Int J Comput Vis (2013) 101:305–328
DOI 10.1007/s11263-012-0573-0

Recognizing Interactive Group Activities Using Temporal Interaction Matrices and Their Riemannian Statistics

Ruonan Li · Rama Chellappa · Shaohua Kevin Zhou

Received: 30 August 2011 / Accepted: 7 September 2012 / Published online: 21 September 2012

Abstract While video-based activity analysis and recognition has received much attention, a large body of existing work deals with activities of a single subject. Modeling and recognition of coordinated multi-subject activities, or group activities, present in a variety of applications such as surveillance, sports, and biological monitoring records, etc., is the main objective of this paper. Unlike earlier attempts which model the complex spatial temporal constraints among multiple subjects with a parametric Bayesian network, we propose a compact and discriminative descriptor referred to as the Temporal Interaction Matrix for representing a coordinated group motion pattern. Moreover, we characterize the space of the Temporal Interaction Matrices using the Discriminative Temporal Interaction Manifold (DTIM), and use it as a framework within which we develop a data-driven strategy to characterize the group motion pattern without employing specific domain knowledge. In particular, we establish probability densities on the DTIM for compactly describing the statistical properties of the coordinations and interactions among multiple subjects in a group activity. For each class of group activity, we learn a multi-modal density function on the DTIM. A Maximum a Posteriori (MAP) classifier on the manifold is then designed for recognizing new activities. In addition, we have extended this model to one with which we can explicitly distinguish the participants from non-participants. We demonstrate how the framework can be applied to motions represented by point trajectories as well as articulated human actions represented by images. Experiments on both cases show the effectiveness of the proposed approach.

Keywords Event analysis · Activity recognition

R. Li (✉)
Harvard School of Engineering and Applied Sciences,
Cambridge, MA 02138, USA
e-mail: ruonanli@seas.harvard.edu

R. Chellappa
Center for Automation Research, UMIACS, and the Department of Electrical and Computer Engineering, University of Maryland,
College Park, MD 20742, USA
e-mail: rama@umiacs.umd.edu

S.K. Zhou
Corporate Research & Technology, Siemens Corporation,
Princeton, NJ 08540, USA
e-mail: shaohua.zhou@siemens.com

1 Introduction

In this work we model and recognize *interactive group activities* involving *multiple subjects* from videos. Human activity analysis and classification has been a research focus of computer vision community for over two decades (Moeslund et al. 2006) and is receiving continuous attention in recent years (Aggarwal and Ryoo 2011; Poppe 2010). However, the majority of previous and ongoing work has focused on activities involving a single subject, where the motion and dynamics of an individual subject are investigated. Activities of multiple subjects exist widely in surveillance, sports, and biological observation records, *etc.*, and consequently modeling and analysis of multi-subject activities will be useful in these applications. Although the multi-subject tracking problem has been extensively studied (Yilmaz et al. 2006), less attention has been paid to *recognize* the tracked motion pattern of a whole group and *detect* the activity by distinguishing its participants from irrelevant moving subjects.

Limited initial attempts considered simple scenarios, where the individuals in a group undergo a structurally fixed