ICPR 2026 Workshop on

Machine Vision for Industrial Inspection (MVI2)

Workshop Program

Date and Time	**August 21, 2026 (Friday)**	**13:30-17:30**
Venue	**Louis Néel Building, Room B1**	**Lyon, France**

Organized by IAPR Technical Committee 8 (TC8) in conjunction with the 28th International Conference on Pattern Recognition (ICPR 2026), MVI2 focuses on advances in machine vision for industrial inspection and intelligent manufacturing.

Program Notes

1. For the ICPR 2026 main conference schedule and venue information, please refer to the official program at the following link: https://icpr2026.org/program.html
   All ICPR workshops will take place in the Louis Néel Building, Click here to view the location on Google Maps.
2. For tea/coffee break and lunch timings for the workshops will be the same as those of the ICPR 2026 main conference. Friendly reminder: please remember to complete your ICPR lunch reservation in advance.
3. For presentation instructions, please refer to the official ICPR 2026 Oral Presentation Instructions at the following link: https://icpr2026.org/oralPresInstructions.html
   Each paper is allocated **20 minutes**, including approximately **15 minutes for the presentation** and **3–5 minutes for questions and answers**.

MVI2 Program Schedule

Time	Program Item	Details
13:30-13:35	**Opening Remarks and Welcome**	Workshop Chairs / IAPR TC8
13:35-15:15	Paper presentations and discussion	Each paper: 15-minute presentation, 3~5-minute Q&A
15:15-15:35	**Break**	Break
15:35-17:15	Paper presentations and discussion	Each paper: 15-minute presentation, 3~5-minute Q&A
17:15-17:25	Open Discussion and Future Directions	Academic-industry exchange and TC8 community discussion
17:25-17:30	**Closing Remarks**	Workshop Chairs

MVI2 Program Schedule Detail

Time	Program
13:30–13:35	**Opening Remarks and Workshop Introduction**
	Session: Multi-Modal Vision, 3D Reconstruction, and Industrial Applications
13:35–13:55	**Paper 2: MARC-V: Multi-Agent Rendering for CAD-VQA** **Abstract:** Although visual question answering for CAD models (CADVQA) serves a wide range of applications, including in-progress design review and manufacturing feature recognition, the existing approaches face two key limitations: **reliance** on images rendered from fixed, predefined viewpoints and the lack of informative geometric annotations in these rendered images. Fixed viewpoints are not designed for arbitrary questions, and may not satisfy the requirements of a given question. To address this issue, we introduce a question-conditioned adaptive viewpoint selection agent, inspired by CAD design review workflows in which designers carefully select viewpoints for targeted evaluation. However, viewpoint selection by itself is not sufficient to accurately answer questions about part shapes and dimensions, because rendered images do not provide explicit geometric measurements. To address this limitation, we augment our approach with an adaptive geometric annotation agent that overlays question-relevant measurements on rendered images in a question-conditioned manner. To accommodate multiple data modalities—namely CAD models and multi-view rendered images—our method is implemented as a composition of specialized agents. Extensive experiments across multiple vision–language models demonstrate consistent performance improvements on both a CAD model dataset and a multiview rendered image dataset.
13:55–14:15	**Paper 3: End-To-End Multi-View Multi-Modal Detection-Driven Image Fusion: One Method to Fuse Them All** **Abstract:** We present EDIF, an end-to-end detection-driven framework designed to unify multi-modal and multi-view image fusion within a single architecture. While most existing fusion methods address either spectral complementarity (multi-modal) or viewpoint variability (multi-view) in isolation, real-world perception systems increasingly require both. EDIF formulates fusion as an object-level alignment problem: heterogeneous images are encoded as sets of keypoints, which are matched and aggregated through a graph attention mechanism to form object-centric representations directly optimized for detection. To stabilize training across heterogeneous components, we introduce a three-stage task-driven strategy that progressively aligns keypoint extraction, object localization, and cross-sensor grouping. In addition, we release the Multi-Modal and Multi-View Object Detection Dataset (MMDOD), a new benchmark designed to study detection-driven fusion under strong modality–view dependencies. MMDOD contains over 10,000 images of transparent objects captured under four complementary modalities (visible, NIR, low-contrast, polarization shift) and six viewpoints, with detailed object-level annotations. Experiments on RGB–thermal, multi-camera, and joint multi-modal multi-view benchmarks show that EDIF achieves performance competitive with recent specialized methods, while uniquely operating within a unified framework. On MMDOD, EDIF significantly outperforms adapted multi-modal multi-view baselines, highlighting the benefits of detection-driven, object-level fusion. The proposed MMDOD dataset is available at https://datasets.liris.cnrs.fr/mmdod-version1.
14:15–14:35	**Paper 9: Mind the Domain Gap: Benchmarking Structure-from-Motion Methods on Paired Industrial Physical-Digital Twin Datasets to Enable Radiance Field Rendering** **Abstract:** As industry transitions toward fully digitalized factories, mirroring physical environments in digital-twins has **become** essential for robotics, monitoring, and process automation, a capability that fundamentally relies on 6D camera pose estimation and 3D scene reconstruction. We propose an evaluation methodology to benchmark these tasks in paired physical and digital environments. Real-world acquisitions are performed on a diverse set of industrial objects - including a mobile robot, a robotic arm, a bike, and a bench drill - captured with a stereo camera, with accurate Ground-Truth (GT) poses provided by a Motion Capture (MoCap) system. To complement these acquisitions, we generate Digital-Twin (DT) counterparts by replaying the same camera trajectories in a similar Unity-based environment using the LOOPER tool. The resulting paired Physical-Twin (PT) and DT sequences share the same camera poses, enabling controlled analysis of the gap between real and simulated domains. We apply the methodology to evaluate several Structure-from-Motion (SfM) pipelines, assessing their pose accuracy. The resulting benchmark highlights the influence of image realism, object shape, and domain shift on pose estimation performance, providing a practical testbed for industrial vision research and supporting future work in 6D camera pose estimation, digital-twins, and 3D scene reconstruction.
14:35–14:55	**Paper 10: Semi-Automated BIM Generation from Multi-Modal Data for Nuclear Facilities Decommissioning** **Abstract:** Building Information Modelling (BIM) is increasingly important for the digital documentation, analysis, and maintenance of complex industrial facilities. In the nuclear sector, and particularly in decommissioning contexts, the creation of accurate BIM models remains a challenging and time-consuming task due to the complexity of the environment, the heterogeneity of available data, and the limited accessibility of many assets. This work proposes an automated methodology for generating BIM-ready 3D representations from multi-modal data, combining point clouds,

	images, and geometric priors. The approach integrates multi-scale unsupervised decomposition, learning-based semantic and instance segmentation, image-to-3D projection, model retrieval, primitive fitting, parametric instantiation, and surface reconstruction for uncategorised objects. A dedicated data generation pipeline is also introduced to support training, validation, and evaluation of the proposed methods. Preliminary results, obtained in the context of our target field of application, highlight both the strengths and limitations of SAM3D-based 3D generation under varying conditions. The study discusses key practical difficulties, including data sparsity, and incomplete geometry. Overall, this work lays the groundwork for a scalable framework to support semi-automated BIM model creation in demanding industrial environments.
14:55–15:15	**Paper 11: XCT-SAM: Sequential Parameter-Efficient Domain Adaptation of SAM for Industrial XCT Defect Segmentation** **Abstract: Defect** segmentation in additive manufacturing (AM) X-ray computed tomography (XCT) images remains challenging due to severe class imbalance and large distribution shifts across scan conditions. Although recent foundation models such as the Segment Anything Model (SAM) provide strong general-purpose segmentation priors, their natural-image pre-training transfers poorly to the AM XCT domain, where defects appear as subtle non-semantic microstructural anomalies. Moreover, adapting SAM to the AM domain is further limited by the large domain gap and scarcity of labeled real XCT data. We present XCT-SAM, a sequential parameter-efficient adaptation framework for AM XCT defect segmentation. Instead of adapting SAM directly from natural images to XCT data, we first fine-tune Conv-LoRA adapters on an alloy-microstructure dataset and subsequently transfer the adapted model to XCT images, progressively bridging the domain gap. Using Conv-LoRA adapters with rank r=2, the framework injects convolutional spatial inductive bias into SAM's backbone while training approximately 4.15M parameters and keeping over 99% of the model frozen. We evaluate XCT-SAM on out-of-distribution CycleGAN-XCT benchmarks and real-world NIST XCT scans. Across both settings, XCT-SAM consistently outperforms zero-shot SAM and other domain-adapted SAM baselines, achieving the best overall IoU and Dice scores. These results demonstrate the effectiveness of intermediate domain adaptation with parameter-efficient adapters for industrial XCT defect segmentation.
15:15–15:35	**Break**
	Session: Industrial Quality Inspection and Anomaly Detection
15:35–15:55	**Paper 1: Diffusion Models for Advances in Manufacturing Quality – Abrasive Grains** **Abstract:** Artificial Intelligence is playing a significant and growing role in manufacturing quality assessment and control. Some examples include methods for real-time condition monitoring and prediction of adverse digressions, online vision based systems that can obviate sample based testing, enabling optimal control that can run processes at near-optimal conditions most of the time, etc. For most of these systems and in particular for sophisticated methods like CNNs, model accuracy and performance depend on the number and representativeness of data samples, especially for samples from faulty conditions which are relatively rare and trigger training data imbalance. Generative Artificial Intelligence can potentially play a major role in creating new samples, especially those of the rarer kind, provided statistical distinction between synthetic and real samples can be minimized. Diffusion based techniques are the current frontline among generative models, and in this work we have developed mechanisms for using vision based Diffusion models to get real time quality indications in the manufacturing process of Abrasive Grains – which are critical from a functional viewpoint and quite challenging from a technical perspective. We have converted Stable Diffusion architectures that take text-based pre-conditions to generate images, into statistics-based quality demand vector conditions and generate new and diverse images of type and merit demanded in the pre-conditions. Generated images contribute significantly to training accurate CNN models that provide quality status of a running production process, allowing not only quality decisions but also potential in-process feedback control of relevant production parameters.
15:55–16:15	**Paper 4: Improving Smartphone-Based Dental Ceramic Manufacturer Classification Across Different Devices** **Abstract:** Driven by suboptimal results in cross-sensor manufacturer classification for dental ceramics, this work proposes performance enhancements by transitioning from traditional texture descriptors to deep learning architectures. Previous studies indicated that using raw images instead of ISP-processed data leads to only minor improvements and fails to fully resolve the domain shift in 1-to-1 cross-sensor scenarios. In contrast, we propose utilizing raw color filter array data, reordered into three channels by averaging the two green components. Our findings suggest that sample-wise normalization further mitigates inter-sensor variability. Furthermore, employing noise and denoising as data augmentation strategies produces measurable accuracy gains by forcing the model to learn sensor-agnostic features. Experimental results demonstrate that these targeted engineering enhancements significantly outperform the existing literature benchmarks in cross-sensor dental material manufacturer identification.

16:15–16:35	**Paper 5: A Convolutional Block Attention and Bidirectional Pyramid Network for Gear Defect Detection in Challenging Industrial Datasets** **Abstract:** Gear defect detection is an important task for condition monitoring and predictive maintenance in industrial transmission systems. However, practical gear inspection remains challenging because defect regions are often small, visually subtle, and affected by complex industrial imaging conditions, such as background noise, uneven illumination, oil contamination, and variations in viewing angles. In addition, gear defect datasets usually suffer from class imbalance, where some failure modes appear more frequently than others, making robust multi-class defect detection difficult. To address these challenges, this paper proposes CABiP-YOLO, a convolutional block attention and bidirectional pyramid network for vision-based gear defect detection. The proposed framework is developed based on a YOLO detection architecture and is designed to detect and classify seven representative gear failure modes, including wear, scuffing, plastic deformation, Hertzian fatigue, cracking, fracture, and bending fatigue. A convolutional block attention mechanism is integrated to enhance spatial and channel-wise defect-related features, allowing the network to focus on critical damaged regions while suppressing irrelevant background information. Meanwhile, a bidirectional pyramid network is incorporated to strengthen multi-scale feature fusion by combining low-level texture details and high-level semantic information. Experimental results demonstrate that the proposed framework can localize and classify multiple types of gear defects under challenging imaging conditions. The analysis further shows that CABiP-YOLO has practical potential for automated visual inspection and gear condition monitoring, while class imbalance and subtle defect appearance remain important research for future improvement.
16:35–16:55	**Paper 6: D-MECAD: Dynamic Multi-Expert Continual Anomaly Detection with Adaptive Architecture Discovery** **Abstract:** In this paper, we propose Dynamic Multi-Expert Continual Anomaly Detection (D-MECAD), a self-organizing framework that autonomously discovers the optimal number of expert networks for sequential anomaly detection tasks. The framework employs an Adaptive Assignment mechanism that dynamically transitions from exploratory expert creation to strategic consolidation, regulating expert creation through feature similarity and capacity constraints. A novel Progressive Monitoring protocol continuously evaluates assignment quality using held-out validation data, identifying victim classes (experiencing performance degradation) and toxic classes (destabilizing expert performance), triggering automatic reassignment to preserve knowledge. Evaluation on the MVTec AD dataset shows that the system autonomously converges to 7 experts, achieving an average Area Under the Receiver Operating Characteristic curve (AUROC) of 0.94 with both cosine similarity and Euclidean distance metrics. Validation across similarity metrics demonstrates robustness of the dynamic discovery mechanism, while generalizability testing on the VisA dataset confirms autonomous expert allocation without manual architecture tuning. Results establish an effective plasticity stability balance in dynamic industrial environments.
16:55–17:15	**Paper 7: Neuro-Symbolic Logical Anomaly Detection with Interpretable Neural Rules** **Abstract:** Identifying logical anomalies in industrial settings remains a significant challenge for conventional anomaly detection methods, as these defects involve subtle violations of component relationships, quantities, and arrangements rather than structural flaws. In this paper, we pro- pose a Neuro-Symbolic Logical Anomaly Detection (NeSy-LAD) frame- work that integrates deep learning-based visual analysis with explicit symbolic reasoning to address these complex constraints. At the core of our approach is the Neural Pattern Discovery (NPD) module, a novel rule-extraction engine that autonomously identifies component-specific visual signatures without manual supervision and learns a comprehensive taxonomy of nine distinct rule types. These symbolic rules are integrated into a Logic Tensor Network (LTN), enabling end- to-end differentiable reasoning that grounds logical constraints in visual data. By combining data-driven feature extraction with symbolic reasoning, NeSy-LAD provides human-readable explanations of detected anomalies through explicit rule- violation analysis—a capability no competing method offers— while achieving competitive detection performance on the MVTec LOCO dataset. Crucially, all rule parameters are discovered automatically from normal training images without manual specification, making the framework directly applicable across diverse product categories.
17:15–17:25	**Open Discussion and Future Directions**
17:25–17:30	**Closing Remarks**

Organizers

IAPR Technical Committee 8 (TC8)
Dr. Zheng Liu (Chair, The University of British Columbia)
Dr. Hiroyuki Ukida (Co-Chair, Tokushima University)
Dr. Ling Bai (TC8 Secretary, The University of British Columbia)