Detection of Abnormal Gait from Skeleton Data

Meng Meng[1], Hassen Drira[1], Mohamed Daoudi[1] and Jacques Boonaert[2]

[1]*Télécom Lille, CRIStAL (UMR CNRS 9189), Lille, France*
[2]*Ecole des Mines de Douai, Douai, France*

Keywords: Joint Distances, Abnormal Gait, Spatio Temporal Modeling.

Abstract: Human gait analysis has becomes of special interest to computer vision community in recent years. The recently developed commodity depth sensors bring new opportunities in this domain.In this paper, we study the human gait using non intrusive sensors (Kinect 2) in order to classify normal human gait and abnormal ones. We propose the evolution of inter-joints distances as spatio temporal intrinsic feature that have the advantage to be robust to location. We achieve 98% success to classify normal and abnormal gaits and show some relevant features that are able to distinguish them.

1   INTRODUCTION

The human activity analysis is a challenging theme due to the motion's complexity and diversity. At the same time, the recent improvement of of low cost depth cameras with real-time capabilities such as Microsoft Kinect have been employed and in a wide range of applications, including human-computer interfaces, smart surveillance, quality-of-life devices for elderly people, assessment of pathologies, rehabilitation, and movement optimization in sport (A. A. Chaaraoui and Flrez-Revuelta, 2012). Current methods for human action classification in this field have moved towards more structured interpretation of complex human activities involving abnormal gait in various realistic scenarios.

Due to the interpretation of human behavior have obtained successfully from (J. Shotton et al., 2011), researchers have explored different compact representations of human actions recognition and detection in recent years. The release of the low-cost RGBD sensor Kinect has brought excitement to the research in computer vision, gaming, gesture-based control, virtual reality, especially gait analysis. There are several works(Omar and Liu, 2013)(R. Slama and Srivastava, 2015)(M. Devanne et al., 2013) relied on skeleton information and developed features based on depth images for human motion. Some works made use of skeleton joint positions to generate their features such as(Dian and Medioni, 2011). They proposed a spatio-temporal model STM to analyze skeleton sequential data. Then using alignment algorithm DMW calcu-

lated the similarity between two multivariate time series. Based on STM and DMW, they achieved view-invariant action recognition on videos. Meanwhile the work (W. Li and Liu, 2010) presented a study on recognizing human actions from sequences of depth maps. They have employed the concept of BOPs in the expandable graphical model framework to construct the action graph to encode the actions. Each node of the action graph which represents a salient postures is described by a small set of of representative 3D points sampled from the depth maps.

Now human abnormal gait detection attracts more concern for earlier detection of human diseases. In the sense, the present research aims to apply the recent improvements in human gait analysis based on low-cost RGB-D devices. The method (A. Paiement et al., 2014) analyzed the quality of movements from skeleton representations of the human body. They used a non-linear manifold learning to reduce the dimensions of the noisy skeleton data. Then building a statistical model of normal movement from healthy subjects, and computing the level of matching of new observations with this model on a frame-by-frame basis following Markovian assumptions. Both of (J. Snoek et al., 2009) and (G. S. Parra-Dominguez and Mihailidis, 2012) analyzed the abnormal gait from skeletons information. (J. Snoek et al., 2009) used monocular RGB images to track of the feet by using a mixed state particle filter, and computed two different sets of features to classify stairs descents using a hidden Markov model. In this work(G. S. Parra-Dominguez and Mihailidis, 2012), they used binary classifiers of

133

harmonic features to detect abnormalities in stairs descents from the lower joints of Kinect's skeletons.

Along similar lines, (C. Alexandros and Flórez-Revuelta, 2015) detected joint motion history feature based on RGB-D devices and used the BagOfKey-Poses to classify temporal feature sequences for abnormal gait detection. They also recorded their own dataset for abnormal gait detection by Microsoft Kinect 2 camera. We use the dataset to test our approach.

We propose in this paper an approach for dynamic abnormal gait detection using frame-based feature and a memory of k previous frames. For $k = 0$, we achieve the recognition without any memory of previous frames. We demonstrate that the use of few distances can be enough for his classification problem.

The rest of the paper is organized as follows. Section 2 presents the spatio-temporal modeling of dynamic skeleton data. The features used for encoding the data, the classification method and an overview of the proposed approach are detailed in this section. Section 3 presents experimental results, discusses the relevant features generated by proposed method, then the conclusion and future works are given in the Section 4.

2 OUR APPROACH

The 3-D humanoid skeleton can be extracted from depth images (via RGB-D cameras, such the Microsoft kinect) in real-time thanks to the work of Shotton et al. (J. Shotton et al., 2011). This skeleton contains the 3-D position of a certain number of joints representing different parts of the human body and provides strong cues to recognize abnormal gait detection. We propose in this paper to recognize human abnormal gait based on the inter skeleton joints distances in each frame.

2.1 Overview of Our Method

An overview of the proposed approach is given in Fig.1. We first use skeleton position information from dataset. Then, a feature vector based on all the skeleton joints is calculated for each frame. Next the feature vectors of n successive frames are concatenated together to build the feature vector for the sliding window. Finally, the abnormal gait detection is performed by Random Forest classifier.
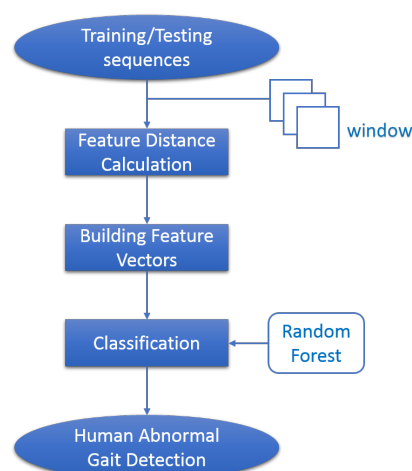

Figure 1: Overview of our method. Three main steps are shown.

2.2 Intrinsic Features

In order to have real time approach, one needs to use frame-based features. These features have to be invariant to pose variation of the skeleton but discriminative for abnormal gait detection. We propose to use the inter-joints distances.

Then we obtained our new skeleton information that is donated as S which contains 20 joints from the original data .

$$S = \{j_1, j_2, ..., j_{20}\}$$

V refers to the set of the pairwise distances between the joint a and joint b from S.

$$V = \{d(a, b)\}, a \in S, b \in S$$

Thus the feature vector is composed by the all pairwise distances between the joints The size of this vector is equal to $m \times (m-1)/2$, with $m = 20 : 20$ joints.

2.3 Dynamic Shape Deformation Analysis

To capture the dynamic of skeleton deformations across sequences, we consider the inter-joint distances computed at n successive frames. In order to make possible to come to the recognition system at any time and make the recognition process possible from any frame of a given video, we consider sub-sequences of n frames as sliding window across the video.

Thus, we chose the first n frames as the first sub-sequence. Then, we chose n-consecutive frames

starting from the second frame as the second sub-sequence. The process is repeated by shifting the starting index of the sequence every one frame till the end of the sequence.

The feature vector for each sub-sequence is built based on the concatenation of individual features of the n frames of the sub-sequence.

Thus, each sub-sequence is represented by a feature vector of size the number of distances for one frame times the size of the window n. For the sliding window of size $n \in [1, L]$ that begins at frame i, the feature vector is:

$$x = [V_i, V_{i+1}, ..., V_{i+n-1}],$$

with L the length of the sequence.

For $n = 1$, our system is equivalent to recognition frame by frame without any memory of previous frames. If $n = L$, the length of the video, our system will provide only one decision at the end of the video. The effect of the size of the window on the performance is studied later in experimental part.

2.4 Random Forest-based Real Time Human Action Interaction

For the classification task we used the Multi-class version of Random Forest algorithm. The Random Forest algorithm was proposed by Leo Breiman in (Breiman, 2001) and defined as a meta-learner comprised of many individual trees. It was designed to operate quickly over large datasets and more importantly to be diverse by using random samples to build each tree in the forest. Diversity is obtained by randomly choosing attributes at each node of the tree and then using the attribute that provides the highest level of learning. Once trained, Random Forest classify a new action from an input feature vector by putting it down each of the trees in the forest. Each tree gives a classification decision by voting for that class. Then, the forest chooses the classification having the most votes (over all the trees in the forest). In our experiments we used Weka Multi-class implementation of Random Forest algorithm by considering 100 trees.

3 HUMAN ABNORMAL GAIT DETECTION ON DGD DATASET (DAI GAIT DATASE)

3.1 Dataset and Experimental Protocol

The DAI gait dataset(C. Alexandros and Flórez-Revuelta, 2015) is collected in recording a front view

of a corridor which contains seven subjects walk towards the camera normally and abnormal gait. This dataset have two types of anomalies which are knees injured and feet dragged performed by the right and the left leg respectively. So there are four different abnormal gait types. With other four normal instances, each of the seven subjects made of 56 sequences totally. We report some skeleton frames from DGD dataset in Fig.2. All of the videos are captured by Kinect 2. The experimental setting is the same as in (C. Alexandros and Flórez-Revuelta, 2015) which employed 20 of the available joints, discarding the fingers and a redundant joint of the torso (SpineShoulder). Fig.3 shows the joints order of DAI gait dataset.

3.2 Experimental Results and Comparison to State-of-the-art

We selected two abnormal and two normal sequences of each subject randomly as training set which contains 28 sequences. The remaining of the dataset is testing set. We compare our approach with the state-of-the-art methods on the cross-subject test setting.

Table. 1 lists the results of our approach and result of the state of the art(C. Alexandros and Flórez-Revuelta, 2015) based on DAI gait dataset. However, our protocol is different from the protocol used in (C. Alexandros and Flórez-Revuelta, 2015). Actually, in (C. Alexandros and Flórez-Revuelta, 2015), the authors did the training on normal sequences and detect the abnormal ones among the test sequences. Whereas, our training set includes also some abnormal sequences.

Using a sliding window of size 37 to 42, Alexandros et al. (C. Alexandros and Flórez-Revuelta, 2015) reported 98% F1-measure. We report 81% recognition rate using only one frame, 94.23% using a sliding window of 42 frames and 98.47% using a sliding window of 60 frames.

3.3 Effect of the Temporal Size of the Sliding Windows

Fig.4 In order to study the effect of size of the sliding window, we report the classification results using several values of the window size as illustrated in Fig.4. One can see that the more frames we use in the window, the better the result is. Using only 37 frames, the recognition rate is 90%. It reaches 94.23% and 97.01% for window size 42 and 50 respectively. Using 60 frames in the sliding window, the recognition rate is 98.47
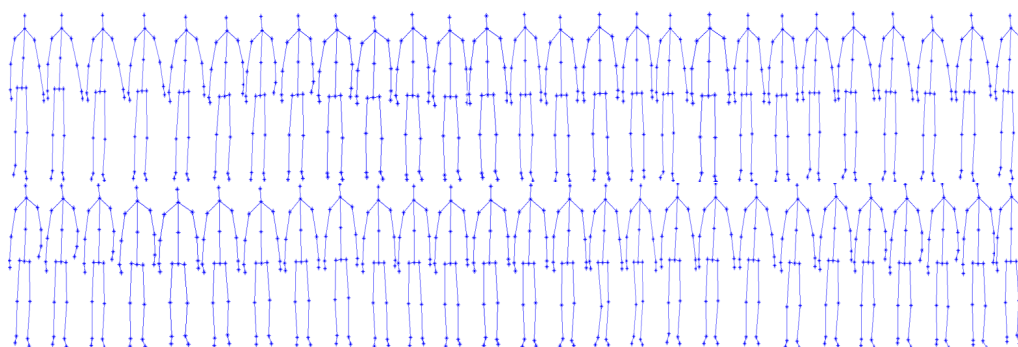
Figure 2: Some skeleton frames of right knee injured abnormal gait from DAI gait dataset.

Table 1: Reported results Comparison to state of the art.

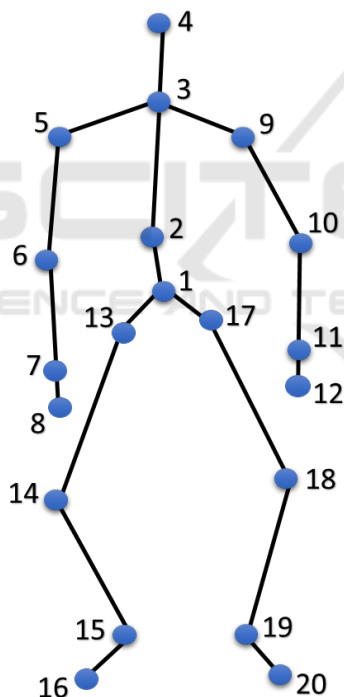| Method	Accuracy
Joint Motion History Features (37,42)(C. Alexandros and Flórez-Revuelta, 2015)	98%
Our approach (without memory)	81%
Our approach (with 42 frames)	94.23%
Our approach (with 60 frames)	98.47%

Figure 3: Joints Order of DAI gait dataset.

3.4 Relevant Features

The proposed approach is based on a feature vector of 190 inter joints distances per frame. One important question is are some distances more relevant than others to classify normal and abnormal gaits? Can one do this binary classification (normal versus abnormal) using only one distance? We investigate these questions and report classification results based on individual distances. We show in Table.2 some classification results based on one distance which has a memory with 40 frames. The distances number 170 is able to classify normal and abnormal gaits with a success of 92.18%. The classification results based on distances 171, 178,188 and 184 are respectively 92%, 89.83%, 93.09% and 97.15%. The distance 179 report a success of 98.46% which is better than the result reported using all the distances together.

Table 2: Selected features recognition rate.

Number of features	recognition rate
No.170	92.18%
No.171	92%
No.178	89.83%
No.179	98.46%
No.184	97.15%
No.188	93.09%

We report in Fig.5 and Fig.6 an illustration of the most relevant distances for normal-abnormal gait classification. As illustrated in these figures, the most relevant distances correspond to the distance between the knee and the foot of the same leg (distance No.171), the distance between knee and the foot of the other leg (distance No.178), the distance between the feet from different legs (distance No.184) and the distance between ankles (distance No.179). This result is in agreement with the data as the 4 abnormal types in the dataset are:

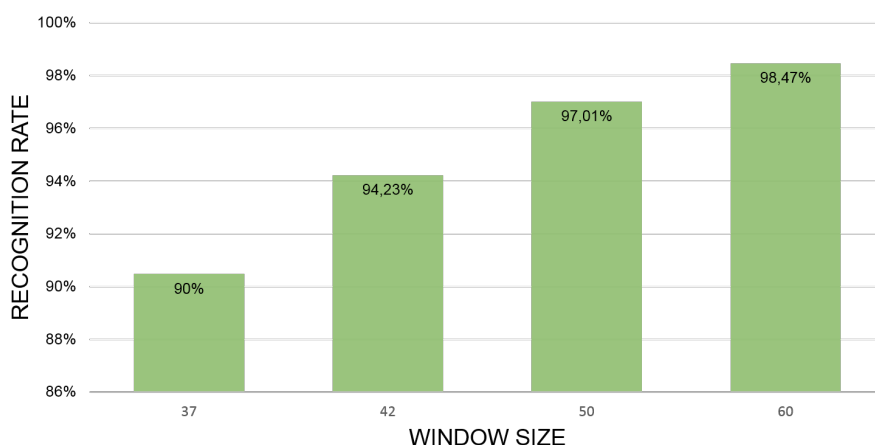- RKI: Right knee injury (cannot bend the right knee, starting with left foot)

Figure 4: Effect of the temporal size of the sliding window on the results. The classification rates increase when increasing the length of the temporal window.
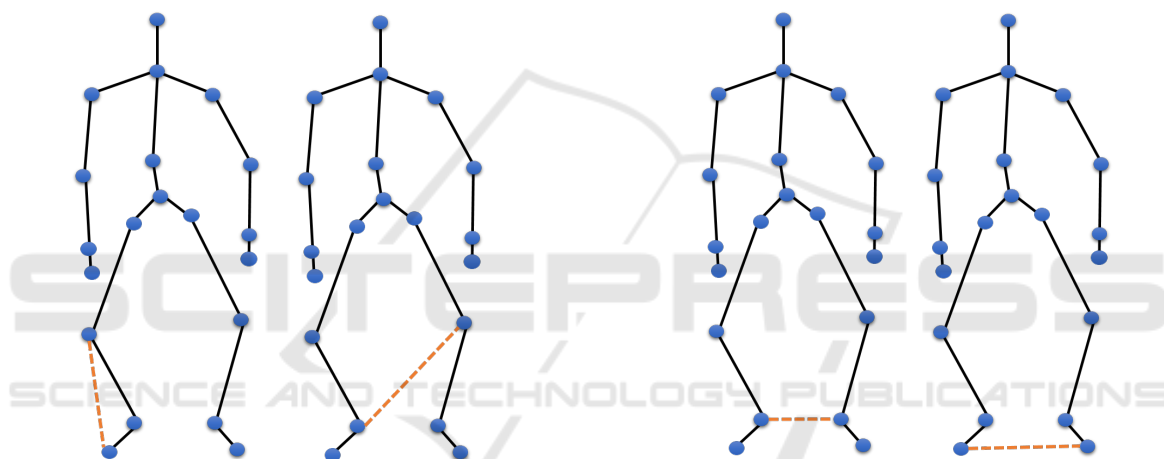

Figure 5: Illustration of distances No.171 and No.178 from DAI gait Dataset.

Figure 6: Illustration of distances No.179 and No.184 from DAI gait Dataset.

- LKI: Left knee injury (cannot bend the left knee, starting with right foot)
- RFD: Right foot dragging (dragging right leg, starting with left foot)
- LFD: Left foot dragging (dragging left leg, starting with right foot)

In order to better understand the behavior of the relevant distances, we computer the mean normal distance No.179 (distance between right and left ankles) and the mean abnormal one. We show previously that the classification of normal-abnormal gaits using only this distance is better than one based on all distances. Fig.7 shows the evolution of the mean of this distance in both normal and abnormal cases. One can see that the mean distance is bigger in general in the abnormal case and it includes more oscillations than the normal one. It is clear that the variation of this distance is more stable in the normal case.

3.5 Leave-one-actor-out Experiments

We did leave-one-actor-out experiments on DAI gait dataset. For each time, we selected all kinds of actions of one subject as testing set and the rest of the dataset as training set. In total, we have 7 subjects so we did the same experiments seven times. At last, we obtained the mean recognition rate of the seven results. Now we can see the performance of each subject without memory, with 42 frames and with 60 frames as shown in Table.3.
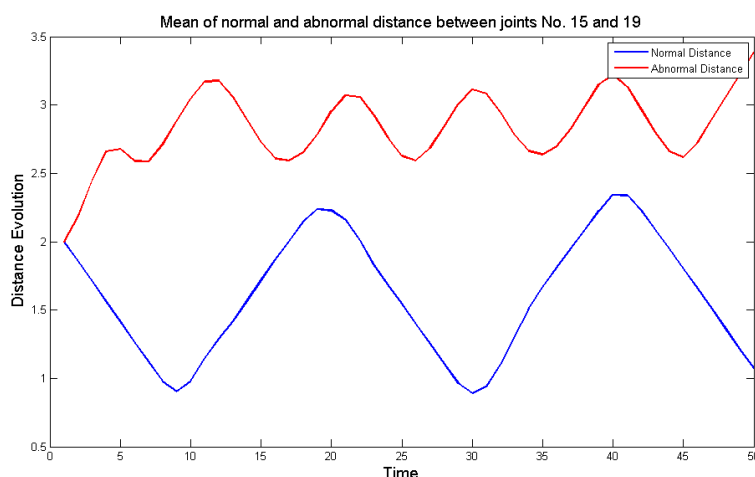
Figure 7: The mean distance of abnormal and normal gait between joints No.15 and 19.

Table 3: Leave-one-actor-out Experiments.

The number of subject	without memory	with 42 frames	with 60 frames
Subject 1	69.9%	88.12%	100%
Subject 2	65.36%	72.99%	88.26%
Subject 3	80.49%	96.58%	100%
Subject 4	66.57%	90.35%	99.12%
Subject 5	77.1%	89.35%	98.43%
Subject 6	77.53%	90.32%	100%
Subject 7	75.41%	93.36%	100%
Mean Recognition Rate	73.19%	88.72%	97.97%

4 CONCLUSIONS

In this paper we propose a spatio-temporal modeling of the skeleton data based on inter-joint distances for normal and abnormal gaits classification. The proposed features are discriminative enough to classify abnormal and normal human actions. We report 98.47% recognition rate and show that some distances related to the knees, ankles and the feet are more relevant than other distances. Future work will be focused on more features to be able to distinguish the different types of abnormal gaits.

REFERENCES

A. A. Chaaraoui, P. C.-P. and Flrez-Revuelta, F. (2012). A review on vision techniques applied to human behaviour analysis for ambient-assisted living. In *Expert Systems with Applications, 39(12):1087310888.*

A. Paiement, L. Tao, S. H., Camplani, M., Damen, D., and M.Mirmehdi (2014). Online quality assessment of human movement from skeleton data. In *Proceedings of British Machine Vision Conference (BMVC).*

Breiman, L. (2001). Random forests. In *Machine Learning, vol. 45, pp 5-32.*

C. Alexandros, J. P.-L. and Flórez-Revuelta, F. (2015). Abnormal gait detection with rgb-d devices using joint motion history features.

Dian, G. and Medioni, G. (2011). Dynamic manifold warping for view invariant action recognition. In *IEEE International Conference on Computer Vision (ICCV).*

G. S. Parra-Dominguez, B. T. and Mihailidis, A. (2012). 3d human motion analysis to detect abnormal events on stairs. In *In International Conference on 3D Imaging, Modeling,Processing, Visualization and Transmission, pages 97103.*

J. Shotton, A. Fitzgibbo, M. C., Sharp, T., Finocchio, M., Moore, R., Kipman, A., and Blake, A. (June 2011). Real-time human pose recognition in parts from single depth images. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition.*

J. Snoek, J. Hoey, L. S., Zemel, R. S., and Mihailidis, A. (2009). Automated detection of unusual events on stairs. In *Image and Vision Computing, vol. 27, no. 1-2, pp. 153 166,.*

M. Devanne, H. Wannous, S. B., Daoudi, M., and Bimbo, A. (2013). Space-time pose representation for 3d human action recognition. In *New Trends in Image Analysis and Processing ICIAP 2013.*

Omar, O. and Liu, Z. (2013). Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

R. Slama, H. Wannous, M. D. and Srivastava, A. (2015). Accurate 3d action recognition using learning on the grassmann manifold. In *Pattern Recognition 48.2 : 556-567*.

W. Li, Z. Z. and Liu, Z. (2010). Action recognition based on a bag of 3d points. In *IEEE Computer Vision and Pattern Recognition Workshops (CVPRW))*.