Pattern Recognition 42 (2009) 1559 – 1571

Contents lists available at ScienceDirect

Pattern Recognition

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

Automatic reconstruction of 3D human motion pose from uncalibrated monocular video sequences based on markerless human motion tracking

Beiji Zou, Shu Chen*, Cao Shi, Umugwaneza Marie Providence

School of Information Science and Engineering, Central South University, Changsha 410083, People's Republic of China

ARTICLE INFO

Article history:
Received 19 December 2007
Received in revised form 3 September 2008
Accepted 17 December 2008

Keywords:
3D human motion reconstruction
Human motion tracking from monocular video sequences
Inverse kinematics
Human skeleton model
Camera model

ABSTRACT

We present a method to reconstruct human motion pose from uncalibrated monocular video sequences based on the morphing appearance model matching. The human pose estimation is made by integrated human joint tracking with pose reconstruction in depth-first order. Firstly, the Euler angles of joint are estimated by inverse kinematics based on human skeleton constrain. Then, the coordinates of pixels in the body segments in the scene are determined by forward kinematics, by projecting these pixels in the scene onto the image plane under the assumption of perspective projection to obtain the region of morphing appearance model in the image. Finally, the human motion pose can be reconstructed by histogram matching. The experimental results show that this method can obtain favorable reconstruction results on a number of complex human motion sequences.

1. Introduction

Human motion contains a wealth of information about actions, intentions, emotions, and personality traits of a person and plays an important role in many application areas, such as surveillance, human motion analysis, and virtual reality. It is a hot topic to track human joint and reconstruct the corresponding 3D human motion posture from an uncalibrated monocular video sequences, the human motion pose reconstruction can be categorized into two groups: (1) using multi-view video sequences, and (2) using monocular video sequences. Reconstruction of human motion pose from monocular video sequences is more attractive because it has many advantages such as convenient to use, conveniently available to general public and less restrictions. The depth value of an object will be lost when the object is projected onto 2D image plane. Therefore, 3D motion reconstruction from 2D motion sequences is still a challenging task. The conventional methods to reconstruct human pose from monocular video sequences may require some restrictions or prior knowledge. Rather than the classical algorithms, in this paper, we propose an approach to reconstruct the 3D human motion pose from uncalibrated monocular video sequences by combining human joint tracking and pose extraction, whose advantages include fewer constraints, without knowing the parameters of camera model, easy to implement and more precise performance of the pose reconstruction.

Human pose reconstruction from monocular video sequences is roughly divided into two categories: machine learning methods and object tracking methods. Researchers propose the use of machine learning methods that exploit prior knowledge in gaining more stable estimates of 3D human body pose [1–5]. However, these methods require a large amount of samples which limit their applications. Object tracking methods commonly follow two sequence steps: first, locating feature of human and tracking them in each frame, then, reconstructing human pose by these obtained features. Many researchers have conducted studies on the first step, and general surveys can be found in recent review papers [6,7], in this step, people always use the configuration in the current frame and a dynamic model to predict the next configuration [8,9]. Most approaches perform prediction by variants of kalman filtering [9,10] and particle filtering [11–13]. Particle filters restrict themselves to predictions returned by a motion model which is hard to construct, such a scheme is susceptible to drift due to imprecise motion model that the predictions were worse. Annealing the particle filter [14] or performing local searches [15] is the ways to attack this difficulty. The second step is human pose reconstruction (i.e., extracting 3D coordinates of feature from its corresponding 2D image coordinates). Some researchers reconstruct the human motion pose from video sequences by using some constrains such as human skeleton proportions based on camera model, these methods can be classified into two classes depending on the camera model adopted: (1) using affine camera model; and (2) using perspective camera model. Affine camera model is only an approximation of the real camera model. Scaled-orthographic camera model is an important instance of this kind and is popularly used

* Corresponding author.
E-mail address: csu_cs@163.com (S. Chen).

0031-3203/$ - see front matter
doi:10.1016/j.patcog.2008.12.024

by many researchers [16,17]. The scale factor *s* has a significant effect on the result of human motion pose reconstruction by using scaled-orthographic camera model [16]. In these methods, the scale factor *s* is estimated by satisfying a constrained formula, but not a ground truth value; so the reconstructed human pose is great different from the real human pose, and these methods can only handle images with very little perspective effects. In addition, there are very limited research efforts working on human pose reconstruction based on perspective camera model [18–20]. Zhao et al. [20] restrict all body segments of the human figure as almost parallel to the image plane in order to acquire accurate human skeleton proportions. Peng requires estimating the virtual scale factor for each frame [19].

The remainder of this paper describes our algorithm in more detail. In Section 2, we explain the diagram of data flow in our system, and in Section 3, we describe the initialization of our system. The detail procedure to reconstruct human motion pose is described in Section 4, while in Section 5, we illustrate results from our system. In the end of this paper, we conclude about this study and point out the future further work.

2. Overview

The basic idea of our algorithm is to reconstruct 3D human pose reconstruction from the corresponding 2D joints on the image plane. The positions of human joints in each frame of the video are located with a local search by using the technique of morphing appearance model matching.

The proposed algorithm is divided into four major steps as shown in Fig. 1. The first step is to initialize models by a simple user interface with the first frame as input, the texture information and space information about the appearance of body segments can be acquired by marking projected landmarks of the subject's body on the image plane, and the relative lengths of body segments in the human skeleton model are also estimated.

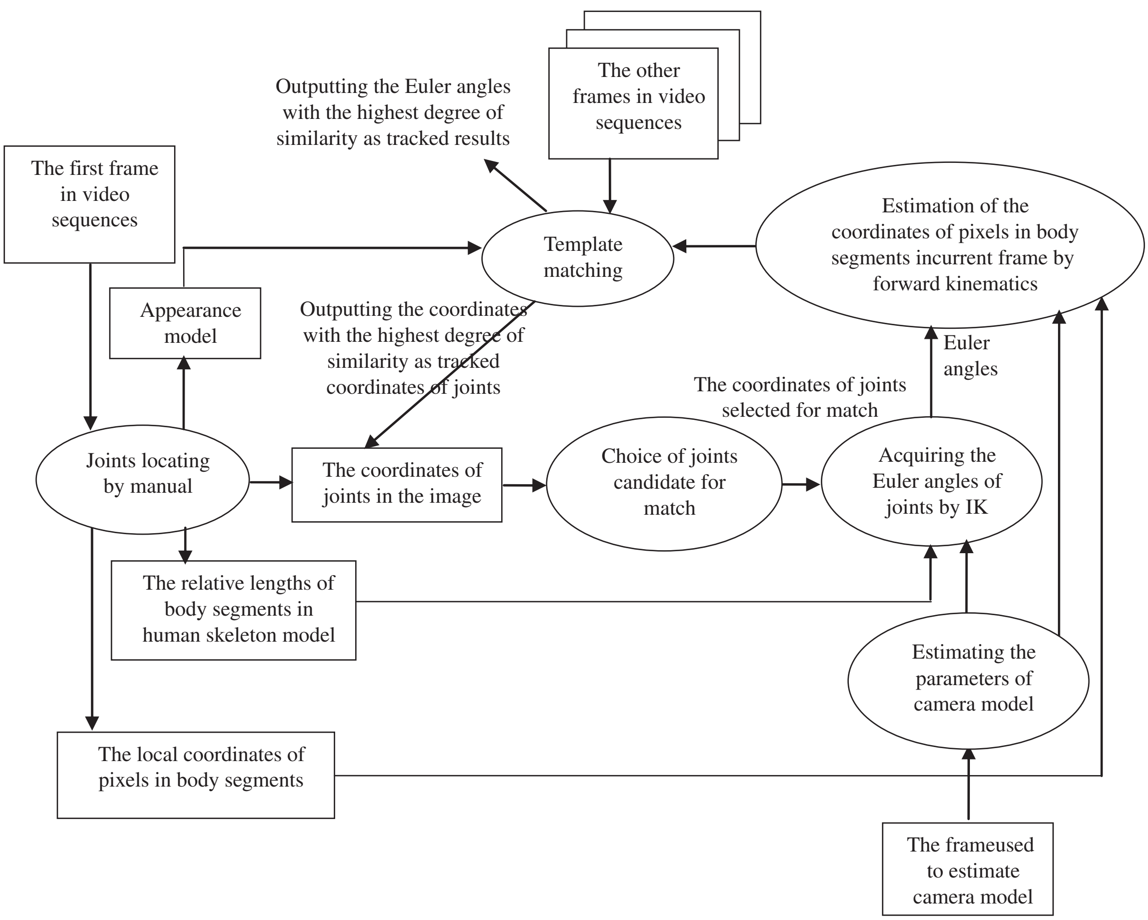

Fig. 1. Diagram of data flow in our system.

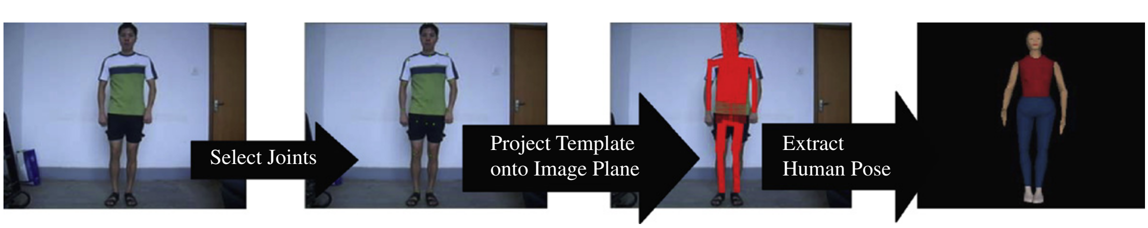

Fig. 2. Morphing appearance model matching to extract human pose.