A General Descriptor for Detecting Abnormal Action Performance
From Skeletal Data

Amr Elkholy¹, Mohamed E. Hussein¹,², Walid Gomaa¹,², Dima Damen³, Emmanuel Saba⁴

Abstract—We propose an action-independent descriptor for detecting abnormality in motion, based on medically-inspired skeletal features. The descriptor is tested on four actions/motions captured using a single depth camera: sit-to-stand, stand-to-sit, flat-walk, and climbing-stairs. For each action, a Gaussian Mixture Model (GMM) trained on normal motions is built using the action-independent feature descriptor. Test sequences are evaluated based on their fitness to the normal motion models, with a threshold over the likelihood, to assess abnormality. Results show that the descriptor is able to detect abnormality with accuracy ranging from 0.97 to 1 for the various motions.

I. INTRODUCTION

Monitoring and assessing the quality of elderly peoples actions are critical for understanding their quality of life, as well as assessing potential mobility risks or degradations. Most elderly people are living alone and they are susceptible to falling, as well as more serious conditions such as strokes. Finding an automated unobtrusive method for detecting abnormal action performance can promise the elderly a safer independent living.

Different methods are proposed for automatic motion quality assessment and abnormality detection using full body sensing (e.g. MoCap data) and wearable sensors [1]. But, these sensors are wearable sensors which should be consistently and frequently charged. The depth sensor is one of the emerging sensing technologies, which provides the skeletal data of the person facing the camera. This technology can be used as an unobtrusive technique for abnormality detection.

Most of the proposed methods using skeletal data depend on extracting common features such as joint positions, pairwise joint distances, pairwise joint angles, and joint velocities. These features represent high dimensional data which needs to be reduced by dimensionality reduction techniques as in [2] which requires extra computational time, and storage. On the other hand, action specific features are extracted such as stride time, stride length, and walking speed, then a classification or regression model is used as in [3], [4]. But, these features are for specific actions (e.g. walking) and generally cannot be applied to other daily activities. Another piece of work exists for quantitatively evaluating musculoskeletal disorder by measuring specific parameters such as step size, postural swing level, arm swing level, and stepping time as proposed in [5], but this method is limited to periodic actions or repeated sequences of non-periodic actions.

In contrast, our method directly extracts medically justified low number of features (hence, no need for dimensionality reduction), which are able to discriminate between normal and abnormal action performances, and can be applied to most, if not all, actions. Our method starts by applying our proposed descriptor on the skeletal data extracted from a depth sensor, and then builds a probabilistic model for the normal cases only, as we cannot restrict the cases of abnormality in action performance; finally, the normalcy of any sequence of the same action can be evaluated. The approach is both simple and fast, and can be applied to any action for detecting abnormal performance. We tested our method on one public dataset [2], which contains four common daily actions, specifically; sit-to-stand, stand-to-sit, flat-walk, and climbing-stairs. The dataset includes normal and abnormal action performances. The abnormal sequences include the following conditions: Parkinson, stroke, freezing, restricted knee, restricted hip flexion, left leg lead, and right leg lead. The results show that our method performs well and gaining high accuracy ranging from 0.97 to 1.

Beside this introduction, our proposed method is described in Section II. In Section III, the experimental results are described. Finally, the paper’s conclusion and future work are drawn in Section IV.

II. PROPOSED METHOD

Preprocessing includes noise removal, action instance cropping, and coordinate scaling. The noisy skeletal data sequence is first cleaned using a uniformly-weighted averaging filter of length 20. Then, the frames of the action instance are cropped from the sequence, which may contain extra frames before or after the action instance ¹. Finally, the 3D joint coordinates are scaled so that each of the three dimensions lies in and covers the interval [0, 1] over the entire sequence. This is to make the features invariant to skeleton size and location with respect to the camera.

After preprocessing, the proposed features are extracted, and a generative probabilistic model (Gaussian Mixture Model in our evaluation) for normal sequences is built in training. For testing, we compute the likelihood for the test

¹The cropping is done automatically based on the velocity of motion for the sitting down and standing up actions and manually for the walking and stairs actions.
sequence; and, based on a learned threshold, the system can decide whether the sequence is normal or not.

A. Feature Extraction

In our proposed method we attempt to extract medically justified low number of features that can capture abnormality in action performance, to get rid of the curse of dimensionality. Through observation and physiatrists consultation we construct a descriptor from concatenating three features that can capture most of the neuromusculoskeletal disorder abnormalities.

The first feature measures the asymmetry between the left and right body parts' movements over the sequence of the action, the second measures the average velocity of performing an action, and the third measures the width of the base of support. The three features are described in detail in the following subsections.

1) Asymmetry Feature: It can be observed that most of the abnormal action performances are lacking symmetry between the movements of the left and the right body parts [6]. The purpose of this feature is to capture the degree of asymmetry between the left and the right body parts’ movements along the whole sequence of an action. We extract this feature by splitting the body joints into two overlapping groups, as shown in Fig. 1. Then, for every pair of joints in each group, we compute the average distance over the sequence. Let \( D_L \) and \( D_R \) be the average pairwise distance matrices for the left and right body parts, respectively. The asymmetry feature is computed as the Euclidean distance between the vectorized upper triangles of the two matrices \( D_L \) and \( D_R \). Letting \( A \) be the asymmetry feature, its computation can mathematically be expressed as

\[
D_{ij}^P = \frac{1}{n} \sum_{t=1}^{N} \sqrt{(x_{it} - x_{jt})^2 + (y_{it} - y_{jt})^2 + (z_{it} - z_{jt})^2} \tag{1}
\]

\[
\forall i, j \in N_P, P \in \{L, R\}
\]

\[
A = ||\text{uvec}(D_L) - \text{uvec}(D_R)|| \tag{2}
\]

where \( D^P \) is the left/right body part's average pairwise distance matrix, \( x_{it}, y_{it}, z_{it} \) are the 3D coordinates of the joint \( i \) at frame \( t \), \( n \) is the number of frames in the action sequence, \( N_P \) is the set of joints in the left/right body part, and the \( \text{uvec}() \) operator outputs the vectorized upper triangle of its operand matrix.

This feature captures the asymmetry between the movements of the two body parts over the sequence, so it should be low for normal action performances, and higher depending on the degree of abnormality in performing an action.

2) Velocity Feature: The velocity of performing an action is the second feature of our descriptor as most neuromusculoskeletal patients perform their daily activities significantly slower. In fact, slowness of movement is one of the Parkinson disease’s symptoms [7]. In addition to slow action performance, this feature is also responsible for capturing motion freezing.

The velocity feature is extracted by computing the displacement magnitude for each body joint between two successive frames, then calculating the average displacement magnitude over all joints between the two successive frames. We consider this average inter-frame joint displacement magnitude as an estimate for the average joint velocity. Therefore, for the whole sequence, we will have \( n - 1 \) average velocities. Then we take the average a second time over frames. This will produce the average velocity during the motion over the entire sequence, which is expected to be low in case of freezing or abnormally slow motion. In mathematical terms, the velocity feature \( V \) is computed as

\[
V = \frac{1}{n-1} \sum_{t=1}^{n-1} \sqrt{(x_{it+1} - x_{it})^2 + (y_{it+1} - y_{it})^2 + (z_{it+1} - z_{it})^2} \tag{3}
\]

where \( N \) is the number of joints, \( n \) is the number of frames.

3) Base of Support (BOS) Feature: This refers to the area of contact between an object or a person and the supporting surface [8]. In many abnormal cases, the base of support is larger than usual as the patient tries to increase the distance between his/her feet to achieve more stability [9]. For simplicity, the horizontal distance between the two feet can be an expressive measure for BOS. In gait, having a large value for such a distance is an indicator of abnormality [9]. We include this feature as the third feature of our descriptor for abnormality detection. We compute the distance between knees instead of feet as the knees are less susceptible to occlusion than feet as observed from the dataset. Although the distance between feet and the distance between knees are different, the averages of such distances over time are correlated.

B. Model

After extracting the above features, we first normalize each feature in the training and testing data by subtracting its sample mean and dividing by its sample standard deviation, which are estimated only from the training data. The sample statistics are estimated using a robust estimate technique to cope with noise and outliers (we used Huber’s M-method [10] in our results). Then, we build a Gaussian Mixture Model (GMM) for each action with the training samples (normal cases only). For testing, we compute the likelihood of a testing sequence by evaluating the trained GMM, and with a learned threshold, the method decides if the sequence is normal or abnormal.
III. EVALUATION AND RESULTS

In this section we illustrate the results obtained by applying our method on a public dataset, released by the University of Bristol [2]. The dataset contains four different common daily actions: gait on stairs, walking on a flat surface, sitting down, and standing up; each partitioned into training and testing sequences.

We use the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) curve and the best accuracy all over the curve as our performance metrics. These metrics are shown for each single feature, and all of them combined, for all actions in Table I.

The details of applying our proposed method on each action and the results obtained are shown in the sub-sections below, as well as a comparison to the work of [2] as the most relevant to ours.

A. Walking On Flat Surface

This action includes 40 sequences by 10 subjects, with 23 normal and 17 abnormal sequences. The sequences were captured by an Asus Xtion Pro depth camera with 15 joints skeleton data. We built the model using 18 normal sequences (45%) of 7 subjects, and tested on 22 sequences including normal and abnormal sequences performed by different subjects. The abnormal cases include 8 Parkinson, and 9 stroke cases. The distribution of the walking samples in the 3D feature space is illustrated in Fig. 3(a). The figure shows that the normal and abnormal sequences are clearly separated in our feature space, and that all sequences were correctly classified by our model. Our method achieved best accuracy and AUC of 1.

From Table I we can see that the most discriminative features are velocity and BOS. Each of the two features alone can distinguish between normal and abnormal cases with high accuracy. We can observe from Fig. 3(a) that the abnormal cases mostly are slower than normal ones. Also, the abnormal gait mostly has a wider BOS than the normal gait.

B. Sitting down and Standing up

The sitting and standing sequences were captured using Microsoft Kinect V2 with 25 skeleton joints recorded. We excluded the extra joints, and used the 15 common joints with Asus Xtion Pro’s skeleton. Overall, there are 109 sequences by 6 subjects, containing 233 action instances, 108 instances for sitting down, and 125 instances for standing up. We built our model using 64 action instances (59%) for sitting down and 77 action instances (61%) for standing up.

For the sitting down and standing up sequences we achieved best accuracy of 1. Fig. 3(b),(c), illustrate the distribution of the sitting down and standing up sequences, respectively, in the 3D feature space. The normal and abnormal sequences are completely separated and our model could achieve zero miss-classification rate.

The most discriminative feature for both actions is the velocity feature and the second discriminative feature is the asymmetry as the patient tends to lean to support her body as observed from the recorded sequences, whereas the normal standing up or sitting down is mostly symmetric. This can be observed from the Motion History Images (MHI) [11] of normal and abnormal samples in Fig. 2(b),(c).

C. Stairs

This action includes 48 sequences by 12 subjects. The sequences were captured by Asus Xtion Pro depth camera. Overall, there are 31 normal and 17 abnormal sequences, which include the cases of left leg lead, right leg lead, and freeze of gait. We built the model using 17 normal sequences (35%) of 6 subjects, and tested on 31 sequences performed by different subjects. The distribution of the sequences in the 3D feature space is illustrated in Fig. 3(d). We achieved 0.97 best accuracy with 1 miss-classified sequence and an AUC of 0.98.
In this action, no single feature is very discriminative. We think this is due to the highly noisy skeletons and occlusion in the sequences of this action. The most discriminative feature is the velocity feature, which can achieve 0.74 best accuracy as illustrated in Table I while the asymmetry and BOS are less discriminative.

D. GMM Components

The number of the GMM components used in all the above experiments is 5, which produced the best results for all actions in our experiments. However, our proposed model is robust to changing the number of GMM components. As shown in Table II, the accuracies for all actions change within a small range for 4 GMM components, and above.

E. Comparison

We compare our results to the work of [2], as it is the most relevant to our work, with the following notes. First, the method of [2] proposed four different models with different tuning parameters and we compared our results with the best result for each action. Second, the results listed for the stairs action by [2] are per abnormal event, while ours are per sequence. So, we assume the sequence to be abnormal if at least one abnormal event is detected in it. For the rest of actions, the listed results by [2] are per sequence, and hence, the results are directly comparable. Table III illustrates the results of our proposed method compared to the results of [2]. For our results, we used the same three features with all actions with the same parameters (5 GMM components). For the results of [2], the Table illustrates the model, features, and manifold dimensions that produced the best results.

Although the results listed on Table III are close except in the Stairs action, our proposed method uses the same features, modeling approach, and parameters for all actions, which makes our method more general without the need for parameter tuning or changing the features. Our method takes approximately 30 seconds for training and less than 1 second for testing a sequence, while the method of [2] takes approximately 2 hours for building the manifold, and approximately 84 seconds for testing a sequence on the same machine using published code by the authors.

### IV. Conclusion and Future Work

In this paper, we proposed a general descriptor, made up of concatenating three efficient to compute features, that is able to distinguish between normal and abnormal action performances for different action types. The features are used to train a probabilistic model, which is a GMM in our implementation. The method is applied to a publicly available dataset that includes different actions and different abnormalities. Our method is simple, fast, doesn’t need dimensionality reduction, and can be applied to any action. In the future, we plan to carry out more experiments on additional actions and various abnormalities.

### References


