2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

3-D Model-Based Pan-Tilt Multi-Camera System for High-Speed Wide-Area Volumetric Shooting

Adam Mail[1], Feiyue Wang[1], Kohei Shimasaki[1], and Idaku Ishii[1]

Abstract— **Stereo matching between different cameras presents significant challenges, particularly for wide-area scanning systems using pan-tilt cameras with continuously changing orientations. This paper proposes a 3-D model-based multi pan-tilt camera system that establishes stereo correspondence between cameras by introducing a shared 3-D coordinate space derived from a reference model. Our approach reduces matching complexity by leveraging 3-D model-based calibration, enabling scalable volumetric capture for real-world measurement and monitoring applications. To validate the proposed system, we conducted wide-area experiments on outdoor Heating, Ventilation, and Air Conditioning (HVAC) units. Each pan-tilt camera captured 25 high-resolution images, stitched into panoramic mosaics. Using 3-D reference points from a laser scanner, we established projective mappings through feature-point correspondence registration and mapped pixel coordinates to pan-tilt angles via linear interpolation. The experimental results demonstrated that multiple camera views can be directly aligned through their shared 3-D model, achieving stereo correspondence without appearance-based matching. Volumetric measurement through stereo DIC on 4000 frames captured at 400 fps successfully show 3-D vibration with sub-millimeter velocity accuracy and identified dominant frequencies, validating the system's effectiveness for large-scale structural monitoring and industrial diagnostics.**

I. INTRODUCTION

High-speed wide-area volumetric shooting is increasingly important for real-world measurement and monitoring across domains such as infrastructure inspection [1], [2], agricultural monitoring [3], [4], and human motion analysis [5], [6]. When combined with Digital Image Correlation (DIC) [7], [8], high-speed imaging enables pixel-level motion estimation over broad frequency ranges, supporting precise detection of rapid events such as structural vibrations, surface deformations, and dynamic displacements [9]. Many practical scenarios, however, demand both high temporal resolution and scalable spatial coverage (e.g., large assets, industrial facilities, or civil infrastructure), motivating systems that can expand the measurable surface area while preserving per-pixel fidelity and measurement accuracy [10].

Traditional stereo camera setups face two fundamental limitations for wide-area applications: (i) the field of view is inherently constrained by static viewpoints, and expanding coverage by adding more fixed cameras results in resolution trade-offs while significantly increasing synchronization and calibration complexity [11]; and (ii) establishing reliable stereo correspondences becomes increasingly difficult as viewpoints, scales, and surface appearances vary—particularly challenging when cameras or their pan-tilt units move during scanning operations [12]. These challenges undermine dense, geometrically consistent multiview reconstruction at scale [13] and render appearance-based matching methods fragile in wide-area, high-speed deployment scenarios.

Conventional stereo correspondence methods include appearance-based and feature-based matching. Appearance-based methods, such as block matching [14] and semi-global matching [15], compare intensity patterns between camera images to identify corresponding pixels. While effective in controlled environments with artificial patterns [16], [17], they degrade significantly on texture-less surfaces common in industrial equipment and infrastructure [18]. Feature-based methods [19], [20] detect distinctive keypoints (e.g., SIFT, SURF, ORB) and match feature descriptors. Although more robust to illumination and viewpoint changes, they depend on salient visual features and require expensive descriptor matching. For n cameras, traditional pairwise calibration requires $O(n^2)$ procedures, with each camera pair calibrated through calibration patterns [21] or feature matching [22], severely limiting scalability.

Recent model-based approaches have explored using 3-D reference models to simplify stereo correspondence. Qin et al. [23] demonstrated that introducing a 3-D model as a shared coordinate system eliminates the need for appearance-based stereo matching in vibration measurement applications. Similarly, Wang et al. [24] showed that model-based calibration enables robust displacement measurement for infrastructure monitoring even with texture-less bridge surfaces. However, these existing model-based methods are designed for static measurement scenarios where camera positions remain fixed. When cameras move—as required in pan-tilt scanning systems for wide-area coverage—the established correspondences become invalid, and calibration must be repeated for each new camera orientation [25]. This limitation prevents existing model-based methods from supporting dynamic wide-area scanning applications.

To overcome these limitations, we introduce a 3-D model-based multi pan-tilt camera framework that fundamentally differs from conventional approaches. We validate our approach on a large HVAC installation, demonstrating that two synchronized pan-tilt cameras achieve accurate stereo correspondence and volumetric reconstruction across high-resolution imagery. The experimental results show reliable cross-view alignment and practical readiness for wide-area volumetric capture, enabling full-surface 3-D velocity mea-

[1]All authors are with the Smart Robotics Laboratory, Hiroshima University, Higashihiroshima, Hiroshima 739-0046, Japan. Email : {adam-mail, feiyue, simasaki, iishii}@hiroshima-u.ac.jp

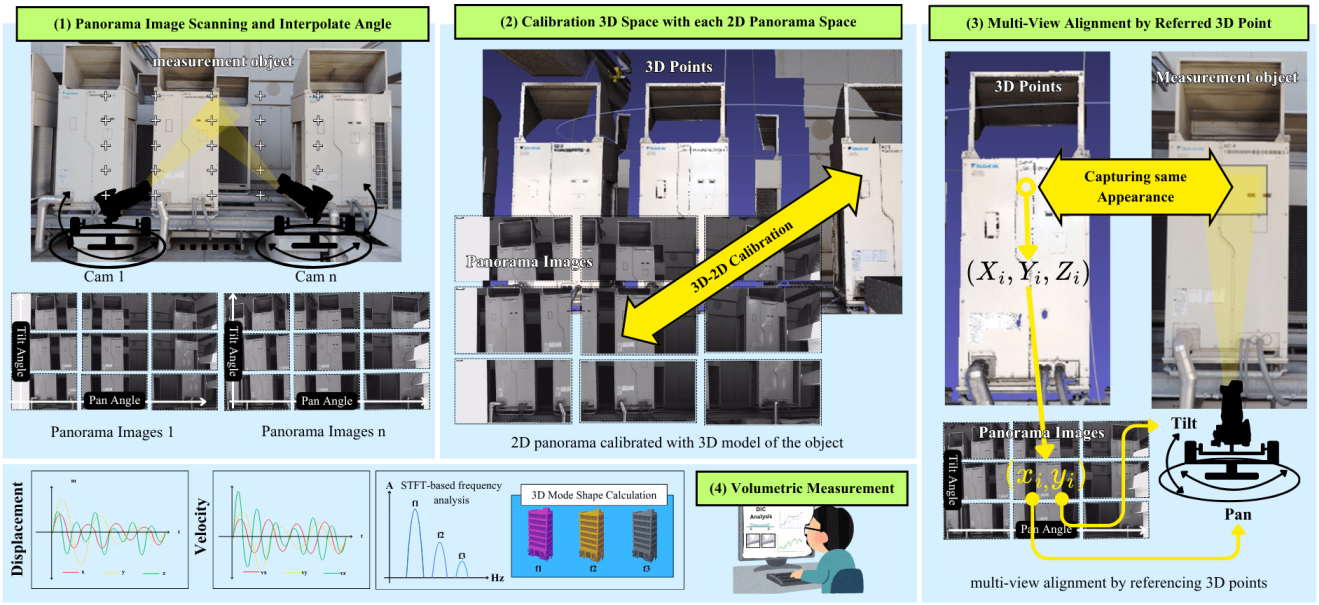


Fig. 1. Pipeline for model-based volumetric measurement system. (1) Pan-tilt cameras capture overlapping images and stitch them into panoramas with interpolated angle fields. (2) Each panorama is calibrated to the 3D model via feature correspondences. (3) Multi-view alignment is achieved by projecting any 3D point to both cameras . (4) Volumetric measurement stereo DIC provides displacement, velocity, frequency response, and 3D mode shapes for vibration analysis.

surement and vibration analysis at thousands of points simultaneously for structural health monitoring, industrial quality control, and equipment vibration diagnostics.

II. 3-D MODEL-BASED PAN-TILT MULTI-CAMERA

A. Concept

Volumetric capturing requires at least two cameras for stereo vision. While two cameras can be managed manually, larger multi-camera systems face several limitations: 1) **Calibration overhead** – Each camera requires individual point-to-point calibration from 2-D image to 3-D reference coordinates, becoming time-consuming as camera count increases. 2) **Orientation control** – Ensuring all cameras target the same measurement point requires precise angle adjustments, inefficient and error-prone when cameras reorient dynamically during scanning. 3) **Stereo correspondence** – Traditional appearance-based matching becomes unreliable when cameras move or measure texture-less industrial surfaces.

To address these limitations, we propose a 3-D model-based calibration framework that reduces complexity by calibrating each camera independently to a shared 3-D reference model. This approach uses 2-D image points from cameras and 3-D points from a laser scanner to establish persistent coordinate mapping valid across all camera orientations. The 3-D to 2-D calibration determines image-plane locations for each 3-D point, enabling pan-tilt angle computation through interpolated angle fields. Once established, both cameras automatically direct to any designated 3-D point for synchronized stereo capture, enabling volumetric reconstruction through DIC without runtime appearance-based matching.

The overall procedure illustrated in Fig. 1 consists of four steps: 1) **Surface scanning** – Controlled pan-tilt sweeps capture the object surface from multiple orientations, with overlapping images stitched into seamless mosaics per camera. 2) **3-D to 2-D correspondence** – Stitched mosaics are registered with laser-scanned 3-D points to estimate calibration matrices mapping each 3-D point to 2-D pixel locations. 3) **Pan-tilt angle mapping** – Pan and tilt angles are computed across mosaics and interpolated into dense angle fields, back-projecting to assign angles to every 3-D point. 4) **Synchronized stereo capture** – Using established pan-tilt mapping, stereo DIC independently computes 2-D displacement fields on each camera, then triangulates into 3-D displacement vectors using model-based correspondence.

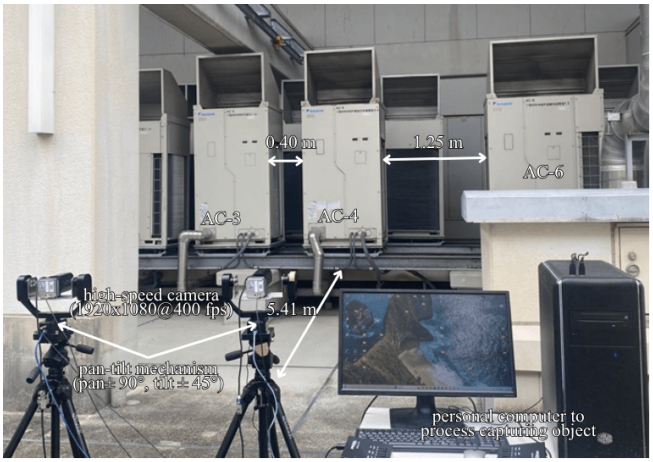
B. Implemented Algorithm

We estimate each pan-tilt camera's orientation over a wide area by combining stitched high-resolution mosaics with a laser-scanned 3-D reference model. Let the laser scan provide a point set $\mathcal{P} = \{(X_i, Y_i, Z_i) \mid i = 0, \ldots, N-1\}$. Each camera c executes a controlled pan-tilt scan, producing frames $\{I_{c,t}\}_{t=0}^{T_c-1}$ and a stitched mosaic I_c^{stitch}. The per-pixel pan and tilt angles are obtained by direct interpolation from the commanded scan limits. The mosaic is then registered to the 3-D model through direct 3-D to 2-D mapping, enabling per-point pan-tilt assignment and synchronized stereo capture for volumetric measurement using DIC.

1) Stitch frames and build an interpolated per-pixel angle field: For each camera c, input frames are warped into a common canvas by homographies $H_{c,t}$ and blended to form the stitched mosaic:

$$I_c^{\text{stitch}}(x', y') = \text{Blend}\Big(\big\{ I_{c,t}\big(H_{c,t}^{-1}[x', y', 1]^\top\big) \big\}_{t=0}^{T_c-1} \Big), \quad (1)$$

where Blend$(\cdot)$ can be average, feathering, or Laplacian compositing [26]. Let the valid mosaic domain be $[x_L, x_R] \times$



(a)

(b)

Fig. 2. (a) Experiment field setup for capturing outdoor HVAC units. (b) 3D model object

$[y_T, y_B]$ and define normalized coordinates $u = \frac{x'-x_L}{x_R-x_L} \in [0,1]$, $v = \frac{y'-y_T}{y_B-y_T} \in [0,1]$. Assuming a left-to-right pan sweep with angles $(\alpha_c^{\min}, \alpha_c^{\max})$ and a top-to-bottom tilt sweep with angles $(\beta_c^{\min}, \beta_c^{\max})$, the per-pixel angles are given by linear interpolation:

$$\begin{aligned} \widehat{\alpha}_c(x', y') &= (1-u)\,\alpha_c^{\min} + u\,\alpha_c^{\max}, \\ \widehat{\beta}_c(x', y') &= (1-v)\,\beta_c^{\min} + v\,\beta_c^{\max}. \end{aligned} \tag{2}$$

2) Register stitched mosaic to the 3-D model via direct 3-D to 2-D mapping: From I_c^{stitch} we select correspondence pairs $\{(x_i, y_i) \leftrightarrow (X_i, Y_i, Z_i)\}_{i=1}^{N}$. We estimate a projective matrix $P_c \in \mathbb{R}^{3\times4}$ that maps homogeneous 3-D points $\tilde{\mathbf{X}}_i = [X_i\ Y_i\ Z_i\ 1]^\top$ to homogeneous image points $\tilde{\mathbf{u}}_i = P_c\,\tilde{\mathbf{X}}_i$. Following the Direct Linear Transform (DLT) formulation, we assemble a linear system $A\,\mathbf{p} = \mathbf{0}$ with $\mathbf{p} = \text{vec}(P_c)$ and two rows per correspondence:

$$\begin{bmatrix} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -x_iX_i & -x_iY_i & -x_iZ_i & -x_i \end{bmatrix},$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -y_iX_i & -y_iY_i & -y_iZ_i & -y_i \end{bmatrix}.$$

The solution $\mathbf{p}$ is obtained as the right singular vector of A associated with the smallest singular value and reshaped to P_c. Reprojection to pixel coordinates uses homogeneous normalization: if $\tilde{\mathbf{u}} = [u\ v\ w]^\top = P_c\,\tilde{\mathbf{X}}$, then $\pi(\tilde{\mathbf{u}}) = (u/w,\ v/w)^\top$.

Fig. 3. Results of 25 surface captures from the left and right pan-tilt cameras.

Fig. 4. Stitched results from 25 surface captures using the left and right pan-tilt cameras.

3) Assign pan-tilt to 3-D points via the angle field: For any $\mathbf{X} \in \mathcal{P}$, project it into the stitched mosaic using the estimated projective matrix:

$$\tilde{\mathbf{u}}'_c = P_c \begin{bmatrix} \mathbf{X}^\top & 1 \end{bmatrix}^\top, \qquad (x', y')^\top = \pi(\tilde{\mathbf{u}}'_c). \tag{3}$$

If (x', y') lies inside the mosaic domain, assign the per-point angles by sampling the interpolated field:

$$\theta_{\text{pan}}(\mathbf{X}; c) = \widehat{\alpha}_c(x', y'), \qquad \theta_{\text{tilt}}(\mathbf{X}; c) = \widehat{\beta}_c(x', y'). \tag{4}$$

4) Synchronized stereo capture for volumetric measurement: With the established pan-tilt mapping, both cameras are automatically oriented to designated 3-D points for synchronized stereo capture. The 2-D displacement is calculated using DIC implemented independently for each camera:

$$\big(u_i^L(t), v_i^L(t)\big) = \text{DIC}\big(I^L(x, y, t), I_0^L(x, y); (x_i^L, y_i^L)\big), \tag{5}$$

$$\big(u_i^R(t), v_i^R(t)\big) = \text{DIC}\big(I^R(x, y, t), I_0^R(x, y); (x_i^R, y_i^R)\big). \tag{6}$$

The 2-D displacements are triangulated into 3-D displacements:

$$\begin{aligned} \big(U_i(t), V_i(t), W_i(t)\big) = F\big((x_i^L + u_i^L(t), y_i^L + v_i^L(t)), \\ (x_i^R + u_i^R(t), y_i^R + v_i^R(t))\big), \end{aligned} \tag{7}$$

where $F(\cdot)$ is a triangulation function determined by the calibrated transformation matrices. The relative 3-D displacement is:

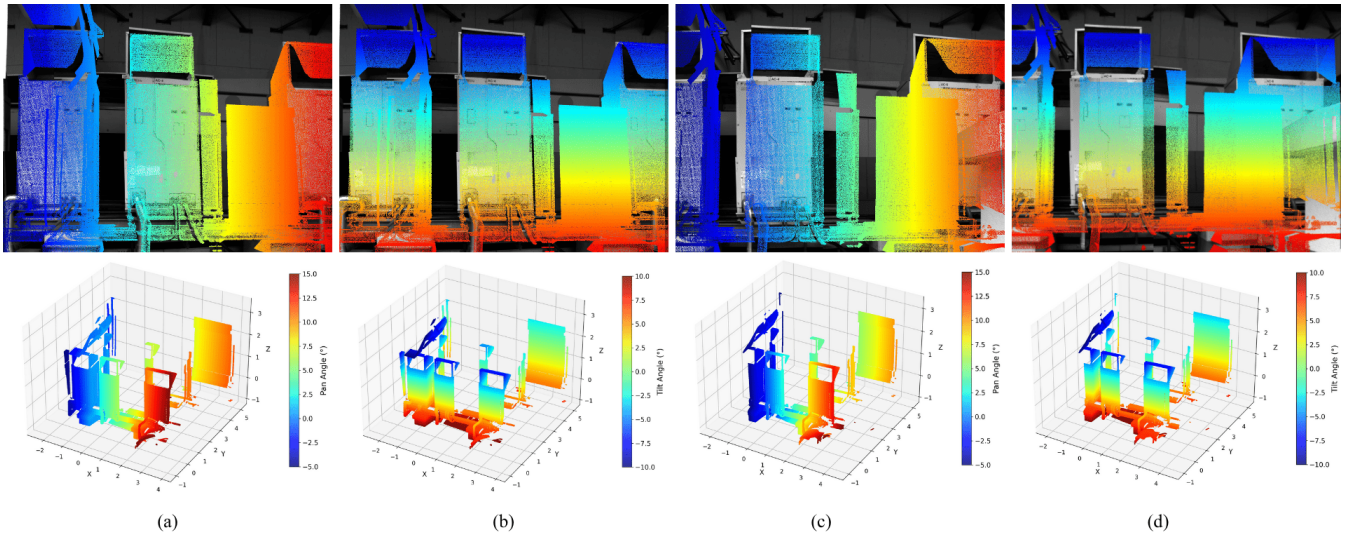
$$\big(\Delta U_i(t), \Delta V_i(t), \Delta W_i(t)\big) = \big(U_i(t) - X_i, V_i(t) - Y_i, W_i(t) - Z_i\big). \tag{8}$$

To analyze vibration frequency distribution, STFT is applied:

$$F_i(t) = \text{STFT}\big(A_i(t), \ldots, A_i(t + (K-1)\tau)\big), \tag{9}$$

$$\overline{F}(t) = \frac{1}{N} \sum_{i=0}^{N-1} F_i(t), \tag{10}$$

where $A_i(t)$ represents $\Delta U_i(t)$, $\Delta V_i(t)$, or $\Delta W_i(t)$.



(a) (b) (c) (d)

Fig. 5. Projection of the 3-D point cloud onto the images and pan-tilt angle color maps. (a) Camera 1 pan angle. (b) Camera 1 tilt angle. (c) Camera 2 pan angle. (d) Camera 2 tilt angle.

C. System Configuration

The proposed system employed two high-speed cameras (BASLER BOA 1936-400 CM, Germany) to capture three outdoor HVAC units, each measuring approximately 4.75 m × 2.31 m. Each camera was fitted with a COMPUTAR S10X10MMP lens (Japan) providing adjustable zoom, focus, and iris control, and mounted on ORION RHM-PA1L200 pan–tilt units (±180° pan, ±90° tilt). Both cameras were placed on 75 cm tripods to ensure stable coverage of the target area.

A terrestrial laser scanner (FARO Focus S150, FARO Technologies, USA) was used to acquire a 3-D point cloud of the environment, providing reference points for calibration and mapping between 2-D images and 3-D world coordinates. Image acquisition and processing were performed on a workstation (Windows 10 Pro, Intel Core i9-9900 @ 3.60 GHz, 32 GB RAM, NVIDIA GeForce RTX 2080 Ti), offering sufficient computational power for high-speed vision algorithms. The overall system configuration is illustrated in Fig. 2.

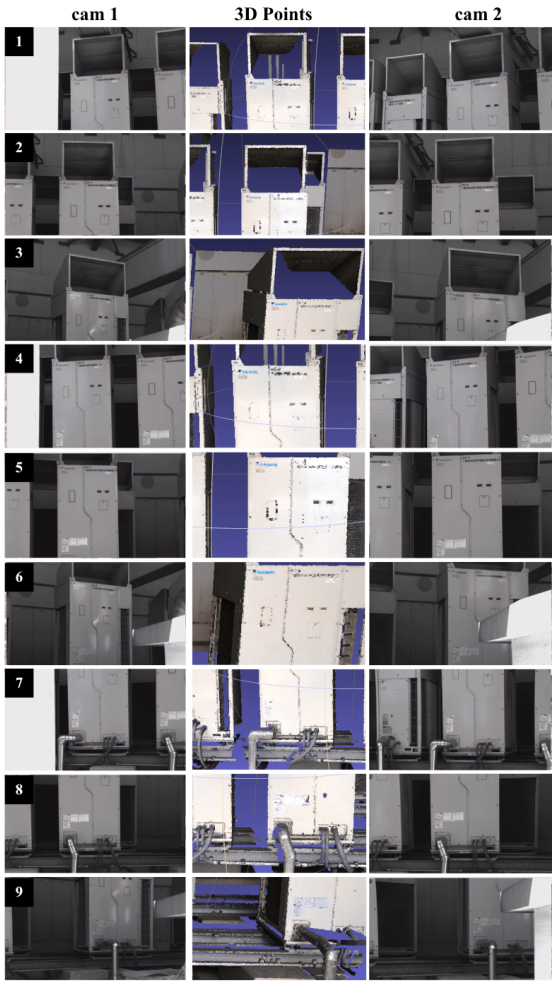
III. EXPERIMENT

To validate the proposed system and algorithm, we conducted an outdoor experiment capturing three HVAC units using two synchronized high-speed cameras. The total target area covered by the HVAC surfaces was approximately 4.75 m in width and 2.31 m in height. The capturing process was performed in a systematic sweep to cover the entire object surface with 5° increments between consecutive captures. Each camera executed a serpentine scanning pattern, starting from the top-left corner and moving horizontally from left to right. After completing one row, the camera tilted downward and continued scanning from right to left, repeating this pattern until the entire surface area was covered. In total, 25 different viewpoints were recorded per camera, providing dense coverage of the HVAC units. Each image was acquired at a resolution of 1920 × 1080 pixels with close zoom to the object surface, ensuring sufficient spatial resolution for high-precision volumetric measurement. The captured dataset of 25 overlapping regions is illustrated in Fig. 3.

All captured frames were stitched into a single panoramic mosaic to obtain a complete surface representation. The resulting stitched image measured 3500 × 2600 pixels, providing sufficient detail across the entire HVAC structure, as shown in Fig. 4. In parallel, a terrestrial laser scanner acquired a dense 3-D point cloud of the same HVAC units containing 560,357 points, serving as the reference geometry. For calibration, approximately 40 correspondence pairs between 3-D model points and 2-D image pixels were manually selected from distinctive features visible in both the stitched mosaic and the laser scan. Using these correspondences, we estimated the projective transformation matrix P_c for each camera, enabling projection of every 3-D point into the stitched image. Through the interpolated angle field derived from the commanded scan limits, each projected pixel location was assigned its corresponding pan-tilt angles. The 3-D to 2-D reprojection illustrated in Fig. 5 exhibits tight alignment between projected 3-D points and corresponding image features, validating the calibration accuracy.

Pan and tilt angles were computed for each 3-D point based on its projected pixel location and visualized as color-coded spatial maps. The spatial distribution of colors aligns well with the expected orientations of the pan-tilt units, clearly illustrating how each camera's field of view maps onto the object surface. The smooth color gradients show interpolation across the entire measurement region. These results demonstrate that the recovered pan and tilt parameters can effectively guide both cameras toward any specific 3-D point for synchronized stereo capture. Fig. 6 presents the experimental validation of multi-view correspondence, where both pan-tilt cameras were commanded to capture nine identical surface regions based on the computed angle

(a)

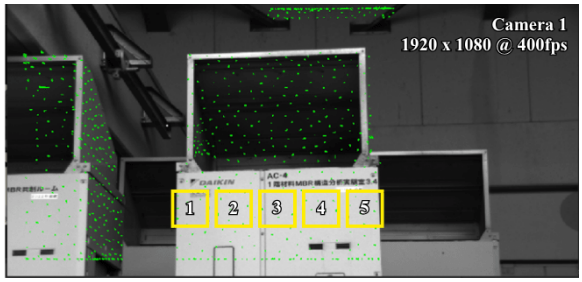


(b)

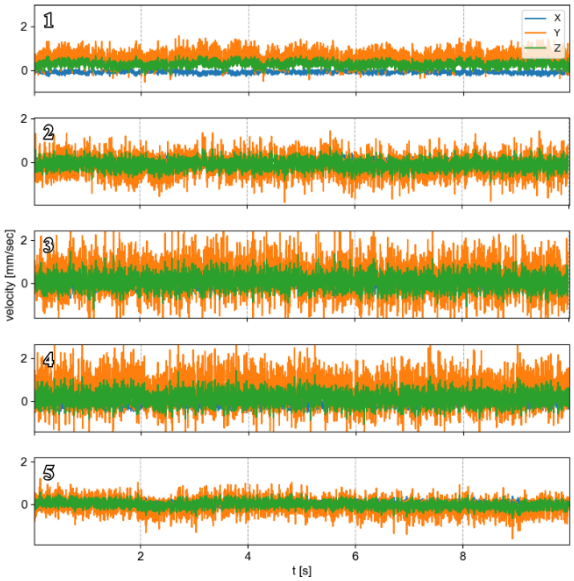
Fig. 6. Experimental setup. A 3D point is selected, and both Cam 1 and Cam 2 capture the same area, showing correspondence of multiple cameras to the same 3D point.

maps. The consistent alignment between Camera 1 and Camera 2 views demonstrates that the pan-tilt angles are reliably connected to all 3-D points through the shared model coordinate system.

To validate the 3-D DIC measurement capability, we performed vibration analysis on the operating HVAC system. Both cameras synchronously captured 4000 frames at 400 fps, enabling high-temporal-resolution velocity tracking over a 10-second period. Multiple measurement points were selected across the HVAC surface from the central compressor unit to peripheral regions. Figure 7 shows the



(a)



(b)

Fig. 7. (a) Five measurement points on HVAC surface. (b) Three-component velocity time histories showing spatial vibration.

velocity time histories with three-component vectors (X, Y, Z directions) revealing vibration characteristics at different locations. Points near the central compressor exhibit larger velocity amplitudes, while distant points show progressively smaller amplitudes, confirming that vibration energy propagates outward and attenuates with distance. Figure 8 presents the full-field 3-D velocity distribution visualized on the HVAC surface at multiple time, with color-coded magnitude illustrating the vibration propagation pattern concentrated around the object and decreasing toward peripheral regions. The frequency domain analysis in Figure 9 reveals the dominant vibration frequency at 10.82, 36.96, 43.87, 47.87, and 54.18 Hz, corresponding to the HVAC compressor operational frequency, with additional harmonic peaks at integer multiples indicating nonlinear vibration characteristics typical of machinery object.

In summary, the combination of stitched high-resolution imagery, laser-scanned 3-D reference geometry, and model-based stereo correspondence enables robust alignment between 3-D points and 2-D images. The system achieves reliable per-point pan-tilt mapping for scalable wide-area volumetric capture, with quantitative validation through vibration measurement demonstrating sub-millimeter velocity accuracy and frequency analysis for structural health monitoring and industrial equipment diagnostics.

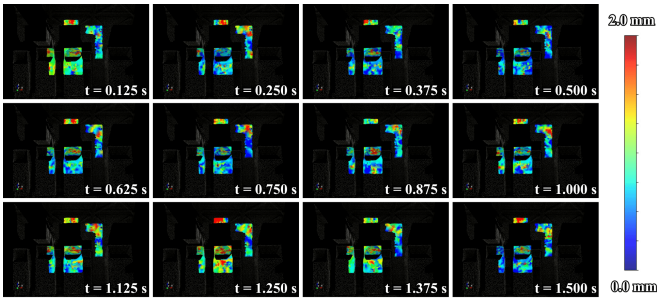


Fig. 8. 3-D velocity distribution evolution during vibration cycle.

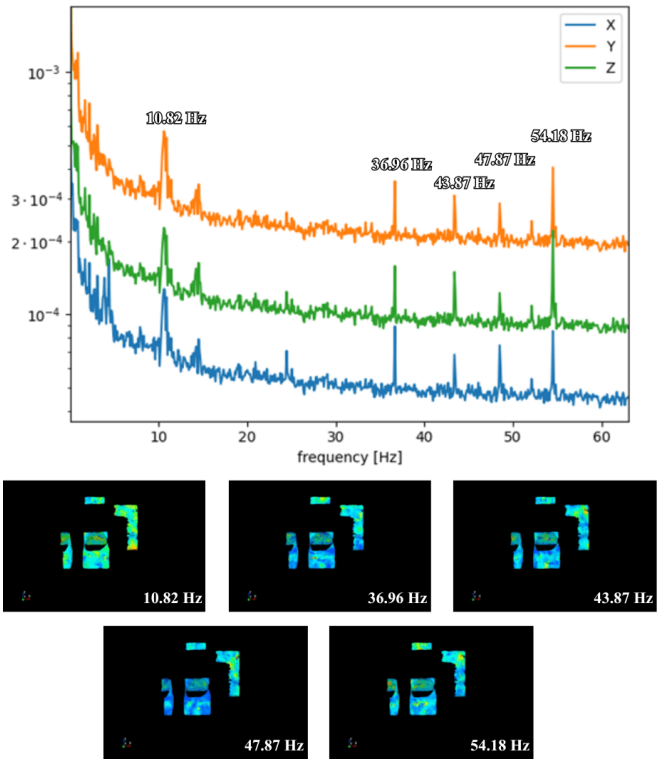


Fig. 9. Frequency response and mode shapes at dominant vibration frequencies

IV. CONCLUSION

This paper demonstrated a 3-D model-based pan-tilt multi-camera system that enables wide-area volumetric capture by establishing stereo correspondence through a shared 3-D reference model. The system reduces calibration complexity while achieving performance on texture-less industrial surfaces where conventional appearance-based methods fail. Experimental validation on HVAC units showed that the system achieves accurate 3-D velocity measurement and vibration analysis, demonstrating practical applicability for structural health monitoring and industrial diagnostics. Future work will focus on real-time implementation with closed-loop camera steering and extension to multi-camera networks for long-term monitoring applications.

REFERENCES

[1] Feiyue Wang, Kohei Shimasaki, Shaopeng Hu, Idaku Ishii, Yoshiyuki Umegaki, Tomohiko Ito, "Displacement Distribution Monitoring of an Unloader Crane in Ironworks Using High-speed and High-resolution Vision", *Tetsu-to-Hagane*, vol. 110, issue 11, pp. 797-809, 2024

[2] Y. Ezaki, Y. Moko, T. Hayakawa, and M. Ishikawa, "Angle of View Switching Method at High-Speed Using Motion Blur Compensation for Infrastructure Inspection," J. Robot. Mechatron., Vol.34 No.5, pp. 985-996, 2022.

[3] J. Li, K. Shimasaki, I. Ishii, M. Ogihara and M. Yoshiyama, "Detection of Honeybee Flower-Visiting Activities in Wide Areas Using High-Speed Vision," in *IEEE Sensors Letters*, vol. 9, no. 6, pp. 1-4, 2025

[4] H. Duan, S. Hu, K. Shimasaki, I. Ishii, and F. Wang, "3-D Motion Analysis of Falling Grains Using a Stereo HFR Camera System," in *Proc. SAMCON2025*, pp. 1–4, Japan, 2024

[5] T. Mori, F. Wang, K. Shimasaki and I. Ishii, "High-Speed Volumetric Capture for Vibration Propagation in the Human Body," in *IEEE Sensors Letters*, vol. 9, no. 5, pp. 1-4, 2025

[6] Yuta ABE, Taishi IRIYAMA, Takashi KOMURO, Kohei SHIMASAKI, Idaku ISHII, "High-Accuracy Acquisition of Both Hands' Postures for Multiple Users Using Ultrafast Mirror-Drive Pan-Tilt Camera", *Journal of the Japan Society for Precision Engineering*, 2025

[7] Timothy J. Beberniss, David A. Ehrhardt, "High-speed 3-D digital image correlation vibration measurement: Recent advancements and noted limitations", *Mechanical Systems and Signal Processing*, vol. 86, part B, 2017

[8] Ghulam Mubashar Hassan, "Deformation measurement in the presence of discontinuities with digital image correlation: A review", *Optics and Lasers in Engineering*, vol. 137, 2021

[9] C. D. B. Ferreira, "Digital image correlation for vibration analysis," M.S. thesis, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, 2021.

[10] Y. Jin, K. Hu, J. Liu, F. Wang, X. Liu, "From Capture to Display: A Survey on Volumetric Video", http://arxiv.org/abs/2309.05658, 2024

[11] S. Soro and W. Heinzelman, "A Survey of Visual Sensor Networks. Advances in Multimedia", 10.1155/2009/640386, 2009

[12] Y. Furukawa and C. Hernández, "Multi-View Stereo: A Tutorial", doi: 10.1561/0600000052, 2015

[13] M. Chen, K. Shimasaki, S. Hu, Q. Gu and I. Ishii, "An Active Tacking and Projection-based 3-D Reconstruction Method for Moving Objects," *2024 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO)*, pp. 167-172, Hong Kong, 2024

[14] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," Int. J. Comput. Vis., vol. 47, no. 1-3, pp. 7-42, 2002.

[15] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 2, pp. 328-341, 2008.

[16] B. Pan, "Digital image correlation for surface deformation measurement: historical developments, recent advances and future goals," Meas. Sci. Technol., vol. 29, no. 8, 2018, Art. no. 082001.

[17] M. A. Sutton, J. J. Orteu, and H. Schreier, Image Correlation for Shape, Motion and Deformation Measurements. Boston, MA: Springer, 2009.

[18] Y. L. Dong and B. Pan, "A review of speckle pattern fabrication and assessment for digital image correlation," Exp. Mech., vol. 57, no. 8, pp. 1161-1181, 2017.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91-110, 2004.

[20] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in Proc. IEEE Int. Conf. Comput. Vis., 2011, pp. 2564-2571.

[21] Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 11, pp. 1330-1334, 2000.

[22] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[23] W. Qin, K. Shimasaki, F. Wang, S. Hu, and I. Ishii, "Model-Based Three-Dimensional Digital Image Vibration Measurement Method Using Stereo High-Speed Cameras," in *Proc. Int. Conf. on Optics-photonics Design & Fabrication (ODF)*, 2024

[24] F. Wang, W. Qin, K. Shimasaki, I. Ishii and H. Matsuda, "HFR-Video-Based 3-D Software Sensor for Bridge Displacement Monitoring," in *IEEE Sensors Letters*, vol. 8, no. 10, pp. 1-4, 2024

[25] S. K. Zhou, H. Greenspan, and D. Shen, Deep Learning for Medical Image Analysis. London, U.K.: Academic Press, 2017.

[26] M. Alomran and D. Chai, "Feature-based panoramic image stitching," *14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 1-6, Thailand, 2016