Guest Editorial: Special issue on Traditional Computer Vision in the Age of Deep Learning

Matteo Poggi[1], Federica Arrigoni[2], Andrea Fusiello[3], Stefano Mattoccia[1], Adrien Bartoli[4], Torsten Sattler[5], Tomas Pajdla[5]

In the last 5-10 years we have witnessed that deep learning has revolutionized Computer Vision, conquering the main scene in most top-tier conferences and journals. However, several problems and topics for which deep-learned solutions are currently not preferable over classical ones exist, that typically involve a strong mathematical model (e.g., camera calibration and structure-from-motion). This special issue collects contributions related to algorithms and methodologies that address Computer Vision problems in a "traditional" or "classic" way, in the sense that analytical/explicit models are deployed, as opposed to learned/neural ones. A particular focus is given to traditional approaches that perform better than neural ones (for instance, in terms of generalization across different domains) or that, although performing sub-par, provide clear advantages with respect to deep learning solutions (for instance, in terms of efforts to collect data, computational requirements, power consumption or model compactness). We hope this special issue can inspire the reader towards critical discussions about preferring a traditional solution rather than a deep learning approach, igniting relevant questions about how to bridge the gap between learning and classic knowledge, as well as ethical implications of deep learning approaches in comparison to traditional ones.

This issue consists of 21 papers that cover various aspects within these themes. The articles have undergone rigorous peer-review according to the journal's high standards. The special issue was preceded by an international workshop on Traditional Computer Vision in the Age of Deep Learning, that was held online in conjunction with ICCV 2021.

We now provide a brief summary of each paper:

The first article, "Finite Aperture Stereo" by Matthew Bailey, Adrian Hilton & Jean-Yves Guillemaut, presents a comprehensive pipeline for modeling scenes using finite aperture cameras, combining stereo and defocus cues for multi-view 3D reconstruction and incorporating pre-trained deep features. It demonstrates improved performance across various materials and geometries with respect to modern multi-view stereo methods.

The second article, "When Multi-Focus Image Fusion Networks Meet Traditional Edge-Preservation Technology" by Zeyu Wang, Xiongfei Li, Libo Zhao, Haoran Duan, Shidong Wang, Hao Liu & Xiaoli Zhang, develops a novel edge-aware layer for multi-focus image fusion, derived from isometric domain transformation and a recursive filter. The study shows improved decision map quality by highlighting edge discrepancies between focused and defocused regions.

The third article, “Perspective-1-Ellipsoid: Formulation, Analysis and Solutions of the Camera Pose Estimation Problem from One Ellipse-Ellipsoid Correspondence” by Vincent Gaudillière, Gilles Simon & Marie-Odile Berger, envisions a novel theoretical framework specific to ellipsoids in computer vision for camera pose estimation with analytical derivations provided. This approach offers advantages such as reducing the estimation problem to position or orientation-only estimation and ultimately simplifying it to a 1 Degree-of-Freedom (1DoF) problem.

The fourth article, “On Making SIFT Features Affine Covariant” by Daniel Barath, presents a fast method to recover affine correspondences (ACs) from orientation- and scale-covariant features using pre-estimated epipolar geometry. It also introduces a minimal solver for estimating the relative pose of a camera, leading to significant speed improvements in various computer vision tasks with comparable accuracy to state-of-the-art methods.

The fifth article, “Blur Invariants for Image Recognition” by Jan Flusser, Matěj Lébl, Filip Šroubek, Matteo Pedone & Jitka Kostková, proposes a novel theory of blur invariants, which enables the description and recognition of blurred images without the need for deblurring or data augmentation. This framework, unlike previous methods, does not require prior knowledge of the blur type and is constructed in the Fourier domain using orthogonal projection operators and moment expansion for efficient computation.

The sixth article, “A Family of Approaches for Full 3D Reconstruction of Objects with Complex Surface Reflectance” by Gianmarco Addari & Jean-Yves Guillemaut, introduces innovative 3D reconstruction methods that leverage Helmholtz Stereopsis to create complete 3D models of objects with unknown surface reflectance, addressing a previous limitation of 2.5D modeling. These approaches are evaluated on various datasets, showcasing their effectiveness in achieving high-quality 3D reconstructions of complex objects.

The seventh article, “Fast and Accurate 3D Registration from Line Intersection Constraints” by André Mateus, Siddhant Ranade, Srikumar Ramalingam & Pedro Miraldo, presents a novel approach to 3D registration using line intersection constraints, which outperforms traditional methods relying on point or plane correspondences. The proposed method involves a two-step process: a coarse estimation with outlier rejection followed by a refinement step using non-linear techniques.

The eighth article, “Refractive Pose Refinement” by Xiao Hu, François Lauze & Kim Steenstrup Pedersen, addresses pose estimation under refraction, providing geometric constraints and efficient optimization algorithms. It also contributes with novel datasets made publicly available.

The ninth article, “A Minimal Solution for Image-Based Sphere Estimation” by Tekla Tóth & Levente Hajder, introduces a novel minimal solver for fitting a sphere from its 2D central projection, which is represented as a special ellipse. This method only requires three contour points, making it efficient, numerically stable, and easy to implement. Experimental results demonstrate its superiority over existing methods in scenarios such as LiDAR-camera calibration.

The tenth article, “Expression-Preserving Face Frontalization Improves Visually Assisted Speech Processing” by Zhiqi Kang, Mostafa Sadeghi, Radu Horaud & Xavier Alameda-Pineda, introduces a

face frontalization method that preserves non-rigid facial expressions, enhancing visually assisted speech processing. It combines rigid and non-rigid transformation estimation, utilizing the Student's t-distribution and a Bayesian filter to handle data errors and dynamic facial deformations. Comparative evaluations demonstrate its efficacy in preserving facial expressions when integrated into speech processing pipelines.

The eleventh article, “Relating View Directions of Complementary-View Mobile Cameras via the Human Shadow” by Ruize Han, Yiyang Gan, Likai Wang, Nan Li, Wei Feng & Song Wang, explores the use of mobile cameras, including drones and person-worn cameras, for enhanced video surveillance. To connect these different perspectives, it introduces a shadow-direction-aware network to estimate angles. The approach is also effective in real-world scenarios.

The twelfth article, “RM3D: Robust Data-Efficient 3D Scene Parsing via Traditional and Learnt 3D Descriptors-Based Semantic Region Merging” by Kangcheng Liu, addresses the challenge of understanding 3D point clouds with limited labeled data. It introduces a framework called Region Merging 3D (RM3D) that excels in various weakly supervised 3D point cloud understanding tasks, outperforming existing methods on multiple benchmark datasets.

The thirteenth article, “Shuffled Linear Regression with Outliers in Both Covariates and Responses” by Feiran Li, Kent Fujiwara, Fumio Okura & Yasuyuki Matsushita, investigates a shuffled linear regression problem that deals with data correspondences. It explores the problem's conditions, NP-hardness, and presents an efficient solution algorithm with global convergence properties. Experimental results validate its effectiveness in various tasks.

The fourteenth article, “Minimal Solvers for Relative Pose Estimation of Multi-Camera Systems using Affine Correspondences” by Banglei Guan, Ji Zhao, Daniel Barath & Friedrich Fraundorfer, presents three new solvers for estimating the relative pose of a multi-camera system from affine correspondences. These solvers are more efficient than existing approaches as they require fewer correspondences and offer superior accuracy in estimating poses.

The fifteenth article, “Poincaré Kernels for Hyperbolic Representations” by Pengfei Fang, Mehrtash Harandi, Zhenzhong Lan & Lars Petersson, introduces valid kernel functions for hyperbolic spaces, addressing the challenges of curved geometry and enhancing performance across various machine learning tasks, including few-shot learning, zero-shot learning, person re-identification, and more.

The sixteenth article, “Improving Domain Adaptation Through Class Aware Frequency Transformation” by Vikash Kumar, Himanshu Patil, Rohit Lal & Anirban Chakraborty, develops Class Aware Frequency Transformation (CAFT) to reduce domain shift in unsupervised domain adaptation (UDA) between source and target domains. CAFT is computationally efficient and can enhance the performance of existing UDA methods.

The seventeenth article, “Benchmarking the Complementary-View Multi-human Association and Tracking” by Ruize Han, Wei Feng, Feifan Wang, Zekun Qian, Haomin Yan & Song Wang, focuses on multi-human tracking using multiple moving cameras with diverse perspectives. It introduces a new dataset containing synchronized videos from drones and wearable cameras, with thorough

annotations and cross-frame associations. The paper also presents a baseline algorithm for multi-view multiple human tracking.

The eighteenth article, “Building 3D Generative Models from Minimal Data” by Skylar Sutherland, Bernhard Egger & Joshua Tenenbaum, presents a method for creating generative models of 3D objects from a single 3D mesh and improving them through unsupervised learning from 2D images. It focuses on constructing 3D morphable models from a single template, particularly in the context of face recognition and unsupervised learning with minimal data.

The twentieth article, “DeepFTSG: Multi-Stream Asymmetric USE-Net Trellis Encoders With Shared Decoder Feature Fusion Architecture for Video Motion Segmentation” by Gani Rahmon, Kannappan Palaniappan, Imad Eddine Toubal, Filiz Bunyak, Raghuveer Rao and Guna Seetharaman, introduces DeepFTSG, a novel deep architecture for robust moving object detection in challenging scenarios, such as cluttered backgrounds, occlusions, and varying environmental conditions. It demonstrates strong cross-dataset generalization capabilities when evaluated on unseen benchmark videos, outperforming existing methods.

The twenty-first article, “ToTem NRSfM: Object-wise Non-Rigid Structure-from-Motion with a Topological Template” by Agniva Sengupta and Adrien Bartoli, presents a Non-Rigid Structure-from-Motion (NRSfM) method for reconstructing objects with a known topology using a Topological Template that weakly resembles the object. Unlike traditional templates, it does not require a feasible object shape or a texture map, making it easier to construct.

Collectively, these 21 papers provide a detailed compilation of the diverse range of issues currently being investigated in the field of Traditional Computer Vision in the Age of Deep Learning.

---

[1] University of Bologna, Department of Computer Science and Engineering (DISI), Viale del Risorgimento 2, 40136, Bologna, Italy

[2] Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Via Ponzio 24/5, 20133 Milano, Italy

[3] University of Udine, Dipartimento Politecnico di Ingegneria e Architettura (DPIA), Via delle Scienze 206, 33100 Udine, Italy

[4] Université Clermont Auvergne, Faculté de Médecine, Place Henri Dunant 28, 63001 Clermont-Ferrand, France

[5] Czech Technical University in Prague, Czech Institute of Informatics, Robotics and Cybernetics (CIIRC), Jugoslávských Partyzánů 1580/3, 160 00 Prague, Czech Republic