Guest Editorial Introduction to the Special Issue on Large-scale Visual Sensor Networks: Architectures and Applications.

LARGE-scale visual sensor networks have become progressively an essential part of our daily lives underpinning many technological, financial, and social advancement today, with applications in smart cities, traffic monitoring, environmental pollution control, public safety and crime prevention.

Sensing technologies have made huge progress in the last two decades, with the emerging opportunity to miniaturize and communicate wirelessly in low power modalities at the edge. Motivated by several societal needs, such as security, natural interfaces, gaming, affective computing and assisted living, there has been an increasing interest in developing visual sensor networks for urban space, smart homes, and environmental monitoring in recent years.

The integration of multiple devices in sensor networks is a hot topic in the scientific community. Thanks to both the decreasing costs of visual sensors and the increasing performance of distributed edge devices, the possibility of performing integration of a forest of both homogeneous and heterogeneous sensors is not only a theoretical study but also a makeable technological solution in a real world context.

On one hand, the capability of processing data from multiple sensors opens up new scenarios, allowing the realization of large-scale sensor infrastructures. In this way, many social activities can be understood from a new point of view. On the other hand, large-scale sensor networks introduce new challenges including data synchronization, communications, security, data fusion, real-time processing, and high-level decision making.

This poses several research issues ranging from data storage, wireless communication coverage that allows self-organization of the topology of the networked sensors, to their coordination in terms of optimizing computation and where to focalize sensing and what to sense.

In a sensor network, a key-role is played by large-scale visual networks: the decreasing cost of smart devices enables sensor networks to scale to a large number of devices. Adding or removing sensors in a seamless way is an important requirement for sensor network architectures of the future, where everybody should be able to modify the network architecture without compromising the operation of the network itself. In addition, the impact of deep learning on multi-source big data from distributed networks need be investigated. The capacity of managing large-scale sensor networks will become strategic in the coming years for most IT organizations.

Sensor network research was initially driven by expensive military applications such as battlefield surveillance and enemy tracking. Nowadays, the diffusion of more accessible fixed and mobile sensors has changed significantly our daily life, e.g. smartphones. Many applications based on distributed sensing systems for civil applications have been developed. These applications can be classified into habitat monitoring, environment observation and forecast systems, human activity monitoring for health, security, and surveillance. An interesting perspective on how to exploit sensor networks has also emerged from the COVID-19 pandemic: The availability of capillary distributed sensor networks can be exploited for a deep monitoring of specific conditions, in an ideal way at a single-person level, by means of mobile phones specific apps. The availability of such detailed information can be used for massive tracing operations, that can be formally considered as borderline in terms of privacy or human rights, but that can be considered as a possibility in specific critical periods, like a pandemic.

This special issue highlights the latest developments in large scale visual sensor networks and their applications, ranging from the design and implementation to the processing and understanding of large amount of video data. It presents contributions on new solutions addressing some of the challenges and difficulties in this field, as well as prototypes, systems, tools and techniques. It also includes general survey papers predicting future directions of research and technology development.

Lu et al. propose an adaptive region proposal scheme with feature channel regularization to facilitate robust object tracking. Tracking is considered as a linear regression problem; an ensemble of correlation filters is trained online to distinguish the foreground target from the background. Authors integrate adaptively learned region proposals into an enhanced two-stream tracking framework based on correlation filters. For the tracking stream, a two-stage cascade correlation filters on deep convolutional features is learned to ensure competitive tracking performance. For the detection stream, adaptive region proposals have been employed, which are effective in recovering target objects from tracking failures caused by heavy occlusion or out-of-view movement. In contrast to traditional tracking-bydetection methods using random samples or sliding windows, target re-detection over adaptively learned region proposals is performed. Since region proposals naturally take the objectness information into account, authors show that the proposed adaptive region proposals can handle the challenging scale estimation problem as well. Authors have extensively tested their approach on OTB, VOT and UAV-123 datasets demonstrating that their method performs favorably against state-of-the-art tracking algorithms.

The contributions of the method proposed by Farinella et

al. are twofold. First, an approach for localization of shopping carts in a retail store from egocentric images is proposed. Second, a new database is presented to the scientific community. Addressing the first task allows to infer information on the behavior of the customers to understand how they move in the store and what they pay more attention to. To study the problem, a large dataset of images collected in a real retail store is proposed. The dataset comprises 19,531 RGB images along with depth maps, ground truth camera poses, as well as class labels specifying the areas of the store in which each image has been acquired. The dataset could surely encourages research in large-scale image-based indoor localization; moreover, it addresses the scarcity of large datasets to tackle the problem. Authors perform a benchmark of several image-based localization techniques exploiting images and depth information on the proposed dataset. In this work authors compare both localization performances and space/time requirements. The results show that, while state-of-the-art approaches allow to achieve good results, there is space for improvement.

Zhang et al. propose an approach based on Siamese networks; these have been successfully introduced into visual tracking, which match the best candidate and a target template via a couple of networks with shared parameters. However, most Siamese network-based trackers (SNTs) are tailored to best match the canonical posture of the template and the search-region images, resulting in inferior performance when the target objects have large-scale pose variations. Besides, SNTs fail to accurately discriminate distractors because they only leverage high-level semantic features as target representations that cannot well tell from different targets of the same category. To address these issues, authors introduce an efficient and effective SNT that is based on feature alignment and aggregation networks. An effective feature alignment network module to calibrate the search-region image is designed. This module results in a more reliable matching response that is robust to severe target pose variations. Then, an effective shallow-level and high-level feature aggregation network module to complement the feature characteristics is developed. This way, the learned feature representation not only well differentiates the target from distractors, but also introduces robustness to target appearance variations. Afterwards, a channel-attention mechanism to further strengthen the discriminative capability of the aggregated feature representation is employed. Finally, both the alignment and the aggregation modules are seamlessly integrated into the Siamese networks for robust tracking. Extensive evaluations on a variety of benchmarks including VOT-2017, OTB-100, UAV123 and GOT-10k demonstrate favorable performance of this tracker against state-of-the-art ones with a speed of 60 fps.

Fang et al. introduce an approach for future frame prediction in video, a traditional problem in computer vision, useful for a range of practical applications, such as intention prediction or video anomaly detection. However, this task is challenging because of the complex and dynamic evolution of scene. The difficulty of video frame prediction is to model the inherent spatio-temporal correlation between frames and pose an adaptive and flexible framework for large motion change or appearance variation. Authors construct a deep multi-branch mask network (DMMNet) which adaptively fuses the advantages of optical flow warping and RGB pixel synthesizing methods. In the procedure of DMMNet, mask layer is added in each branch to adaptively adjust the magnitude range of estimated optical flow and the weight of predicted frames by optical flow warping and RGB pixel synthesizing, respectively. In other words, a more flexible masking network for motion and appearance fusion on video frame prediction is provided. Exhaustive experiments on Caltech pedestrian and UCF101 datasets show that this model can obtain favorable video frame prediction performance compared with the state-of-the-art methods.

Paolo Spagnolo
National Research Council of Italy
Via Monteroni - Lecce, Italy
email: paolo.spagnolo@cnr.it

George Bebis
University of Nevada, USA
Reno, NV 89557
email: bebis@cse.unr.edu

Hamid Aghajan
Department of Electrical Engineering
Sharif University of Technology, Iran
email: aghajan@ee.sharif.edu

Shaogang Gong
School of Electronic Engineering and Computer Science
Queen Mary University of London, UK
email: s.gong@qmul.ac.uk

Amy Loutfi
School of Science and Technology
Orebro University, Sweden
email: amy.loutfi@oru.se

Leonid Sigal
Department of Computer Science
University of British Columbia
Vancouver, Canada
email: lsigal@cs.ubc.ca

Wei-Shi Zheng
School of Data and Computer Science
Sun Yat-sen University, China
email: wszheng@ieee.org

Paolo Spagnolo received the engineering degree in computer science from the University of Lecce, Lecce, Italy, in 2002. Since then he has been with the Italian National Research Council. He has been working on several research topics regarding Artificial Intelligence and Computer Vision. He has studied techniques and methodologies for multidimensional digital signal processing; linear and nonlinear signal characterization; signal features extraction; supervised and unsupervised classification of signals; deep neural network (CNN). He is author of over 80 papers on Artificial Intelligence. He also acts as a reviewer for several international journals. He participated in a number of international projects in the area of image and video analysis. He has been regularly invited to take part in the Scientific Committees of national and international conferences.

Hamid Aghajan received his B.S. degree from Sharif University of Technology, Tehran, Iran, in 1989, and M.S. and Ph.D. degrees from Stanford University in 1991 and 1995, respectively, all in electrical engineering. After gaining industry experience in the silicon valley from 1995 to 2002 in the fields of semiconductors, wireless and optical communications, and genetics, he established and served as director of a research lab on wireless sensor networks and ambient intelligence at Stanford University in 2003, where he supervised the research of numerous students and international visitors until 2013 on topics involving methods and applications in smart homes, multi-camera networks, behavior learning and adaptive services in home and office, smart transportation, and elderly monitoring. He served as founder and steering committee member of the Int. Conf. on Distributed Smart Cameras from 2007 to 2019. He co-founded Journal of Ambient Intelligence and Smart Environments in 2009 and has been serving as its co-editor-in-chief. He has been associated with the department of telecommunication and information processing in Gent University since 2009, Belgium, and with the department of electrical engineering at Sharif University of Technology, Iran, since 2013. His recent research interests are in the field of neuroscience with an emphasis on characterizing the brain's response to external stimuli for understanding network deficiencies caused by neurodegenerative diseases such as Alzheimer's and designing stimulation techniques for therapeutic brain entrainment.

George Bebis received the B.S. degree in mathematics and the M.S. degree in computer science from the University of Crete, Crete, Greece, in 1987 and 1991, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Central Florida, Orlando, FL, USA, in 1996. He is currently a Foundation Professor with the Department of Computer Science and Engineering (CSE), University of Nevada, Reno (UNR), Reno, NV, USA, and the Director of the Computer Vision Laboratory. From 2013 to 2018, he served as a Department Chair of CSE, UNR. His research has been funded by NSF, NASA, ONR, NIJ, and Ford Motor Company. His research interests include computer vision, image processing, pattern recognition, machine learning, and evolutionary computing. Dr. Bebis is an Associate Editor of the Machine Vision and Applications Journal and serves on the Editorial Board of the International Journal on Artificial Intelligence Tools, and the Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization. He has served on the program committees of various national and international conferences and the Founder/main Organizer of the International Symposium on Visual Computing (ISVC) and the International Symposium on Mathematical and Computational Oncology (ISMCO).

Shaogang Gong received the DPhil degree in computer vision from Keble College, Oxford University, in 1989. He has been the Professor of Visual Computation with the Queen Mary University of London since 2001. He is an elected fellow of the Institution of Electrical Engineers, a fellow of the British Computer Society, a member of the UK Computing Research Committee, and a Turing Fellow of the Alan Turing Institute. He served on the Steering Panel of the UK Government Chief Scientific Advisor's Science Review. His research interests include computer vision, machine learning, and video analysis. He has published over 400 research papers and 7 books on topics including Person Re-Identification, Visual Analysis of Behaviour, Video Analytics for Business Intelligence, Dynamic Vision from Images to Face Recognition, Analysis and Modelling of Faces and Gestures. He is the inventor of 42 international patents. He won the Institution of Engineering and Technology 2020 Achievement Medal for Vision Engineering, for outstanding achievement and superior performance in contributing to public safety.

Amy Loutfi is a Professor in Information Technology at Örebro University and leads the research group - AASS Machine Perception and Interaction Lab. Amy Loutfi received her Ph.D. in Computer Science in 2006 with a research topic in machine perception. Specifically, she researched about how gas sensors could be integrated onto robotic platforms and how these robots can interact with humans in order to solve a range of problems that required sensing and perception. She has since broadened her research interests to include general research directions within machine perception, where AI methods like Machine learning are used for the interpretation of sensor data. She also has broadened her research in the area of Human Robot Interaction where she has studied HRI in various platforms that include fully autonomous robots, but also teleoperated robots. She has a long experience working with industry and the public sector on research projects dealing with AI, robotics and human-robot interaction.

Leonid Sigal is an Associate Professor in the Department of Computer Science at the University of British Columbia. He is also a Canada Research Chair (CRC II) in Computer Vision and Machine Learning and a remote Faculty Member of the Vector Institute for AI in Toronto. In addition, he serves as an Academic Advisor to Borealis AI. His research focuses on problems of visual understanding and reasoning. This includes object recognition, scene understanding, articulated motion capture, motion modeling, action recognition, motion perception, manifold learning, transfer learning, character and cloth animation and a number of other directions on the intersection of computer vision, machine learning, and computer graphics

Wei-Shi Zheng Dr. Wei-Shi Zheng is now a Professor with Sun Yat-sen University. His research interests include person/object association and activity understanding in visual surveillance, and the related large-scale machine learning algorithm. He has ever served as area chairs of ICCV, CVPR, BMVC, IJCAI and AAAI. He is an IEEE MSA TC member. He is an associate editor of Pattern Recognition. He has ever joined Microsoft Research Asia Young Faculty Visiting Programme. He is a recipient of Excellent Young Scientists Fund of the National Natural Science Foundation of China, and a recipient of Royal Society-Newton Advanced Fellowship of United Kingdom.