Guest Editorial
Special Issue on Wearable and Ego-Vision Systems for Augmented Experience

RAPID progress in the development of low-level component technologies such as wearable sensors, wearable displays, and wearable computers is making our digital lives grow, connect, and play a relevant role in reality. To name a few examples, body-mounted sensors and displays help athletes in training by presenting real-time performance metrics such as speed, distance, and heart rate. Wearable systems allow medical staff in hospitals to consult specialists located anywhere in the world, in real time, providing optimal patient care. And within the context of assistive technologies, a head-mounted camera can be used to identify and convey the presence of objects, people, or text to a visually impaired user. Given the huge potential of wearable technologies, high-tech companies have hastened to release devices, such as the Google Glasses, Microsoft Hololens, Fitbit Flex, Facebook's Oculus, and Epson's Moverio glasses.

These recent technological developments have introduced considerable challenges and opened new lines. Most wearable and egocentric vision systems generate multimodal data streams (e.g., video, audio, motion, and emotions). Motivated by the processing of heterogeneous data, relatively young research communities such as Computer Vision, Multimedia, Augmented Virtual Reality, Human Computer Interaction, and Pervasive Computing are permanently developing new ideas to understand human activities, enhance our capabilities, and augment perception.

Wearable and egocentric data are unstructured and continuous, without evident boundaries, and most of the time datasets are of huge size. Information such as relevant events and personal experiences is not easy to retrieve. There is a clear need for automated tools that assist users in accessing long lasting stream of information. Regarding recognition and segmentation of human daily activities, several research studies have explored object detectors, video motion, and biomechanical features [items 1), 2), and 3) in the Appendix]. Moreover, methods that produce a compact summary of a day of the wearer have been suggested. They predict involvement or attention by detecting the most salient objects, people with whom the camera wearer interacts [items 4) and 5) in the Appendix], or by investigating physiological data such as brain waves [item 6) in the Appendix] and gaze [item 7) in the Appendix]. Egocentric vision systems can also provide insight into social interaction by estimating, for example, the three-dimensional (3-D) position and gaze direction of faces around a recorder [item 8) in the Appendix].

Tactile sensors on the skin have also been explored to communicate emotions, thus augmenting the user perception of selected experiences. Researchers have presented techniques to generate illusory tactile sensations by electrodermal activity processing and vibration feedback devices [items 9), 10), and 11) in the Appendix].

Recent advances in head-mounted displays opened doors to applications that use bare hands for augmented experiences, particularly those related to interaction with virtual objects. Methods have been presented for the estimation of the position of fingertips or palm center [item 12) in the Appendix], hand segmentation [item 13) in the Appendix], and gesture recognition [item 14) in the Appendix]. [items 15), 16), and 17) in the Appendix] present comprehensive reviews on such topics.

Although we have witnessed impressive progress in several specific applications, our opinion is that the field is only at its beginning. More impressive developments will come in the short and mid-term, and we believe that the most impressive will be ones we cannot imagine now. Based on the multidisciplinary nature of this field, one of the next steps is calling for a converged effort of communities such as psychology, cognitive science, and computer science to understand and identify emerging challenges and opportunities of wearable devices augment human lives.

This Special Issue is one of the first attempts to gather recent advances of different research communities in enhancing human performance through egocentric sensing. It contains 13 papers that we have divided into three different groups, according to their content. The first part of the Special Issue presents several relevant works on wearable and egocentric vision systems aiming to scene understanding and summarize user interaction with the world.

In item 18) of the Appendix, the authors study the recognition of personal locations from egocentric videos. Contextual awareness is a relevant factor in many wearable computing applications, and location stands out among the different context features. GPS positioning is not enough, as it only provides outdoor locations. The authors suggest a multiclass classification algorithm followed by an entropy-based negative rejection mechanism to shape temporal coherence. Their experimental results show deep convolutional networks outperform more traditional classification schemes and that most of the false positives are due to negative classes, hence the negative rejection scheme is of key importance.

Item 19) of the Appendix presents a smartwatch platform based on an ultralow power heterogeneous system. It consists

Digital Object Identifier 10.1109/THMS.2016.2646600

of a TI MPS430 microcontroller and a set of four different sensors: camera, microphone, accelerometer, and thermistor. The proposed hardware enables the implementation of complex vision algorithms based on convolutional neural network. In particular, they present a visual context classification technique to infer whether the smartwatch user is in one of the following contexts: *morning preparation*, *walking outdoors*, *public transportation*, *in the car*, and *in the office*.

Item 20) of the Appendix presents an approach for fall detection based on a wearable camera that can be applied for monitoring the activities of elderly people; avoiding a more expensive and complex setup based on static sensors covering home spaces. It uses a modified version of histogram of oriented gradients (HOG) features and gradient local binary patterns. The threshold of the fall detection is then learned by a training set using a relative entropy approach which is a member of Ali-Silvery distance measures.

Item 21) of the Appendix proposes an approach for a multimodal target identification task relevant to egocentric applications with wearable audio–visual devices. The authors define time-dependent target models for each modality, which are adapted in an unsupervised fashion using the complementary of the other modality when a new observation is available. Experiments on two real challenging audio–visual datasets show the robustness of their solution even in presence of mild mismatches.

Item 22) of the Appendix introduces a real-time face recognition solution to assist visually impaired individuals. The system uses a Microsoft Kinect sensor to acquire RGB-D images. The author solution uses a variation of the K-nearest neighbors algorithm over HOG descriptors dimensionally reduced by principal component analysis. Once the system recognizes a face, it uses its location to generate 3-D audio signals coming from a face location.

Item 23) of the Appendix provides a comprehensive survey of the state of the art on egocentric video summarization. First, the authors illustrate a wide range of applications that may benefit from a robust summarization. They describe the specific differences between first-person view videos and third-person videos. The authors finally compare all relevant approaches, highlighting their main strengths and limitations and present the datasets and evaluation procedures.

Item 24) of the Appendix presents an overview of the state of visual life-logging analysis organizing the literature around the key task-related question, including: was the user interacting with somebody?, how?, where was he/she?, when did the event occur?, and, what was the person wearing the camera doing? The review presents relevant solutions for acquiring, organizing, summarizing, and browsing large collection of egocentric video data.

The second part of the Special Issue presents wearable systems posing stimuli to the human body that can be used to understand a user's internal state.

Item 25) of the Appendix presents a study about the effects of affective haptic stimuli on autonomic nervous systems. In particular, the authors suggest a convex optimization technique to automatically determine force and velocity of caressing stimuli estimated through the analysis of the Electrodermal Activity (EDA). A solid experimental evaluation is presented on a wearable haptic system, which conveyed to subjects caress-like stimuli by means of two motors.

Item 26) of the Appendix examines whether and how out-of-body illusions (e.g., through user-held external objects) can be extended to improve the feedback experience of a user who is interacting with augmented virtual objects. The authors focus their study on the most notable out-of-body tactile illusions, funneling, and saltation, in order to minimize the number of tactile actuators required.

The third part of the Special Issue is dedicated to techniques and interfaces for controlling and interacting with virtual and augmented realities.

Item 27) of the Appendix presents a natural user interface framework for a virtual reality environment, called Meta-Gesture, which can simultaneously recognize static and dynamic gestures for holding and manipulating the tools of daily use. The authors suggest an approach that combines a novel voxel coding schema, called Layered Shape Pattern, and a variation of Random Forest classifier.

Item 28) of the Appendix presents *TunnelSlice* a freehand interaction system to select and attain a distant 3-D region specifically designed for head-mounted displays. In particular, the authors present a new slicing solution to determine the cuboid transform via a pinch-tip gesture. TunnelSlide was tested in six scenarios involving central objects status and different levels of occlusions showing better performance than recent and related techniques.

Item 29) of the Appendix proposes an augmented navigation system designed to assist people in indoor environments. During navigation, the route was presented to participants via visual and auditory channels. The proposed solution can be installed in both head-mounted displays and smartphones. The paper presents comparative tests between these two settings analyzing human factors and usability issues such as the feasibility and reliability of the system, subjective comfort, perceived accuracy, and navigation time.

Item 30) of the Appendix presents a framework to support the automatic generation of speech interfaces for controlling virtual and augmented reality applications. The generated interface allows the user to activated functionalities, by a set of voice commands, which are also represented visually in a head-mounted display. An extensive evaluation is presented by comparing the automatically generated interface with a user-generated interfaces in terms of the impact of visual cues, semantic processing, and first-time user experience.

Acknowledgment

The Guest Editors would like to thank all the authors for their contributions and all the reviewers who took part in the evaluation of manuscripts, and provided insights which resulted in a solid Special Issue on Wearable and Ego-Vision Systems for Augmented Experience. The Guest Editors would also like to thank Prof. D. Kaber, Editor-in-Chief, and Prof. E. J. Bass, previous Editor-in-Chief, for giving us the chance to organize this Special Issue and for the exceptional support and effort they provided throughout the entire process.

GIUSEPPE SERRA, *Guest Editor*
Department of Mathematics,
Computer Science and Physics
University of Udine
33100 Udine, Italy

RITA CUCCHIARA, *Guest Editor*
Department of Engineering "Enzo Ferrari"
University of Modena and Reggio Emilia
41121 Modena, Italy

KRIS M. KITANI, *Guest Editor*
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213 USA

JAVIER CIVERA, *Guest Editor*
Department of Computer Science and
System Engineering
University of Zaragoza
50009 Zaragoza, Spain

APPENDIX
RELATED WORK

1) Y. Poleg, C. Arora, and S. Peleg, "Temporal segmentation of egocentric videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2537–2544.
2) H. Pirsiavash and D. Ramanan, "Detecting activities of daily living in first-person camera views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2847–2854.
3) S. Alletto, G. Serra, and R. Cucchiara, "Motion segmentation using visual and bio-mechanical features," in *Proc. ACM Multimedia Conf.*, 2016, pp. 476–480.
4) Z. Lu and K. Grauman, "Story-driven summarization for egocentric video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2714–2721.
5) Y. Su and K. Grauman, "Detecting engagement in egocentric video," in *Proc. Eur. Conf. Comput. Vis.*, 2016.
6) H. W. Ng, Y. Sawahata, and K. Aizawa, "Summarization of wearable videos using support vector machine," in *Proc. IEEE Int. Conf. Multimedia Expo.*, 2002, pp. 325–328.
7) J. Xu, L. Mukherjee, Y. Li, J. Warner, J. M. Rehg, and V. Singh, "Gazeenabled egocentric video summarization via constrained submodular maximization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 2235–2244.
8) A. Fathi, J. Hodgins, and J. Rehg, "Social interactions: A first-person perspective," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1226–1233.
9) M. P. Kilgard and M. M. Merzenich, "Anticipated stimuli across skin," *Nature*, vol. 373, 1995, Art. no. 663.
10) R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affective Comput.*, vol. 1, no. 1, pp. 18–37, Jan. 2010.
11) C. Pacchierotti, D. Prattichizzo, and K. J. Kuchenbecker, "Displaying sensed tactile cues with a fingertip haptic device," *IEEE Trans. Haptics*, vol. 8, no. 4, pp. 384–396, Oct. 2015.
12) T. Ha, S. Feiner, and W. Woo, "Wearhand: Head-worn, RGB-D camerabased, bare-hand user interface with visually enhanced depth perception," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2014, pp. 219–228.
13) C. Li and K. M. Kitani, "Pixel-level hand detection in ego-centric videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 3570–3577.
14) H. Cheng, L. Yang, and Z. Liu, "Survey on 3D hand gesture recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1659–1673, Sep. 2016.
15) A. Betancourt, P. Morerio, C. S. Regazzoni, and M. Rauterberg, "The evolution of first person vision methods: A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 744–760, May 2015.
16) M. A. Eid and H. A. Osman, "Affective haptics: Current research and future directions," *IEEE Access*, vol. 4, pp. 26–40, 2016.
17) S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, Jan. 2015.
18) A. Furnari, G. M. Farinella, and S. Battiato, "Recognizing personal locations from egocentric videos," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 6–18, Feb. 2017.
19) F. Conti, D. Palossi, R. Andri, M. Magno, and L. Benini, "Accelerated visual context classication on a low-power smartwatch," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 19–30, Feb. 2017.
20) K. Ozcan, S. Velipasalar, and P. K. Varshney, "Autonomous fall detection with wearable cameras by using relative entropy distance measure," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 31–39, Feb. 2017.
21) A. Brutti and A. Cavallaro "Online cross-modal adaptation for audio–visual person identification with wearable cameras," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 40–51, Feb. 2017.
22) L. B. Neto *et al.*, "A Kinect-based wearable face recognition system to aid visually impaired users," *IEEE Trans. Human-Machine Syst.*, vol. 47, no. 1, pp. 52–64, Feb. 2017.
23) A. G. del Molino, C. Tan, J-H. Lim, and A-H. Tan, "Summarization of egocentric videos: A comprehensive survey," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 65–76, Feb. 2017.
24) M. Bolaños, M. Dimiccoli, and P. Radeva, "Towards storytelling from visual lifelogging: An overview," *IEEE Transactions on Human-Mach. Syst.*, vol. 47, no. 1, pp. 77–90, Feb. 2017.
25) A. Greco, G. Valenza, M. Nardelli, M. Bianchi, L. Citi, and E. P. Scilingo, "Force–velocity assessment of caress-like stimuli through the electrodermal activity processing: Advantages of a convex optimization approach," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 91–100, Feb. 2017.

26) J. Lee, Y. Kim, and G. J. Kim, "Effects of visual feedback on out-of-body illusory tactile sensation when interacting with augmented virtual objects," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 101–112, Feb. 2017.
27) Y. Jang, I. Jeon, T.-K. Kim, and W. Woo, "Metaphoric hand gestures for orientation-aware VR object manipulation with an egocentric viewpoint," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 113–127, Feb. 2017.
28) H. Lee, S.-T. Noh, and W. Woo, "TunnelSlice: Freehand subspace acquisition using an egocentric tunnel for wearable augmented reality," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 128–139, Feb. 2017.
29) U. Rehman and S. Cao, "Augmented-reality-based indoor navigation: A comparative analysis of handheld devices versus Google Glass," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 140–151, Feb. 2017.
30) F. Lamberti, F. Manuri, G. Paravati, G. Piumatti, and A. Sanna, "Using semantics to automatically generate speech interfaces for wearable virtual and augmented reality applications," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 152–164, Feb. 2017.

Giuseppe Serra received the Ph.D. degree in computer engineering, multimedia, and telecommunications from the University of Florence, Florence, Italy, in 2010.

He is currently with the University of Udine, Udine, Italy. He was a Visiting Scholar at Carnegie Mellon University, Pittsburgh, PA, USA, and at Telecom ParisTech/ENST, Paris, France, in 2006 and 2010, respectively. His research interests include egocentric vision, image and video analysis, multimedia ontologies, and multiple view geometry. He has published more than 40 publications in scientific journals and international conference proceedings.

Dr. Serra received the Best Paper Award at the IEEE International Conference on Intelligent Technologies for Interactive Entertainment in 2015 with the paper "Wearable Vision for Retrieving Architectural Details in Augmented Tourist Experiences." He was a Technical Program Committee member of several workshops and conferences. He regularly serves as a Reviewer for international conferences and journals such as CVPR, ECCV, ACM Multimedia, and IEEE Transactions on Multimedia, and IEEE Transactions on Information Forensics and Security and Pattern Recognition.

Rita Cucchiara is currently a Full Professor at the University of Modena and Reggio Emilia, Modena, Italy, where she is the Head of the Imagelab Laboratory and the Director of the Research Center in ICT SOFTECH-ICT. Her research interests include wearable and egocentric vision, computer vision, multimedia, and pattern recognition. She has been the Coordinator of several projects on vision and video surveillance, mainly for people detection, tracking, and behavior analysis. In the field of multimedia, she works on mobile vision, annotation, retrieval, and human-centered searching in images and video big data for cultural heritage. She is the author of more than 200 papers published in journals and international conference proceedings, and is a Reviewer for several international journals.

Prof. Cucchiara has been a Fellow of ICPR since 2006. She is a member of the Editorial Boards of *Multimedia Tools and Applications* and *Machine Vision and Applications* and the Chair of several workshops and conferences. She served as the Workshop Chair for ACM MM 2010, the Demo Chair for ECCV 2012, the Tutorial Chair for ECCV 2016, the Track Chair for ICPR 2012, the Area Chair for ACMMM 2013, and the Area Chair for CVPR 2014 and 2015. She is a member of the IEEE Computer Society, ACM, GIRPR, and AI*IA. Since 2006, she has been a Fellow of IAPR.

Kris M. Kitani received the B.S. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1999, and the M.S. and Ph.D. degrees in information science and technology from The University of Tokyo, Tokyo, Japan, in 2005 and 2008, respectively.

He is currently an Assistant Research Professor at the Robotics Institute, Carnegie Mellon University (CMU), Pittsburgh, PA, USA, working in the area of computer vision. During his time at CMU, he has worked primarily in the area of first-person vision and human activity forecasting.

Javier Civera received the Ph.D. degree from the University of Zaragoza, Zaragoza, Spain, in 2009.

He is currently an Associate Professor at the University of Zaragoza, teaching computer vision, control, and machine learning courses. He has participated in several EU-funded, national and technology transfer projects related to vision and robotics and has been funded for research visits to Imperial College (London, U.K.) and ETH (Zürich, Switzerland). He has coauthored about 30 publications published in top conference proceedings and journals, receiving about 2200 references (GoogleScholar). His research interests include use of 3-D vision, cloud architectures and learning algorithms to produce robust and real-time vision technologies for robotics, wearables, and AR applications.

Dr. Civera has been an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING since 2015, and for the IEEE ROBOTICS AND AUTOMATION LETTERS since 2016. From 2014 to 2016, he was an Associate Editor for the IEEE/RSJ International Conference on Intelligent Robots and Systems.