J. Vis. Commun. Image R. 32 (2015) 10–19

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

A survey of human pose estimation: The body parts parsing based methods ☆

Zhao Liu *, Jianke Zhu, Jiajun Bu, Chun Chen

College of Computer Science, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

Article history:
Received 15 May 2014
Accepted 22 June 2015
Available online 2 July 2015

Keywords:
Human pose estimation
Articulated object detection
Survey
Body parts parsing
Motion capture
Feature Extraction
Appearance models
Structure models

ABSTRACT

Estimating human pose from videos and image sequences is not only an important computer vision problem, but also plays very critical role in many real-world applications. Main challenges for human pose estimation are variation of body poses, complicated background and depth ambiguities. To solve these problems, considerable research efforts have been devoted to the related fields. In this survey, we focus our attention on the recent advances in vision-based human pose estimation. We first present a general framework of human pose estimation, and then go through the latest technical progress on each stage. Finally, we discuss the limitations of the existing approaches and foresee the future directions to be explored.

1. Introduction

Human pose estimation (HPE) is the process of inferring the 2D or 3D human body part positions from still images or videos. Conventional HPE methods usually employ extra hardware devices to capture human poses and construct a human skeleton based on the captured body joints. These methods are either expensive or inefficient. During the past decade, considerable research efforts have been devoted to HPE problem in computer vision domain.

Although having investigated the issues of human body part configuration, human body detection and human motion [1] in the previous studies, there still lacks a survey to summarize the most recent progress on body pose estimation. In this survey, we mainly review the recent advances in vision-based human pose estimation. Human pose estimation includes nearly all the human-related problems in computer vision, ranging from the whole human body pose parsing to the detailed body parts localization. As it is hard to cover all these fields within a single survey, we mainly focus on the body part parsing methods. For better comparison of different body part parsing methods, we divide them into four parts, including 2D single person parsing in images, 2D multi-person parsing in images, 2D single person parsing in videos and 3D single person parsing in images and videos. Moreover, we discuss the limitations of the existing approaches and foresee the future trend.

Human pose estimation techniques become more and more mature in the past decades. Being the great interest of different domains, new applications constantly emerge along with the technological evolutions. Human pose estimation is not only an important computer vision problem, but also plays critical role in a variety of real-world applications in the following.

Video Surveillance. Video surveillance aims at tracking and monitoring the locations and motions of pedestrians in special circumstances. It is the earliest application area that HPE technologies have been used. The common scenes are the supermarket and airport passageway.

Human–Computer Interaction (HCI). Advanced human computer interaction systems with human pose estimation have been developed rapidly. In these systems, instructions can be analyzed accurately by capturing the human body poses. In recent years, intelligence driving emerges as a novel practical application.

Digital Entertainment. Digital entertainment, including computer games, computer animation and films, has become a huge industry and an active domain in recent years. For instance, People enjoys the pleasure the body sensor games give to them. Also, In the pre-production of the special effects for movie Avatar [2], actors wear the special equipments to animate the activities of Avatars.

Medical Imaging. Human pose estimation has been widely used in the automatic medical field. A specific instance is that HPE can

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail addresses: liuzhao@zju.edu.cn (Z. Liu), jkzhu@zju.edu.cn (J. Zhu), bjj@zju.edu.cn (J. Bu), chenc@zju.edu.cn (C. Chen).

http://dx.doi.org/10.1016/j.jvcir.2015.06.013
1047-3203/

Fig. 1. Various applications fields for human pose estimation.

Table 1
Main commercial systems for human pose estimation.

Systems	Principle	Application areas	Institution	Related URLs
Kinect Sensor	Structure light capture and machine learning	Motion Capture Multi-View Pose estimation [9]	Microsoft	http://www.xbox.com/en-US/kinect
Leap Motion	Double sensors Infrared light Vision different	Gesture Recognition [10]	Leap Motion Inc.	https://www.leapmotion.com/
		Industrial Robot [11]		
Vicon	Reflected light based system	Animation, Military Remote sensing, Bioinformation	Oxford Metrics Limited	http://www.vicon.com/
Wii	Bluetooth communication Infrared light	Games Physical treatment [12]	Nintendo	http://www.nintendo.com/wii

be used to assist doctors to check patients' activities from the remote monitor, which greatly simplifies the therapeutic process.

Sports Scenes. In sports news and live broadcast, human pose estimation is employed to track athletes' locations and activities. Moreover, the estimated poses can be used to employed the detailed movements of their actions.

Other applications include military, children mental development, virtual reality, and so on. The related application fields of HPE are shown in Fig. 1.

In recent years, various devices and commercial systems have been released accompany with HPE technology, including Microsoft Kinect sensor [3,4], Leap Motion [5], body mounted camera [6], 3D laser scanner [7] and infrared light source [8]. These commercial systems have quite different implementation principles and application fields, as shown in Table 1.

2. Related surveys and overview

During the last decade, several surveys have been published to summarize the related work on human pose estimation. 3D HPE has attracted lots of attentions in computer vision. For instance, Hen and Paramesran [13] summarize the single camera 3D pose estimation from images and Sminchisescu [14] aims to reconstruct 3D human poses from monocular video sequences. Wearable equipments make it possible to estimate the depth in motion capture, Helten et al. [15] review the depth camera based motion capture work. Compared with methods relying on the specific hardwares, the vision-based approaches are more efficient and economic, which have been rapidly developed in these years. Poppe [1] reviews the vision-based methods for human pose estimation on the marker-less data. Moeslund et al. [16] summarize the research work on visual analysis of human, which covers various topics including pose estimation, human recognition and their applications.

Most recently, we witness the rapid development in HPE field. For instance, the model based methods, especially the pictorial structure model, have played an important role in human body parsing [17,18].

On the other hand, since video play an more important role in recent applications, action/activity recognition have been paid much attention, although closely related with HPE, the technique used in action/activity recognition are quite different. We refer the readers to [19–22] to see more details of action/activity recognition. Moreover, deep learning based methods have attracted lots of research attentions [23,24]. As all of these technologies will continue to be the focus for a few years in future, there is a need for the discussion in hot topics.

To clearly illustrate the recent studies on human pose estimation, we categorize them into two stages: preprocessing and body parts parsing. Fig. 2 summarizes the whole process of the common human pose estimation. The preprocessing stage includes feature extraction, camera calibration, body detection and foreground

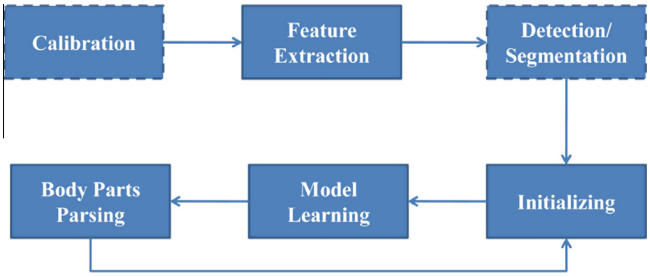

Fig. 2. The framework of the common human pose estimation system, stages in the boxes with dash lines means that they may be removed in some methods.