Int J Comput Vis
DOI 10.1007/s11263-013-0683-3

Max-Margin Early Event Detectors

Minh Hoai · Fernando De la Torre

Received: 30 December 2012 / Accepted: 27 November 2013

Abstract The need for early detection of temporal events from sequential data arises in a wide spectrum of applications ranging from human-robot interaction to video security. While temporal event detection has been extensively studied, early detection is a relatively unexplored problem. This paper proposes a maximum-margin framework for training temporal event detectors to recognize partial events, enabling early detection. Our method is based on Structured Output SVM, but extends it to accommodate sequential data. Experiments on datasets of varying complexity, for detecting facial expressions, hand gestures, and human activities, demonstrate the benefits of our approach.

Keywords Early detection · Event detection · Structured output learning

1 Introduction

The ability to make reliable early detection of temporal events has many potential applications in a wide range of fields, including security (e.g., pandemic attack detection), environmental science (e.g., tsunami warning), healthcare (e.g., risk-of-falling detection), entertainment (e.g., gaming), and robotics (e.g., affective computing). A temporal event has a duration, and by early detection, we mean to detect the event as soon as possible, *after it starts but before it ends*, as illustrated in Fig. 1. To see why it is important to detect events before they finish, consider a concrete example of building a robot that can affectively interact with humans. Arguably, a key requirement for such a robot is its ability to accurately and rapidly detect a human's emotional states from facial expression so that appropriate responses can be made in a timely manner. More often than not, a socially acceptable response is to imitate the human's current behavior. This requires facial events such as smiling or frowning to be detected even before they are complete; otherwise, the imitation response would be out of synchronization.

Despite the importance of early detection, few machine learning formulations have been explicitly developed for early detection. Most existing methods for event detection and modeling (e.g., Ke et al. 2005; Smith et al. 2005; Gorelick et al. 2007; Oh et al. 2008; Satkin and Hebert 2010; Klaser et al. 2010; Ali and Shah 2010; Niebles et al. 2010; Shi et al. 2010; Pei et al. 2011; Liu et al. 2011; Marin-Jiménez et al. 2011; Lan et al. 2011; Hoai and De la Torre 2012b; Amer et al. 2012; Yang and Shah 2012) are designed for offline processing. They have a limitation for processing sequential data as they are only trained to detect complete events. But for early detection, it is necessary to recognize partial events, which are ignored in the training process of existing event detectors.

This paper proposes Max-Margin Early Event Detectors (MMED), a novel formulation for training event detectors that recognize partial events, enabling early detection. MMED is based on structured output SVM (SOSVM) (Taskar et al. 2003; Tsochantaridis et al. 2005), but extends it to accommodate the nature of sequential data. In particular, we simulate the sequential frame-by-frame data arrival for training time series and learn an event detector that correctly classifies partially observed sequences. Figure 2 illustrates the key idea behind MMED: partial events are simulated and used as positive training examples. It is important to emphasize that we train a *single* event detector to recognize *all*

M. Hoai (✉) · F. De la Torre
Robotics Institute, Carnegie Mellon University,
5000 Forbes Ave., Pittsburgh, PA15213, USA
e-mail: minhhoai@robots.ox.ac.uk; minhhoai@cmu.edu

F. De la Torre
e-mail: ftorre@cs.cmu.edu

Published online: 13 December 2013