Scene Coordinate Regression–Based Visual Localization for Advanced Multi-Crane Safety Monitoring in Construction Sites

Tran Dang Khoa Vo[1], **Tan-Dat Pham**[1], **Sujin Jin**[1], **Pa Pa Win Aung** [1], **Solmoi Park**[1] **and Seunghee Park**[1,2]

[1] Department of Global Smart City, Sungkyunkwan University, South Korea
[2]School of Civil, Architectural Engineering and Landscape Architecture, Sungkyunkwan University, South Korea

19980407@skku.edu, phamtandat@skku.edu, kane0113@skku.edu,
ppwinaung06@skku.edu, solmoipark@skku.edu, shparkpc@skku.edu

Abstract -

Crane operations pose significant safety risks on construction sites, particularly when multiple mobile cranes operate concurrently in confined and dynamic environments. This paper presents a vision-based framework for real-time multi-crane safety monitoring using Scene Coordinate Regression (SCR). By directly regressing 3D scene coordinates from monocular images, the proposed approach enables accurate six-degree-of-freedom crane tip localization without requiring repeated camera calibration during deployment. Visual data acquired from crane-mounted cameras and aerial drone surveys are jointly leveraged to construct scene-specific spatial representations for SCR model training. The estimated crane-end poses are integrated into a BIM-aligned digital twin to support real-time visualization, operational envelope analysis, and automated near-collision detection. Experimental results demonstrate consistent centimeter-level localization accuracy, with median errors of approximately 1–1.3 cm, confirming the suitability of the proposed framework for interaction-aware multi-crane safety monitoring in construction environments.

Keywords -

Visual Localization; Crane Safety Monitoring; Computer Vision; Digital Twins

1 Introduction

Construction remains one of the most hazardous industries worldwide. In 2023, the construction sector recorded 1,075 fatal occupational injuries, the highest among all industries, according to the U.S. Bureau of Labor Statistics (BLS) [1]. Construction-related accidents also impose significant economic burdens, with workplace injuries costing nearly $1 billion per week as estimated by the Occupational Safety and Health Administration (OSHA) [2]. These statistics highlight the urgent need for proactive and technology-driven safety monitoring on construction sites.

On modern construction sites, multiple mobile cranes often operate simultaneously in confined and dynamic environments, where collision risks arise from interactions among cranes, surrounding structures, and workers. Traditional safety approaches, such as manual supervision and rule-based controls, are increasingly insufficient under these conditions, motivating the development of real-time monitoring and proactive collision avoidance systems [3].

Sensor-based approaches have been widely adopted to enhance crane operation management by integrating planning data, operational information, and monitoring technologies [4, 5]. These include laser scanning for visibility analysis [6] and multi-sensor systems for localization and collision prevention [7], often extended through IoT frameworks for coordinated safety management [8]. However, such approaches rely on dedicated hardware, calibration, and sensor fusion, resulting in high deployment complexity and cost.

In parallel, vision-based approaches have emerged as a lightweight alternative. Crane-mounted cameras enable continuous visual perception for tasks such as worker detection and hazard awareness [9], while image-based methods have been extended to 3D spatial understanding using multi-view reconstruction or depth estimation [10]. Hybrid systems combining cameras with additional sensors have also been proposed for crane localization and workspace analysis [11, 12]. Nevertheless, several limitations remain. First, many vision-based localization methods rely on explicit 2D–3D feature matching, making them sensitive to occlusions, illumination changes, and visual noise. Second, both vision- and sensor-based systems often require explicit geometric modeling and repeated calibration, limiting adaptability to dynamic construction environments. Finally, existing approaches provide limited support for unified spatial representation and interaction-aware safety reasoning in multi-crane scenarios.[12].

To address these challenges, this study proposes a vision-based crane localization framework based on Scene Coordinate Regression (SCR). By directly regressing 3D scene coordinates from monocular images, the proposed approach avoids explicit feature matching and enables efficient 6-DoF camera pose estimation from a single image.

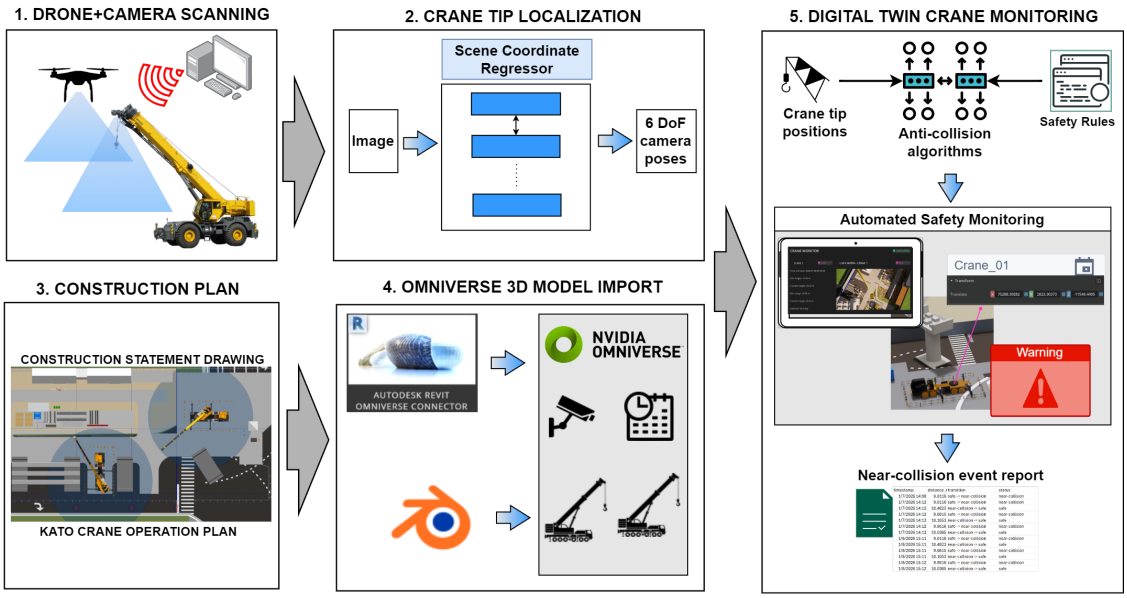

Figure 1. Overall framework of digital-twin crane collision monitoring.

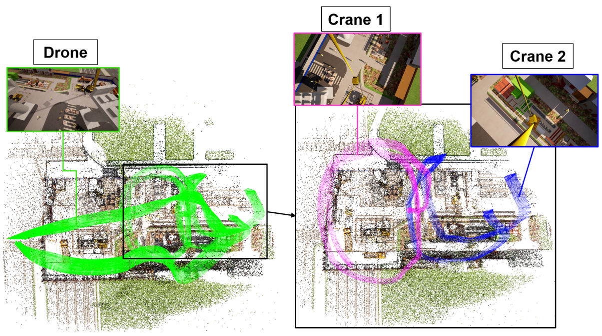

Figure 2. Drone and mobile crane operation path.

Visual data from crane-mounted cameras and drone surveys are jointly leveraged to construct scene-specific spatial representations without requiring additional sensing hardware. The estimated crane-end poses are integrated into a BIM-aligned digital twin to support real-time visualization and interaction-aware safety monitoring in multi-crane environments.

The main contributions of this study are summarized as follows:

- A real-time multi-crane safety monitoring framework that integrates vision-based localization with a BIM-aligned digital twin, enabling interaction-aware analysis of crane operations in shared workspaces.
- A unified spatial representation for multi-crane environments, where crane tip poses estimated from monocular images are directly mapped into a common BIM coordinate system for consistent safety reasoning.
- A lightweight deployment strategy using crane-mounted cameras and drone imagery, eliminating the need for additional sensing infrastructure while supporting continuous monitoring in dynamic construction sites.
- A digital twin–driven safety monitoring pipeline that enables real-time visualization, event logging, and near-collision detection for multiple cranes operating simultaneously.

2 Background

2.1 Advanced Crane Monitoring

The increasing complexity of construction sites has driven extensive research on advanced crane monitoring systems to improve operational safety and situational awareness [4, 5]. Early approaches relied on geometric and rule-based models to define crane working envelopes and collision constraints [3, 6], but their dependence on accurate modeling and frequent manual updates limits adaptability to dynamic site conditions. To improve localization accuracy, sensor-based systems incorporating GNSS, IMUs, UWB, and laser scanners have been widely adopted [7, 11], with multi-sensor fusion and IoT-based frameworks enabling real-time crane monitoring and fleet-level safety management [8]. However, these approaches rely on dedicated sensing infrastructure and precise calibration, resulting in high deployment cost and limited scalability. As a result, their applicability to rapidly changing construction environments and heterogeneous crane fleets remains limited. Vision-based crane monitoring has therefore emerged as a lightweight alternative, leveraging camera data and deep learning to support crane operation awareness [9]. While recent studies have extended vision-based methods toward three-dimensional spatial reasoning through image-based reconstruction or depth-assisted perception [10], many existing approaches remain constrained by explicit geometric assumptions or hybrid sensor dependencies [12, 11]. Moreover, most prior work focuses on single-crane scenarios, with limited consideration of interaction-induced safety risks in multi-crane environments [4].

2.2 Scene Coordinate Regression

Scene Coordinate Regression (SCR) is a deep-learning-based camera localization paradigm that directly regresses 3D scene coordinates from image pixels, enabling camera pose estimation without explicit 2D-3D feature matching [13]. SCR methods learn a mapping from image appearance to scene-specific spatial representations, from which six-degree-of-freedom (6-DOF) camera poses can be efficiently inferred [14]. By implicitly learning scene geometry, SCR demonstrates improved robustness to texture ambiguity, repetitive structures, and illumination variations compared with feature-based localization pipelines [15]. In addition, SCR-based localization can be operated using a single RGB image, enabling efficient fast inference for time-critical applications [16]. Although these characteristics make SCR well suited to dynamic and visually challenging environments such as construction sites, the application of SCR to construction equipment localization remains largely unexplored. In crane operation monitoring, SCR offers clear advantages. Crane-mounted and aerial cameras naturally observe consistent site geometry from diverse viewpoints, providing suitable visual input for learning scene coordinate representations. Unlike conventional vision-based crane localization methods, SCR does not require explicit camera-to-crane calibration or handcrafted geometric constraints, enabling flexible deployment across heterogeneous cranes and evolving site layouts. Furthermore, SCR-based localization can provide a unified spatial reference for interaction-aware safety analysis in multi-crane environments.

3 Proposed Approach

This section presents an end-to-end SCR-Digital Twins framework for real-time multi-crane safety monitoring, covering visual data acquisition, scene coordinate learning, crane-end localization, and digital twin–based collision analysis. Starting with multi-camera scanning, data from drone and crane-mounted cameras are then used to train the model for fully scene understanding. The trained model, which is adopted from [17] then is able to output the camera positions, which are also crane-end positions. Meanwhile, the 3D BIM model is built in Omniverse and the crane operation areas are defined based on construction plan. Finally, the anti-collision algorithms are used to analyze mobile crane-end positions with safety rules, then perform automated anti-collision monitoring and critical hazard event reports. The whole process is illustrated as in Figure. 1.

3.1 Data Processing

In order enable robust multi-crane monitoring with minimal site instrumentation, we construct a scene-specific sparse 3D representation by jointly leveraging crane-end mounted cameras that provide continuous close-range observations of the working area and a drone survey that ensures global coverage and mitigates blind spots as shown in Figure. 2. The drone flight can be conducted following site visit schedule to capture wide-baseline images of the entire operation zone and update on-site changes, while the crane-end cameras continuously record image streams during lifting operations. In the next step, a Structure-from-Motion (SfM) pipeline is then executed to recover a sparse point cloud and camera poses. Specifically, keypoints and descriptors are extracted for each image, feature correspondences are established across views, and a bundle-adjusted reconstruction yields a set of 3D points with associated multi-view feature tracks. This SfM model provides reliable 2D-3D associations that are later used as supervision for scene coordinate learning.

3.2 SCR-based Visual Localization

This subsection describes the D2S-based scene coordinate regression architecture adopted for real-time crane-end visual localization. The pipeline is divided into 2 parts: (1) Offline training where Scene Coordinate Regression model is trained with input images and their corresponding point cloud and (2) Online inference where the live-streaming data is feed directly into trained model to predict camera positions, as shown in Figure. 3.

3.2.1 Offline Training

Given the SfM reconstruction aligned with the BIM coordinate frame, each training image $\mathbf{p}_i \in \mathbf{P}$ provides a set of verified descriptor–scene coordinate pairs $\mathbf{V} = \{(d_i, \mathbf{X}_i)\}$ obtained from multi-view feature tracks. Following the D2S formulation [17], a scene coordinate regressor is trained to learn a direct mapping from local image descriptors to 3D scene coordinates. The architecture consists of a scene-agnostic local feature extractor and a scene-specific regression module equipped with self-attention to capture global descriptor context and suppress

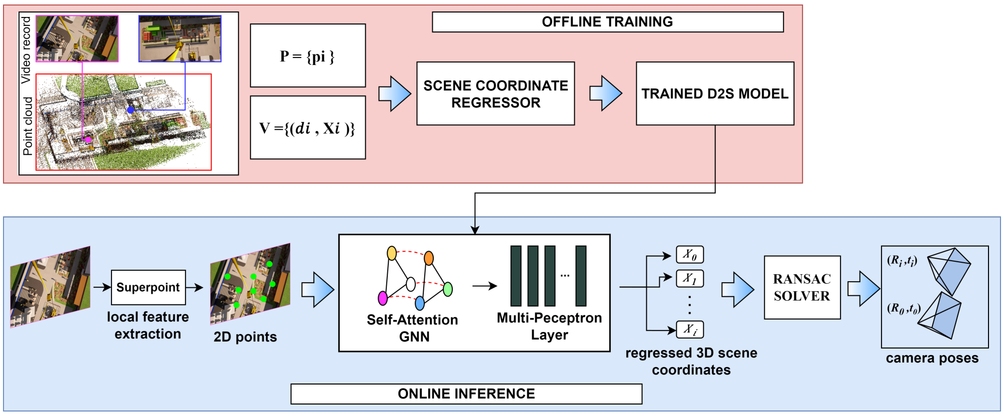

Figure 3. Overall SCR-based visual localization pipeline.

unstable regions such as moving objects or textureless areas. A shared MLP head outputs the regressed scene coordinates $\hat{\mathbf{X}}_i$ together with an optional reliability score $\hat{p}_i$. Once trained, the model is able to predict the new 3d feature coordinates from a single image without any SfM-guided descriptors.

3.2.2 Online Inference

During operation, each incoming crane-end camera frame is processed independently in a purely feed-forward manner. Sparse local features are first extracted from the image and passed through the trained D2S model, which directly regresses a set of 3D scene coordinates $\{\hat{\mathbf{X}}_i\}$ from the corresponding descriptors without explicit feature matching against a global 3D map. This regression-based formulation avoids costly nearest-neighbor search and scale-dependent matching, making the inference computationally efficient.

To ensure robustness, only the most reliable correspondences are retained based on the predicted confidence scores $\hat{p}_i$ (or by selecting the top-N correspondences). These sparse 2D–3D pairs are then fed into a RANSAC-based PnP solver to estimate the 6-DoF camera pose $(\mathbf{R}, \mathbf{t})$, followed by a lightweight nonlinear refinement. Since the camera is rigidly mounted at the crane tip, the estimated camera pose is directly transformed into the crane-end pose in the BIM coordinate frame using a pre-calibrated extrinsic transformation. This transformation is defined once and remains fixed during operation, eliminating the need for repeated calibration procedures in dynamic construction environments.

By avoiding dense reconstruction and global map matching, the proposed inference pipeline maintains low computational overhead and stable runtime behavior suitable for real-time deployment. The use of sparse descriptors, lightweight regression networks, and a single PnP optimization per frame enables real-time performance, allowing continuous crane-end pose tracking without reliance on dense reconstruction or iterative map matching. This real-time localization capability forms the basis for online multi-crane anti-collision monitoring and hazard analysis within the digital twin environment.

3.3 Digital Twin Visualization and Real-Time Monitoring

Based on the BIM-aligned scene coordinate system, a USD-based digital twin is constructed in NVIDIA Omniverse [18] to enable real-time visualization and monitoring of multi-crane operations. Each mobile crane is represented as a parametric articulated model with a fixed base and a dynamically updated arm. At each time step t, the estimated crane-end position $\mathbf{p}_i^t$ from the visual localization module is integrated by updating the corresponding USD transforms, rendering the crane arm as a line segment connecting the base $\mathbf{b}_i$ and the tip $\mathbf{p}_i^t$.

Following Algorithm 1, frame-wise crane operational metrics—including height, horizontal working radius, and azimuth angle—are computed directly in the BIM coordinate frame and compared against historical maxima to provide intuitive feedback on operating envelopes. All updates are performed in a frame-synchronous manner to ensure temporal consistency between geometry updates, metric computation, and visualization.

To assess collision risk between multiple cranes, the system evaluates the horizontal separation distance between crane tips at each frame,

$$d_t = \left\|(\mathbf{p}_1^t - \mathbf{p}_2^t)_{xz}\right\|,$$

where only the ground-plane components are considered to reflect practical collision risk during lifting operations. When the distance falls below a predefined safety threshold D_{th}, the system transitions from a *safe* to a *near-collision* state. These state transitions are logged with timestamps and frame indices, enabling post-event analysis and safety auditing. In this study, the safety threshold D_{th} is set to 10 m as a conservative safety margin to account for crane motion uncertainties, load swing, and localization errors in dynamic construction environments, following general safety principles outlined in ISO 12480-1 [19].

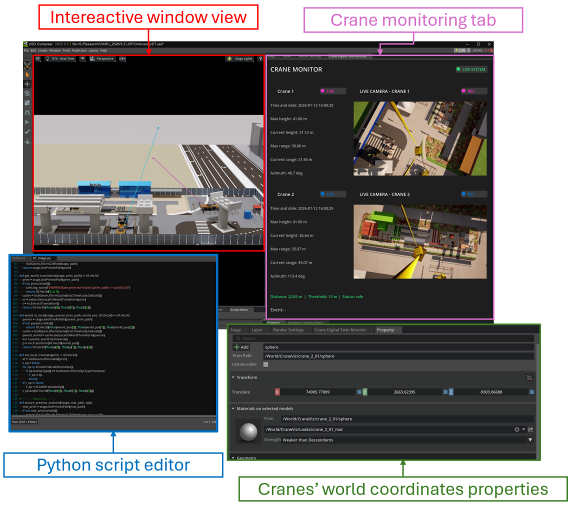

Figure 4. NVIDIA Omniverse window for DT-based crane safety monitoring

In addition to geometric visualization, synchronized live camera streams from crane-end cameras are displayed alongside the digital twin, providing operators with complementary visual context. All visualization elements—including crane geometry, distance indicators, operational statistics, and safety status—are updated in real time within a unified dashboard interface, as illustrated in Figure. 4. The use of sparse crane-end pose updates, lightweight geometric primitives, and event-driven state evaluation allows the system to operate at interactive frame rates without requiring dense reconstruction or physics-based simulation. This tightly coupled integration of real-time visual localization, digital twin visualization, and rule-based safety assessment enables continuous multi-crane monitoring and forms the foundation for automated near-collision detection in complex construction environments.

Algorithm 1 Real-Time Multi-Crane Digital Twin Monitoring and Near-Collision Detection

Require: Frame-wise crane tip positions $\mathbf{p}_i^t$, crane base transforms $\mathbf{T}_i$, synchronized camera images, safety threshold D_{th}

Ensure: Real-time crane visualization, operational metrics, near-collision events

Initialize USD-based digital twin environment and crane geometries
Initialize event logger and playback parameters
for each frame t **do**
  for each crane i **do**
    Retrieve crane base position $\mathbf{b}_i$
    Update crane arm geometry from $\mathbf{b}_i$ to $\mathbf{p}_i^t$
    Compute height, horizontal range, and azimuth angle
    Update historical maximum height and range
  end for
  Compute horizontal distance

$$d_t = \|(\mathbf{p}_1^t - \mathbf{p}_2^t)_{xz}\|$$

  if $d_t \leq D_{\text{th}}$ **then**
    Set system state to *near-collision*
  else
    Set system state to *safe*
  end if
  Log state transition events with timestamp if detected
  Update dashboard indicators and synchronized camera views
end for

Figure 5. Plan view of experimental construction site

4 Implementation and Results

This section illustrates the experimental setup to evaluate model performance and results of proposed method. First, a simulated construction site is built by Twin-motion platform replicating the real construction site at Central Florida Boulevard, USA, as shown in Figure. 5. In the next step, drone and mobile crane working activities are simulated to capture data for model training with image resolution of 1080x1920. The dataset consists of approximately 2000 images collected from drone and crane-mounted cameras. The data are split into training and testing sets at the sequence level to avoid frame overlap. Meanwhile, after real-time video stream is fed into train D2S model to predict mobile crane tips, the output are inputs in NVIDIA Omniverse to perform digital-twin visualization. The whole pipeline is implemented on a computer with NVIDIA RTX 4090 GPU, Intel i9-139000KF Chipset and 64GB RAM.

4.1 Visual localization results

In order to evaluate the accuracy of SCR-based D2S model in estimating spatial information of crane tips, we we establish a testbed from videos recording two KATO cranes operating and the camera poses extracted from COLMAP serving as ground-truth value to estimate the errors of model predictions. The SfM-derived camera poses are used only for evaluation and are not directly used to supervise camera pose learning.

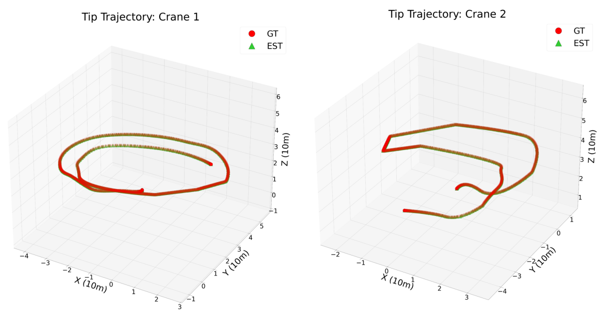

Figure 6. Qualitative results of SCR-based crane tip localization of Crane 1 (Left) and Crane 2 (Right). Axes are shown in units of 10 m for visualization clarity.

Table 1. Quantitative results of SCR-based crane tip localization.

Error threshold	Crane 1	Crane 2
$< 1\,cm$	48.89%	48.80%
$< 2\,cm$	81.89%	82.10%
$< 3\,cm$	93.11%	92.69%
$< 5\,cm$	96.89%	97.60%
Median error (cm)	**1.30**	**1.01**

The Figure. 6 compares the estimated crane tip trajectories with the reference trajectories for both cranes. The results show that the estimated positions closely follow the ground truth over the entire operation sequence, maintaining stable alignment during slewing motions, height changes, and direction transitions. In addition, the quantitative results are summarized in Table.1. For Crane 1, 81.89% of the estimated crane tip positions exhibit errors below 2 cm, increasing to 96.89% within a 5 cm threshold. Crane 2 achieves comparable performance, with 82.10% of estimates below 2 cm and 97.60% within 5 cm. The median localization errors are 1.30 cm and 1.01 cm for Crane 1 and Crane 2, respectively. These results demonstrate that the proposed SCR-based approach can provide reliable centimeter-level crane tip localization, which is sufficient to support downstream safety monitoring and near-collision detection tasks.

4.2 Digital Twin Crane Safety Monitoring

The Figure. 7 presents the digital twin–based real-time multi-crane safety monitoring results. By continuously integrating the estimated crane-end poses into a BIM-aligned digital twin, the system enables synchronized visualization of crane geometry, operational metrics, and live top-views for multiple cranes operating concurrently. During normal operation, a safe state is maintained when the inter-crane distance exceeds the predefined safety threshold, while real-time feedback on crane height, working radius, and azimuth angle is provided. After that, when the horizontal separation between crane tips falls below the safety threshold, the system automatically transitions to a near-collision state, which is immediately reflected in the digital twin visualization and monitoring dashboard. All state transitions are logged in a structured CSV format with timestamps and inter-crane distances, enabling post-operation analysis and safety auditing. The system achieves approximately 21.58 FPS on an RTX 4090 GPU, demonstrating near real-time performance for multi-crane monitoring .These results demonstrate that the proposed framework can effectively integrate vision-based localization with digital twin–based safety monitoring for real-time multi-crane operations.

Table 2. Runtime analysis of the proposed framework on i9-13900KF, RTX4090

Module	Time (ms)
Feature extraction + SCR inference	28.92
PnP + refinement	15.28
Digital Twin update	2.14
Total per frame	**46.34 ms**
Average FPS	**21.58 FPS**

5 Discussion

The proposed framework assumes a fixed rigid transformation between the camera and crane tip, as well as prior alignment between the SfM reconstruction and the BIM coordinate system. These assumptions may introduce deployment constraints in real construction sites. The results demonstrate that the SCR-based framework can support near real-time multi-crane safety monitoring using monocular visual inputs, achieving centimeter-level localization accuracy sufficient for inter-crane distance evaluation and safety threshold enforcement. Compared with feature-based localization methods, the proposed approach avoids explicit feature matching and shows improved robustness to illumination changes and occlusions. In contrast to sensor-based systems, it reduces hardware dependency and

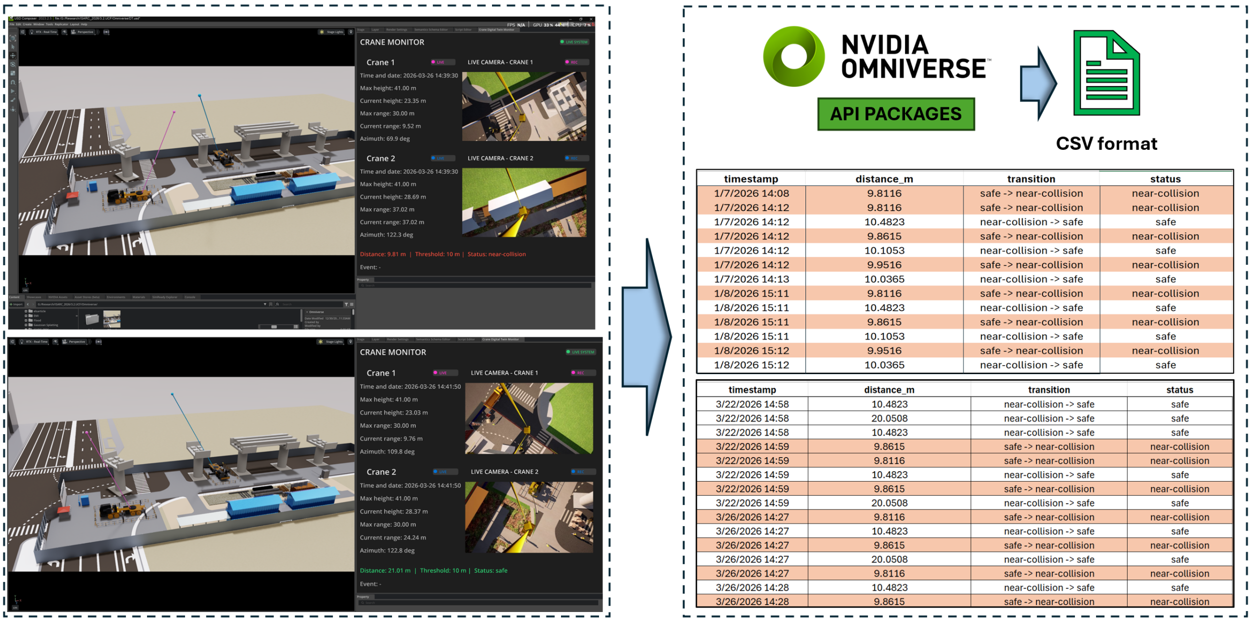

Figure 7. Digital twin–based real-time visualization of multi-crane safety monitoring.

calibration complexity. From a methodological perspective, scene coordinate regression enables a unified scene-centric representation without explicit geometric modeling or repeated calibration, making it suitable for dynamic construction environments. The integration with a BIM-aligned digital twin further supports consistent spatial reasoning, real-time visualization, and rule-based safety assessment for interaction-aware monitoring. Nevertheless, the current implementation focuses on crane-tip separation and does not explicitly model dynamic effects such as load swing or interactions with workers. In addition, the evaluation is conducted in a simulated environment, and real-world conditions such as illumination variation, occlusions, and weather may affect performance. These limitations will be addressed in future work.

6 Conclusion

This paper presented a vision-based framework for real-time multi-crane safety monitoring using Scene Coordinate Regression (SCR) integrated with a digital twin. By directly regressing 3D scene coordinates from monocular images, the proposed approach enables accurate crane tip localization without relying on explicit feature matching or additional sensing hardware, thereby reducing system complexity, calibration requirements, and deployment overhead. Experimental results demonstrate consistent centimeter-level localization accuracy across different crane motions and configurations, indicating robust performance under dynamic operational conditions and varying viewpoints. When integrated into a BIM-aligned digital twin, the estimated crane-end poses support real-time visualization, operational envelope analysis, and automated near-collision detection among multiple cranes operating within shared workspaces, enabling interaction-aware safety assessment in complex site layouts. Furthermore, the unified spatial representation provided by the digital twin facilitates consistent safety rule enforcement and traceable event logging without manual reconfiguration. Overall, the proposed framework offers a lightweight and scalable alternative to conventional sensor-heavy crane monitoring systems and provides a practical foundation for proactive, data-driven, and interaction-aware multi-crane safety management in complex and evolving construction environments.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used GPT-5.2 in order to improve the readability and language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) [RS-2024-00336270] and [RS-2025-02223612].

References

[1] U.S. Bureau of Labor Statistics. National census of fatal occupational injuries in 2022. BLS Census of Fatal Occupational Injuries, 2022. Accessed Feb. 19, 2025.

[2] U.S. Occupational Safety and Health Administration (OSHA). Business Case for Safety and Health. `https://www.osha.gov/businesscase/costs`, 2024. (accessed Jan. 2, 2024).

[3] Ziqing Yang, Jian Yang, and Enliu Yuan. Safety monitoring of construction equipment based on multi-sensor technology. In *ISARC*, pages 677–684, October 2020. doi:10.22260/ISARC2020/0095.

[4] Ali Khodabandelu and JeeWoong Park. *Integrating BIM and ABS for Multi-Crane Operation Planning through Enabling Safe Concurrent Operations*, pages 1128–1135. ASCE, 2021. doi:10.1061/9780784483893.138.

[5] Ala Nekouvaght Tak, Hosein Taghaddos, Ali Mousaei, Anahita Bolourani, and Ulrich Hermann. Bim-based 4d mobile crane simulation and onsite operation management. *Automation in Construction*, 2021. doi:https://doi.org/10.1016/j.autcon.2021.103766.

[6] Tao Cheng and Jochen Teizer. Crane operator visibility of ground operations. In *ISARC*, pages 699–705, June 2011. doi:10.22260/ISARC2011/0131.

[7] Jean-Pierre Sleiman, Emile Zankoul, Hiam Khoury, and Farook Hamzeh. *Sensor-Based Planning Tool for Tower Crane Anti-Collision Monitoring on Construction Sites*, pages 2624–2632. ASCE, 2016. doi:10.1061/9780784479827.261.

[8] Dexing Zhong, Hongqiang Lv, Jiuqiang Han, and Quanrui Wei. A practical application combining wireless sensor networks and internet of things: Safety management system for tower crane groups. *Sensors*, 2014. doi:10.3390/s140813794.

[9] Tanittha Sutjaritvorakul, Axel Vierling, and Karsten Berns. Data-driven worker detection from load-view crane camera. In *ISARC*, pages 864–871, October 2020. doi:10.22260/ISARC2020/0119.

[10] Yuntae Jeon, Dai Quoc Tran, Khoa Tran Dang Vo, Jaehyun Jeon, Minsoo Park, and Seunghee Park. Neural radiance fields for construction site scene representation and progress evaluation with bim. *Automation in Construction*, 2025. doi:https://doi.org/10.1016/j.autcon.2025.106013.

[11] Jingdao Chen, Yihai Fang, and Yong K. Cho. Real-time 3d crane workspace update using a hybrid visualization approach. *Journal of Computing in Civil Engineering*, 2017. doi:10.1061/(ASCE)CP.1943-5487.0000698.

[12] Yu Wang, Hiromasa Suzuki, and Yutaka Ohtake. Spatial maps with working area limit line from images of crane's top-view camera. *Automation in Construction*, page 104475, 2022. doi:https://doi.org/10.1016/j.autcon.2022.104475.

[13] Xiaotian Li, Jakob Verbeek, and Juho Kannala. Hierarchical joint scene coordinate classification and regression for visual localization. *CoRR*, abs/1909.06216, 2019. URL `http://arxiv.org/abs/1909.06216`.

[14] Eric Brachmann, Alexander Krull, Sebastian Nowozin, Jamie Shotton, Frank Michel, Stefan Gumhold, and Carsten Rother. DSAC - differentiable RANSAC for camera localization. *CoRR*, abs/1611.05705, 2016. URL `http://arxiv.org/abs/1611.05705`.

[15] Eric Brachmann and Carsten Rother. Learning less is more - 6d camera localization via 3d surface regression. *CoRR*, abs/1711.10228, 2017. URL `http://arxiv.org/abs/1711.10228`.

[16] Thuan B Bui, Dinh-Tuan Tran, and Joo-Ho Lee. Fast and lightweight scene regressor for camera relocalization. *arXiv preprint arXiv:2212.01830*, 2022.

[17] Bach-Thuan Bui, Huy-Hoang Bui, Dinh-Tuan Tran, and Joo-Ho Lee. D2s: Representing sparse descriptors and 3d coordinates for camera relocalization. *IEEE Robotics and Automation Letters*, 2024.

[18] *USD Composer*. NVIDIA Corporation, Santa Clara, CA, U.S., 2024. URL `https://docs.omniverse.nvidia.com/composer/latest/index.html`.

[19] Iso-12480-1:2024, Sep 2023. URL `https://www.iso.org/standard/57300.html`.