Gait characterization using dynamic skeleton acquisition

This is the author's manuscript

Original Citation:

Availability:
This version is available http://hdl.handle.net/2318/140701 since 2015-12-11T15:06:49Z

Publisher:
IEEE

Published version:
DOI:10.1109/MMSP.2013.6659329

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)
This is an author version of the contribution published on:


The definitive version is available at:

Abstract—Human gait is an important biometric feature for automatic people recognition. Biometric methodologies are generally intrusive and require the collaboration of the subject in order to perform accurate data acquisition. Gait, instead, can be captured at a distance and without collaboration. This makes it an unobtrusive method for recognizing people in video surveillance systems. In this paper we propose a method to characterize walking gait using three-dimensional skeleton information acquired by the Microsoft Kinect sensor. A set of static and dynamic features correlated to human gait are extracted by the estimated skeleton joint positions. Moreover, we proposed to describe joints positions in a coordinate reference system oriented according to the walking direction to better represents the movement of human body. Using unsupervised clustering over a set of 20 subjects we analyze the effectiveness of the selected features in discriminating people gaits. It turns out that a few dynamic parameters involving the movement of knees, elbows and head are good candidates for robust gait characterization.

I. INTRODUCTION

Systems for automatic identification of persons have important applications in many fields, such as access control, surveillance and forensic evaluations. Human recognition can be performed using biometric techniques based on physiological characteristics like iris, hand, face, fingerprint and on typical behaviors like handwriting or gait. Most of biometric techniques require expensive technologies, besides the cooperation of the subject. On the other hand, in surveillance application or forensic analysis, collaboration from the subject may not be possible or expected; in the forensic case his/her awareness is not even desired to avoid biasing the acquisition. Therefore, the design of gait recognition methods using non intrusive data acquisition, e.g. not requiring the person to wear markers or follow particular examination procedures, are of paramount interest. Moreover, especially in video surveillance application, it is very important to perform data acquisition remotely using cheap sensors.

As a consequence, in this study we selected the Microsoft Kinect sensor as a widespread and cheap technology able to monitor body motion and gestures and satisfying the requirements discussed above. The Microsoft Kinect, originally conceived in 2010 as a peripheral for the Xbox 360 gaming console, has recently attracted a lot of interest from the scientific community. Given its potential, in 2012 Microsoft has produced a new sensor expressly devoted to scientific research.

Kinect is a low-cost, non-intrusive system provided with an RGB camera, an infrared (IR) emitter and an IR depth sensor allowing to capture depth image that measures the distance between an object and the sensor. It also includes a multi-array microphone to record audio as well as to find the location of the sound source and the direction of the audio wave. Moreover, it comprises a tilt motor that changes of $\pm 27^\circ$ the vertical range of the field of view (with a viewing angle of $43^\circ$ vertical by $57^\circ$ horizontal). Exploiting depth information one is able to extract in real time the skeleton data [1] in a map of 20 body joints as shown in Fig. 1.

The major contributions of our work are the definition and analysis of a set of features act to describing human gait. Such features can be extracted using the position of the above mentioned skeleton joints. Moreover, a proper transformation of the sensor reference system is proposed to get more immediate representation and characterization of joints trajectories. Our experimental analysis on a set of 20 subjects show that a few important gait characteristics are very effective in discriminating among different walking behaviors. Indeed, using a simple unsupervised clustering algorithm we show that it is possible to characterize people with reasonable accuracy. In particular, we found out that some gait parameters related to the dynamic behavior of the knees, elbows and head play a key role for gait characterization.

The paper is organized as follows. Sect. II presents an overview of related work in gait recognition systems. In Sect. III the proposed approach is presented in details, while in Sect. IV our experiments and the achieved results are discussed. Our conclusion and future work are outlined in Sect. V.

II. RELATED WORK

Gait recognition is a relatively new biometric technique [2], [3] that gained in recent years more interest from Computer Vision community, due to its advantages compared to other
biometrics [4], [5]. There are two main approaches for recognizing gait: model-based and model-free [6]. Model-based approaches describe human body by using a stick figure and a set of joints. During walking it is possible to get some parameters like trajectories of joints and angles between them, length of limbs, and other data useful for classification. This is a robust approach but suffers from some drawbacks, due to the difficulty in defining the model and estimating parameters, which may require high computational efforts [4], [5].

The model-free approaches, instead, process the silhouette of the walking person in order to extract useful features for characterizing gait. Thanks to low computational costs and easy implementation, this is a less consuming technique, when compared to model-based approaches. On the other hand, this method gives good results only if the input data are noiseless.

In our work, we propose a model-based technique using data captured by the Microsoft Kinect sensor, that overcomes some drawbacks typical of these approaches. In fact, creates a body model represented by a skeleton made up of 20 joints, tracked in real time. Thanks to these features, in recent years the use of Kinect for gait recognition purposes has gained considerable success. In particular, we compute the length of limbs (arms and legs) calculating the distance between the relevant joints. We estimate the length of the arm $\alpha$, averaging all the left and right arm lengths for each subject acquisition. The length of limbs, and other data useful for classification, can be defined thanks to the knowledge of the positions of the 20 joints of the skeleton model shown in Fig 1; the joints coordinates are given in the 3-D coordinate system $xyz$ centered on the sensor, with $x$, $y$ and $z$ representing the horizontal, vertical and depth direction as shown in Fig. 2;

- the floor-clipping-plane vector, which contains the coefficients of an estimated floor-plane equation in the coordinate system of the sensor. The plane is given in the canonical form $Ax + By + Cz + D = 0$, where $(A, B, C)$ is the normal vector to the plane and $D$ is the height of the camera center with respect to the floor. This plane is an estimate of the floor plane where the subject is walking, whose equation clearly depends on the sensor tilt;
- a time function, that represents the corresponding frame number, being the frame rate approximatively 30 frames per second (fps).

In the following the joints will be referred to using the notation reported in Tab. I. A rich set of biometric measurements can be defined thanks to the knowledge of the positions of the 20 joints. Being the joint positions the output of complex computer vision system, the obtained estimates are potentially affected by measurement noise. To limit noise effects the time series corresponding to the positions of a given joint is processed by a median filter. The size of the median filter has been selected so as to smooth out anomalous points without impacting on the dynamic features of the time series.

In the following we will distinguish between static biometric parameters, e.g., height, and dynamic gait parameters that depend on the dynamic and mutual positions of one or more joints.

A. Static parameters

Using skeleton data, we extract standard biometric measures. In particular, we compute the length of limbs (arms and legs) calculating the distance between the relevant joints. We use the angle $\theta$ to each acquisition trial.
Furthermore, we estimate the average value of the subject height \( \gamma \), evaluated as the summation of all the mutual distances between pairs of joints going from the head to the foot.

B. Dynamic parameters

Since our objective is gait characterization we clearly need to extract a set of measures related to the dynamic behavior of the joints.

First of all, we extract standard gait features such as stride length \( \delta \) and the walking speed \( \nu \). The stride length is defined as the distance between two stationary positions of the same foot while walking. To this end we have implemented a simple stride detection mechanism based on the detection of left (and right) foot movements. The foot is considered in movement if its position has changed (with a certain tolerance \( d_{tol} \)) with respect to its position in previous \( fn \) frames. The stride length is estimated as the distance between two stationary positions of a left (or right) foot. The details of the procedure followed to estimate the stride length are given in Algorithm 1. Based on our experimental analysis we set \( fn = 3 \) and \( d_{tol} = 5 \) cm. The gait feature that in the following will be referred to as \( \delta \) is computed as the average of all the available estimates on both right and left stride lengths. The walking speed \( \nu \) is then obtained as the ratio between the distance covered by the center of mass during a single stride and its duration. Also in this case we compute a unique gait feature averaging all the acquired speed estimates.

Beside standard stride length and speed characterization is quite intuitive that gait can be characterized by a number of different behaviors involving the movement of the whole body, e.g. how much one oscillates with the head or the shoulders, the distance between the legs, etc. To investigate the impact of such features we define a proper transformation of Kinect coordinates system so as to be able to project the joint movements onto 3 planes that are respectively parallel and normal with respect to the walking direction. Given the limited depth range allowed by the Kinect any subject can be tracked only for a limited amount of time (typically corresponding to three or four strides). It is quite reasonable to assume that, during such limited observation period, the subject walks straight. This allows us to define a unique walking direction per each acquisition.

As a consequence, a novel system of coordinates where the plane \( xz \) coincides with the floor-clipping-plane estimated by the Kinect, axis \( y \) is the normal vector to the floor and axis \( z \) points in the walking direction, is defined. The unit vector on axis \( y \) is therefore \( n_y = (A,B,C) \). Then we define the walking direction as the line described by the movement of the center of mass. To define the unit vector \( n_z \) of axis \( z \) we project such direction on the floor plane. To make the estimation of the walking direction automatic and robust a few positions of the center of mass have been averaged at the beginning and end of the observed walk, respectively.

Finally, the missing orthogonal \( x \) axis is found defining the normal unit vector \( n_x = n_y \times n_z \). Such new orthonormal triple defines the rotation matrix

\[
R = \begin{bmatrix}
  n_x \\
  n_y \\
  n_z
\end{bmatrix}
\]

from the Kinect coordinates system to the gait coordinate system.

From now on we consider all the joints positions expressed in the novel gait coordinates reference system. During experimentation we noted that different subjects exhibit different patterns of certain joint positions observed on the plane normal to the walking direction, i.e. \( xy \) plane. As an example, on the plane \( xy \) one can observe the lateral oscillation of the head or the movement performed with the hands. To characterize the dynamic behavior on plane \( xy \) we computes a set of features corresponding to average value and variance of the joint positions along \( x \) and \( y \) axis as follows:

\[
\mu_{k,x} = \frac{1}{n} \sum_{i=1}^{n} J_{k,i}(x)
\]

\[
\mu_{k,y} = \frac{1}{n} \sum_{i=1}^{n} J_{k,i}(y)
\]

\[
\sigma_{k,x}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (J_{k,i}(x) - \mu_{k,x})^2
\]

Fig. 2. The Kinect 3D coordinate system.
TABLE II
GAIT FEATURES NOTATION.

<table>
<thead>
<tr>
<th>Label</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Length of the arms</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Length of the legs</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Height of the person</td>
</tr>
<tr>
<td>$\delta$</td>
<td>The strides length</td>
</tr>
<tr>
<td>$\nu$</td>
<td>The walking speed</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Mean distance between the elbows</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>Mean distance between the knees</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Mean distance between the hands</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Mean distance between the ankles</td>
</tr>
<tr>
<td>$\sigma_{k,x}^2$</td>
<td>Variance of Head movement in $x$ direction</td>
</tr>
<tr>
<td>$\sigma_{k,y}^2$</td>
<td>Variance of Head movement in $y$ direction</td>
</tr>
<tr>
<td>$\sigma_{k,x}^2$</td>
<td>Variance of right Shoulder movement in $x$ direction</td>
</tr>
<tr>
<td>$\sigma_{k,y}^2$</td>
<td>Variance of right Elbow movement in $y$ direction</td>
</tr>
<tr>
<td>$\sigma_{k,x}^2$</td>
<td>Variance of left Shoulder movement in $x$ direction</td>
</tr>
<tr>
<td>$\sigma_{k,y}^2$</td>
<td>Variance of left Elbow movement in $y$ direction</td>
</tr>
<tr>
<td>$\sigma_{k,x}^2$</td>
<td>Variance of right Knee movement in $x$ direction</td>
</tr>
<tr>
<td>$\sigma_{k,y}^2$</td>
<td>Variance of right Knee movement in $y$ direction</td>
</tr>
</tbody>
</table>

$$\sigma_{k,y}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (J_{k,i}(y) - \mu_{k,y})^2$$

where $J_{k,i}$ is the position of the $k$-th joint at time $i$ with $i = 1, \ldots, n$.

Finally, we extract as other possible characteristics of human gaits the mutual distance between corresponding left and right joints, such as elbows, hands, knees and ankles. We decided not to use feet positions because we noticed that these are very critical points to be tracked due to occlusion and noise. To get a single measurement per acquisition we compute global averages of such distances.

All the features analyzed and evaluated in the following experimental section have been collected in Tab. II to easy the task of the reader.

IV. EXPERIMENTAL results

The goal of our experimental analysis is to find out which of the parameters defined in Sect. III are more appropriate for gait description and classification.

To this end, we acquired the skeleton model of 20 subjects using Kinect for Windows and SDK 1.6. In order to rank the effectiveness of the gait features for recognition application, we selected 5 pairs of people having compatible heights. The subjects comprises 12 males and 8 females. The Kinect sensor has been placed at about one meter from the ground and every subject has been asked to walk towards the camera in a natural manner. Every acquisition has been repeated 10 times, thus recording a total of 200 gait samples. Each sample has been analyzed according to method described in Sect. III. In particular joints positions have been filtered with median filter with size 7 and all the gait features shown in Tab. II have been calculated.

In Fig. 3 a sample acquisition is shown where the trajectories described by the skeleton joints during time are projected on the $xy$ (front) and $zy$ (sagittal) planes, respectively.

Clustering experiments have been worked out using the $K$-means unsupervised algorithm [22] with Euclidean distance. Given $K$ desired clusters, $K$-means defines $K$ centroids and refines them iteratively. At each iteration data samples are clustered according to their distance with respect to the centroid. In turn centroids are updated based on the resulting clusters.

First of all, we tested the ability to discriminate between two subjects in a pair using $K$-means with $K = 2$. To this end we have formed different sets of features starting with all static characteristics, e.g. height, arms and legs and then we have included more dynamic parameters. The goal is to isolate those features that play an important role in gait characterization. In Tab. III the results achieved using different combinations of features are shown. The tested sets of features have been selected by taking into account the human locomotion. Producer accuracy, i.e., the ratio between correctly classified samples over number of trials, has been employed as a metric of clustering performance. The third column of Tab. III show the average producer accuracy obtained on all the 5 pairs of subjects. It can be noted that feature sets 1 to 7, that use different combinations of static and dynamic characteristics, exhibit generally lower clustering accuracy than previous combinations. Adding ankles and hands average distances (see sets 9, 10, 11) does not help because hands may be less important gait characteristics whereas Kinect ankles tracking is not precise. Finally, all other attempts aiming at improving pair clustering accuracy based on other dynamic features, e.g. average and variance of selected joints, exhibit minimal variations (visible only on fractional part) with respect to combination 8.

Fig. 4 shows an example of the dispersion of 20 gait samples of one pair of subjects with respect to the corresponding centroids in the plane $\epsilon \kappa$, corresponding to the best clustering features. The top figure refers to a typical case, whereas bottom figure is the worst experimented situation. It can be noted that the two pairs can be correctly discriminated in such a plane.

In the second set of experiments we have followed the same clustering strategy on the whole dataset, comprising 200 samples of 20 different subjects. Executing $K$-means with $K = 20$, the clustering accuracy has been obtained for various features combinations as shown in Tab. III (fourth column). In this trickier clustering task the best producer accuracy turns out to be about 50%, that is significantly better than chance (5%). The behavior of the different combinations of gait features is quite similar to that observed for pair classification. In particular, it is worth pointing out that again the domain $\epsilon, \kappa$ yields a high clustering accuracy. Furthermore, also the vertical and side variance of the head (set 12) and vertical variance of left and right knees (set 16, 20) seem to improve the results.
V. CONCLUSIONS AND FUTURE WORK

In this paper we have introduced a rich set of features allowing one to characterize human gait by means of a tracked skeleton model. Our experiments, worked out with the Kinect sensor, show that a few features play a very important role in gait characterization. In particular, we have found out that some dynamic characteristics of gait related to the positions of the knees, elbows and head are very helpful in distinguishing different walking subjects. On going work is based on the application of mutual information [23] for the selection of most important gait characteristics. Future works are the design of a reliable gait recognition system for video-surveillance and forensic applications. Moreover, for reproducibility of the experiments and comparison of research results we plan to build a public database of gait samples.

REFERENCES


Fig. 4. Sample dispersion in plane $\epsilon_\kappa$ and corresponding centroids of two pairs: typical (top) and worst (bottom) case.


