

Preface

Welcome to the 2015 ICCV workshop of *3D Reconstruction and Understanding with Video and Sound*.

We believe the availability of new types of input sensors and algorithms to process the sensor data has been the major driving factor for innovation in computer vision. With the proliferation of commodity RGBD cameras, many new techniques have been developed that take advantage of the additional depth channel for different applications, including reconstruction, segmentation, and understanding.

In this workshop, we aim to investigate (or reintroduce) the use of another readily available, but often ignored, source of information— *audio* or *acoustic* sensors—that can be combined with visual cameras. The resulting combination of aural and visual sensors can result in considerable advancement in terms of scene reconstruction and understanding. We need to develop a new set of algorithms that can utilize both types of sensors for different applications.

This workshop deals with new challenges and developments that arise with the use of such combination of sensors. In particular, we address two major issues. The first is the multimodal analysis of videos with sound for enhanced recognition accuracy, early vs. late fusion and end-to-end training of models. This can be used for different applications such as audiovisual speech recognition, video categorization or classification, and event detection in videos. The second major focus of the workshop is to explore the use of acoustic sensors to facilitate the reconstruction and understanding of 3D objects/models beyond the capability of current 3D RGBD sensors. This could include robust handling of scenes with specular/transparent objects, or even reconstruction around corners (i.e. non line-of-sight) and through obstacles or capturing other material characteristics (e.g. acoustic material properties for aural rendering).

To foster this new and exciting research direction, we have invited six distinguished speakers:

- [Dan Ellis](#), Columbia University
- [Ivan Dokmanic](#), EPFL/UIUC
- [Ming C. Lin](#), UNC Chapel Hill
- [Radu Patrice Horaud](#), INRIA
- [Ramesh Raskar](#), MIT
- [Zhengyou Zhang](#), Microsoft Research

These speakers represent a broad spectrum of research areas, from traditional multimodal visual analysis, to digital signal processing, to acoustic simulation. In addition, we have organized an industrial panel with leading companies in 3D sensors, including Intel, Microsoft, and Oculus/Facebook, that will discuss many challenges with respect to use the technologies in the next generation productions .

Our program also includes three strong papers that explore the combination of audio and video for various tasks. All papers were peer-reviewed and are included in the official ICCV Workshop proceedings and IEEE Digital Library.

We hope the program in this workshop, including invited talks, paper presentation, and panel discussions will result in the development of new ideas and applications based on aural and visual sensors.

Invited Speakers



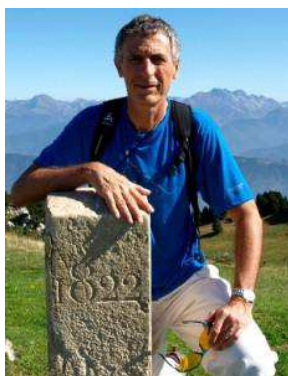
Dan Ellis

Daniel P. W. Ellis received the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, where he was a Research Assistant in the Machine Listening Group of the Media Lab. He spent several years as a Research Scientist at the International Computer Science Institute, Berkeley, CA. Currently, he is a Full Professor with the Electrical Engineering Department, Columbia University, New York. His Laboratory for Recognition and Organization of Speech and Audio (LabROSA) is concerned with all aspects of extracting high-level information from audio, including speech recognition, music description, and environmental sound processing. He also runs the AUDITORY email list of 2100 worldwide researchers in perception and cognition of sound.



Ivan Dokmanić

Ivan Dokmanić received a doctorate in computer and communication science from Ecole Polytechnique Fédérale de Lausanne (EPFL) in 2015. Previously he was a teaching assistant at the University of Zagreb, a codec developer for MainConcept AG, and a digital audio effects designer for Little Endian Ltd. During summer 2013 he was with Microsoft Research, Redmond. He works on a medley of inverse problems, audio and acoustics, signal processing for sensor networks, and fundamental signal processing. For his work on room shape reconstruction using sound he received the best student paper award at ICASSP 2011. In 2014 he received a Google PhD Fellowship. He is currently a postdoctoral researcher with Laurent Daudet (Institut Langevin), Stéphane Mallat (ENS), and Martin Vetterli (EPFL), and an adjunct professor at UIUC where he will start as an assistant professor next fall. He is also the singer and the lead guitarist of Ivan and the Terribles, featuring Martin Vetterli on bass, Paolo Prandoni on everything, and Marta Martinez-Cámara on saxophone.



**Radu Patrice
Horaud**

Radu Patrice Horaud holds a position of director of research at INRIA Grenoble Rhône-Alpes, France. He is the founder and director of the PERCEPTION team. Radu's research interests cover computational vision, audio signal processing, audio-visual scene analysis, machine learning, and robotics. He is the author of over 200 scientific publications. Radu pioneered work in computer vision using range data (or depth images) and developed a number of principles and methods at the cross-roads of computer vision and robotics. In 2006, he started to develop audio-visual fusion and recognition techniques in conjunction with human-robot interaction. He is an area editor for the Computer Vision and Image Understanding (Elsevier), a member of the advisory board for the International Journal of Robotics Research (Sage), and an associated editor for the International Journal of Computer Vision (Kluwer-Springer). In 2001 he was program co-chair of the IEEE Eighth International Conference on Computer Vision (ICCV'01). In 2015 he is program co-chair of ACM International Conference on Multimodal Interaction (ICMI'15).



Ming C. Lin

Ming C. Lin received her B.S., M.S., Ph.D. degrees in Electrical Engineering and Computer Science respectively from the University of California, Berkeley. She is currently the John R. & Louise S. Parker Distinguished Professor of Computer Science at the University of North Carolina (UNC), Chapel Hill and an honorary Chair Professor (Yangtze Scholar) at Tsinghua University in China. She has received several honors and awards, including the NSF Young Faculty Career Award in 1995, Honda Research Initiation Award in 1997, UNC/IBM Junior Faculty Development Award in 1999, UNC Hettleman Award for Scholarly Achievements in 2002, Beverly W. Long Distinguished Term Professor 2007-2010, Carolina Women's Center Faculty Scholar in 2008, Carolina's WOWS Scholar 2009-2011, IEEE VGTC VR Technical Achievement Award 2010, and 9 best paper awards. She is a Fellow of ACM and IEEE. Her research interests include computer graphics, robotics, and human-computer interaction, with focuses on physically-based modeling, sound rendering, haptics, algorithmic robotics, virtual environments, interactive techniques, geometric computing, and distributed interactive simulation. She has (co-)authored more than 250 refereed scientific publications, co-edited/authored four books.



Ramesh Raskar

Ramesh Raskar is an associate professor at MIT, leading the camera culture group. In 2004, Raskar received the TR100 Award from Technology Review, presented to top young innovators under the age of 35, and in 2003, the Global Indus Technovator Award, instituted at MIT to recognize the top 20 Indian technology innovators worldwide. In 2009, he was awarded a Sloan Research Fellowship. In 2010, he received the DARPA Young Faculty award. He holds more than 40 US patents, and has received four Mitsubishi Electric Invention Awards. He is currently co-authoring a book on computational photography.



Zhengyou Zhang

Zhengyou Zhang is a Fellow of the Institute of Electrical and Electronic Engineers (IEEE) (2005, for contributions to robust computer vision techniques) and a Fellow of the Association of Computing Machinery (ACM) (2013, for contributions to computer vision and multimedia). He is the Founding Editor-in-Chief of the newly established IEEE Transactions on Autonomous Mental Development (IEEE T-AMD), and is on the Editorial Board of the International Journal of Computer Vision (IJCV), the Machine Vision and Applications, and the Journal of Computer Science and Technology (JCST). He was on the Editorial Board of the IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE T-PAMI) from 1999 to 2005, the IEEE Transactions on Multimedia (IEEE T-MM) from 2004 to 2009, the International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI) from 1997 to 2008, among others. He is listed in Who's Who in the World, Who's Who in America and Who's Who in Science and Engineering.

Invited Panelists



Greg Leeming

Greg is currently a Member of the University Collaborative Research Group in Intel Labs where he is the Program Director of the Intel Science and Technology Centers (ISTC) for Visual Computing and Pervasive Computing, and of the Intel Visual Computing Institute. Greg joined Intel in 2000. He has 26 years of experience in high technology, mostly in research organizations. Greg has a BS-Math from Bates College, Lewiston ME, BSEE from Northeastern University, Boston MA and an MSEE from Brown University, Providence RI.



Ravish Mehra

Ravish Mehra is a Research Scientist at the Oculus Research lab. He completed his PhD in Computer Science at the University of North Carolina at Chapel Hill under the supervision of Prof. Dinesh Manocha and Prof. Ming Lin. He received his M.S. degree from UNC Chapel Hill in 2011 and Bachelor's degree in Computer Science and Engineering from Indian Institute of Technology (IIT) Delhi in 2008.



Avner Sander

Avner Sander is a member of the AIT (Advanced Imaging Technology) group at Microsoft R&D Israel. He has actively worked on 3D sensing and image processing.



Giora Yahav

Dr. Giora Yahav is the General Manager of AIT (Advanced Imaging Technology) group at Microsoft R&D Israel. He joined Microsoft in June 2009, as part of the acquisition of 3DV Systems. Dr. Yahav previously served as Senior VP R&D and CTO of 3DV Systems, and was also one of 3DV's founders and the Inventor of its basic technology. Dr. Yahav is a leading expert in imaging technologies and is regarded as thought leader in advance imaging technologies especially in novel CMOS based imagers. He is an active member of IEEE and SPIE, lecturing on various international conferences and academic conventions. He has published more than 20 papers and contributed to several fundamental researches in the field of imaging with an emphasis on active 3D imaging. Prior to embarking on a management career. Dr. Yahav spent eight years in academic research at the Hebrew University in Jerusalem, the Technion in Haifa and the Max-Planck Institute in Germany. Dr. Yahav holds a Ph.D. in Physical Chemistry, a M.Sc. in Inorganic Chemistry and a B.Sc. in Chemistry & Physics.



Jie Yang (tentative)

Jie Yang is a program director at NSF. His research interests include multimodal human-computer interaction (modeling and learning), computer vision, and pattern recognition. The main research goal is to enable computers to sense the world from multiple cues, to understand these cues from learning, and to interact with humans in a natural way. He recently has focused on developing technologies for three scenarios: (1) to support multimodal human computer interaction (e.g., Interact Project) and to serve humans who focus on interacting with other humans (e.g., CHIL Project); (2) to support geriatric care (e.g., CareMedia Project); and (3) to support human collaborations (e.g., Gesture Project and Large Scale Collaboration Project).