sejong.ac.kr

***Keywords :** object 6d pose estimation, aerosol environment, nuclear accidents*

1. Introduction

In extreme environments inaccessible to humans, such as severe accidents at nuclear power plants, a rapid initial response by teleoperated robots is essential to prevent the further escalation of the incident. Robots deployed to the environment must successfully perform precise tasks, such as operating valves or inspecting leaks, which necessitates the ability to accurately perceive and localize objects for manipulation [1]. Specifically, 6D pose estimation, which involves determining an object's full 3D position and orientation, is a core technology for enabling robot systems to precisely manipulate equipment. Among various extreme conditions, the aerosol environment is relevant because accidents often release smoke, steam, or dust into the air, reducing visibility. In this work, we explicitly assume such environments in our experiments to evaluate pose estimation robustness under degraded sensing conditions. As shown in Fig. 1, aerosol environments cause light scattering and absorption, which not only degrade the semantic information of RGB images but also make depth measurements noisy, ultimately leading to significant challenges in pose estimation.

To address these limitations, we propose a robust 6D pose estimation pipeline based on step-by-step information restoration. First, the proposed pipeline dehazes the degraded RGB image to clarity via a dehazing module [2] specialized for aerosol removal. Next, using the dehazed RGB image and the original sparse and noisy depth measurements, it generates a dense depth map via a foundation model-based zero-shot depth completion module [3]. Finally, we apply a 6D pose estimation algorithm [4] on the dehazed RGB and completed depth map, achieving robust pose estimation even in aerosol environments.

Furthermore, we analyze how 6D pose estimation performance varies with aerosol density, defined by SSIM between aerosol and normal images. This demonstrates that our step-by-step design remains effective across different aerosol density levels, thereby experimentally verifying the robustness of our pipeline in practical aerosol environments.

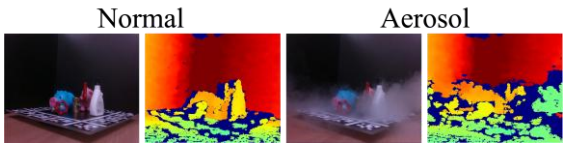

Fig. 1. Effect of aerosol on RGB-D sensor data

2. Methodology

High-density aerosol generated during severe accidents at nuclear power plants severely limits visibility by causing light scattering and absorption. Under such conditions where direct human access is impossible, the visual perception of teleoperated robots becomes essential for stable task performance. This study addresses the challenge of aerosol-induced visual degradation. To this end, we propose a robust RGB-D based 6D pose estimation pipeline.

As illustrated in Fig. 2, the proposed pipeline consists of three main modules. First, a dehazing module specialized for aerosol removal is applied to dehaze the RGB image. Second, a pre-trained zero-shot depth completion module completes the sparse and noisy depth measurements, which are corrupted by the aerosol, to generate a dense depth map. Finally, the dehazed RGB image and the completed depth map are fused and input into a 6D pose estimation module to estimate the final translation (t) and rotation (R) of the target object. The following sections will detail the specific implementation and role of each module within the pipeline.

2.1. Dehazing module

The first stage of the proposed pipeline, the dehazing module, produces dehazed RGB images from aerosol degraded inputs. The aerosol-induced visual degradation is physically defined by the following atmospheric scattering model:

$$I(x) = J(x)t_a(x) + A(x)(1 - t_a(x)) \quad (1)$$

Where I(x) denotes the observed hazy image, J(x) is the scene radiance (clean image), $t_a(x)$ is the medium

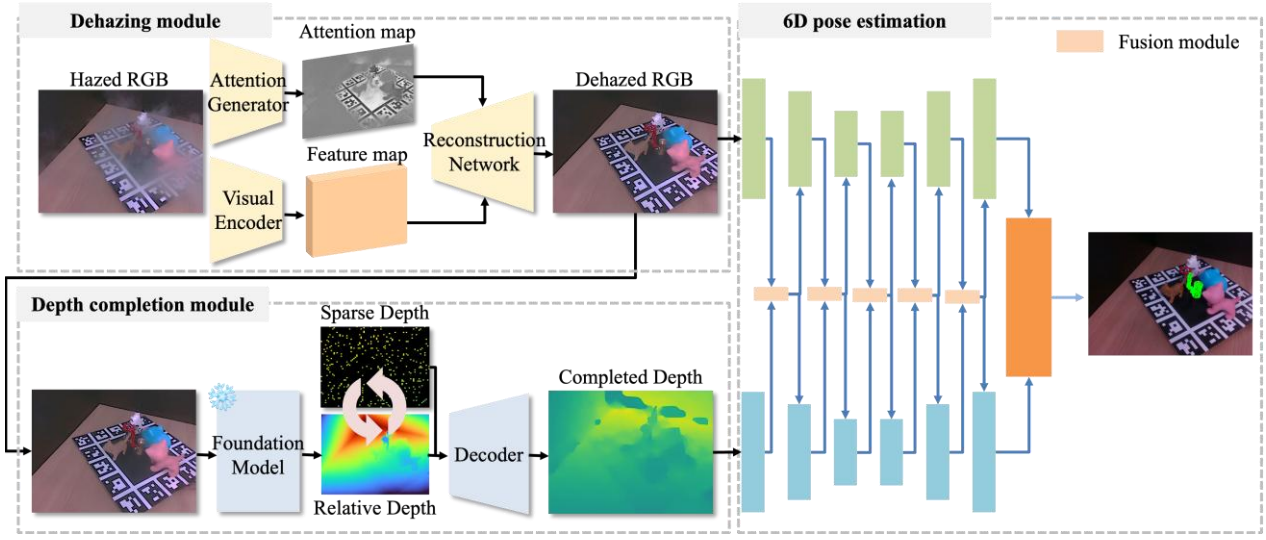

Fig. 2. Overview of the proposed step-by-step information restoration pipeline for 6D pose estimation under aerosol environments. The pipeline consists of two main stages: (1) the Dehazing module, which reconstructs dehazed RGB images from hazed inputs using an attention-guided network; (2) the Depth completion module, which generates completed depth maps from sparse or noisy depth using the guidance of dehazed RGB. Finally, both outputs are fused in the 6D pose estimation network to produce accurate object poses.

transmittance, and A(x) represents the global atmospheric light. This model explains that the observed image is formed as a combination of the attenuated scene radiance and the scattered atmospheric light. However, since the values of $t_a(x)$ and A(x) vary across the image and their distribution is non-uniform, the model parameters also become spatially variant. To effectively handle this non-uniformity, we employ a dehazing module, a state-of-the-art deep learning approach.

This module first identifies regions with high-density aerosol within the image to generate an attention map. Subsequently, the module then uses the attention map to selectively dehaze the visual information in those areas, thereby reconstructing a high-quality, clear image.

2.2. Depth completion module

Although the RGB image has been dehazed by the dehazing module, the depth measurements remain unreliable due to aerosol-induced light scattering, leading to numerous noisy values. Therefore, the second stage of the pipeline, the depth completion module, generates a dense and accurate depth map for all image pixels.

To achieve this, the module adopts a foundation model-based zero-shot depth completion module. The core idea is to exploit two data sources with different characteristics. The first is the structural prior inferred from the dehazed RGB image from the previous stage, which accurately predicts the scene's 3D geometry and relative depth ordering but lacks the absolute metric scale. This is a critical limitation, as 6D pose estimation for robotic manipulation ultimately requires absolute depth values at every pixel.

This limitation is addressed by using a small number of metric depth measurements obtained directly from the sensor. Using the structural prior as guidance and anchoring the absolute scale with metric measurements, the module completes and refines the depth map into a dense and accurate representation.

2.3. 6D pose estimation module

The final stage of the pipeline, the 6D Pose Estimation module, estimates the 6-DoF pose of an object by integrating the dehazed RGB image and the dense depth map from the preceding dehazing and depth completion stages. This module relies on a fusion framework that integrates both types of data effectively.

The core of this module lies in the complementary fusion of the two types of information. Specifically, it generates a feature representation for accurate pose estimation by fusing the detailed color and texture information from the dehazed RGB image with the 3D geometric structure from the completed depth map. Finally, based on this rich fused feature, the module predicts the object's 3D key points and converts them into the precise translation (t) and rotation (R) using a Least-Squares Fitting algorithm.

3. Experiments

To validate the performance of the proposed pipeline under aerosol conditions defined in this study, we used the 'Aerosol 6D Pose Estimation Benchmark Dataset' [5]. This dataset is suitable for evaluating the robustness of our methodology, as it includes aerosol conditions. not considered in previous 6-DoF pose estimation research. The dataset consists of four objects, including household

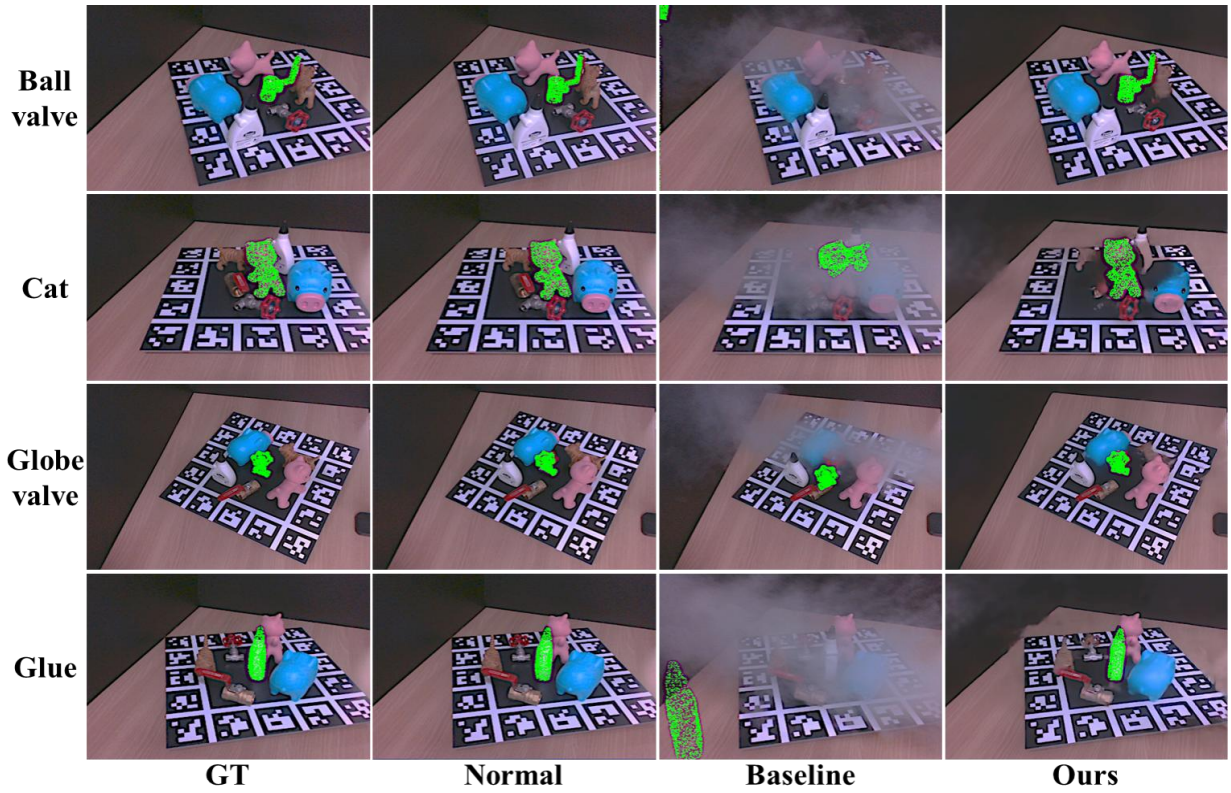

Fig. 3 Qualitative comparison of 6D pose estimation results across four object classes (Ball valve, Cat, Globe valve, and Glue). Columns represent the ground truth (GT), the Normal condition, the Baseline under aerosol degradation, and the proposed pipeline (Ours).

objects and industrial components. The inclusion of two types of valves is particularly relevant to our research objective, as they are key targets that a robot would need to manipulate precisely in a real nuclear power plant accident scenario.

Furthermore, a key feature of the dataset is that normal and aerosol state images are provided as a pair for every scene. In this study, we leverage this feature to quantitatively define the density of aerosol and to conduct an in-depth analysis of its impact on performance.

3.2. Evaluation Metrics

To comprehensively validate the performance of the proposed pipeline, this study employs metrics that evaluate the result and the performance of each component module. To evaluate the final 6D pose estimation accuracy, we used the standard evaluation metrics ADD and ADD-S. ADD is the metric for asymmetric objects, which calculates the mean distance error between the set of 3D model vertices, denoted as $\mathcal{M}$, transformed by the predicted pose ($[R|t]$) and the ground truth pose ($[R^*|t^*]$) as follows:

$$ADD = \frac{1}{|\mathcal{M}|}\sum_{x\in\mathcal{M}}\|(Rx+t)-(R^*x+t^*)\|_2 \quad (2)$$

ADD-S, the metric for symmetric objects, calculates the mean distance to the closest point on the model surface:

$$ADD-S = \frac{1}{|\mathcal{M}|}\sum_{x_1\in\mathcal{M}}\min_{x_2\in\mathcal{M}}\|(Rx_1+t)-(R^*x_2+t^*)\|_2 \quad (3)$$

Here, x_1 denotes a sampled vertex from the predicted pose, and x_2 represents a candidate vertex from the ground-truth model surface used to find the nearest match. This nearest-point comparison is necessary for symmetric objects, where multiple vertices may correspond to the same physical location.

In this study, a pose is considered correct if the value calculated by these two metrics is within 10% of the object's diameter, and the final accuracy is calculated based on this criterion. Furthermore, to analyze the performance of each component of the pipeline, the image restoration quality of the dehazing module was evaluated using PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure). PSNR indicates the degree of quality loss between the original and dehazed images, while SSIM measures the structural similarity between the two images. The depth information error of the depth completion module was quantitatively evaluated using RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) against the ground truth.

3.3. 6D Results and Analysis

Object	Baseline	Ours
	ADD(-s) < 0.1 (%)	
Ball valve*	10	**46.25**
Cat	51	**73.81**
Globe valve*	8.75	**42.5**
Glue*	30	**57.5**
Average	24.94	**55.02**

Table I: Quantitative comparison of 6D pose estimation accuracy (%). The asterisk (*) denotes symmetric objects evaluated with the ADD-S metric.

To quantitatively evaluate the proposed pipeline, we adopted the ADD and ADD-S metrics as described in Section 3.2. The performance comparison between the baseline and our proposed method is summarized in Table I. As shown in Table I, the proposed pipeline achieved substantial improvements across all objects, significantly increasing the 6D pose estimation accuracy compared to the baseline. In particular, the average accuracy improved from 24.94% to 55.02%, corresponding to a gain of +30.08%, more than a 2.2 times improvement.

In particular, the valve objects, which are the core target of this study, are relatively smaller than other objects, and thus the pixel information available for identification is inherently limited. As this already limited visual information was additionally degraded by the aerosol, it was difficult for the baseline model to extract even the minimum features necessary for object recognition. This is analyzed to have led to the low performance of only 10.00% (Ball valve) and 8.75% (Globe valve), respectively. However, when the proposed pipeline was applied, the performance for these valve objects more than four times, reaching 46.25% and 42.50%, respectively.

In addition to the quantitative results, Fig. 3 provides a qualitative comparison across 4 objects. The baseline model often fails to estimate correct poses, producing misaligned or incomplete predictions. In contrast, our pipeline, by applying RGB dehazing and depth completion, restores the degraded RGB information and the noisy depth cues. As a result, the predictions are markedly more accurate than those obtained under aerosol conditions. This qualitative evidence complements the quantitative results and confirms the effectiveness of our pipeline in enhancing 6D pose estimation performance.

3.2. 6D Analysis by Aerosol Density

In this section, we conduct an in-depth analysis of the proposed pipeline's robustness under varying aerosol densities by classifying the test dataset into three difficulty levels. The density was quantified using SSIM. This metric reflects perceptual degradation caused by aerosol scattering and retains the structural information of objects that is crucial for 6D pose estimation. Since

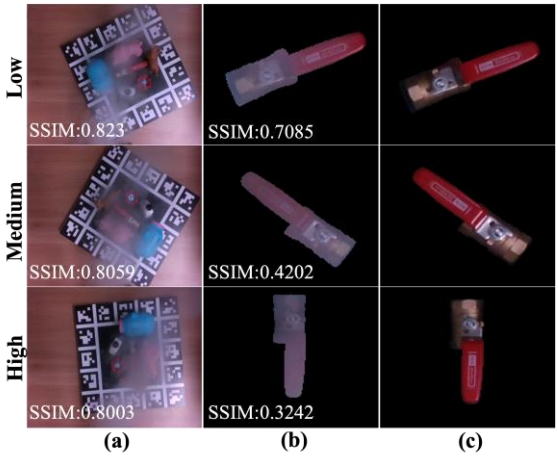

Fig. 4. Comparative analysis of SSIM between the full scene and the object instance. Each row represents a Low, Medium, and High severity case. Columns show (a) the full scene with its SSIM, (b) the masked object from the aerosol image with its SSIM, and (c) the ground truth masked object.

Level	ADD(-s) < 0.1 (%)	
	Baseline	Ours
Low	40.00	**60.74**
Medium	16.67	**56.66**
High	13.13	**45.45**
Average	25.3	**54.93**

Table II: Performance comparison according to aerosol density levels (Low, Medium, High). Accuracy (%) is measured by the ADD(-s) < 0.1 metric, and the density levels are based on the SSIM values defined in Sec. 3.2.

6D pose estimation is ultimately evaluated at the object instance level, SSIM was calculated only within the object instance mask area rather than across the entire scene, providing a more direct and consistent measure of visibility degradation.

Fig. 4. visually demonstrates the validity of this approach. As can be seen in the Fig. 4, while the SSIM, which represents the similarity of the entire scene, remains nearly constant at approximately 0.8 across all three scenes, the SSIM drops significantly from 0.7085 (Low) to 0.3242 (High). Based on the analysis above, we defined the aerosol density level: Low (SSIM $\geq$ 0.5), Medium (0.4 $\leq$ SSIM < 0.5), and High (SSIM < 0.4). Table II presents the results comparing the 6D pose estimation accuracy rates of the baseline and the proposed pipeline under each of these classified levels.

As shown in Table II, the baseline performance decreases considerably with increasing aerosol density, reaching only 13.13% accuracy at the High level. By contrast, our pipeline achieved 45.45% under the same condition—approximately 3.5 times higher than the baseline. These results indicate that our approach enhances the ability to sustain performance even under high-density aerosol conditions.

3.3. Module-by-Module Performance Analysis

Object	Dehazing				Depth Completion			
	Baseline		Ours		Baseline		Ours	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	RMSE↓	MAE↓	RMSE↓	MAE↓
Ball valve*	18.84	0.7359	**28.11**	**0.9013**	1.0801	0.5739	**0.6075**	**0.3491**
Cat	18.96	0.7426	**25.95**	**0.8881**	1.1851	0.6444	**1.0164**	**0.4838**
Globe valve*	17.95	0.7172	**27.69**	**0.8993**	1.194	0.6442	**0.6692**	**0.3608**
Glue*	17.82	0.6811	**26.95**	**0.8931**	1.2572	0.7483	**0.6944**	**0.4385**
Average	18.39	0.7192	**27.17**	**0.8954**	1.1791	0.6527	**0.7469**	**0.4081**

Table III: Quantitative evaluation results of each module in the proposed pipeline.

The performance of each module was evaluated both quantitatively and qualitatively. Table III presents the quantitative results, including the step-by-step analysis of each module, whereas Fig. 5 and Fig. 6 provide qualitative evidence for the effectiveness of dehazing and depth completion, respectively.

First, we evaluated the dehazing module using PSNR and SSIM. As shown in Table III, the average PSNR improved from 18.39 dB to 27.17 dB, and the SSIM increased from 0.7192 to 0.8954, indicating substantial restoration of image quality. The most notable PSNR improvement was observed for the valve objects, indicating that they were dehazed well. In addition, the qualitative comparison in Fig. 5 demonstrates that the dehazing stage effectively dehazes images across the previously defined density levels, yielding perceptually sharper and structurally consistent results.

Second, we evaluated the depth completion module using RMSE and MAE. As shown in Table III, the RMSE decreased from 1.1791 mm to 0.7469 mm, and the MAE decreased from 0.6527 mm to 0.4081 mm. This improvement can be attributed to the foundation model, which more accurately inferred the scene's structural prior by leveraging the dehazed RGB images from the previous stage as guidance. In addition, the qualitative results in Fig. 6 show that the depth completion stage works consistently well across different aerosol density levels.

3.4 Ablation Study

Results of the ablation study are summarized in Table IV, where we evaluate the contribution of each module in the proposed pipeline. When only the dehazing module was added (+Dehazing), performance increased substantially across all object classes, confirming that acquiring a clear RGB image is a critical prerequisite for 6D pose estimation. In contrast, adding only the depth completion (+DC) module yielded only marginal improvements, indicating that depth completion alone has limited effect without the guidance of dehazed RGB images. By integrating both modules, the full pipeline (Ours (Final)) achieved the highest accuracy across most classes, including the key objects. These results confirm that the proposed step-by-step design enables robust 6D pose estimation under aerosol conditions.

Method	ADD(-s) < 0.1 (%)			
	Ball valve*	Cat	Globe valve*	Glue*
Baseline	10	51	8.75	30
+ DC	12.50	54.76	15	33.75
+ Dehazing	43.75	72.62	40	**63.74**
Ours (Final)	**46.25**	**73.81**	**42.5**	57.50

Table IV: Presents the ablation study results. Starting from the baseline, we incrementally add the depth completion (+ DC) and dehazing modules, and finally combine the, into our full pipeline

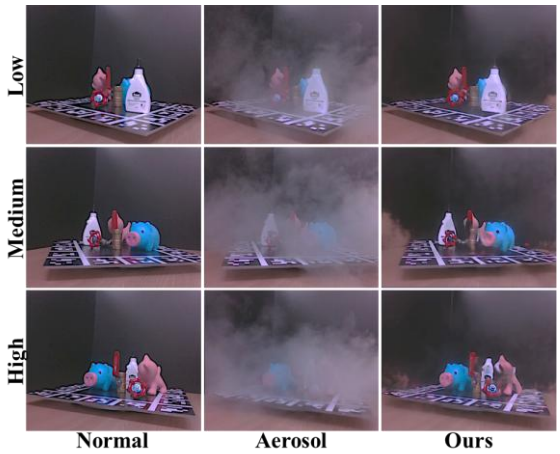

Fig. 5. Qualitative results of the Dehazing module. Each row corresponds to one of the defined aerosol density levels (Low, Medium, High), and the columns represent Normal, Aerosol, and Dehazed conditions.

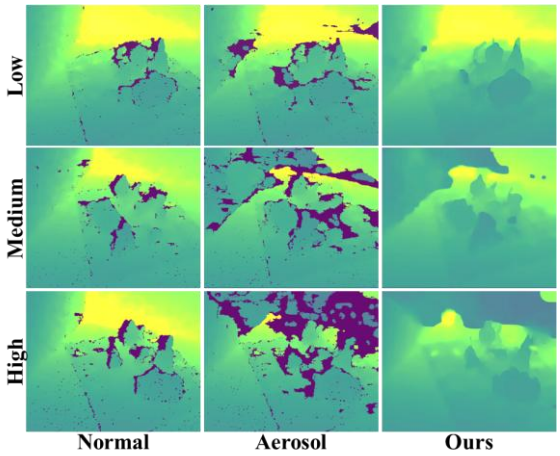

Fig. 6. Qualitative results of the Depth Completion module. Each row corresponds to one of the defined aerosol density levels (Low, Medium, High), and the columns represent Normal, Aerosol, and the output of our method.

4. Conclusions

This study addressed the challenge of 6D pose estimation failures in aerosol environments, where visual information is severely degraded by high-density scattering. To overcome this issue, we proposed a step-by-step information restoration pipeline that sequentially dehazes RGB images and completes depth data before

estimating the final pose. Experimental results on the Aerosol Benchmark Dataset demonstrated that the proposed pipeline achieved up to a four times improvement in pose estimation accuracy compared to the baseline. Notably, the largest gains were observed for valve objects, critical in aerosol scenarios. Even under the most challenging "High" aerosol level, it maintained an accuracy 3.5 times higher than the baseline.

In conclusion, this work demonstrated that the step-by-step information restoration approach enables 6D pose estimation in extreme environments.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(Ministry of Science and ICT)(No. RS-2022-00144385), and by Institute of Information & communications Technology Planning & Evaluation (IITP) under the metaverse support program to nurture the best talents (IITP-2025-RS-2023-00254529) grant funded by the Korea government(MSIT), and by the IITP(Institute of Information & Communications Technology Planning & Evaluation)-ICAN(ICT Challenge and Advanced Network of HRD) grant funded by the Korea government(Ministry of Science and ICT) (IITP-2025-RS-2022-00156345).

REFERENCES

[1] K. Nagatani, S. Kiribayashi, Y. Okada, K. Otake, K. Yoshida, S. Tadokoro, T. Nishimura, T. Yoshida, E. Koyanagi, M. Fukushima, and S. Kawatsuma, Emergency response to the nuclear accident at the Fukushima Daiichi Nuclear Power Plants using mobile rescue robots, *Journal of Field Robotics*, Vol. 30, No. 1, pp. 44-63, 2013.

[2] Y. Guo, H. Chen, Q. Fan, C. Xu, and J. Gu, SCANet: Self-paced semi-curricular attention network for non-homogeneous image dehazing, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1885-1894, 2023.

[3] H. Lee, K. S. Kim, B.-K. Kim, and T.-H. Oh, Zero-shot depth completion via test-time alignment with affine-invariant depth prior, in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39, No. 4, pp. 3877-3885, 2025.

[4] Y. He, H. Huang, H. Fan, Q. Chen, and J. Sun, FFB6D: A full flow bidirectional fusion network for 6D pose estimation, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3003-3013, 2021.

[5] H. Yang, S. Lee, T. Kim, and Y. Choi, Benchmark dataset and baseline for 6-DoF object pose estimation under aerosol conditions, *Journal of Institute of Control, Robotics and Systems* (in Korean), Vol. 30, No. 6, pp. 614-620, 2024.