BMVC 2015 Keynote: Action and Attention in First-person Vision

Kristen Grauman   University of Texas at Austin

1 Speaker Biography

Kristen Grauman is an Associate Professor in the Department of Computer Science at the University of Texas at Austin. Her research in computer vision and machine learning focuses on visual search and object recognition. Before joining UT-Austin in 2007, she received her Ph.D. in the EECS department at MIT, in the Computer Science and Artificial Intelligence Laboratory. She is an Alfred P. Sloan Research Fellow and Microsoft Research New Faculty Fellow, a recipient of NSF CAREER and ONR Young Investigator awards, the Regents' Outstanding Teaching Award from the University of Texas System in 2012, the PAMI Young Researcher Award in 2013, the 2013 Computers and Thought Award from the International Joint Conference on Artificial Intelligence, and a Presidential Early Career Award for Scientists and Engineers (PECASE) in 2013. She and her collaborators were recognized with the CVPR Best Student Paper Award in 2008 for their work on hashing algorithms for large-scale image retrieval, and the Marr Best Paper Prize at ICCV in 2011 for their work on modeling relative visual attributes.

sual observations received. First, she will show how to exploit egomotion when learning image representations. Cognitive science tells us that proper development of visual perception requires internalizing the link between"how I move" and "what I see" - yet today's best recognition methods are deprived of this link, learning solely from bags of images downloaded from the Web. Prof. Grauman introduces a deep feature learning approach that embeds information not only from the video stream the observer sees, but also the motor actions he simultaneously makes. She will demonstrate the impact for recognition, including a scenario where features learned from ego-video on an autonomous car substantially improve large-scale scene recognition. Next, she will present their work exploring video summarization from the first person perspective. Leveraging cues about ego-attention and interactions to infer a storyline, the work automatically detects the highlights in long videos. Prof. Grumman will show how hours of wearable camera data can be distilled to a succinct visual storyboard that is understandable in just moments, and examine the possibility of person- and scene-independent cues for heightened attention. Overall, whether considering action or attention, the first-person setting offers exciting new opportunities for large-scale visual learning.

2 Abstract

A traditional third-person camera passively watches the world, typically from a stationary position. In contrast, a first-person (wearable) camera is inherently linked to the ongoing experiences of its wearer. It encounters the visual world in the context of the wearer's physical activity, behavior, and goals. This distinction has many intriguing implications for computer vision research, in topics ranging from fundamental visual recognition problems to high-level multimedia applications.

Prof. Grauman will present their recent work in this space, driven by the notion that the camera wearer is an active participant in the vi-