

# Automatic Understanding of the Visual World

Cordelia Schmid  
INRIA  
Montbonnot, France  
Cordelia.Schmid@inria.fr

## ABSTRACT

One of the central problems of artificial intelligence is machine perception, *i.e.*, the ability to understand the visual world based on input from sensors such as cameras. In this talk, I will present recent progress of my team in this direction. I will start with presenting results on how to generate additional training data using weak annotations, motion information and synthetic data. Next, I will discuss our results for action recognition in videos, where human tubelets have shown to be successful. Our tubelet approach moves away from state-of-the-art frame based approaches and improves classification and localization by relying on joint information from several frames. We show how to extend this type of method to weakly supervised learning of actions, which allows us to scale to large amounts of data with sparse manual annotation. Finally, I will present recent work on grasping with a robot arm based on learning long-horizon manipulations with a hierarchy of RL and imitation-based skills.

## CCS CONCEPTS

• **Computing methodologies** → **Computer vision.**

## KEYWORDS

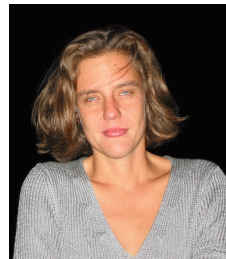
Machine perception, tubelets, action recognition

### ACM Reference Format:

Cordelia Schmid. 2019. Automatic Understanding of the Visual World. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19)*, July 21–25, 2019, Paris, France. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/3331184.3340264>

## BIOGRAPHY

Cordelia Schmid holds a M.S. degree in Computer Science from the University of Karlsruhe and a Doctorate, also in Computer Science, 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. Dr. Schmid has been an Associate Editor for IEEE PAMI (2001–2005) and for IJCV (2004–2012), editor-in-chief for IJCV (2013–2018), a program chair of IEEE CVPR 2005 and ECCV 2012 as well as a general chair of IEEE CVPR 2015 and ECCV 2020. In 2006, 2014 and 2016, 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. She was awarded an ERC advanced grant in 2013, the Humbolt research award in 2015 and the Inria & French Academy of Science Grand Prix in 2016. She was elected to the German National Academy of Sciences, Leopoldina, in 2017. In 2018 she received the Koenderink prize for fundamental contributions in computer vision that have withstood the test of time. Starting 2018 she holds a joint appointment with Google research.



Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGIR '19, July 21–25, 2019, Paris, France

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6172-9/19/07.

<https://doi.org/10.1145/3331184.3340264>