



UNIVERSITY OF CENTRAL FLORIDA
CENTER FOR RESEARCH IN COMPUTER VISION

Dr. Josef Sivic

INRIA, Paris

**“From paintings to neural networks:
learning mid-level image representations for visual recognition”**

Monday, April 28, 2014 · 2:00PM · HEC 101



ABSTRACT

In this talk, I will describe our recent work on developing learnable mid-level representations for instance-level and category-level visual recognition.

First, we have developed a representation of 3D scenes where an entire architectural site is summarized by a set of scene parts learnt in a discriminative fashion from rendered views of its 3D model. We demonstrate recognizing 3D scene instances in challenging historical and non-photographic imagery, such as paintings and drawings, where standard local invariant features fail.

Second, using a similar approach we show that an object category can be non-parametrically modeled by a large collection of 3D CAD models explicitly representing the variation in style and viewpoint. Object detection in images is posed as a type of 2D to 3D alignment accomplished by matching mid-level object parts learnt from synthesized views. We demonstrate detection and alignment of “chairs” in challenging Pascal VOC 2012 images using a reference library of 1,394 CAD models downloaded from the Internet.

Finally, we investigate learning and transferring mid-level image representations using convolutional neural networks. We demonstrate that an image representation learnt on a task with a large amount of fully labelled imagery can significantly improve visual recognition performance on related tasks where supervision is scarce. The proposed model achieves state-of-the-art results on the Pascal VOC image classification and action recognition challenge.

The talk is based on recent papers with M. Aubrey, L. Bottou, A. Efros, I. Laptev, D. Maturana, M. Oquab and B. Russell.

BIOGRAPHY

Josef Sivic received a degree from the Czech Technical University, Prague, in 2002 and PhD from the University of Oxford in 2006. His thesis dealing with efficient visual search of images and videos was awarded the British Machine Vision Association 2007 Sullivan Thesis Prize and was short listed for the British Computer Society 2007 Distinguished Dissertation Award. His research interests include visual search and object recognition applied to large image and video collections. After spending six months as a postdoctoral researcher in the Computer Science and Artificial Intelligence Laboratory at the Massachusetts Institute of Technology, he currently holds a permanent position as an INRIA researcher at the Departement d’Informatique, Ecole Normale Supérieure, Paris. He has published over 40 scientific publications and serves as an Associate Editor for the International Journal of Computer Vision. He has been awarded an ERC Starting grant in 2013.