Int J Comput Vis (2012) 97:276–304
DOI 10.1007/s11263-011-0487-2

Learning Behavioural Context

Jian Li · Shaogang Gong · Tao Xiang

Received: 3 December 2009 / Accepted: 28 July 2011 / Published online: 27 August 2011
© Springer Science+Business Media, LLC 2011

Abstract We propose a novel framework for automatic discovering and learning of behavioural context for video-based complex behaviour recognition and anomaly detection. Our work differs from most previous efforts on learning visual context in that our model learns multi-scale spatio-temporal rather than static context. Specifically three types of behavioural context are investigated: *behaviour spatial context*, *behaviour correlation context*, and *behaviour temporal context*. To that end, the proposed framework consists of an activity-based semantic scene segmentation model for learning behaviour spatial context, and a cascaded probabilistic topic model for learning both behaviour correlation context and behaviour temporal context at multiple scales. These behaviour context models are deployed for recognising non-exaggerated multi-object interactive and co-existence behaviours in public spaces. In particular, we develop a method for detecting subtle behavioural anomalies against the learned context. The effectiveness of the proposed approach is validated by extensive experiments carried out using data captured from complex and crowded outdoor scenes.

Keywords Visual context · Behavioural context · Video-based behaviour recognition · Activity-based scene segmentation · Cascaded topic models · Anomaly detection

J. Li (✉) · S. Gong · T. Xiang
School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK
e-mail: jianli@eecs.qmul.ac.uk

S. Gong
e-mail: sgg@eecs.qmul.ac.uk

T. Xiang
e-mail: txiang@eecs.qmul.ac.uk

1 Introduction

Visual context is the environment, background, and settings within which objects and associated events are observed visually. Humans employ visual context extensively for both object recognition in a static setting and behaviour recognition in a dynamic environment. For instance, for object recognition we can differentiate and recognise whether a hand-held object is a mobile phone or calculator by its relative position to other body parts (e.g. closeness to the ears), even though they are visually similar and partially occluded by the hand. Similarly for behaviour recognition, the arrival of a bus can be detected/inferred just by looking at the passengers' behaviour at a bus stop. Indeed, extensive cognitive, physiological and psychophysical studies have shown that visual context plays a critical role in human visual perception (Palmer 1975; Biederman et al. 1982; Bar and Ullman 1993; Bar and Aminof 2003; Bar 2004). Motivated by these studies, there is an increasing interest in exploiting contextual information for computer vision tasks such as object detection (Heitz and Koller 2008; Murphy et al. 2003; Kumar and Hebert 2005; Carbonetto et al. 2004; Wolf and Bileschi 2006; Rabinovich et al. 2007; Gupta and Davis 2008; Galleguillos et al. 2008; Zheng et al. 2009), action recognition (Marszalek et al. 2009) and tracking (Yang et al. 2008; Ali and Shah 2008).

Previous studies on visual context are predominantly focused on static visual context particularly regarding the scene background, scene category, and other co-existing objects in a scene. However, for understanding object behaviour in a crowded space, the most relevant visual context is no longer static due to the non-stationary background and non-rigid relationships among co-existing objects in a public space. In particular, a meaningful interpretation of object behaviour depends largely on knowledge of spatial