Online quality assessment of human movement from skeleton data

Adeline Paiement
csatmp@bristol.ac.uk
Lili Tao
lili.tao@bristol.ac.uk
Sion Hannuna
sh1670@bristol.ac.uk
Massimo Camplani
massimo.camplani@bristol.ac.uk
Dima Damen
dima.damen@bristol.ac.uk
Majid Mirmehdi
majid@cs.bris.ac.uk

Visual Information Laboratory
Department of Computer Science
University of Bristol
Bristol, UK

Abstract

We propose a general method for online estimation of the quality of movement from Kinect skeleton data. A robust non-linear manifold learning technique is used to reduce the dimensionality of the noisy skeleton data. Then, a statistical model of normal movement is built from observations of healthy subjects, and the level of matching of new observations with this model is computed on a frame-by-frame basis following Markovian assumptions. The proposed method is validated on the assessment of gait on stairs.

1 Introduction

The analysis of human movement through visual information has attracted huge interest due to applications in several areas, from assessment of pathologies, rehabilitation, to movement optimisation in sport [2]. In particular, the discrimination of anomalies has been a strong focus, as illustrated by the comprehensive survey in [13]. Anomalies are often detected by comparison against two models of normal and abnormal movements, e.g. as in [17]. Considering that abnormal movements may have highly variant representations, a single model is unlikely to be sufficient to define and represent them. It is therefore preferable to detect deviations from a model of normal movements, e.g. as in [9] which uses hierarchical appearance and action models of normal movements to detect falls from RGB silhouettes, and [11] which uses binary classifiers of harmonic features to detect abnormalities in stairs descents from the lower joints of a Kinect skeleton.

The work that is most closely related to that proposed here was presented by Snoek et *al*. who used monocular RGB images to detect unusual events during stairs descent using a single hidden Markov model (HMM) framework [15]. Foot position and velocity, together

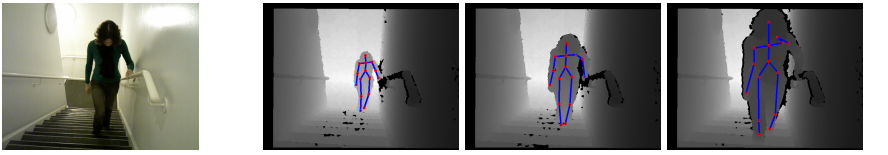

Figure 1: RGB-D data and skeletons at bottom, middle, and top of the stairs. Note, only the skeleton is used in our proposed method.

with optical flow features, were used to build a continuous observation space. Their system relies on a feet tracker, and thus has a significant risk of failure due to occlusion, possibly leading to the sequence being wrongly marked as abnormal. Note that although the lower joints of the body can represent the most discriminating information for walking movement, the use of all joints may enhance the analysis, e.g. by allowing a better assessment of the balance of the subject, and it provides vital cues in many other applications.

Although many works have been proposed to detect abnormality in video sequences, the problem of assessing the quality of human movement has rarely been addressed. Recently, Wang et *al*. presented a method for quantitatively evaluating musculoskeletal disorders on patients who suffer from the Parkinson disease [18]. However, the method is restricted as it is only designed for periodic movement (walking), and the features used (step size, arms and postural swing levels, and stepping time) make it difficult to generalise to other applications. Moreover, the method requires observing a complete gait cycle before being able to classify it.

Joint positions are commonly used to analyse human movements, e.g. [15], however their high dimensionality, especially for full body skeleton, and their often high amount of noise, make it imperative to reduce their dimensionality in a noise-robust fashion. Manifold learning techniques have become increasingly popular for reducing dimensionality of data that contain redundant information [7, 8, 19]. Nonetheless, reducing the dimensionality of noisy data is still a challenging problem. Gerber et *al*. introduced an extension of Laplacian Eigenmaps to cope with noisy input data [6], but Eigenmap representation depends on the density of the points on the manifold, which may not be suitable for non-uniformly sampled data, such as skeleton data.

The contributions of this work may be summarised as follows. We present a novel, general approach which not only detects abnormal events, but also provides an assessment of movement quality, defined as a measure of how much a movement deviates from normal. Such continuous quantification of abnormality aims at allowing clinicians to better establish diagnosis and also to assess the evolution of the condition of patients. The proposed method is based on a continuous statistical representation of the movement, which, contrary to HMM methods, avoids having to divide the movement into segments whose number would have to be determined. Further, it can cope well with different types of movements, including both periodic and non-periodic ones, due to the use of full body skeleton information (see Fig. 1, for example skeletons). This is made possible by a non-linear manifold learning technique that can reduce its high dimensionality, for which we use diffusion maps [3] which we adapt to deal with noise and outliers using the robust extension of Gerber et *al*. [6]. Both individual body poses and dynamics are assessed on a frame-by-frame basis which makes the method suitable for online applications, and allows alerts to be triggered in case of abnormal events before the end of the movement.

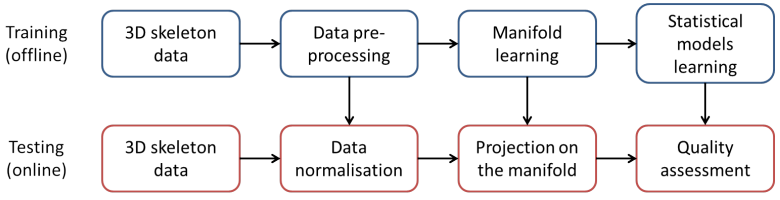

Figure 2: Overview of the proposed method

2 Methodology

An overview of our proposed method is illustrated in Figure 2. We first use a skeleton tracker to extract a full body skeleton from Kinect's depth data [10, 14], from which we derive normalised features. Their dimensionality is then reduced using a non-linear manifold technique. Statistical models of normal pose and dynamics are learnt off-line, and new observations are tested against these models for quality assessment.

2.1 Low level feature description

Skeleton data are view-invariant[1] and, in our application, are derived from depth information which alleviates the effect of human appearance differences and of lighting variations. We used skeleton trackers from Microsoft Kinect SDK [14] or OpenNI SDK [10], and found that the choice between these two trackers did not significantly change the results in our experiments, so we only present results using the OpenNI tracker in Section 3. Since the skeletons tend to be very noisy, we reduce the noise by applying an averaging filter on the joint coordinates.

Given a pose $\hat{\mathbf{S}} = [\hat{\mathbf{s}}_1 \cdots \hat{\mathbf{s}}_P]^T \in \mathbb{R}^{3J\times 1}$ made of the 3D positions $\hat{\mathbf{s}}_i$ of J joints[2], a normalised pose $\mathbf{S} = g(\hat{\mathbf{S}})$ is computed to compensate for global translation and rotation of the view point, and for scaling due to varying heights of the subjects. This allows comparison with the model based on poses that will be described in Section 2.3. We propose two normalisation methods for the computation of $g(.)$. When the apparent vertical size of the body does not change significantly during the movement, standard Procrustes analysis can be used. Alternatively, angles between individual joints and the hip centre may be used instead of scaled joint positions. The use of both normalisation methods and resulting feature types did not significantly change the results in our experiments, and in Section 3 we present results using Procrustes based normalisation. Other features may be used, provided that the associated normalisation function $g(.)$ solves the aforementioned alignment and scaling issues.

2.2 Robust diffusion maps

We reduce the dimensionality of the features $\mathbf{S}$ using manifold learning. We select diffusion maps [3], which is a graph-based technique with quasi-isometric mapping Φ from original higher space $\mathbb{R}^N$ to a reduced low-dimensional diffusion space $\mathbb{R}^{N'}$, where $N' \ll N$. This method shows advantages over conventional dimensionality reduction methods [3]: it can deal with non-uniformly sampled data that lie on non-linear manifolds, and it preserves

[1] Although the skeleton trackers that we use only work well when the subject is facing the camera.
[2] J is 20 and 15 for skeletons of the Microsoft Kinect SDK and OpenNI SDK respectively.

the local structure of the data. Given a set of training data with M normalised samples $\mathcal{S} = \{\mathbf{S}_1 \dots \mathbf{S}_M\} \in \mathbb{R}^N$, the intrinsic geometry of the data can be found depending on the similarity of the samples measured by the diffusion distance $L = d(\Phi(\mathbf{S}_i), \Phi(\mathbf{S}_j))$, where $d(.)$ is the Euclidean distance in reduced space.

In addition to having a high dimensionality, skeleton data acquired with a Kinect sensor tend to suffer from a relatively large amount of noise, and contain outliers, especially when parts of the body are occluded. Our filtering of the skeleton data in Section 2.1 fails to remove the outliers, thus we propose to modify the original diffusion maps by adding the extension of [6] for dealing with them. Building a diffusion map as in [3] requires computing a weighted adjacency matrix $\mathbf{W}$ that contains the distances between neighbouring points weighted by a Gaussian kernel K_G:

$$w_{i,j} = K_G(\mathbf{S}_i, \mathbf{S}_j) \ . \tag{1}$$

We modify the entries of the matrix as

$$w_{ij} = (1-\beta)K_G(\mathbf{S}_i, \mathbf{S}_j) + \beta I(\mathbf{S}_i, \mathbf{S}_j), \ \text{ with } I(\mathbf{S}_i, \mathbf{S}_j) = \begin{cases} 1, \ \mathbf{S}_i \in \mathcal{N}_i \text{ or } \mathbf{S}_j \in \mathcal{N}_j \\ 0, \ \text{otherwise} \end{cases} \tag{2}$$

$I(.)$ is an indicator function that was introduced in [6] to avoid disconnected components in Laplacian eigenmaps, thus reducing the influence of outliers. Then, as in the original diffusion map, the optimal mapping Φ is obtained from the eigenvalues λ and the corresponding eigenvectors φ of the Laplace-Beltrami operator $\mathbf{P}$ [3],

$$\Phi(\mathbf{S}_i) \mapsto [\lambda_1 \varphi_1(\mathbf{S}_i), \cdots, \lambda_{N'} \varphi_{N'}(\mathbf{S}_i)]^T \ . \tag{3}$$

An approximation of $\mathbf{P}$ is computed following [4] from the matrix $\mathbf{W}$.

Mapping testing data - The Nyström extension [1] allows to extend the low dimensional representation computed from a training set to new samples, by computing the mapping of a new pose $\mathbf{S}'$ as

$$\Phi'(\mathbf{S}') = \sum_{\mathbf{S}_j \in \mathcal{S}} \mathbf{P}(\mathbf{S}', \mathbf{S}_j) \varphi_m(\mathbf{S}_j), \ \ \forall m \in 1 \dots M \tag{4}$$

where $\mathbf{P}(\mathbf{S}', \mathbf{S}_j)$ is computed in the same fashion as in [4], but based on our new definition of $w_{i,j}$ with the added indicator function $I(.)$. We use this mapping $\mathbf{Y} = \Phi'(\mathbf{S}')$ as our high level feature for building our statistical model in the next section.

2.3 Statistical model of movement learning

We assess the quality of movements by comparing new pose description vectors $\mathbf{Y}$ with a statistical model of normal movements. We introduce the multivariate random variable Y that takes as value y our high level pose description vector $y = \mathbf{Y}$. Our model comprises *two components that describe respectively the pose and the dynamics of the skeleton during a given movement*.

The first model represents normal poses by their probability density function (pdf) $f_Y(y)$. We obtain this pdf from training data, made up of all the (successive) poses of normal movements, using a Parzen window estimator implemented with the Python library Scikit-learn [12].

The second model is a dynamics model, which is required to take into account the temporal dimension of the data. Thus, we use the conditional pdf $f_{Y_t}(y_t | y_1, \dots, y_{t-1})$ that considers

the sequence of poses from the first pose y_1 at the beginning of the video sequence to the pose y_t at the current frame t. This provides the likelihood of this sequence of poses being represented by the dynamics model.

In order to compute this likelihood, we introduce X_t, a random variable that takes values $x_t \in [0,1]$ and which represents the proportion of movement completion at frame t. X_t may be seen as the continuously evolving stage of the movement, and in the case of periodic movements it is analogous to the movement's phase. Its value x_t increases with t from 0 at the start of the movement to 1 at the completed movement stage. This increase is steady for normal movements. For periodic movements, x_t increases within one cycle of the movement, and then returns to 0 in order to analyse the next cycle, while for non-periodic movements x_t simply increases from 0 to 1. An advantage of using this continuous variable is that, contrary to HMM methods, the movement does not have to be discretised into a number of segments, and the problem of choosing an optimal segment number becomes irrelevant. The value of X_t is considered to be known in the training data, where it is set linearly between 0 and 1 from the first to the last frame of the movement or movement cycle. This assumes that each training movement is performed at a regular speed, although this speed can vary between training samples. For testing data, the value of X_t will need to be estimated, as described in Section 2.4. During this estimation, the hypothesis that the movement speed is stable is not enforced in order to be able to describe the testing data at best, but instead it is used to detect abnormality. For brevity, we denote $\{X_0,\dots,X_t\}$ as $\mathbb{X}^t$ (X_0 is the initial stage of the movement before the first observation Y_1), and $\{Y_1,\dots,Y_t\}$ as $\mathbb{Y}^t$. $f_{Y_t}(y_t|y_1,\dots,y_{t-1})$ may be computed as

$$f_{Y_t}(y_t|y_1,\dots,y_{t-1}) = \frac{f_{\mathbb{Y}^t}(y_1,\dots,y_t)}{f_{\mathbb{Y}^{t-1}}(y_1,\dots,y_{t-1})}\ , \tag{5}$$

with

$$f_{\mathbb{Y}^t}(y_1,\dots,y_t) = \int_{\{x_0,\dots,x_t\}\in\Omega_{\mathbb{X}^t}} f_{\mathbb{Y}^t,\mathbb{X}^t}(y_1,\dots,y_t,x_0,\dots,x_t)\ , \tag{6}$$

and $\Omega_{\mathbb{X}^t}$ being the domain of the possible values for $\{x_0,\dots,x_t\}$. We propose to use the two following Markovian assumptions to compute $f_{Y_t}(y_t|y_1,\dots,y_{t-1})$:

$$\begin{cases} f_{Y_t}(y_t|y_1,\dots,y_{t-1},x_0,\dots,x_t) = f_{Y_t}(y_t|x_t) & , \\ f_{X_t}(x_t|x_0,\dots,x_{t-1}) = f_{X_t}(x_t|x_{t-1}) & , \end{cases} \tag{7}$$

i.e. an observation at frame t is completely defined by the proportion of movement completion X_t at that frame, and the proportion of movement completion X_t at frame t depends only on the proportion of movement completion at the previous frame X_{t-1}. Then,

$$\begin{aligned} f_{\mathbb{Y}^t,\mathbb{X}^t}(y_1,\dots,y_t,x_0,\dots,x_t) &= f_{Y_t}(y_t|y_1,\dots,y_{t-1},x_0,\dots,x_t)\, f_{\mathbb{Y}^{t-1},\mathbb{X}^t}(y_1,\dots,y_{t-1},x_0,\dots,x_t) \\ &= f_{Y_t}(y_t|x_t)\, f_{X_t}(x_t|y_1,\dots,y_{t-1},x_0,\dots,x_{t-1}) \\ &\quad f_{\mathbb{Y}^{t-1},\mathbb{X}^{t-1}}(y_1,\dots,y_{t-1},x_0,\dots,x_{t-1}) \\ &= f_{Y_t}(y_t|x_t)\, f_{X_t}(x_t|x_{t-1})\, f_{\mathbb{Y}^{t-1},\mathbb{X}^{t-1}}(y_1,\dots,y_{t-1},x_0,\dots,x_{t-1}) \\ &\vdots \\ &= f_{X_0}(x_0)\prod_{i=1}^{t} f_{Y_i}(y_i|x_i)\, f_{X_i}(x_i|x_{i-1})\ , \end{aligned} \tag{8}$$

and (6) becomes

$$f_{\mathbb{Y}^t}(y_1,\ldots,y_t) = \int_{\{x_0,\ldots,x_t\}\in\Omega_{\mathbb{X}^t}} f_{X_0}(x_0)\prod_{i=1}^{t} f_{Y_i}(y_i|x_i)\,f_{X_i}(x_i|x_{i-1}) \ . \tag{9}$$

It follows, according to (5), that

$$f_{Y_t}(y_t|y_1,\ldots,y_{t-1}) = \frac{\int_{\{x_0,\ldots,x_t\}\in\Omega_{\mathbb{X}^t}} f_{X_0}(x_0)\prod_{i=1}^{t} f_{Y_i}(y_i|x_i)\,f_{X_i}(x_i|x_{i-1})}{\int_{\{x_0,\ldots,x_{t-1}\}\in\Omega_{\mathbb{X}^{t-1}}} f_{X_0}(x_0)\prod_{i=1}^{t-1} f_{Y_i}(y_i|x_i)\,f_{X_i}(x_i|x_{i-1})} \ . \tag{10}$$

We denote as $\hat{\mathbb{x}}^t = \{\hat{x}_0,\ldots,\hat{x}_t\}$ the optimal value of $\mathbb{X}^t$ that minimises $f_{\mathbb{X}^t}(x_0,\ldots,x_t|y_1,\ldots,y_t)$:

$$\begin{aligned}\hat{\mathbb{x}}^t &= \arg\max_{\{x_0,\ldots,x_t\}} f_{\mathbb{X}^t}(x_0,\ldots,x_t|y_1,\ldots,y_t) = \arg\max_{\{x_0,\ldots,x_t\}} \frac{f_{\mathbb{Y}^t,\mathbb{X}^t}(y_1,\ldots,y_t,x_0,\ldots,x_t)}{f_{\mathbb{Y}^t}(y_1,\ldots,y_t)} \\ &= \arg\max_{\{x_0,\ldots,x_t\}} f_{X_0}(x_0)\prod_{i=1}^{t} f_{Y_i}(y_i|x_i)\,f_{X_i}(x_i|x_{i-1}) \ .\end{aligned} \tag{11}$$

The last equivalence of (11) uses (8) and the fact that $f_{\mathbb{Y}^t}(y_1,\ldots,y_t)$ is a constant for varying values of $\mathbb{X}^t$. Under the assumption that $\hat{\mathbb{x}}^t$ is the only acceptable value for $\mathbb{X}^t$ given our strong constraint that X_t increases steadily during a normal movement, then other values for $\mathbb{X}^t$ have negligible weights in the integrals in (10), and (10) may be simplified as

$$f_{Y_t}(y_t|y_1,\ldots,y_{t-1}) \approx \frac{f_{X_0}(\hat{x}_0)\prod_{i=1}^{t} f_{Y_i}(y_i|\hat{x}_i)\,f_{X_i}(\hat{x}_i|\hat{x}_{i-1})}{f_{X_0}(\hat{x}_0)\prod_{i=1}^{t-1} f_{Y_i}(y_i|\hat{x}_i)\,f_{X_i}(\hat{x}_i|\hat{x}_{i-1})} \approx f_{Y_t}(y_t|\hat{x}_t)\,f_{X_t}(\hat{x}_t|\hat{x}_{t-1}) \ . \tag{12}$$

Note that this approximation is a lower bound of $f_{Y_t}(y_t|y_1,\ldots,y_{t-1})$. This is appropriate in our case, since it is preferable to have false alerts in a health monitoring system when the likelihood of a sequence to be normal is under-estimated, rather than to miss true alerts.

The dynamics model is built from our training data by estimating $f_{Y_t}(y_t|x_t) = \frac{f_{X_t,Y_t}(x_t,y_t)}{f_{X_t}(x_t)}$ using the same Parzen window estimator as previously. To compute $f_{X_t}(x_t|x_{t-1})$, since we assume a constant speed of the movement, we enforce $x_t - x_{t-1}$ to be proportional to the elapsed time, i.e. $x_t - x_{t-1} = \alpha(\tau_t - \tau_{t-1})$, with τ_t the time-stamp of frame t and α a proportionality constant. $f_{X_t}(x_t|x_{t-1})$ is then modelled as a Gaussian distribution around a perfect match between $x_t - x_{t-1}$ and $\alpha(\tau_t - \tau_{t-1})$:

$$f_{X_t}(x_t|x_{t-1}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{(x_t - x_{t-1}) - \alpha(\tau_t - \tau_{t-1})}{\sigma}\right)^2\right) \ . \tag{13}$$

σ is the standard deviation of the Gaussian distribution and its choice will be discussed next in Section 2.4. α is estimated at each new frame. Thus, the dynamics model adapts to the personal speed of the subject. However, when the speed varies significantly during a movement, this results in a low likelihood both for $\mathbb{X}^t$ and for this movement to be normal according to (12).

2.4 Movement quality assessment

The quality of a newly observed movement is assessed using two quality measures of pose and dynamics that are computed from the two models presented in Section 2.3. These measures represent the likelihoods of the movement to be described by the two models, and they

are computed on a frame-by-frame basis so that alerts may be triggered as early as possible when the observed movement drops below a threshold in its level of normality.

For each new frame t of an observation, the pose quality measure is computed for the new pose y_t, regardless of the previous frames, by computing its log-likelihood according to the pose model:

$$llh_{pose} = \log f_Y(y_t) \ . \tag{14}$$

llh_{pose} provides a continuous measure of the level of normality of the pose, distinguished by a threshold $thresh_{pose}$.

The dynamics quality measure is obtained by computing the log-likelihood of the sequence $(y_1,\ldots,y_{t-1})$ of poses from frame 1 to t to follow the dynamics model according to (12):

$$llh_{seq} = \log f_{Y_t}(y_t|y_1,\ldots,y_{t-1}) \approx \log\left(f_{Y_i}(y_t|\hat{x}_t)\, f_{X_i}(\hat{x}_t|\hat{x}_{t-1})\right) \ . \tag{15}$$

Similarly to llh_{pose}, llh_{seq} provides a continuous measure of the level of normality of the sequence of poses, *i.e.* the movement, again distinguished by threshold $thresh_{seq}$. The computation of llh_{seq} requires estimating the value of $\mathbb{X}^t$ first. This may be done, according to (11), by maximising $f_{\mathbb{Y}^t,\mathbb{X}^t}(y_1,\ldots,y_t,x_0,\ldots,x_t)$. In order to reduce computation time, we can use the fact that an estimate for $\{x_0,\ldots,x_{t-1}\}$ was already computed at the previous iteration. We note that, after a few iterations, the estimated value x_i at any previous frame i does not change significantly any more. Thus, we may consider that the optimal value for X_i has been found and stop re-estimating it. Following this idea, we define a temporal window of variable size ω that contains all the frames i for which x_i has not yet converged:

$$\omega = t - t_{min} + 1 \ , \tag{16}$$

with t_{min} the oldest frame that requires a re-estimation of $x_{t_{min}}$. In our implementation, x_i is considered to have converged when its change is $< 10^{-3}$ during 2 consecutive iterations. Thus, ω is set at each iteration. For convenience and efficiency, we limit ω to a maximum of 15 frames, although it rarely goes above 10 frames. All the values of X_i within this window, denoted as $\varkappa^\omega = \{x_{t_{min}},\ldots,x_t\}$, are estimated by solving the following modification of (11):

$$\hat{\varkappa^\omega} = \arg\max_{\varkappa^\omega} f_{X_{t_{min}-1}}(x_{t_{min}-1}) \prod_{i=t_{min}}^{t} f_{Y_i}(y_i|x_i)\, f_{X_i}(x_i|x_{i-1}) \ . \tag{17}$$

The estimated values x_i may be kept between 0 and 1 by using the modulo operator in the case of periodic movements such as gait, since under the condition of (13) they would tend to keep increasing during consecutive cycles together with the time τ.

In (12), $\hat{\varkappa^t}$ is considered to be the best and only acceptable value for $\mathbb{X}^t$. In our case, the value of $\mathbb{X}^t$ converges progressively, and at iteration t all its values within the window ω are re-estimated. Thus, in order to take into account the confidence in the newly estimated values of $\mathbb{X}^\omega$ and not only in x_t, we modify (15) as

$$llh_{seq} \approx \frac{1}{\omega} \sum_{i=t_{min}}^{t} \log\left(f_{Y_i}(y_i|x_i)\, f_{X_i}(x_i|x_{i-1})\right) \ . \tag{18}$$

For the computation of $f_{X_i}(x_i|x_{i-1})$ in (17) and (18) using (13), the value of α needs to be estimated. We simply use the average proportionality between x_i and τ_i inside the temporal window:

$$\alpha = \frac{1}{\omega - 1} \sum_{i=t_{min}}^{t-1} \frac{x_i - x_{i-1}}{\tau_i - \tau_{i-1}} \ . \tag{19}$$

The value of σ affects the flexibility of the estimation of $\mathbb{X}^{\omega}$ and should therefore be chosen with care. A low value for σ enforces a strong stable speed constraint on $\mathbb{X}^{\omega}$ that, in the case of an abnormal movement with significant speed variations, may prevent the model to match the observed data properly. On the contrary, a higher value for σ would provide more flexibility to better describe the movement, at the cost of a lesser penalisation of movements with irregular speeds. As a compromise, we use two distinct values for σ, with a relatively high value σ_{estim} during the estimation of $\hat{\varkappa^{\omega}}$, and a lower value σ_{assess} for the computation of llh_{seq}.

3 Experimental results

We evaluate our proposed method on walking-up-stairs movement against manual detection of abnormalities by a physiotherapist. Analysing such gait has obvious relevance for several clinical applications [5, 16, 18].

We build our model from 17 sequences, using 6 healthy subjects having no injury or disability, from which we extract 42 gait cycles[3]. We empirically determined and set $\beta = 0.01$, $\sigma_{estim} = 7e{-}3$, $\sigma_{assess} = 10^{-3}$, $thresh_{pose} = -2.5$, and $thresh_{seq} = 2$.

We first prove the ability of the proposed method to generalise to movements of new subjects by assessing the normal gait of 6 subjects who were not involved in the training phase. In the majority of the normal sequences, the gaits of the new subjects are judged as normal by our method, with only one false detection of anomaly in 13 normal sequences[4].

Next, we evaluate the ability of our method to generalise to various types of abnormality. A qualified physiotherapist (who was not included in the model training phase) simulated three standard scenarios of knee injury that are illustrated in Fig. 3, and labelled the abnormal frames manually (blue shaded areas in Fig. 3). Five other subjects simulated the same range of anomalies under his guidance. The top row of Fig. 3 presents the values of the first dimension of the reduced pose vector $\mathbf{Y}$, which clearly embodies the periodicity of the data. The second row displays the estimated values of the movement stage x_t. The third and bottom rows present llh_{pose} and llh_{seq} respectively. Superimposed in colour are the decisions of our system: in green, the frames that are judged sufficiently close to the normal model, while in red, the frames for which llh_{pose} or llh_{seq} are below the acceptable thresholds $thresh_{pose}$ and $thresh_{seq}$ respectively. In orange are refined detections of deviations of the movement from normal, just before llh_{seq} drops below $thresh_{seq}$ and an alert is triggered. These frames are found by examining the derivative of llh_{seq} and detecting its sudden change. In our implementation, we simply detect decrease rates of llh_{seq} higher than 0.3. Similarly, in blue are the refined detections of frames that are back to normal, with increase rate of llh_{seq} higher than 0.3. This refinement strategy attempts to compensate for the delay in changes of llh_{seq} that is due to the computation of llh_{seq} over the window w.

In our first two tests, the subjects walk up the stairs always initially using their right leg (see the "Right leg lead" or RL test in Fig. 3) or left leg ("Left leg lead" or LL test in Fig. 3) to move to the next upper step. In both cases, the pose of the subjects does not deviate significantly from the pose model, thus llh_{pose} remains above $thresh_{pose}$. On the contrary,

[3] Our training and testing sequences, along with the ground-truth, are available on our project webpage at www.irc-sphere.ac.uk/work-package-2/movement-quality.

[4] We have not compared against Mihailidis and co-workers [11, 13] - the only works that we know of which analyse gait on stairs - as their codes and labelled groundtruth data was not available.

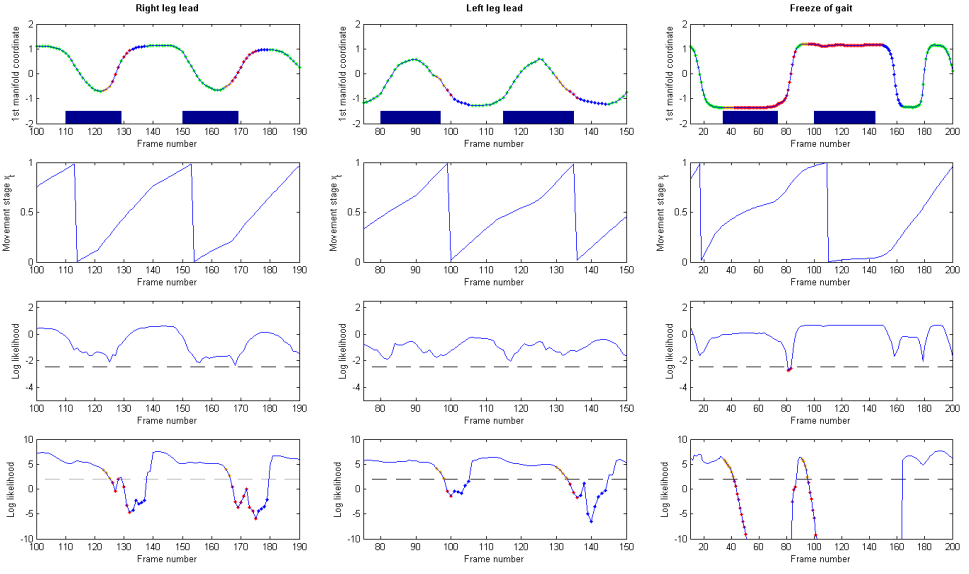

Figure 3: Analysis of abnormal gaits on stairs from three new subjects. For each test: top row: first dimension of $\mathbf{Y}$, second row: estimated movement stage x_t, third row: llh_{pose}, bottom row: llh_{seq}. Dotted lines: thresholds for normal/abnormal classification. Superimposed in colour are the online movement assessment of our system: green: acceptably normal movement, red: abnormal movement, orange: refined detections of abnormal frames, blue: refined detection of normal frames. The blue shaded areas are the manual labelling of anomalous frames. Note, in the right and left leg lead cases, these detections by our method are delayed due to the method responding to the use of the wrong leg in the cycle rather than to the early slowing down of the other leg.

Type of event	No. of occurrences	FP	TP	FN	Proportion missed
RL	25	0	23	2	0.08
LL	21	0	19	2	0.10
Freeze	12	2	12	0	0
All	**58**	**2**	**54**	**4**	**0.07**

Table 1: Results on detection of abnormal events

Type of event	No. frames with event	FP frames	TP frames	FN frames	TN frames	False positive rate	Proportion missed
RL	500	144	223	276	363	0.28	0.55
LL	435	117	108	327	263	0.31	0.75
Freeze	658	164	536	122	791	0.17	0.19
All	**1593**	**425**	**867**	**725**	**1417**	**0.23**	**0.46**

Table 2: Results on classification of frames

the repeated use of the same leg makes part of each cycle in disagreement with the dynamics model and is detected in every cycle.

In our last test (right of Fig. 3), the subjects freeze at some stage of the movement. Note that this type of anomaly deviates more strongly from the normal model, due to variable X_t not evolving any more, which is in contradiction with the dynamics embodied in (13). Thus, this triggers a much stronger response from the system as seen in the bottom part of Fig. 3. When the subject freezes then resumes a normal gait, the freeze is correctly detected by the system, as well as the return to normality. In Fig. 3, this ability to resume the analysis after the gait returns to normal allows the system to detect a second freeze that happens immediately, i.e. within one gait cycle, after the first one.

Table 1 presents the true, false, and missed detections of abnormal events in all the sequences, and Table 2 provides similar measures regarding the classifications of individual frames. The results show all three types of events detected with a rate of 0.93, with only 2 false positive detections out of 58 events. The frame classification is less satisfactory, with overall false positive rate at 0.23 and proportion of missed abnormal frames at 0.46 . The frame classification is especially difficult for the RL and LL anomalies, where the detections are often in phase opposition with the ground-truth, resulting in a high amount of false positive and false negative classifications. This is mostly due to the alarm being triggered late by the use of the unexpected leg rather than by the premature stopping of the previous leg. Indeed, the computation of α over a local window makes the method able to adapt to the decrease of the movement speed to some extent. Similarly, for the freeze of gait events, the alarm is frequently delayed until α has changed significantly.

To demonstrate the flexibility of our method, e.g. non-periodic movements, we also applied it to boxing and sitting-and-standing movements. These results are reported elsewhere (see last footnote) due to lack of space here.

4 Conclusion

We have presented a method for analysing the quality of movements from skeleton representations of the human body. The method makes use of a robust manifold technique to reduce the dimensionality of the noisy skeleton data, and then compares the resulting features with a pose and dynamics model learnt from normal occurrences of a movement. We tested the method on gait on stairs, and demonstrated its ability to generalise to the movements of unknown subjects and to detect a range of abnormality types. Future work include further assessment on both periodic and non-periodic movements and comparison against other methods, as well as evaluating the benefit from the continuous measure of movement quality that our likelihoods provide, against a binary classification of normal vs. abnormal.

Acknowledgements

This work was performed under the SPHERE IRC funded by the UK Engineering and Physical Sciences Research Council (EPSRC), Grant EP/K031910/1.

References

[1] P. Arias, G. Randall, and G. Sapiro. Connecting the out-of sample and pre-image problems in kernel methods. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8. IEEE, 2007.

[2] A. A. Chaaraoui, P. Climent-Pérez, and F. Flórez-Revuelta. A review on vision techniques applied to human behaviour analysis for ambient-assisted living. Expert Systems with Applications, 39(12):10873–10888, 2012.

[3] R. R. Coifman and S. Lafon. Diffusion maps. Applied and computational harmonic analysis, 21(1):5–30, 2006.

[4] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. Proceedings of the National Academy of Sciences of the United States of America, 102(21):7426–7431, 2005.

[5] C. P. S. Commission. National injury surveillance system. Online, 2005.

[6] S. Gerber, T. Tasdizen, and R. Whitaker. Robust non-linear dimensionality reduction using successive 1-dimensional Laplacian eigenmaps. In Proceedings of the 24th international conference on Machine learning, pages 281–288. ACM, 2007.

[7] D. Gong and G. Medioni. Dynamic manifold warping for view invariant action recognition. In IEEE International Conference on Computer Vision, pages 571–578. IEEE, 2011.

[8] L. Mei, J. Liu, A. Hero, and S. Savarese. Robust object pose estimation via statistical manifold modeling. In IEEE International Conference on Computer Vision, pages 967–974. IEEE, 2011.

[9] F. Nater, H. Grabner, and L. Van Gool. Exploiting simple hierarchies for unsupervised human behavior analysis. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2014–2021. IEEE, 2010.

[10] OpenNI User Guide. OpenNI organization, November 2010. URL http://www.openni.org/documentation.

[11] G. S. Parra-Dominguez, B. Taati, and A. Mihailidis. 3D human motion analysis to detect abnormal events on stairs. In International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, pages 97–103. IEEE, 2012.

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

[13] O. P. Popoola and K. Wang. Video-based abnormal human behavior recognition a review. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 42(6):865–878, 2012.

[14] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from a single depth image. In IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[15] J. Snoek, J. Hoey, L. Stewart, R. S. Zemel, and A. Mihailidis. Automated detection of unusual events on stairs. Image and Vision Computing, 27(1):153–166, 2009.

[16] J. A. Templer. The Staircase: Studies of Hazards, Falls, and Safer Design. MIT Press, 1994.

[17] M. Z. Uddin, J. T. Kim, and T. S. Kim. Depth video-based gait recognition for smart home using local directional pattern features and hidden Markov model. Indoor and Built Environment, 23(1):133–140, 2014.

[18] R. Wang, G. Medioni, C. J. Winstein, and C. Blanco. Home monitoring musculoskeletal disorders with a single 3D sensor. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 521–528. IEEE, 2013.

[19] F. Zhou, F. De la Torre, and J. K. Hodgins. Hierarchical aligned cluster analysis for temporal clustering of human motion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(3):582–596, 2013.