Kinect-based Non-intrusive Human Gait Analysis and Visualization

Nguyen-Luc Dao, Yuzhe Zhang, Jianmin Zheng and Jianfei Cai

School of Computer Engineering, Nanyang Technological University, Singapore

{nldao,zhang.yz,asjmzheng,asjfcai}@ntu.edu.sg

Abstract—Home healthcare becomes more and more important with the increased aging population. The advent of various low-cost sensing devices makes it tempting to develop low-cost, non-intrusive systems to monitor the variations of human being. In this paper, we describe a Kinect-based gait analysis and visualization system as a case study in this direction. The system uses depth images and skeleton captured by the Kinect to generate a BVH file recording the motion information, extracts features of gait for detecting abnormal gait, and customizes the 3D body model for personalized motion visualization. Compared to previous work in this area, the proposed system has advantages since it integrates gait classification and visualization which may bring new possibilities in healthcare. The experiments show that our proposed system achieves accurate gait classification as well as flexible personalized 3D visualization.

I. INTRODUCTION

Home healthcare is supportive care provided for helping adults, seniors and pediatric clients in their home. Distinctly from medical care or custodial care, home healthcare provides professional services such as recovery after a hospital stay and support to remain healthy at home, which avoid unnecessary hospitalization for clients. Integration of technologies such as sensors, networks, digital media makes in-home healthcare a digitalized system. The conceptualization of home healthcare system can provide doctor’s advices online, use various sensors to monitor behavior of clients, analyze captured data to estimate health condition and notice doctor if necessary. The core techniques lie in data capture, analysis and visualization. Some companies such as BioSensics [1], for example, collect motion data through portable wear devices or sensors and provide solutions to do fast quantitative gait analysis. For the purpose of pragmactically, developing low-cost, non-intrusive sensing systems to monitor the health conditions of human being is in a growing need.

This paper takes gait joint motion, which is related to biometrics and clinical science [2], [3], [4], [5], [6], as a case study to demonstrate how new technologies with low-cost sensing devices could provide new possibilities and potential for home healthcare applications. In particular, we are interested in Kinect, a low-cost commodity depth sensor, which can capture RGB-D images and motion data of human body. There are already some Kinect-based healthcare systems being developed. For example, in [7], the data captured by Kinect is used to evaluate walking quality of clients. In [8], [9], Kinect is used for in-door fall detection of elderly group.

While the existing Kinect-based healthcare systems mainly use motion data for motion classification, in this paper we present a Kinect-based system for not only gait analysis but also visualization, which is able to detect abnormal gait of a subject and meanwhile provides a personalized 3D visualization of the walking motion for further diagnosis. Technically, the system consists of two components: motion data analysis and motion visualization. Once the motion of the subject is captured by Kinect, the motion data analysis component processes the data and encodes the data into a Biovision Hierarchy (BVH) file [10]. Then the gait features are extracted and the classification is performed via a learning process. The motion visualization component uses 3D graphics and animation techniques to reconstruct the 3D body model of the subject from the captured data. The reconstructed model combined with the BVH data can be used to visualize the walking of the subject from any view and can be sent to medical specialists for further examination. Such personalized 3D visualization could provide more perspectives and information than a single RGB video.

The main contribution of this work lies in the integration of Kinect-based motion analysis and motion visualization. The existing works study either Kinect-based motion analysis or motion visualization, but few of them integrates both. Another contribution is the constructed RGB-D dataset in BVH format for gait analysis, for which we are unaware of any similar publicly available dataset. We will release our dataset to the public.

II. RELATED WORK

Utilizing Kinect to aid home healthcare applications has emerged in recent years. In [11], Gabel et al. use skeleton provided by Kinect to analyze human gait. They extract a large set of features including the direction of progress and the direction tangent at each joint of the skeleton. The features are fed to a regression model to predict stride duration and angular velocities. Compared with wearable sensors, the method obtains the results with the mean difference less than 1%.
In [12], Stone et al. present a method to measure stride length with Kinect depth images and compare the result with Vicon marker-based capture system. They achieve a maximum absolute percentage difference of 9.4%. In this research, we also measure foot location, but we use both depth images and skeleton from Kinect. The experimental results show that our method obtains a lower maximum absolute difference.

In [7], Arai and Asmara propose a method to classify disable gait quality using a single Kinect. Their method captures the left view of a walking subject and encodes the motion in BVH file format. The left knee angle features are extracted and fed to Support Vector Machine (SVM) to classify gait into five categories ranging from normal gait to neuropathy gait. They report a classification accuracy of 86.63%. This is the most related work to the analysis component of our system. Different from it, our method captures the front view of subjects, which allows us to extract features from both left and right legs at the cost of more limited in capturing area.

Meanwhile, a lot of work has been done for 3D body reconstruction based on Kinect captured information. Tong et al. [13] use multiple Kinects to build a full body scanning system that deforms a template model with successive frames. The system requires calibration and complex setup. Weiss et al. [14] propose a single Kinect-based body reconstruction application using the SCAPE model [15] by estimating model parameters and meanwhile minimizing the distance between the projected 3D body contour and the image silhouette. The low-resolution and noisy properties of RGB-D data captured by Kinect restrict the reconstruction quality. Different from these works, the anthropometry based body adaptation method [16] constructs customized body models by adapting a template based on partial measurements, which is flexible in practice. In this research, we use the simple method of [17] to obtain measurements from body point cloud captured by Kinect and then incorporate the results into [16] to generate customized 3D body model.

III. OVERVIEW OF THE PROPOSED SYSTEM

Our proposed system needs only one Kinect 1.0. It can be placed on a chair or a coffee table in a space with a dimension of about 4 meter by 4 meter, e.g. living room, as shown on the top of Fig. 1. In particular, the user is requested to stand in front of the Kinect in T-pose and four views (front, left, back and right) are captured. The captured data are used to build the 3D model of the user, which only needs to be done once in an offline manner. During online monitoring or testing, the user is requested to walk a few steps towards the Kinect, where the motion data are captured for gait analysis and visualization. Our system provides a non-intrusive way for healthcare of gait, i.e. detecting whether the gait of the user is normal or abnormal and generates a personalized 3D visualization of the motion for further diagnosis. Fig. 1 shows the proposed system consisting of two major components: motion analysis and motion visualization, which integrate various techniques such as data capturing, data filtering, SVM, body reconstruction, and animation.

The motion analysis component is to process motion data captured by Kinect, extract feature of the gait and assess whether the gait is normal or not. The motion of a user is captured through depth and skeleton streams of Kinect. The motion data are then filtered and encoded in BVH format which is used for both feature extraction and motion visualization. Feature data include the bounding box of hip movement, foot steps, and X-rotation angles of the knees. Based on the extracted features, the trained SVM is applied to classify the gait into either normal or abnormal.

The motion visualization component accounts for visualizing the walking motion personalized for the user. For this purpose, we extract the anthropometric measurements of the body from the RGB-D images of the T-posed user in the four views. The measurements are used to generate a customized 3D body model for the user. The reconstructed 3D body model and the generated BVH file can be used together to visualize walking motions. These outputs can also be sent to medical specialists for further diagnosis.

In the next two sections we elaborate the technical details of the two major components.

IV. MOTION ANALYSIS

A. Motion capturing and filtering

Referring to Fig.1, while the user is walking from location 1 to location 2, depth frames and skeleton frames are captured by the Kinect. The skeleton frames record the joints of the skeleton, which are mapped into the depth frames. We can calculate the 3D position \((X_P, Y_P, Z_P)\) of a joint point \(P\) in
each skeleton frame as

\[
\begin{align*}
X_P &= (x - c_x) \times d(x, y)/f_x \\
Y_P &= (y - c_y) \times d(x, y)/f_y \\
Z_P &= d(x, y)
\end{align*}
\]

where \(x, y\) are the indices of the skeleton point in the corresponding depth image \(d\), \((c_x, c_y)\), \(f_x\) and \(f_y\) are the optical center and focal lengths of the Kinect which can be estimated via calibration [18]. The skeleton data are then encoded into a BVH file, a popular file format for character animation, containing two parts: skeleton information and motion data.

Considering Kinect data are usually noisy, we filter the skeleton joint values including translation and rotation information simply by cubic B-spline filter: \(q'_i = \frac{1}{6}q_{i-1} + \frac{2}{6}q_i + \frac{1}{6}q_{i+1}\), where \(q_i\) is a skeleton joint value in frame \(i\) and \(q'_i\) is the filtered value. In this way, we can achieve real-time denoising with only one frame delay.

**B. Gait feature extraction**

To analyze gait characteristics, we extract three gait features from the recorded BVH file, including hip progression line, foot steps, and X-rotation angles of the knees.

**Foot step:** During the walking period, there is always one foot that serves as a pivot for the other foot to move. Since the pivot foot moves very little in a short time interval, if we can classify the input motion data into left or right pivot time intervals, for a left (right) pivot interval we can compute the location of the left (right) pivot foot by averaging their locations in the particular interval. Particularly, to classify a frame, we extract: (1) the Euclidean distances for the left and right foot joints respectively between the current frame and the two previous frames, (2) the angles formed by hip center, hip and knee, and (3) the angles formed by hip, knee and ankle for both left and right legs. Here we train a kernel based SVM with radial basis functions on a set of manually labelled frames.

To reduce the errors caused by misclassification, given a set of classified frames in one motion, a majority vote in a range of ten neighboring frames is further applied for each frame. As a result, the motion is segmented into left and right pivot intervals. The location of the pivot foot in an interval is obtained by averaging its locations over all the frames in the interval. Finally, the average and the variance of \(x, y\) distances between the left and the corresponding right pivot foot locations over different time intervals are used as the foot step features for gait analysis (see Fig. 2).

**X-rotation angle of the knees:** A BVH file starts with the position of the root and other joints are represented by translation and rotation information relative to their father joints hierarchically. For example, the left hip is the father joint of the left knee. Thus, the position \(P\) of the left knee can be computed from the position \(P'\) of the left hip as follows:

\[
P = \begin{bmatrix} Z & Y & X \\ Y & X & 0 \\ Z & 0 & 1 \end{bmatrix} P'
\]

where \(X, Y, Z\) are the matrices of the rotations about the \(x, y, z\) axes, respectively, and \(T\) is the translation vector.

\]

The X-rotation angles of knees refer to the rotation angles of the matrix X, which can be extracted from the BVH file. It is observed that the left and right knees usually have similar X-rotation angles but with a shift in time if the motion is balanced (see Fig. 3). Thus we compute the average X-rotation angles of the left and right knees. Then, the average and the variance of peak height and the distances between adjacent peaks are used as features for gait analysis.

**Hip progression line:** The hip point is the root of skeleton structure in BVH format. Therefore, 3D positions of the hip point during the motion can be directly extracted from the BVH file. All these positions form the hip progression line. We choose the size of the bounding box that bounds the hip progression line as a feature for gait analysis.

**C. Gait analysis for healthcare**

Once the gait feature data are extracted, what is left is just a binary classification problem, i.e. classifying each feature vector into a normal or abnormal gait, for which we use linear SVM. We train the SVM model with a set of abnormal and normal walking motions. Individual dimensions of the feature data are normalized to the range of \([0,1]\) so that each of them contributes equally to the distance in the SVM model.
Body reconstruction is to construct a triangular mesh model for the body, which is convenient for animation and visualization. The basic idea of our algorithm is to adapt a feature embedded body template under the control of full measurement set $M$. The template mesh $T$ as shown in Fig. 4(a) is represented by a triangular mesh $\{V, E, F\}$ attached with several features including feature curves and joints, an embedded skeleton as well as girth and length measurements. The skeleton is built from a subset of feature joints according to the structure used in the Kinect. A girth is defined as the circumstance of a cross section in the plane perpendicular to the skeleton of the body, while the sequence of intersection points $p_k$ for the cross section form feature points of a girth. A length measurement refers to the length of a skeleton segment, which is controlled by two corresponding feature joints. According to the structure of skeleton, the lengths could be organized hierarchically and girths could be grouped according to the corresponding bones.

During initial length and girth adjustment, the joints and feature points of girths are updated according to input measurements. For length adjustment, the end joint of the bone is translated and the transformation of a parent bone is easily propagated to all its descendants. Meanwhile, any translation, scale and rotation of the bone will cause a transformation of $p_k$ on related girth measurements. The girth adjustment is to scale the sequence of feature points into $\tilde{p}_k$ to make the circumstance match the required value. The body adaptation is then achieved by computing the new position of vertices $\{\tilde{v}_i\}$ that minimize the following objective function to make the mesh interpolate all updated feature points in $P_G$ while minimizing the change of the Laplacian of each vertex:

$$\arg\min_{\{\tilde{v}_i\}} \|\delta(\tilde{v}_i) - \delta(v_i)\|^2$$

$$\text{st.} \ (1 - \lambda)\tilde{v}_i + \lambda\tilde{v}_j = \tilde{p}_k \quad \text{for} \quad p_k \in P_G$$

where $\delta$ is a discrete Laplacian operator and a point $p_k$ in $P_G$ is on the edge connecting mesh vertices $v_i$ and $v_j$, which can be represented as a linear combination of the two vertices.

C. Motion visualization

During the body reconstruction, the joints of the template are relocated. As a result, we could obtain the skeleton of the customized body model simultaneously. Moreover, since the topology of the customized model is the same as the template, the skinning information could inherit from the template with minor adjustment using the method of [19].

Based on the skinning information and the adapted skeleton, the motion data could be used to animate the customized model using the standard linear blend skinning method. That is, for every frame, we calculate the transformed position of each mesh vertex $v_j$ using the following equation:

$$v_j' = \sum_i w_{ij}^i T^i(v_j)$$

where $T^i$ is the transformation of the $i$-th bone and $w_{ij}^i$ is the weight of the $i$-th bone for vertex $v_j$. In this way, the animation of the customized model can be generated accordingly (see Fig.5 for an example).
### TABLE I

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Pivot</td>
<td>0.846</td>
<td>0.238</td>
<td>0.765</td>
<td>0.846</td>
<td>0.803</td>
<td>0.804</td>
</tr>
<tr>
<td>Right Pivot</td>
<td>0.762</td>
<td>0.154</td>
<td>0.844</td>
<td>0.762</td>
<td>0.800</td>
<td>0.804</td>
</tr>
<tr>
<td>Weighted Avg.</td>
<td>0.802</td>
<td>0.194</td>
<td>0.806</td>
<td>0.802</td>
<td>0.802</td>
<td>0.804</td>
</tr>
</tbody>
</table>

### VI. EXPERIMENTAL RESULTS

We conduct some experiments to evaluate the proposed techniques and system.

#### A. Evaluation of footsteps

Our training dataset for left and right pivot classification consists of 415 left pivot frames and 453 right pivot frames. These frames are manually labelled based on the BVH skeleton visualization. Table I shows the results of left and right pivot classification with 10-fold cross validation. It can be seen that our method correctly classifies 80.2% of frames.

In order to examine the accuracy of foot step locations, we drew marks on the floor and asked two volunteers to step on these marks when they walked. Each volunteer walked three times with different sets of marks. The distances between corresponding marks are considered as the ground truth for stride length. The experimental results show that, compared with the ground truth, our method achieves an average absolute distance difference of 2.8 cm and an maximum absolute difference of 4.3 cm. The percentage for the maximum difference is 5.18%, which is lower than 9.4% reported in [12].

#### B. Evaluation of abnormal gait detection

To the best of our knowledge, there is no publicly available BVH dataset, especially Kinect-captured dataset, for gait analysis. Thus, in this research we create a dataset for the purpose, which will be released to public.

Our dataset consists of 20 normal and 30 abnormal walking motions. Ten volunteers were asked to walk five times in different ways: (1) normal walk, (2) normal walk with shorter steps, (3) waddling walk [20], (4) walk with left foot pain, and (5) walk with right foot pain. We captured the data at two different locations with different illumination conditions, and the distances between the Kinect and the floor were 43 cm and 58 cm for the two locations respectively.

Due to the limited data, to evaluate the accuracy of gait classification, we use leave-one-out cross validation in which the SVM model predicts each motion after being trained with 49 remaining motions. Table II shows the gait classification results. It can be seen that 88% of motions are classified correctly by our method, which is higher than 86.63% reported in [7].

#### C. Motion visualization

Fig.5 shows five frames of a normal walking body model in the front view (top row) and in the side view (bottom row). This animation is generated by applying the BVH file to a customized body. Fig.6 and Fig.7 show the visualization results of the same body walking with left foot pain and with a waddling, respectively.

### VII. CONCLUSION

This paper has described a Kinect-based home healthcare system, which integrates motion analysis and motion visualization, for detecting abnormal gait and visualizing gait motion. The experimental results show that the proposed system calculates stride length with the error percentage less than 5.18% and achieves a gait classification accuracy of 88%. The skeleton information encoded in BVH files can be used to animate personalized 3D body models for flexible viewing.
REFERENCES