21st International Conference on Computer Vision Theory and Applications (VISAPP 2026)

Volume 3

Marbella, Spain
9-11 March 2026

ISBN: 979-8-3313-3353-9

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2026) by SCITEPRESS – Science and Technology Publications, Lda.
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

For permission requests, please contact SCITEPRESS – Science and Technology Publications, Lda. at the address below.

SCITEPRESS – Science and Technology Publications, Lda.
Avenida de S. Francisco Xavier, Lote 7 Cv. C,
2900-616 Setúbal, Portugal

Phone: +351 265 520 185
Fax: +351 265520 186

info@scitepress.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

CONTENTS

INVITED SPEAKERS

KEYNOTE SPEAKERS

Rethinking Multimodal AI Models: Beyond Accuracy, Towards Trust 3
Elisa Ricci

Self-Supervised Learning Is Dead, Long Live Self-Supervised Learning (in the Age of MLLMs) 5
Yuki Asano

Seeing, Speaking, and Reasoning in a Visual World 7
Cees G. Snoek

3D VISION, MOTION, ROBOTICS, APPLICATION & SYSTEMS

FULL PAPERS

Gamifying Social Robotics: The Impact of Connect 4 on Human-Robot Interaction 13
Giuseppe De Simone, Luca Greco, Alessia Saggese and Mario Vento

3D Shrinkage Estimation for Printed Ceramics 23
Ulugbek Alibekov, Nicholas Baraghini, Tamas Magyar, Pablo Eugui, Laurin Ginner and Nicole Brosch

Consistent Multi-Lane Tracking with Temporally Recursive Spline Modeling 32
Sanghyeon Lee, Donghun Kang and Min H. Kim

Visual Autoregressive Modelling for Monocular Depth Estimation 44
Amir El-Ghoussani, André Kaup, Nassir Navab, Gustavo Carneiro and Vasileios Belagiannis

MH-Flow: Multi-View and Occlusion-Robust Dense Correspondence Estimation Based on Homographic Decomposition 55
Thomas Vincent Chang, Simon Seibt, Kay Hartmann, Bartosz von Rymon Lipinski, Konrad Kluwak and Marc Erich Latoschik

A Comparative Study of 3D Person Detection: Sensor Modalities and Robustness in Diverse Indoor and Outdoor Environments 66
Malaz Tamim, Andrea Matic-Flierl and Karsten Roscher

Place-NeRFs: A Smart Approach to Divide Large and Complex Scenes into Multiple Regions of Locally Related Views 75
Jose L. Huillca, Horácio Henriques, Andre Luiz da S. Pereira, Raphael dos S. Evangelista, Christian Erik Condori, Robinson Luiz Souza Garcia, Michelle Soares Pereira Facina, Maikon Bressani, Marcos Amaral de do Almeida, Lucas Bertelli Martins, Esteban Walter Gonzales Clua and Leandro A. F. Fernandes

Towards 3D Visualization of Insect Trajectories Using Stereo Event Camera Data and RGB-Based SfM Point Clouds 87
Regina Pohle-Fröhlich, Tobias Bolten, Colin Gebler and Felix Lögler

FutrTrack: A Camera-LiDAR Fusion Transformer for 3D Multiple Object Tracking 98
Martha Teiko Teye, Ori Maoz and Matthias Rottmann

Semantic Segmentation of Point Clouds for Autonomous Driving: An Approach for Imbalanced Classes 109
Beatriz Pinheiro de Lemos Lopes, Matheus Leonel de Andrade, Analucia Schiaffino Morales, Iwens Gervasio Sene Junior and Lucas Araújo Pereira

BladeR: Scalable, Shape-Guided 3D Reconstruction of Wind Turbine Blades from UAV Imagery 119
Jonathan Sterckx, Michiel Vlaminck and Hiep Luong

DiGS-3D: Diffusion Transformer as Unstructured 3D Gaussian Splatting Generator 131
Baptiste Engel, Florian Chabot, Mohamed Chaouch and Quoc-Cuong Pham

Eyes on the Grass: Biodiversity-Increasing Robotic Mowing Using Deep Visual Embeddings 141
Lars Beckers, Arno Waes, Aaron Van Campenhout and Toon Goedemé

Accuracy Improvement of 3D Point Cloud Segmentation Using Attention with Coordinate-Based Clustering 149
Konan Sasaki, Junya Komori and Kazuhiro Hotta

U3D: Unified Landmark-Displacement Framework for Real-Time Multi-Modal Emotion-Controllable 3D Facial Animation 159
Laxmi Narayen Nagarajan Venkatesan, Rittik Panda, Rahulraj B. R., Dinesh Babu Jayagopi, Raj Tumuluri and Magnus Revang

3DThermoScan: Simultaneous 3D Reconstruction and Temperature Measurement with a Single Camera 170
Tom Fevrier, Yvain Quéau, Thierry Sentenac and Florian Bugarin

MVTOP: Multi-View Transformer-Based Object Pose-Estimation 181
Lukas Ranftl, Felix Brendel, Bertram Drost and Carsten Steger

Towards AI-Assisted Kidney Visual Biopsy: Automatic Estimation of Banff Score and IFTA from the Analysis of Kidney Ultrasound Images 195
Arturo Fuentes, Valentina M. Blanco-Martínez, Javier Juega-Mariño and Jorge Bernal

MAIA-D: A Multi-Patch Anomaly Inspection Approach Based on DINOv2 206
Juan F. Flores Jarrin, Fabian Sturm and Tobias Bocklet

PGT-NeRF: Physics-Guided Thermal Neural Radiance Fields 216
Deepika Rani Kaliappan Mahalingam, Ajay Kumar Sigatapu, Deep Doshi and Ganesh Sistu

SHORT PAPERS

3D Gaussian Splatting with Fisheye Images: Field of View Analysis and Depth-Based Initialization 229
Ulas Gunes, Matias Turkulainen, Mikhail Silaev, Juho Kannala and Esa Rahtu

Early Detection of Colour Blindness Using Video Games 237
James O'Connell and Kenneth Dawson-Howe

A Simple and Efficient 3D Gaussians Storage Format 242
Denis Chaplygin, Juho Kannala and Esa Rahtu

MST Based Ordering of Point Clouds and Curve Fitting 250
Christoph Dalitz and Alexander Jongbloed

PairAnom: A Benchmark Dataset for Pairwise Fine-Grained Multi-Class Video Anomaly Detection 258
Brajesh Kumar, Kamakshya Prasad Nayak, Kamalakar Vijay Thakare, Priti Pal, Manvendra Singh and Debi Prosad Dogra

Multi-View Instance Segmentation of 3D Repeating Fine-Scale Structure for Precision Mapping of Coral Reefs 266
Fabio Ganovelli, Gaia Pavoni, Elena Cenni, Massimiliano Corsini, Somnath Dutta and Paolo Cignoni

ROI-NeRFs: Hi-Fi Visualization of Objects of Interest Within a Scene by NeRFs Composition 275
Quoc-Anh Bui, Gilles Rougeron, Géraldine Morin and Simone Gasparini

GenPlan: Generation Vector Residential Plans Based on the Textual Description 283
Egor Bazhenov, Stepan Kasai, Viacheslav Shalamov and Valeria Efimova

Talking Motion Signatures for Identity Recognition 291
Ivan Samarskyi and Jan Čech

Inpainting-Driven Depth Completion for Safe and Adaptive Collaborative Robotics 299
Ambra Vandone, Andrea Quattrini, Carlotta Pizzolato and Anna Valente

Robust Multi-View Self-Calibration from Dense Matches 307
Johannes Hägerlind, Bao-Long Tran, Urs Waldmann and Per-Erik Forssén

A Comprehensive Analysis of Monocular Depth Estimation Models Performance in Automotive Scenarios 319
Frederike Hörcher, Mikel García, Martí Sánchez, Nagore Barrena and Unai Elordi

Predicting Driver Maneuvers with Driver's Gaze and Vehicle Dynamics 327
Farzan Heidari, Taufiq Rahman and Michael A. Bauer

A Framework for Vehicular Instrumentation and Dataset Creation for Designing ADAS and Driver Intention Prediction Based on Eye Data 337
Farzan Heidari, Farhad Dalirani, Taufiq Rahman, Daniel Singh Cheema and Michael A. Bauer

PM25Vision: A Large-Scale Benchmark Dataset for Visual Estimation of Air Quality 348
Han Yang

Bridging the Dimensionality Gap: A Taxonomy and Survey of 2D Vision Model Adaptation for 3D Analysis 353
Akshat Pandya and Bhavuk Jain

Hierarchical Generation of Robot Programs Based on the Integration of Visual Affordance Recognition and Language Model 365
Kazuki Yamada, Iori Kokabu, Fumiya Murase, Kotaro Kato, Koki Maruyama, Shuichi Akizuki and Manabu Hashimoto

Robust Optical Flow Computation: A Higher-Order Differential Approach 373
Chanuka Algama and Kasun Amarasinghe

Development of an Intelligent AI-Powered Cooking Assistant 381
Manel Hentati, Raida Hentati, Ines Chaabane and Rim Ben Salah

Enhancing QDTrack with Self-Attention in Autonomous Driving Environments 389
Diego Gragnaniello, Antonio Greco, Antonio Parziale and Mario Vento

Fast-HaMeR: Boosting Hand Mesh Reconstruction Using Knowledge Distillation 398
Hunain Ahmed Jillani, Ahmed Tawfik Aboukhadra, Ahmed Elhayek, Jameel Malik, Nadia Robertini and Didier Stricker

Semantic Segmentation of LiDAR Point Clouds: A Hybrid Approach Combining Geometric Processing with Deep Learning 408
Matheus Leonel de Andrade, Beatriz Pinheiro de Lemos Lopes, Savio Salvarino Teles De Oliveira and Lucas Araujo Pereira

Neuro-Symbolic Multi-Head Learning for Spatial Relation Classification 415
Katia Abdou, Ala Mhalla and Olivier Caspary

Yolo-Key-6D: Single Stage Monocular 6D Pose Estimation with Keypoint Enhancements 424
Kemal Alperen Çetiner and Hazım Kemal Ekenel

R3GW: Relightable 3D Gaussians for Outdoor Scenes in the Wild 432
Margherita Lea Corona, Wieland Morgenstern, Peter Eisert and Anna Hilsmann

Almost 28 Years later: Mobile Visual Search Using Large Datasets on Handheld Hardware 444
Timon Kanik, Philipp Fleck, Thomas Kernbauer and Clemens Arth

Transformer-Based Framework for 3D Human Pose Estimation Using YOLO Backbone 456
Miguel F. Lima, Ana Filipa Rodrigues Nogueira, Cláudia D. Rocha, Luís F. Teixeira and Hélder Oliveira

2VA: Vessel-Aware Retinal Image Anonymization that Preserves Diabetic Retinopathy Utility 464
Rishabh Shukla, Pansuriya Uttambhai Bharatbhai and Harkeerat Kaur

Geometry-Based Differentiable Camera Placement for Optimal 3D Coverage 473
A. Moraza, X. Lin, D. Mejia-Parra and N. Barrena

Track-UGV-SYNC: A Framework for Synchronizing Sports User-Generated Videos Using Multi-Object Tracking Features 484
Elton Alencar and Rosiane de Freitas

Multi-Pathology Segmentation in Lumbar Spine MRI: A Comparative Deep Learning Approach 492
Claudio Leite, Samuel Felipe dos Santos and Jurandy Almeida

PreVolE: A Robust Data-Driven Framework for Text-Guided Food Volume Estimation 503
Umair Haroon, Ahmad AlMughrabi, Ricardo Marques and Petia Radeva

Decentralized Privacy-Preserving Federated Learning of Computer Vision Models on Edge Devices 512
Damian Harenčák, Lukáš Gajdošech and Martin Madaras

MExECON: Multi-View Extended Explicit Clothed Humans Optimized via Normal Integration 520
Fulden Ece Uğur, Rafael Redondo, Albert Barreiro, Stefan Hristov and Roger Marí

RIGATING: Radar-Image Gated Attention Fusion for Multimodal Semantic Segmentation in Autonomous Mobile Systems 529
Patrick Ziegler, Jonas Mehl, Jörg Franke and Sebastian Reitelshöfer

From Captions to Queries: A VLM Pipeline for Queryable Metadata in Synthetic Environments 538
Christopher May, Abhijith Raghavan Shaji, Jörg Franke and Sebastian Reitelshöfer

Towards Understanding 3D Vision: The Role of Gaussian Curvature 548
Sherlon Almeida da Silva, Davi Geiger, Luiz Velho and Moacir Antonelli Ponti

Adversarial Domain Adaptation for Penile Cancer Diagnosis in Histopathological Images 560
Rick Eick Vieira dos Santos, Victor José Beltrão Almajano Martinez, Marcos Gabriel Mendes Lauande and Geraldo Braz Júnior

Estimating the 3D Position of Hidden Humans Using Reflections on Vehicle Bodies 568
Hiroto Kozawa, Fumihiko Sakaue and Jun Sato

3D Reconstruction of Hidden Objects from Simultaneous Recovery of Light Source and Environment 576
Yuma Matsubara, Fumihiko Sakaue and Jun Sato

Binocular Visual Distortion Correction on Display 584
Yan Yingxin, Fumihiko Sakaue and Jun Sato

GGD-Tex: Geometry-Guided Diffusion for 3D Texture Synthesis from Single RGB Image 592
Jit Chatterjee, Gijs Fiten and Maria Torres Vega

Caption-Matching: A Multimodal Approach for Cross-Domain Image Retrieval 600
Lucas Iijima, Nikolaos Giakoumoglou and Tania Stathaki

Automatic Seal Quality Inspection Using Deep Learning in Mono Material Flexible Packaging 608
Tajmout Ilyas, Hamid Ladjal, Fayez Shakil Ahmed, Hammouri Hassan and Roumanet Frédéric

Detecting 3D Line Segments for 6DoF Pose Estimation with Limited Data 616
Matej Mok, Lukáš Gajdošech, Michal Mesároš, Martin Madaras and Viktor Kocur

PDNet: Exploring Pruning and Distillation for Real-Time Worker Activity Recognition in Industrial Assembly 624
Malek Baba, Thinesh Thiyakesan Ponbagavathi and Alina Roitberg

POM-NeRF: Perception-Oriented Modification of Neural Radiance Fields for Enhanced 3D Volumetric Rendering from Multi-View Images 631
Joy Datta, Rawhatur Rabbi, Md. Shaleh Islam Tonmoy, Puja Saha and Chad Mourning

Non-invasive Growth Monitoring of Small Freshwater Fish in Home Aquariums via Stereo Vision 639
Clemens Seibold, Anna Hilsmann and Peter Eisert

An Agentic Framework for Context-Guided Industrial Label Inspection 647
Ravneet Kaur, Sai Rama Raju Penmatsa and Joydeep Acharya

Simultaneous Reconstruction of Scene Information and Boundary Surface Geometry from Images Captured through Dynamic Refractive Interfaces 655
Mao Okamura, Fumihiko Sakaue and Jun Sato

Light Field Representation Using a DLP Projector and Focus Tunable Lens 664
Kizuku Miyazawa, Fumihiko Sakaue and Jun Sato

Monocular Image-Based 3D Seated Posture Estimation Integrated with a Seat-Angle Adjustable Smart Chair 672
Hitoshi Shimomae and Noriko Takemura

Towards GNSS-Denied Geo-Positioning Using Search Area Refinement 680
Signe Møller-Skuldbøl, Kasper Kjærsgaard Lauritsen, Maria Daliana Moga, Omkar Arun Korgaonkar, Peter Guld Leth and Andreas Møgelmose

Fasciculation Detection in Ultrasound Images for ALS Diagnosis Based on Black Box Variational Inference 688
Turrnum Shahzadi, Junfeng Zhou, Kota Bokuda and Norio Tagawa

Neural BRDF Decomposition for Consistent Material Recovery in Inverse Rendering 696
Anant Kumar and Parag Chaudhuri

SkipFusion: Multimodal Process-Aware Segmentation for Steel Surface Inspection 704
Lars De Pauw and Toon Goedemé

Face2Cam: A User-to-Webcam Distance Estimation Image Dataset 712
Yixuan Zhai and Arryn Robbins

AI-Driven Enhancement of Depth Map Consistency 720
Jakub Kit, Dominika Klóska, Dawid Mieloch and Adrian Dziembowski

Exploring Multimodal LMMs for Online Episodic Memory Question Answering on the Edge 728
Giuseppe Lando, Rosario Forte and Antonino Furnari

A Learnable 3D Perlin-Noise Representation for Image-Domain 3D Reconstruction 736
Yuchen Liu, Eiji Kamioka and Phan Xuan Tan

AUTHOR INDEX 749