Recognizing Human Action in the Wild

Ivan Laptev

INRIA Paris - Rocquencourt / ENS, France
`ivan.laptev@inria.fr`

Abstract. Automatic recognition of human actions is a growing research topic urged by demands from emerging industries including (i) indexing of professional and user-generated video archives, (ii) automatic video surveillance, and (iii) human-computer interaction. Most applications require action recognition to operate reliably in diverse and realistic video settings. This challenging but important problem, however, has mostly been ignored in the past due to several issues including (i) the difficulty of addressing the complexity of realistic video data as well as (ii) the lack of representative datasets with human actions “in the wild”. In this talk we address both problems and first present a supervised method for detecting human actions in movies. To avoid a prohibitive cost of manual supervision when training many action classes, we next investigate weakly-supervised methods and use movie scripts for automatic annotation of human actions in video. With this approach we automatically retrieve action samples for training and learn discriminative visual action models from a large set of movies. We further argue for the importance of scene context for action recognition and show improvements using mining and classification of action-specific scene classes. We also address the temporal uncertainty of script-based action supervision and present a discriminative clustering algorithm that compensates for this uncertainty and provides substantially improved results for temporal action localization in video. We finally present a comprehensive evaluation of state-of-the-art methods for actions recognition on three recent datasets with human actions.