Int J Comput Vis
DOI 10.1007/s11263-013-0655-7

Detecting People Looking at Each Other in Videos

M. J. Marin-Jimenez · A. Zisserman · M. Eichner · V. Ferrari

Received: 5 October 2012 / Accepted: 14 August 2013

Abstract The objective of this work is to determine if people are interacting in TV video by detecting whether they are looking at each other or not. We determine both the temporal period of the interaction and also spatially localize the relevant people. We make the following four contributions: (i) head detection with implicit coarse pose information (front, profile, back); (ii) continuous head pose estimation in unconstrained scenarios (TV video) using Gaussian process regression; (iii) propose and evaluate several methods for assessing whether and when pairs of people are looking at each other in a video shot; and (iv) introduce new ground truth annotation for this task, extending the TV human interactions dataset (Patron-Perez et al. 2010) The performance of the methods is evaluated on this dataset, which consists of 300 video clips extracted from TV shows. Despite the variety and difficulty of this video material, our best method obtains an average precision of 87.6 % in a fully automatic manner.

M. J. Marin-Jimenez (✉)
Department of Computing and Numerical Analysis,
Maimonides Institute for Biomedical Research (IMIBIC),
University of Cordoba, Cordoba, Spain
e-mail: mjmarin@uco.es

A. Zisserman
Department of Engineering Science,
University of Oxford, Oxford, UK
e-mail: az@robots.ox.ac.uk

M. Eichner
ETH Zurich, Zurich, Switzerland
e-mail: marcin.eichner@vision.ee.ethz.ch

V. Ferrari
School of Informatics, University of Edinburgh,
Edinburgh, UK
e-mail: vferrari@staffmail.ed.ac.uk

Keywords Person interactions · Video search · Action recognition · Head pose estimation

1 Introduction

If you read any book on film editing or listen to a director's commentary on a DVD, then what emerges again and again is the importance of eyelines. Standard cinematography practice is to first establish which characters are looking at each other using a medium or wide shot, and then edit subsequent close-up shots so that the eyelines match the point of view of the characters. This is the basis of the well known 180° rule in editing.

The objective of this paper is to determine whether eyelines match between characters within a shot—and hence understand which of the characters are interacting. The importance of the eyeline is illustrated by the three examples of Fig. 1—one giving rise to arguably the most famous quote from *Casablanca*, and another being the essence of the humour at that point in an episode of *Fawlty Towers*. Our target application is this type of edited TV video and films. It is very challenging material as there is a wide range of human actors, camera viewpoints and ever present background clutter.

Determining whether characters are interacting using their eyelines is another step towards a fuller video understanding, and complements recent work on automatic character identification (Everingham et al. 2006; Cour et al. 2009; Sivic et al. 2009), human pose estimation (Ferrari et al. 2009; Andriluka et al. 2009; Bourdev et al. 2010; Sapp et al. 2010;Yang et al. 2012), human action recognition (Laptev et al. 2008; Liu et al. 2009; Marín-Jiménez and Pérez de la Blanca 2012; Raptis et al. 2012; Sadanand and Corso 2012), and social (Fathi et al. 2012) and specific interaction recognition (e.g. hugging,

Published online: 14 September 2013