Pattern Recognition Letters 49 (2014) 48–54

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Activity-based methods for person recognition in motion capture sequences ☆

Eftychia Fotiadou *, Nikos Nikolaidis

Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece

ARTICLE INFO

Article history:
Received 27 January 2014
Available online 27 June 2014

Keywords:
Person recognition
Motion capture
Classification
Dimensionality reduction

ABSTRACT

In this paper we present two algorithms for efficient person recognition operating upon motion capture data, depicting persons performing various everyday activities. The first approach is driven from the assumption that, if two motion sequences depict a certain activity performed by the same person, then, consecutive frames (poses) of one sequence are expected to be similar to consecutive frames of the other. The proposed method constructs a pose correspondence matrix to represent the similarity between poses and utilizes an intuitive method for estimating a similarity score between two motion capture sequences, based on the structure of the correspondence matrix. The second algorithm is based on a Bag of Words model (BoW), where histograms are extracted from motion sequences, based on the frequency of occurrences of characteristic poses. This method is combined with the application of Locality Preserving Projections (LPP) on the data, in order to reduce their dimensionality. Our methods achieved more than 98% correct person recognition rate, in three different datasets.

1. Introduction

Motion capture has long been used in film and games industry in order to allow for more realistic renditions of human and animal motion. It is also useful in virtual reality applications, as well as in various disciplines that involve the study of human motion, such as ergonomic analysis, sports biomechanics and rehabilitation. The result of motion capture is a skeletal animation sequence, i.e. a series of skeleton configurations (poses) over time. Today, motion capture is becoming more affordable. As an example, the Kinect sensor with the accompanying human skeleton tracking software can deliver fairly good motion capture data with minimal cost. As a consequence, motion capture techniques gain more and more ground, giving boost to the development of diverse kinds of applications.

Person recognition refers to the process by which the identity of a person is recognized by a system, based on information that he or she carries. Examples of applications involving person recognition include security or surveillance systems, access control, patient monitoring, as well as a wide range of systems involving human–computer interaction. Traditionally, recognition is performed by means of credentials, supplied by the person, in form of IDs, smart cards or passwords. However, in the last decades, an increasing use of biometric features is observed [9]. These features may include physiological characteristics of a person, such as fingerprints, face/iris characteristics, palm prints or DNA, as well as behavioural characteristics [23], such as gait or style when performing a certain action, keyboard typing, and voice. Biometric characteristics show advantage over the aforementioned credentials, with respect to counterfeiting or loss risk.

To our knowledge, there do not exist methods for activity-based person recognition from motion capture data, other than those related to human gait analysis. Although those algorithms apply mostly on video motion data, several methods have been recently proposed, for gait-based person recognition (usually referred to simply as gait recognition) from motion capture sequences. In Tanawongsuwan and Bobick [19], the proposed method for gait recognition is based on the analysis of the trajectories of lower body joint angles, projected onto the sagittal plane. At the first step of the method, the joint angles are estimated by fitting a skeleton model to the sensor measurements. Thereafter, the trajectories are normalized with respect to duration and walk cycles, by means of a segmentation technique and Dynamic Time Warping. Finally, the trajectories are classified using a nearest-neighbor classifier, based on Euclidean distance. In Lin et al. [12], motion capture data are combined with measurements from force plates, in order to combine both kinematic and kinetic data. As a result, the features representing the data include joint angles and angular velocities, as

☆ This paper has been recommended for acceptance by A. Fernandez-Caballero.

* Corresponding author. Tel./fax: +30 2310996361.
E-mail address: eftifot@aiia.csd.auth.gr (E. Fotiadou).

http://dx.doi.org/10.1016/j.patrec.2014.06.005
0167-8655/

well as forces applied on joints. Classification of gait data is performed using Self Organizing Maps (SOMs). Additionally, the importance and contribution of each feature is investigated, in order for the factors that cause differences in gait to be determined.

The method presented in Das et al. [3] is suitable for both classification of gait type (walking, running, jogging and limping) and person recognition. Deriving from the fact that the perception of human motion by an observer relies on the detection of specific "motion features", representing relative motion of body parts, a two-stage PCA scheme is applied on the motion data. The first stage of PCA is applied on the data, represented by joint angles and velocities, in order for a trajectory on a low-dimensional manifold to be extracted. The second stage of PCA detects the variability in the shape of this manifold across individuals or gait types. In Preis et al. [17], gait recognition is performed on motion data acquired from recordings with the Kinect sensor. The 3D positional joint data are used to extract both static (e.g. height, body part lengths) and dynamic (speed, step length) features. Gait data are classified by three different types of classifiers, namely Naive Bayes, 1R and C4.5. Also, the influence of the different features on the recognition rate is investigated. In Świtoński et al. [21] the trajectories of specific parts of the human body during walking, referred to as gait paths, are used for the extraction of features suitable for person recognition. From the motion capture data, four different kinds of features are extracted and subsequently classified using a Naive Bayes and a k-Nearest Neighbor approach.

The two methods proposed in this paper follow a more general approach to the person recognition problem, in comparison to the aforementioned methods. Recognition is based on motion capture data representing a repertory of different classes of human actions, such as waving, sitting down or standing up, and not solely on gait. Thus, the proposed approaches broaden the applicability of movement-based person recognition methods, to cover a large set of actions. As a matter of fact, the proposed approaches indicate that, other human actions apart from walking bear significant person-specific characteristics, that allow person recognition with high recognition rates. The first proposed algorithm is based on the hypothesis that, motion sequences of the same action performed by the same person, exhibit strong similarity between successive frames in one or more segments within them, which is expressed through specific patterns. In order to classify motion capture sequences to distinct humans, we developed a scheme for similarity estimation between such sequences. The second algorithm we propose consists a Bag of Words (BoW) approach, that combines dimensionality reduction of the motion data as a pre-processing step.

2. Correspondence-based person recognition

As already mentioned, the first proposed method for activity-based person recognition is based on the similarity between two motion capture sequences and comprises of two distinct steps:

1. The construction of a correspondence matrix, that describes which frame in the second sequence is the most similar to each frame in the first sequence.
2. A process of calculating a similarity score between two sequences from the correspondence matrix.

Person recognition is subsequently performed, by classifying test motion sequences, using an 1-Nearest Neighbor classifier. The workflow of the correspondence-based method is summarized in Fig. 1. In the following subsections the aforementioned steps are described in detail.

2.1. Correspondence matrix construction

Let us consider two motion capture sequences denoted with $\mathbf{X}_s = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$ and $\mathbf{Y}_s = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$, consisting of M and N frames respectively. A frame in a sequence consists of the rotation angles of each joint and describes the pose of the human body at a certain time instance. Some frames from motion sequences describing the activities "deposit floor" and "sit down chair" are illustrated in Fig. 2. In order to construct the correspondence matrix, the distances from each pose $\mathbf{y}_i$ of the sequence $\mathbf{Y}_s$ to every pose $\mathbf{x}_i$ in sequence $\mathbf{X}_s$ are calculated.

For this purpose, we use a distance measure, based on the logarithmic representation of the rotational data of each joint. Let us assume that a pose is represented by J unit quaternions ($\hat{q}_k$), each of them describing the rotation on one of the J joints. A unit quaternion can be projected to the tangent space at some reference point of the 3-sphere that the unit quaternions lie on. The reference point we selected for the projection of the data, was the sample mean $\hat{q}_m$ of quaternions for each joint calculated over a number of frames, which was estimated as the quaternion that minimizes the sum of its squared distances from the other quaternions, as described in Johnson [10]. The aforementioned projection is performed by applying a logarithmic mapping:

$$\log^{(\hat{q}_m)}(\hat{q}) = \ln(\hat{q}_m^* \times \hat{q}), \tag{1}$$

where $\times$ denotes the quaternion multiplication and $\hat{q}_m^*$ is the conjugate of the unit quaternion representing the sample mean. In this way, quaternions can be mapped to 3D points in Euclidean space. Consequently, the distance between two rotations represented by quaternions, can be approximated by the Euclidean distance between two points in $\mathbb{R}^3$. Therefore, each joint rotation can be represented by a 3D point $\mathbf{P} = \{p_1, p_2, p_3\}$, and a pose of a skeleton consisting of J joints can be denoted as $\mathbf{x} = \{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_J\} = \{p_1, p_2, \ldots, p_{3J}\}$, i.e. as a vector of $3J$ elements. The distance between two such poses $\mathbf{x}$ and $\mathbf{y}$ can then be estimated by the Euclidean distance between the two pose vectors:

$$d^{Log}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{3J} (p_i - q_i)^2} \tag{2}$$

The calculated distances between all pairs of poses in the two sequences $\mathbf{X}_s$, $\mathbf{Y}_s$ are used to construct a correspondence matrix of dimensionality $M \times N$, denoted with $\mathbf{C}$. The rows/columns of $\mathbf{C}$ correspond to poses of sequence $\mathbf{X}_s$, $\mathbf{Y}_s$ respectively. For each pose $\mathbf{x}_i$ of $\mathbf{X}_s$, the nearest pose $\mathbf{y}_j$ of $\mathbf{Y}_s$ is found and the element (i, j) of $\mathbf{C}$ is set to one, whereas all other elements (i, k), $k \neq j$ of the ith row are set to zero.

The result of this process is, that $\mathbf{C}$ exhibits distinct structures depending on the similarity between the two sequences under examination. When the two compared sequences describe movements of the same class (e.g. two walking sequences) the correspondence matrix contains diagonal segments of ones, of various lengths, either continuous or interrupted, since successive poses from one sequence are in general most similar to successive poses from the other. Specifically, these diagonal segments extend from the upper left to the bottom right of the matrix. In case that the two sequences depict the same movements performed by different subjects, these diagonal segments tend to be smaller in length and weaker in terms of slope. When the sequences describe motions of different classes, irrespective of whether they come from the same subject or not, there are two possibilities: First, there may exist long vertical lines, implying that many poses in sequence $\mathbf{X}_s$ are matched to the same pose in $\mathbf{Y}_s$. This is often the case, when the