VISIGRAPP 2019

14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications

PROCEEDINGS

Volume 3: IVAPP

Prague, Czech Republic

25-27 February, 2019

EDITORS
Andreas Kerren
Christophe Hurter
Jose Braz

http://www.visigrapp.org/

SPONSORED BY

INSTICC

PAPERS AVAILABLE AT

SCITEPRESS DIGITAL LIBRARY

VISIGRAPP 2019

Proceedings of the
14th International Joint Conference on
Computer Vision, Imaging and Computer Graphics
Theory and Applications

Volume 3: IVAPP

Prague - Czech Republic

February 25 - 27, 2019

Sponsored by
INSTICC - Institute for Systems and Technologies of Information, Control and Communication

In Cooperation with
EUROGRAPHICS - European Association for Computer Graphics
AFIG - French Association for Computer Graphics

Edited by Andreas Kerren, Christophe Hurter and Jose Braz

Printed in Portugal
ISSN: 2184-4321
ISBN: 978-989-758-354-4
Depósito Legal: 450998/19

http://www.ivapp.visigrapp.org
ivapp.secretariat@insticc.org

BRIEF CONTENTS

INVITED SPEAKERS

Daniel McDuff
Microsoft
United States

Diego Gutierrez
Universidad de Zaragoza
Spain

Jiri Matas
Czech Technical University in Prague, Faculty of Electrical Engineering
Czech Republic

Dima Damen
University of Bristol
United Kingdom

ORGANIZING COMMITTEES

CONFERENCE CHAIR

Jose Braz, Escola Superior de Tecnologia de Setúbal, Portugal

PROGRAM CO-CHAIRS

Andreas Kerren, Linnaeus University, Sweden
Christophe Hurter, French Civil Aviation University (ENAC), France

SECRETARIAT

Bruno Encarnação, INSTICC, Portugal
Ana Beatriz Filipe, INSTICC, Portugal

GRAPHICS PRODUCTION AND WEBDESIGNER

André Poeira, INSTICC, Portugal

WEBMASTER

João Francisco, INSTICC, Portugal
Carolina Ribeiro, INSTICC, Portugal

PROGRAM COMMITTEE

Vladan Babovic, National University of Singapore, Singapore

Juhee Bae, University of Skovde, Sweden

Till Bergmann, Max Planck Society, Germany

David Borland, University of North Carolina at Chapel Hill, United States

Massimo Brescia, Istituto Nazionale di AstroFisica, Italy

Ross Brown, Queensland University of Technology, Brisbane, Australia

Maria Beatriz Carmo, BioISI, Faculdade de Ciências da Universidade de Lisboa, Portugal

Guoning Chen, University of Houston, United States

Yongwan Chun, University of Texas at Dallas, United States

António Coelho, INESC-TEC, Faculdade de Engenharia, Universidade do Porto, Portugal

Danilo B. Coimbra, Federal University of Bahia, Brazil

Christoph Dalitz, Niederrhein University of Applied Sciences, Germany

Mihaela Dinsoreanu, Technical University of Cluj-Napoca, Romania

Georgios Dounias, University of the Aegean, Greece

Achim Ebert, University of Kaiserslautern, Germany

Danilo Eler, São Paulo State University, Brazil

Maria Ferreira De Oliveira, University of São Paulo, ICMC, Brazil

Chi-Wing Fu, The Chinese University of Hong Kong, Hong Kong

Mohammad Ghoniem, Luxembourg Institute of Science and Technology, Luxembourg

Randy Goebel, University of Alberta, Canada

Martin Graham, University of Edinburgh, United Kingdom

Naeemul Hassan, University of Mississippi, United States

Torsten Hopp, Karlsruhe Institute of Technology, Germany

Jie Hua, University of Technology Sydney, Australia

Jusufi Ilir, Linnaeus University, Sweden

Stefan Jänicke, Leipzig University, Germany

Mark Jones, Swansea University, United Kingdom

Bijaya Karki, Louisiana State University, United States

Jörn Kohlhammer, Fraunhofer Institute for Computer Graphics Research, Germany

Martin Kraus, Aalborg University, Denmark

Haim Levkowitz, University of Massachusetts, Lowell, United States

Chun-Cheng Lin, National Chiao Tung University, Taiwan

Giuseppe Liotta, University of Perugia, Italy

Rafael Martins, Linnaeus University, Sweden

Krešimir Matkovic, VRVis Research Center, Austria

Kazuo Misue, University of Tsukuba, Japan

Steffen Oeltze-Jafra, Otto-von-Guericke-University Magdeburg, Germany

Benoît Otjacques, Luxembourg Institute of Science and Technology (LIST), Luxembourg

Jinah Park, KAIST, Korea, Republic of

Fernando Paulovich, Dalhousie University, Canada

Torsten Reiners, Curtin University, Australia

Philip Rhodes, University of Mississippi, United States

Patrick Riehmann, Bauhaus-Universitaet Weimar, Germany

Maria Riveiro, University of Skövde, Sweden

Adrian Rusu, Fairfield University, United States

Filip Sadlo, Heidelberg University, Germany

Beatriz Santos, University of Aveiro, Portugal

Giuseppe Santucci, University of Roma, Italy

Angel Sappa, ESPOL Polytechnic University (Ecuador) and Computer Vision Center (Spain), Spain

Falk Schreiber, University of Konstanz, Germany and Monash University Melbourne, Germany

Hans-Jörg Schulz, Aarhus University, Denmark

Celmar Silva, University of Campinas, Brazil

Juergen Symanzik, Utah State University, United States

Sidharth Thakur, Renaissance Computing Institute (RENCI), United States

Roberto Theron, Universidad de Salamanca, Spain

Günter Wallner, University of Applied Arts Vienna, Austria

Jinrong Xie, eBay Inc, United States

Anders Ynnerman, Linköping University, Sweden

Hongfeng Yu, University of Nebraska - Lincoln, United States

Lina Yu, Intel Corporation, United States

Jianping Zeng, Microsoft, United States

Yue Zhang, Oregon State University, United States

Jianmin Zheng, Nanyang Technological University, Singapore

AUXILIARY REVIEWERS

Aris Alisandrakis, Linnaeus University, Sweden

Xiaopei Liu, ShanghaiTech University, China

Benjamin Mora, Swansea University, United Kingdom

Wei Zeng, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

SELECTED PAPERS BOOK

A number of selected papers presented at IVAPP 2019 will be published by Springer in a CCIS Series book. This selection will be done by the Conference Chair and Program Co-chairs, among the papers actually presented at the conference, based on a rigorous review by the IVAPP 2019 Program Committee members.

FOREWORD

This book contains the proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019) which was organized and sponsored by the Institute for Systems and Technologies of Information, Control and Communication (INSTICC), in cooperation with ACM SIGCHI, ACM SIGGRAPH, AFIG, Eurographics and UXPA International.

The proceedings here published demonstrate new and innovative solutions and highlight technical problems in each field that are challenging and worthwhile being disseminated to the interested research audiences.

VISIGRAPP 2019 was organized to promote a discussion forum about the conference's research topics between researchers, developers, manufacturers and end-users, and to establish guidelines in the development of more advanced solutions.

We received a high number of paper submissions for this edition of VISIGRAPP, 396 in total, with contributions from all five continents. This attests to the success and global dimension of VISIGRAPP. To evaluate each submission, we used a double-blind evaluation method where each paper was reviewed by two to six experts from the International Program Committee (IPC).

The IPC selected for oral presentation and for publication as full papers 12 papers from GRAPP, 6 for HUCAPP, 12 papers for IVAPP, and 36 papers for VISAPP, which led to a result for the full-paper acceptance ratio of 17% and a high-quality program. Apart from the above full papers, the conference program also features 88 short papers and 115 poster presentations. We hope that these conference proceedings, which are submitted for indexation by Thomson Reuters Conference Proceedings Citation Index, SCOPUS, DBLP, Semantic Scholar, Google Scholar and EI, will help the Computer Vision, Imaging, Visualization and Computer Graphics communities to find interesting research work. Moreover, we are proud to inform that the program also includes four plenary keynote lectures, given by internationally distinguished researchers, namely Daniel McDuff (Microsoft, United States), Diego Gutierrez (Universidad de Zaragoza, Spain), Jiri Matas (Czech Technical University in Prague, Faculty of Electrical Engineering, Czech Republic) and Dima Damen (University of Bristol, United Kingdom), thus contributing to increase the overall quality of the conference and to provide a deeper understanding of the conference's interest fields.

Furthermore, a short list of the presented papers will be selected to be expanded into a forthcoming book of VISIGRAPP Selected Papers to be published by Springer during 2019 in the CCIS series. Also, a short list of presented papers will be selected for publication of extended and revised versions in a special issue of the Open Access Information Science Journal (IVAPP) and in a special issue of the Pattern Recognition and Artificial Intelligence Journal (VISAPP). All papers presented at this conference will be available at the SCITEPRESS Digital Library. Three awards are delivered at the closing session, to recognize the best conference paper, the best student paper and the best poster for each of the four conferences.

We would like to express our thanks, first of all, to the authors of the technical papers, whose work and dedication made possible to put together a program that we believe to be very exciting and of high technical quality. Next, we would like to thank the Area Chairs, all the members of the program committee and auxiliary reviewers, who helped us with their expertise and time. We would also like to thank the invited speakers for their invaluable contribution and for sharing their vision in their talks. Special thanks should be addressed to the INSTICC Steering Committee whose invaluable work made this event possible.

We wish you all an exciting conference and an unforgettable stay in Prague, Czech Republic. We hope to meet you again for the next edition of VISIGRAPP, details of which are available at http://www.visigrapp.org.

Andreas Kerren
Linnaeus University, Sweden

Christophe Hurter
French Civil Aviation University (ENAC), France

Jose Braz
Escola Superior de Tecnologia de Setúbal, Portugal

CONTENTS

INVITED SPEAKERS

KEYNOTE SPEAKERS

Building Emotionally Intelligent AI: From Sensing to Synthesis

Daniel McDuff
Microsoft, United States

Abstract: Emotions play an important role in our everyday lives. They influence memory, decision-making and well-being. In order to advance the fundamental understanding of human emotions, build smarter affective technology, and ultimately help people, we need to perform research in-situ. Leveraging exciting advances in machine learning and computer vision it is now possible to quantify emotional and physiological responses in new ways and on a large scale using webcams and microphones in everyday environments. I will present novel methods for physiological and behavioral measurement via ubiquitous hardware and show data from the largest longitudinal data collection of its kind. Then I will present state-of-the-art approaches for emotion synthesis (both audio and visual) that can be used to create rich human-agent or robot interactions. Finally, I will show examples of new human-computer interfaces that leverage behavioral and physiological signals, including emotion-aware natural conversational systems.

BRIEF BIOGRAPHY

Daniel McDuff (http://alumni.media.mit.edu/ djmcduff/) is a Researcher at Microsoft where he leads research and development of affective computing technology, with a focus on scalable tools to enable the automated recognition and analysis of emotions and physiology. He is also a visiting scientist at Brigham and Women's Hospital in Boston where he works on deploying these methods in primary care and surgical applications. Daniel completed his PhD in the Affective Computing Group at the MIT Media Lab in 2014 and has a B.A. and Masters from Cambridge University. Previously, Daniel was Director of Research at Affectiva and a post-doctoral research affiliate at the MIT Media Lab. During his Ph.D. and at Affectiva he built state-of-the-art facial expression recognition software and lead analysis of the world's largest database of facial expression videos. His work in machine learning, AR and affective computing has received nominations and awards from Popular Science magazine as one of the top inventions in 2011, South-by-South-West Interactive (SXSWi), The Webby Awards, ESOMAR and the Center for Integrated Medicine and Innovative Technology (CIMIT). His projects have been reported in many publications including The Times, the New York Times, The Wall Street Journal, BBC News, New Scientist, Scientific American and Forbes magazine. Daniel was named a 2015 WIRED Innovation Fellow and has spoken at TEDx Berlin and SXSW. We is a member of the ACM Future of Computing Academy.

McDuff, D.
Building Emotionally Intelligent AI: From Sensing to Synthesis.
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019) - Volume 3: IVAPP*, page 5
ISBN: 978-989-758-354-4

Reinventing Movies: How Do We Tell Stories in VR?

Diego Gutierrez
Universidad de Zaragoza, Zaragoza, Spain

Abstract: Traditional cinematography has relied for over a century on a well-established set of editing rules, called continuity editing, to create a sense of situational continuity. Despite massive changes in visual content across cuts, viewers in general experience no trouble perceiving the discontinuous flow of information as a coherent set of events. However, Virtual Reality (VR) movies are intrinsically different from traditional movies in that the viewer controls the camera orientation at all times. As a consequence, common editing techniques that rely on camera orientations, zooms, etc., cannot be used. In this talk we will investigate key relevant questions to understand how well traditional movie editing carries over to VR, such as: Does the perception of continuity hold across edit boundaries? Under which conditions? Do viewers' observational behavior change after the cuts? We will make connections with recent cognition studies and the event segmentation theory, which states that our brains segment continuous actions into a series of discrete, meaningful events. This theory may in principle explain why traditional movie editing has been working so wonderfully, and thus may hold the answers to redesigning movie cuts in VR as well. In addition, and related to the general question of how people explore immersive virtual environments, we will present the main insights a second, recent study, analyzing almost 2000 head and gaze trajectories when users explore stereoscopic omni-directional panoramas. We have made our database publicly available for other researchers.

BRIEF BIOGRAPHY

Diego Gutierrez is a Professor at the Universidad de Zaragoza in Spain, where he leads the Graphics and Imaging Lab. His areas of interests include physically based global illumination, perception, computational imaging, and virtual reality. He's published many papers in top journals and conferences, being Papers Chair of Eurographics (2018), the Rendering Symposium (2012), and the Symposium on Applied Perception (2011). He's been Editor in Chief of ACM Transactions on Applied Perception, and is currently an Associate Editor of four other journals. He has received many awards including a Google Faculty Research Award in 2014, or an ERC Consolidator Grant in 2016.

Gutierrez, D.
Reinventing Movies: How Do We Tell Stories in VR?.
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019) - Volume 3: IVAPP*, page 7
ISBN: 978-989-758-354-4

Robust Fitting of Multiple Models in Computer Vision

Jiri Matas
Czech Technical University in Prague, Faculty of Electrical Engineering, Czech Republic

Abstract: Many computer vision problems can be formulated seen as multi-class multi-instance fitting, where the input data is interpreted as a mixture of noise and observations originating from multiple instances of multiple model types, e.g. as lines and circles in edge maps; as planes, cylinders and point clusters in 3D laser scans; as multiple homographies or fundamental matrices consistent with point correspondences in multiple views of a non-rigid scene. I will review properties of three popular data fitting methods - RANSAC, the Hough transform and Isack's and Boykov's PEARL, which disposes of the assumption of independent data errors. I will then present a novel method, called Multi-X, for general multi-class multi-instance model fitting. The proposed approach combines a random sampling strategy like RANSAC, lobal energy minimization using alpha-expansion as PEARL, and mode-seeking in the parameter domain like the Hough Transform. Multi-X outperforms significantly the state-of-the-art on standard datasets, runs in time approximately linear in the number of data points, an order of magnitude faster than available implementations of commonly used methods.

BRIEF BIOGRAPHY

Jiri Matas is a full professor at the Center for Machine Perception, Czech Technical University in Prague. He holds a PhD degree from the University of Surrey, UK (1995). He has published more than 200 papers in refereed journals and conferences. His publications have approximately 34000 citations registered in Google Scholar and 13000 in the Web of Science. His h-index is 65 (Google scholar) and 43 (Clarivate Analytics Web of Science) respectively. He received the best paper prize e.g. at the British Machine Vision Conferences in 2002 and 2005, at the Asian Conference on Computer Vision in 2007 and at Int. Conf. on Document analysis and Recognition in 2015. J. Matas has served in various roles at major international computer vision conferences (e.g. ICCV, CVPR, ICPR, NIPS, ECCV), co-chairing ECCV 2004, 2016 and CVPR 2007. He is on the editorial board of IJCV and was the Associate Editor-in-Chief of IEEE T. PAMI. He served on the computer science panel of ERC. His research interests include visual tracking, object recognition, image matching and retrieval, sequential pattern recognition, and RANSAC- type optimization metods.

Matas, J.
Robust Fitting of Multiple Models in Computer Vision.
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019) - Volume 3: IVAPP*, page 9
ISBN: 978-989-758-354-4

A Fine-grained Perspective onto Object Interactions from First-person Views

Dima Damen[a]
Department of Computer Science, University of Bristol, Bristol, U.K.
Dima.Damen@bristol.ac.uk

Keywords: Egocentric Vision, First-Person Vision, Action Recognition, Fine-Grained Recognition, Object Interaction Recognition, Skill Determination, Action Completion, Action Anticipation, Wearable Cameras, First-Person Datasets, EPIC-Kitchens.

Abstract: This extended abstract summarises the relevant works to the keynote lecture at VISAPP 2019. The talk discusses understanding object interactions from wearable cameras, focusing on fine-grained understanding of interactions on realistic unbalanced datasets recorded in-the-wild.

1 INTRODUCTION

Humans interact with tens of objects daily, at home (e.g. cooking/cleaning), during working (e.g. assembly/machinery) or leisure hours (e.g. playing/sports), individually or collaboratively. The field of research, within computer vision and machine learning, that focuses on the perception of object interactions from a wearable cameras is commonly referred to as 'first-person vision'. In this extended abstract, we cover novel research questions, particularly related to the newly released largest dataset in object interactions, recorded in people's native environments: EPIC-Kitchens.

2 DEFINITIONS

Object interactions could be perceived from different ordinal-person viewpoints - where 'ordinal' is used to generalise between *first-*, *second-* and *third*-person views. A view is referred to as a first-person view, if the interaction is captured by a wearable sensor, worn by the actor performing the interaction itself. Conversely, a second-person view is when the interaction is captured by a camera of a co-actor, or a recepient of the action. Finally, a third-person view, common in remote static cameras, is when the interaction is captured by an observer not relevant to the interaction or the actor during that interaction.

[a] https://orcid.org/0000-0001-8804-6238

3 DATASETS AND EPIC-Kitchens

For years, Computer Vision has focused on capturing videos from a third-person view, with the majority of action recognition datasets using a remote camera observing the action or interaction (Marszalek et al., 2009; Kuehne et al., 2011; Caba Heilbron et al., 2015; Carreira and Zisserman, 2017).

Increasingly, first-person vision datasets have been recorded, capturing full body motion such as sports (Kitani et al., 2011), social interactions (Alletto et al., 2015; Fathi et al., 2012a; Ryoo and Matthies, 2013) and object interactions (De La Torre et al., 2008; Fathi et al., 2012b; Pirsiavash and Ramanan, 2012; Damen et al., 2014; Georgia Tech, 2018; Sigurdsson et al., 2018).

In 2018, the largest dataset on wearable cameras was released through a collaboration led by the University of Bristol alongside the University of Catania and the University of Toronto - http://epic-kitchens.github.io/. EPIC-Kitchens (Damen et al., 2018) offers more than 11.5M frames, captured using a head-mounted camera in 32 different kitchens, with over 55 hours of natural interactions from cooking to washing the dishes (Fig 1).

4 FINE-GRAINED OBJECT INTERACTIONS

Datasets, such as EPIC-Kitchens, can offer unique opportunities to studying previously unexplored pro-

Damen, D.
A Fine-grained Perspective onto Object Interactions from First-person Views.
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019) - Volume 3: IVAPP*, pages 11-13
ISBN: 978-989-758-354-4

Figure 1: Sample frames from EPIC-Kitchens.

blems in fine-grained object interactions. A few of these opportunities are highlighted here.

- *Overlapping Object Interactions:* Defining the temporal extent of an action is fundamentally an ambiguous problem (Moltisanti et al., 2017; Sigurdsson et al., 2017). This is usually resolved through multi-labels, i.e. allowing a time-segment to belong to multiple classes of actions. However, actual understanding of interaction overlapping requires an space of action labels that captures dependencies (e.g. filling a kettle requires opening the tap). Models that capture and predict overlapping interactions are needed for a finer-understanding of object interactions.
- *Object Interaction Completion/Incompletion:* Beyond classification and localisation, action completion/incompletion is the problem of identifying whether the action's goal has been successfully achieved, or merely attempted. This is a novel fine-grained object interaction research question proposed in (Heidarivincheh et al., 2016). This work has been recently extended to locating the moment of completion (Heidarivincheh et al., 2018) - that is the moment in time beyond which the action's goal is believed to be completed by a human observer.
- *Skill Determination from Video:* Even when an interaction is successfully completed, further understanding of 'how well' the task was completed would offer knowledge beyond pure classification. In this leading work (Doughty et al., 2018a), a collection of video could be ordered by the skill exhibited in each video, through deep pairwise ranking. This method has been recently extended to include rank-aware attention (Doughty et al., 2018b) - that is a novel loss function capable of attending to parts of the video that exhibit higher skill as well as parts that demonstrate lower skill including mistakes or hesitation.
- *Anticipation and Forecasting:* Predicting upcoming interactions has recently gathered additional attention, triggered by the presence of first-person datasets (Furnari et al., 2018; Rhinehart and Kitani, 2017). Novel research on uncertainty in anticipating actions (Furnari et al., 2018), or relating forecasting to trajectory prediction (Rhinehart and Kitani, 2017) have recently been proposed.
- *Paired Interactions:* One leading work has attempted capturing both the action and its counteraction (or reaction), both from a wearable camera (Yonetani et al., 2016). This is a very exciting area of research, still under-explored.

5 CONCLUSION

Recent deep-learning research has only scratched the surface of potentials for finer-grained understanding of object interactions. As new hardware platforms for first-person vision emerge (Microsoft's Hololens, Magic Leap, Samsung Gear, $\cdots$), applications of fine-grained recognition will be endless.

REFERENCES

Alletto, S., Serra, G., Calderara, S., and Cucchiara, R. (2015). Understanding social relationships in egocentric vision. In *Pattern Recognition*.

Caba Heilbron, F., Escorcia, V., Ghanem, B., and Carlos Niebles, J. (2015). Activitynet: A large-scale video benchmark for human activity understanding. In *CVPR*.

Carreira, J. and Zisserman, A. (2017). Quo vadis, action recognition? a new model and the kinetics dataset. In *CVPR*.

Damen, D., Doughty, H., Farinella, G. M., Fidler, S., Furnari, A., Kazakos, E., Moltisanti, D., Munro, J., Perrett, T., Price, W., and Wray, M. (2018). Scaling egocentric vision: The EPIC-KITCHENS Dataset. In *ECCV*.

Damen, D., Leelasawassuk, T., Haines, O., Calway, A., and Mayol-Cuevas, W. (2014). You-do, I-learn: Discovering task relevant objects and their modes of interaction from multi-user egocentric video. In *BMVC*.

De La Torre, F., Hodgins, J., Bargteil, A., Martin, X., Macey, J., Collado, A., and Beltran, P. (2008). Guide to the Carnegie Mellon University Multimodal Activity (CMU-MMAC) database. In *Robotics Institute*.

Doughty, H., Damen, D., and Mayol-Cuevas, W. (2018a). Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. In *CVPR*.

Doughty, H., Mayol-Cuevas, W., and Damen, D. (2018b). The Pros and Cons: Rank-aware temporal attention for skill determination in long videos. In *Arxiv*.

Fathi, A., Hodgins, J., and Rehg, J. (2012a). Social interactions: A first-person perspective. In *CVPR*.

Fathi, A., Li, Y., and Rehg, J. (2012b). Learning to recognize daily actions using gaze. In *ECCV*.

Furnari, F., Battiato, S., and Farinella, G. (2018). Leveraging uncertainty to rethink loss functions and evaluation measures for egocentric action anticipation. In *ECCVW*.

Georgia Tech (2018). Extended GTEA Gaze+. http://webshare.ipat.gatech.edu/coc-rim-wall-lab/web/yli440/egtea_gp.

Heidarivincheh, F., Mirmehdi, M., and Damen, D. (2016). Beyond action recognition: Action completion in RGB-D data. In *BMVC*.

Heidarivincheh, F., Mirmehdi, M., and Damen, D. (2018). Action completion: A temporal model for moment detection. In *BMVC*.

Kitani, K. M., Okabe, T., Sato, Y., and Sugimoto, A. (2011). Fast unsupervised ego-action learning for first-person sports videos. In *CVPR*.

Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., and Serre, T. (2011). HMDB: A large video database for human motion recognition. In *ICCV*.

Marszalek, M., Laptev, I., and Schmid, C. (2009). Actions in context. In *CVPR*.

Moltisanti, D., Wray, M., Mayol-Cuevas, W., and Damen, D. (2017). Trespassing the boundaries: Labeling temporal bounds for object interactions in egocentric video. In *ICCV*.

Pirsiavash, H. and Ramanan, D. (2012). Detecting activities of daily living in first-person camera views. In *CVPR*.

Rhinehart, N. and Kitani, K. M. (2017). First-person activity forecasting with online inverse reinforcement learning. In *ICCV*.

Ryoo, M. S. and Matthies, L. (2013). First-person activity recognition: What are they doing to me? In *CVPR*.

Sigurdsson, G. A., Gupta, A., Schmid, C., Farhadi, A., and Alahari, K. (2018). Charades-ego: A large-scale dataset of paired third and first person videos. In *ArXiv*.

Sigurdsson, G. A., Russakovsky, O., and Gupta, A. (2017). What actions are needed for understanding human actions in videos? In *ICCV*.

Yonetani, R., Kitani, K. M., and Sato, Y. (2016). Recognizing micro-actions and reactions from paired egocentric videos. In *CVPR*.

BRIEF BIOGRAPHY

Dima Damen: Associate Professor in Computer Vision at the University of Bristol, United Kingdom. Received her PhD from the University of Leeds, UK (2009). Dima's research interests are in the automatic understanding of object interactions, actions and activities using wearable and static visual (and depth) sensors. She has contributed works to novel research questions including fine-grained object interaction recognition, understanding the completion of actions, skill determination from video, semantic ambiguities of actions and the robustness of classifiers to actions temporal boundaries. Her work is published in leading venues: CVPR, ECCV, ICCV, PAMI, IJCV, CVIU and BMVC. In 2018, she led on releasing the largest dataset in first-person vision to date (EPIC-KITCHENS) - 11.5M frames of non-scripted recordings with full ground truth. Dima co-chaired BMVC 2013, is area chair for BMVC (2014-2018), associate editor of Pattern Recognition (2017-). She was selected as a Nokia Research collaborator in 2016, and as an Outstanding Reviewer in ICCV17, CVPR13 and CVPR12.