KEYNOTE TALK Tuesday, July 17, 2012 1:30PM – 2:30 PM / MEGAS ALEXANDROS

ISVC 2012: 8th International Symposium on Visual Computing Crete, Greece, July 16 - 18, 2012

Actions recognition from videos: some recent results

Dr. Cordelia Schmid INRIA, France

Abstract

The amount of digital video content available is growing daily, on sites such as YouTube. Recent statistics on the YouTube website show that around 48 hours of video are uploaded every minute. This massive data production calls for automatic analysis. In this talk we present some recent results for action recognition in videos. Bag-of-features have shown very good performance for action recognition in videos. We briefly review the underlying principles and introduce trajectory-based video features, which have shown to outperform the state of the art. These trajectory features are obtained by dense point sampling and tracking based on displacement information from a dense optical flow field. Trajectory descriptors are obtained with motion boundary histograms, which are robust to camera motion. We, then, show how to integrate temporal structure into a bag-of-features based on an actom sequence model. Action sequence models localize actions based on sequences of atomic actions, i.e., represent the temporal structure by sequences of histograms of actom-anchored visual features. This representation is flexible, sparse and discriminative. The resulting actom sequence model is shown to significantly improve performance over existing methods for temporal action localization. Finally, we show how to move towards more structured representations by explicitly modeling human-object interactions. We learn how to represent a human actions as an interactions between persons and objects. We localize in space and track over time both the object and the person, and represent an action as the trajectory of the object with respect to the person position, i.e., our human-object interaction features capture the relative trajectory of the object with respect to the human.

This is joint work with A Gaidon, V. Ferrari, Z. Harchaoui, A. Klaeser, A. Prest, H. Wang.



Speaker Bio-Sketch: Cordelia Schmid holds a M.S. degree in Computer Science from the University of Karlsruhe and a Doctorate from the Institut National Polytechnique de Grenoble (INPG). Her doctoral thesis received the best thesis award from INPG in 1996. Dr. Schmid was a post-doctoral research assistant in the Robotics Research Group of Oxford University in 1996--1997. Since 1997 she has held a permanent research position at INRIA Grenoble Rhone-Alpes, where she is a research director and directs the INRIA team called LEAR for LEArning and Recognition in Vision. Dr. Schmid is the author of over a hundred technical publications. She has been an Associate Editor for the IEEE Transactions on Pattern Analysis and Machine Intelligence (2001--2005) and for the International Journal of

Computer Vision (2004---), and she has been program chair of the 2005 IEEE Conference on Computer Vision and Pattern Recognition and of the 2012 European Conference on Computer Vision. In 2006, she was awarded the Longuet-Higgins prize for fundamental contributions in computer vision that have withstood the test of time. She is a fellow of IEEE.