

Advances in Tracking and Recognition of Human Motion

A**Niki Aifanti***Informatics & Telematics Institute, Greece***Angel D. Sappa***Computer Vision Center, Spain***Nikos Grammalidis***Informatics & Telematics Institute, Greece***Sotiris Malassiotis***Informatics & Telematics Institute, Greece*

INTRODUCTION

Tracking and recognition of human motion has become an important research area in computer vision. In real world conditions it constitutes a complicated problem, considering cluttered backgrounds, gross illumination variations, occlusions, self-occlusions, different clothing and multiple moving objects. These ill-posed problems are usually tackled by making simplifying assumptions regarding the scene or by imposing constraints on the motion. Constraints such as that the contrast between the moving people and the background should be high and that everything in the scene should be static except for the target person are quite often introduced in order to achieve accurate segmentation. Moreover, the motion of the target person is often confined to simple movements with limited occlusions. In addition, assumptions such as known initial position and posture of the person are usually imposed in tracking processes.

BACKGROUND

The first step towards human tracking is the segmentation of human figures from the background. This problem is usually addressed either by exploiting the temporal relation between consecutive frames (e.g., background subtraction (Sato & Aggarwal, 2001), optical flow (Okada, Shirai & Miura, 2000)), by modeling the image statistics of human appearance (Wren, Azarbayejani, Darrell & Pentland, 1997) or by exploiting the human shape (Leibe, Seemann & Schiele, 2005). Efficient texture-based methods for modeling the background and detecting moving objects from a video sequence have been developed as well (Heikkila & Pietikainen, 2006), while

some other recent research copes with the problem of occlusions (Capellades, Doermann, DeMenthon & Chellappa, 2003). The output of the segmentation, which could be edges, silhouettes, blobs, and so forth, comprises the basis for feature extraction.

Feature correspondence is established in order to track the subject. Tracking through consecutive frames commonly incorporates prediction of movement, which ensures continuity of motion especially when some body parts are occluded. For example, when a person is walking there are some moments when one of the legs occludes the other. Furthermore, there are scenes with multiple persons occluding one another. Depending on the scene and the chosen methodology, some techniques try to determine the precise movement of each body part (Sidenbladh, Black, & Sigal, 2002), while other techniques focus on tracking the human body as a whole (Okada, Shirai & Miura, 2000). Tracking may be classified as 2D or 3D. 2D tracking consists in following the motion in the image plane either by exploiting low-level image features or by using a 2D human model. 3D tracking aims at obtaining the parameters, which describe body motion in three dimensions. The 3D tracking process, which estimates the motion of the body parts, is inherently connected to 3D human pose recovery.

3D pose recovery aims at defining the configuration of the body parts in the 3D space and estimating the orientation of the body with respect to the camera. This work will mainly focus on model-based techniques, since they are usually used for 3D reconstruction. Model-based techniques rely on a mathematical representation of human body structure and motion dynamics. The 3D pose parameters are commonly estimated by iteratively matching a set of image features extracted from the current frame with the projection of the

model on the image plane. Thus, 3D pose parameters are determined by means of an energy minimization process.

Instead of obtaining the exact configuration of the human body, human motion recognition consists in identifying the action performed by a moving person. Most of the proposed techniques focus on identifying actions belonging to the same category. For example, the objective could be to recognize several aerobic exercises or tennis strokes or some everyday actions such as sitting down, standing up, walking, running, or skipping.

Next, some of the most recent approaches addressing human motion tracking and 3D pose recovery are presented, while the following subsection introduces some whole-body human motion recognition techniques. Previous surveys of vision-based human motion analysis have been carried out by C  dras and Shah (1995), Aggarwal and Cai (1999), Gavrilu (1999), Moeslund and Granum (2001), and Moeslund, Hilton, and Kruger (2006).

This overview presents briefly some of the techniques developed during the last years. The outline of this work is as follows. Firstly, a survey about human motion tracking and 3D pose recovery is given. Next, human motion recognition is introduced; following, a summary of some application works is presented. Finally, a section with future trends and conclusions is introduced.

HUMAN MOTION TRACKING AND 3D POSE RECOVERY

Tracking relies either on monocular or multiple camera image sequences. Using *monocular* image sequences is quite challenging due to occlusions of body parts and ambiguity in recovering their structure and motion from a single perspective view (different configurations have the same projection). On the other hand, single camera views are more easily obtained and processed than multiple camera views. In the following table, some recent techniques using only one camera are presented.

In contrast to single-view approaches, *multiple camera* techniques are able to overcome occlusions and depth ambiguities of the body parts, since useful motion information missing from one view may be recovered from another view. The following table presents some recent approaches using multiple cameras.

Some currently published papers specifically tackle the *pose recovery* problem using multiple sensors. In Mikic, Trivedi, Hunter, and Cosman (2001) a 3D voxel reconstruction of the person's body is computed from silhouettes extracted from four cameras. Body parts are located sequentially from the voxel data. Consistency with an articulated body model is guaranteed by feeding measurements to

Table 1. Monocular systems

<i>Authors</i>	<i>Description</i>
Sminchisescu and Triggs (2001)	A 3D human body model, consisting of tampered superellipsoids, is fitted on the image features by means of an iterative cost function optimization scheme. A multiple-hypothesis approach with the ability of escaping local minima in the cost function is proposed.
Sidenbladh, Black, and Sigal (2002)	A probabilistic approach for modeling 3D human motion for synthesis and tracking, where learning of state transition probabilities is replaced with efficient probabilistic search in a large training set.
Ning, Tan, Wang, and Hu (2004)	An approach to tracking walking human based on both body model and a motion model (learnt from semiautomatically acquired training data), in a CONDENSATION framework. A pose evaluation function combining boundary and region information is proposed. Automatic acquisition of initial model pose and recovery from severe failures are addressed.
Antonini, Martinez, Bierlaire, and Thiran (2006)	Image processing methods are combined with behavioral models for pedestrian dynamics in a detection/tracking system. Behavioral models are used to find globally coherent trajectories. Pedestrian detection is based on the behavior rather than appearance.
Liu and Chellappa (2007)	An articulated model of a real human body consisting of ellipsoids connected by joints is used. Motion in a video sequence is observed using optical flow. Optical flow using scaled orthographic projection relates the spatial-temporal intensity change of the sequence to the motion parameters.
Caillette, Galata, and Howard (2007)	A 3D human body tracker, build upon the Monte-Carlo Bayesian framework, is capable of handling fast and complex motions in real-time. Novel prediction and evaluation methods improving the robustness and efficiency of the tracker are proposed. The tracker is also capable of automatic initialisation and self-recovery.

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/advances-tracking-recognition-human-motion/13550

Related Content

Intraorganizational Versus Interorganizational Uses and Benefits of Electronic Mail

Denise J. McManus, Chetan Sankar, Houston H. Carrand F. Nelson Ford (2002). *Information Resources Management Journal* (pp. 5-13).

www.irma-international.org/article/intraorganizational-versus-interorganizational-uses-benefits/1223

Information and Communication Overload: Can DigComp Help?

Margarida Lucasand António A. F. G. Moreira (2017). *Information and Communication Overload in the Digital Age* (pp. 157-175).

www.irma-international.org/chapter/information-communication-overload/176569

An Advanced Apriori Algorithm Technology to Enhance Sports Data Mining

Maojun Cao (2024). *Information Resources Management Journal* (pp. 1-18).

www.irma-international.org/article/an-advanced-apriori-algorithm-technology-to-enhance-sports-data-mining/361709

Cost Optimization for Dynamic Content Delivery in Cloud-Based Content Delivery Network

S. Sajitha Banuand S. R. Balasundaram (2021). *Journal of Information Technology Research* (pp. 18-32).

www.irma-international.org/article/cost-optimization-for-dynamic-content-delivery-in-cloud-based-content-delivery-network/289854

The Role of User Characteristics in the Development and Evaluation of E-Learning Systems

Dianna Newmanand Aikaterini Passa (2007). *Information Resources Management: Global Challenges* (pp. 333-350).

www.irma-international.org/chapter/role-user-characteristics-development-evaluation/23048