Toward an Efficient Body Expression Recognition Based on the Synthesis of a Neutral Movement

Arthur Crenn, Alexandre Meyer, Rizwan Ahmed Khan, Hubert Konik, Saïda Bouakaz

Toward an Efficient Body Expression Recognition Based on the Synthesis of a Neutral Movement

Arthur Crenn
Université de Lyon, CNRS, France
Université Lyon 1, LIRIS, UMR5205, F-69622,, France
arthur.crenn@liris.univ-lyon1.fr

Alexandre Meyer
Université de Lyon, CNRS, France
Université Lyon 1, LIRIS, UMR5205, F-69622,, France
alexandre.meyer@liris.univ-lyon1.fr

Rizwan Ahmed Khan
Université de Lyon, CNRS, France
Université Lyon 1, LIRIS, UMR5205, F-69622, France
Faculty of IT, Barrett Hodgson University, Karachi,, Pakistan

Hubert Konik
Université de Saint-Etienne, LHC, UMR5516, F-42000,, France
hubert.konik@univ-st-etienne.fr

Saida Bouakaz
Université de Lyon, CNRS, France
Université Lyon 1, LIRIS, UMR5205, F-69622,, France
saida.bouakaz@liris.univ-lyon1.fr

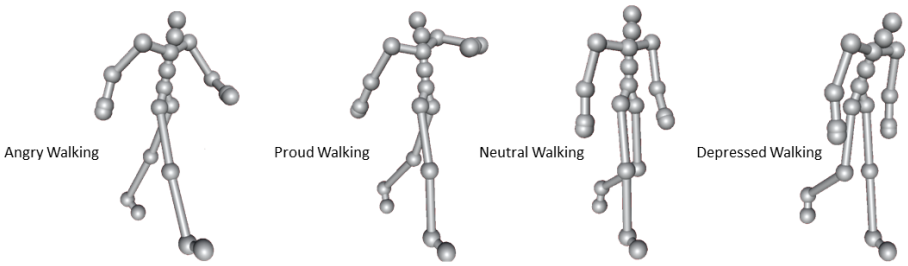

Figure 1: Our method recognizes the expression of a human 3D skeleton animation.

ABSTRACT

We present a novel framework for the recognition of body expressions using human postures. Proposed system is based on analyzing the spectral difference between an expressive and a neutral animation. Second problem that has been addressed in this paper is formalization of neutral animation. Formalization of neutral animation has not been tackled before and it can be very useful for the domain of synthesis of animation, recognition of expressions, etc. In this article, we proposed a cost function to synthesize a neutral motion from expressive motion. The cost function formalizes a neutral motion by computing the distance and by combining it with acceleration of each body joints during a motion. We have evaluated our approach on several databases with heterogeneous movements and body expressions. Our body expression recognition results exceeds state of the art on evaluated databases.

CCS CONCEPTS

• **Computing methodologies** → **Activity recognition and understanding**; *Motion capture*;

KEYWORDS

Computer Vision;Body Expression;Automatic Recognition;3D Skeleton;Classification

ACM Reference Format:
Arthur Crenn, Alexandre Meyer, Rizwan Ahmed Khan, Hubert Konik, and Saida Bouakaz. 2017. Toward an Efficient Body Expression Recognition Based on the Synthesis of a Neutral Movement . In *Proceedings of 19th ACM International Conference on Multimodal Interaction (ICMI'17).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3136755.3136763

1 INTRODUCTION

Many applications would benefit from the ability to understand human emotional state in order to provide more natural interaction, e.g. video games, video surveillance, human-computer interaction, artistic creation, etc. Emotion is a complex phenomena that is difficult to formalize. Expression of an emotion could be personal and subjective as two persons could perceive and interpret differently the same expression. Its perception changes from one culture to another [18]. Furthermore, human expresses emotional information through several channels, like facial expression, body movement and sound. Several studies from various domains have shown that body expressions are as powerful as those of the face to express emotion [17]. Whereas facial expression recognition was widely studied [7, 13, 15, 16], body expression recognition is more an emerging area. Early proposed methods for body movement analysis [14, 19] were limited to specific movements or expressions. Nevertheless,

ICMI'17, November 13–17, 2017, Glasgow, UK

with the growth and ease of accessibility of devices that track 3-dimensional body like the Kinect [4, 22] or accelerometer [26] based motion capture system, many applications will benefit from analysis of body movements.

This paper presents a method to detect and classify expression through a sequence of 3D skeleton-based poses. The challenge is to have a system that recognize the expression invariant to body movement. For example, expression of happiness could be shown while running, jumping, kicking, etc.

In the rest of the paper, depending on the case involved we use the term "Body Expressions" to refer to emotion expressed by body posture while term of style refers to the domain of synthesis of animation. Style is an important component of character animation, as that includes aesthetic, original, and emotional characteristics. For instance, an old walking, a depressed kicking, a character with a wooden leg, etc. are different way to express how a motion differs from the neutral movement.

Inspired by style transfer in animation synthesis [33], we have used the difference between an expressive and corresponding neutral animation for classification. The difference efficiently makes the purpose system independent of the movement. Computing this residue requires a neutral motion. Since real applications capture the whole movement including the expression, we proposed to retrieve a neutral motion from an expressive motion. We compute the residue in the frequency domain between the neutral motion produced by our method and the input motion containing the expression. This residue is used by classifier in order to recognize the body expression. We have evaluated our work on four databases that contains heterogeneous movements and expressions.

The paper is organized as follows. Section 2 presents the state of the art on emotion analysis. Section 3 details our method. In Section 4 we present the results obtained on different databases. We compared our method with state-of-the-art approaches. Finally, Section 5 concludes the paper and presents future work.

2 RELATED WORK

In recent years, researchers have mainly focused on automatic facial expression recognition (FER) [7, 13, 16]. However, the automatic detection of expression based on human posture is an emerging topic motivated by the recent ease to capture human motions; without marker and with low price devices. Furthermore, psychological studies show that the human posture is as powerful as facial expressions in conveying emotions [21]. First, we present the classical approaches based on features extraction, then we continue with the different representation of motions in order to motivate our approach.

Many existing approaches focus on very specific actions, as often done in psychology studies. Many papers have focused on the locomotion [5, 14, 23, 24, 24], some on knocking actions [6, 11], on artistic performance [25], or on talking persons [29]. Body movements are characterized by a high dimensional configuration space with many interrelated degrees of freedom. Psychologists' studies have sought to understand body expression according to body movements, i.e. form and movement. Two main levels of body descriptors have emerged: high- and low-level descriptors. The low-level descriptors provide features that quantifies the movements (joint angles, 3D positions, distance between joints, velocity, acceleration, etc.). We note a lack of common vocabulary in different papers whereas they use similar descriptors. The high-level descriptors are very context-dependent and thus, are difficult to compare in a general point of views. The most popular formalism is certainly the one used to analyze dance movements proposed by Laban and Ullmann [30]. The Laban analysis provides a consistent representation of gestures expressibility. This model consists of characterizing the body motion in terms of a fixed number of qualities. The Laban model analysis includes five major qualities: body, relationship, space, effort, and shape which allows summarizing and interpreting movements with a small set of parameters.

Few state of the art papers tackle the analysis of heterogeneous body movement. Kleinsmith et al. [18] proposed the UCLIC database featuring 13 participants from different cultural regions, portraying four emotions (anger, fear, happiness and sadness). Wang et al. [31] proposed a real-time system that recognizes emotions from body movements. They used a combination of low-level 3D postural features and high level kinematic and geometrical features. They obtained a classification rate of 78% with a Random Forest classifier on the UCLIC database, we will compare our method with their results in the Section 4 for comparison. Truong et al. [27] proposed a new set of 3D gesture descriptors based on a generalization of the Laban descriptor model for gestures expressiveness. They obtained a recognition rate for action recognition of 97% on the Microsoft Research Cambridge-12 dataset [10]. They also tested their classification approach on their own database which contains 882 gestures and achieved best F-Score of 56.9%. The F-score is a measure of a test accuracy. Finally, Crenn et al. [9] proposed a new set of two-levels 3D descriptors based on psychological studies. They proposed low-level features which are based on visual cues, and the high-level features are statistic operator in order to reduce the feature vector size and compact the information. They have evaluated their approach on different databases with heterogeneous movements and body expression. They obtained a classification rate of 93% on a synthetic database and results at par with state of the art for another database.

We have presented different methods based on features extraction in order to recognize body expression. The main issue with that kind of approaches is that they are not invariant to the movement performed whereas it is very important given the high degrees of freedom of the body. Indeed, most of the features used by the different methods mentioned above are highly dependent on the movement realized (speed, form, distance, etc.). From this conclusion, we have modelized a style as an enrichment of a gesture. We argue that separating the gesture from the expression is probably an important point in the analysis of expressions in an movement. Our method is inspired by the domain of the Computer Graphics which used different representations of body movement and gesture expressiveness to edit and generate new animations. These editions of an animation can be done by time warping, editing speed and spatial amplitude of the motions of body part joints [1, 12, 32]. Early in Computer Graphics, researchers [8, 28] proposed to edit an animation as a signal using Fourier transform. Recently, with the goal of transferring expressions between two animations, Yumer and Mitra [33] proposed an approach which succeeds to transfer a

style based on the residue obtains from the difference on the spectral intensity representations of reference and source styles for an arbitrary action. They showed that the difference in the frequency domain between an expressive motion and neutral one is highly correlated even though the actions are different. Our approach is inspired by their method as we believe that the spectral residue is more invariant to the movement than methods based on feature extraction.

3 PROPOSED METHOD

Our main focus in this work is to recognize body expressions independently of the movement performed. Our method is based on a neutral movement synthesis from a given expressive one. Our framework consists of two steps, the synthesis of the neutral motion and the classification based on the residue. As pointed out above, we assume that spectral difference between two movements contains the information for the extraction of body expression. The first problem consists to obtain a neutral animation from an expressive movement. The overview of our approach is given in Figure 2.

3.1 Neutral Animation Synthesis

The notion of neutral versus expressive movement is always linked to a context. Often understood as a sequence of actions without an emotional mark, a neutral movement may eventually be confused with a robotic movement. Our approach is based on a filtering of the trajectories of each articulation in order to reduce the oscillations in the expressive movement. The second step used inverse kinematics to produce a robotic motion without expression. Finally, we introduce a cost function that will compromise between joint filtering and inverse kinematics.

3.1.1 Body Joint Trajectory Smoothing. The input motion data are temporal samples of joint angles that represent the skeleton expressive animation. Our first step is to compute the joint trajectories in 3D and convert them in a cubic B-spline formalism 1. The B-spline curves are well-adapted for the operations of simplification we will apply on the trajectories to derive a neutral motion. The Equation of a B-splines is:

$$\mathbf{S}(t) = \sum_{i=0}^{m-n-1} b_{i,n}(t).\mathbf{P}_i, t \in [0,1] \tag{1}$$

where n is the order of our B-spline. In our case, we use the Cubic B-spline so we have $n = 2$. m is the number of control points. The points $\mathbf{P}_i$ form a polygon called the control polygon: the number of points making up this polygon is $m - n$. These $m - n$ functions B-spline of degree n are defined by recurrence on the lower degree:

$$\begin{aligned} b_{j,0}(t) &= \begin{cases} 1, & \text{if } t_j \le t \le t_{j+1} \\ 0, & \text{otherwise} \end{cases} \\ b_{j,n}(t) &= \frac{t - t_j}{t_{j+n} - t_j} b_{j,n-1}(t) + \frac{t_{j+n+1}}{t_{j+n+1} - tj + 1} b_{j+1,n-1}(t) \end{aligned} \tag{2}$$

As a first step, we smooth the original trajectories of each body joints represented by B-spline. We decimate each B-spline by removing one control point from each curve. We reconstruct the B-spline with the remaining control points. Our method for smoothing the trajectory of body joint is presented is the Figure 3. As one can see from Figure 3, we remove one control point per curve, i.e. on each $m - n$) functions. The position of the new control points is interpolated on each $m - n$ functions with a parameter t. On the Figure 3, we used $t = 0.5$.

Our iterative method has two parameters that control the number of iterations and the position of the new points by a cubic interpolation. The different levels of details which can be generated from the input trajectory are presented on the Figure 4. This simplification reduced the variation of the motion for each body joint trajectory. By analyzing lots of expressive movements in different databases, we have categorized two types of expressive trajectories: energetic (happy, proud, anger, etc.) and moderate (depressed, old, etc.) movements. A neutral trajectory is more "flatten" that an energetic trajectory, i.e. joint has less variations during the motion and can be clearly produced by this smoothing step. For the moderate movements the smoothing step does not change much the trajectory but the IK step during the optimization process presented in the two next Sections will edit the movement toward a neutral one.

Our method to smooth an input motion is presented on a the Figure 5. The input motion is represented by the black curve, the B-spline obtained by our method is represented by the green curve and the ground truth B-spline is represented by the red curve. The ground truth curve is the neutral trajectory of the joint as provided in the databases including neutral movement. These curves show the right foot during a kicking action. With a simple pass of smoothing, we synthesize a new motion similar to the neutral motion. To obtain this neutral trajectory from the input expressive one, we have manually tweaked the simplifications parameters of the B-spline. Finding efficient parameters working on any kind of movements is manually unreachable. In the Section 3.1.3 we propose an optimization process based on a function which characterize a neutral motion. In addition to these parameters setting, the trajectories generated individually by this method do not guaranty that the constraints of the animation are respected like the constant length of the bones or the feet that do not slide on the ground.

3.1.2 Inverse Kinematics. Smoothing the B-spline of the joint trajectories is not sufficient to produce a neutral movement. First, the motion generated from the smoothing step does not respect the constant length of the bones and suffer from artifacts as feet sliding on the ground. To solve these issues, we apply an Inverse Kinematics (IK) technique. IK is defined as the problem of determining a set of appropriate joint configurations for which the end effectors move to the desired position as smoothly, rapidly, and as accurately as possible. We use the Fabrik algorithm proposed by Aristidou and Lasenby in [3]. Fabrik has the advantages to converge in few iterations, produces visually realistic poses and other constraints like non-foot-sliding can easily be included. In our method, we used Fabrik with Model Constraints [2] in order to represent a human-like model. The Model Constraints is an extension of the Fabrik method that proposed a human-like model using Fabrik.

The Algorithm 1 describes the different steps to synthesize a neutral motion from an expressive one and the different parameters used during the computation. This algorithm with an efficient set of parameters allows to synthesize a neutral motion from any kind of expressive motion. Even if it may produce a neutral motion which

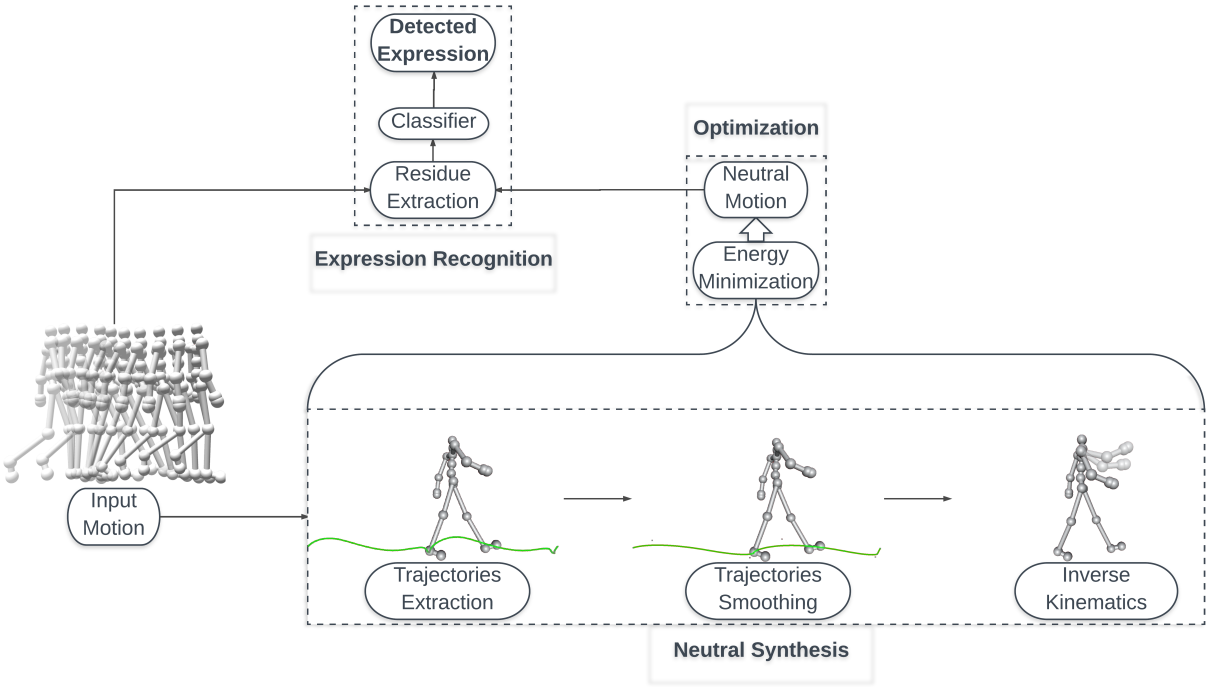

Figure 2: Overview of our framework.

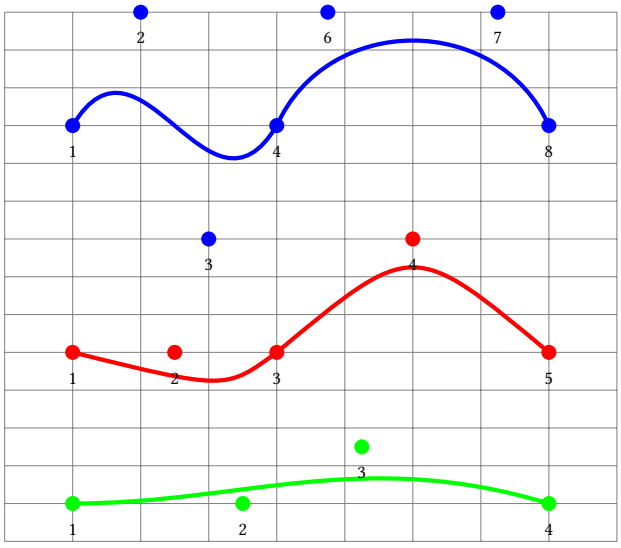

Figure 3: Our method for smoothing a body joint trajectory based on a decimation of B-spline control points. Blue B-spline represents the original motion. Red B-spline shows the smoothed curve after one iteration, green B-spine after two iterations.

could be a bit "robotic", it is nevertheless efficient for the classification part, as confirmed by the recognition rate in Section 4. The algorithm has a set of 16 parameters in order to produce a neutral motion. Finding efficient values for these different parameters is not manually tractable for different motions on different databases.

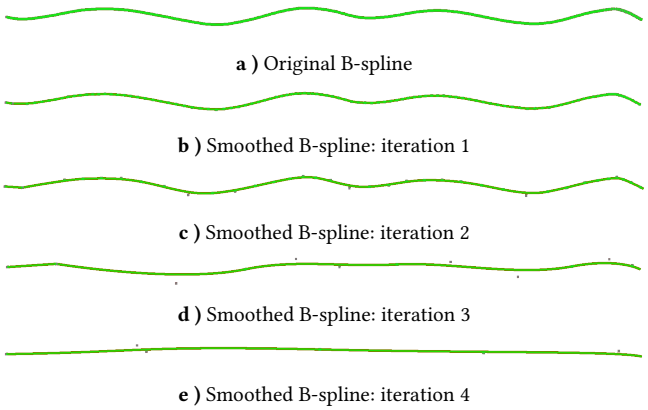

Figure 4: Different level of detail of an input trajectory generated by our smoothing tool.

We have proposed an energy-based term which describes a neutral motion in the next Section and an optimization process provide parameters to generate a neutral animation.

3.1.3 Cost Function for generating a neutral motion. The several parameters of the trajectories simplification and the IK step are determined by a minimization process. For this purpose, we propose a function that characterizes a coarse neutral animation. For the purpose of an intuitive formalism we have designed the function as more the value returned by this function is small, more "neutral" the animation is. A neutral motion costs little. The minimization

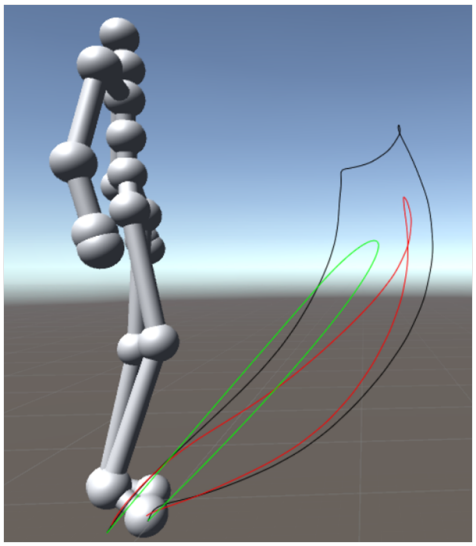

Figure 5: Comparison of different trajectories for a kicking action. The input motion is in black (angry expression). The "neutral" trajectory generated by our method from the input one is in green. And in red, the neutral trajectory provided by the database may serve as a comparison.

process seeks to find the best parameters of the trajectories simplifications and the IK step that produce the lowest value by the cost function as possible for any kind of input action as jump, run, walk, etc. This cost function that characterizes the neutrality of a motion is relatively simple and produce coarse neutral animations but these coarse animations are good enough for our purpose of classification. We based our criteria on the distance covered by each joints and their acceleration during an action. We assume that a neutral movement is related to the muscular energy produced by the person. The person aims to rest his body while reaching the gesture he wants to do. Thus, we minimize the distance and the acceleration done by the joints. Denoting $D_s(j)$ (respectively $D_o(j)$) the distance covered by the joint j during an action computed on the synthesized movement (respectively the original movement). Given $\mathbf{A_s}(j)$ (respectively $\mathbf{A_o}(j)$) the acceleration of the joint j during an action computed on the synthesized movement (respectively on the original movement). The cost function is defined as the sum of the difference between original distance and acceleration for each joint and computed one. The minimization of the cost function provides a coarse neutral animation used in Section 3.2 to compute the residue between the original motion and the synthesized neutral one.

$$Cost = \sum_{j \in \theta} |(1-\lambda)(D_s(j) - D_o(j)) + \lambda(\mathbf{A_s}(j) - \mathbf{A_o}(j))|^2 \quad (3)$$

with j represent a body joint, θ is the set of body joints of the input skeleton and $\lambda \in [0, 1]$ is a weight parameter. The influence of the factor λ, for the classification rate on the Siggraph database, is shown in Figure 6. With $\lambda = 0$, we are only using the distance term: and with $\lambda = 1$, we are only using the acceleration term.

Algorithm 1: Algorithm for neutral movement synthesis

Input: $M_i i$: Input Motion
Data: joints: a table of all body joints
trajectories: a table of all trajectories for each joint
trajectoriesSmooth: a table of all trajectories smooth for each joint
Result: M_n: Neutral Motion
Parameters: samplingValue: temporal parameter determining anchoring postures for the IK step
weightTargets: a table of weight that balance the IK goal (1 = at the goal and 0 = idle pose of our skeleton) for the rotation and position.
weightHints: a table of weight that balance the IK hint (1 = at the goal and 0 = idle pose of our skeleton) for the rotation and position.

```
1  foreach joint, j_i, in joints do
2      trajectories_i ← computeTrajectory(j_i);
3      trajectoriesSmooth_i ← smoothTrajectory(trajectories_i);
4  end
5  foreach EndEffector, end_i, in joints do
6      indiceHint ← end_i.getParent();
7      for i =0; i < endTime; i += samplingValue do
8          target_{end_i} ← trajectoriesSmooth_i ;
9          hint_{end_i} ← trajectoriesSmooth_{indiceHint} ;
10         IK_Step( target_{end_i}, hint_{end_i}, weightTargets,
             weightHints);
11     end
12 end
```

The distances $D_o(j)$ and $D_s(j)$ covered by a joint j during an action are given by the length of our B-spline. These distances are approximating by sampling multiple points on our spline and then calculate the distance between these points. The accelerations $\mathbf{E}_o(j)$ and $\mathbf{E}_s(j)$ of a joint j gives us information about the energy spend by joint during the action. To find these accelerations, we are calculating the second derivative of each B-spline per joint.

Implementation Details The cost function is minimized iteratively using Particle Swarm Optimization (PSO). PSO has the advantage to require few or no assumptions about the function being optimized and can search in very large spaces of candidate solutions. The main issue with PSO is that the solution founded is not guaranteed to be the optimal solution. Based on our tests, finding a good solution instead of the optimal one is sufficient, since the coarse neutral animation is used for the differentiation with the expressive one before the classification. We have experimented different value of λ (see Figure 6) and we have empirically found that $\lambda = 0.2$ is a good default setting for all databases.

3.2 Residue Between Neutral and Expressive Motion

At this stage, we have the original expressive motion and a coarse neutral motion obtained by optimizing the cost function. In order to extract the expression from the original input motion, our idea is to analyze the difference between the neutral and the expressive motion. This difference extracts the body expression invariant to

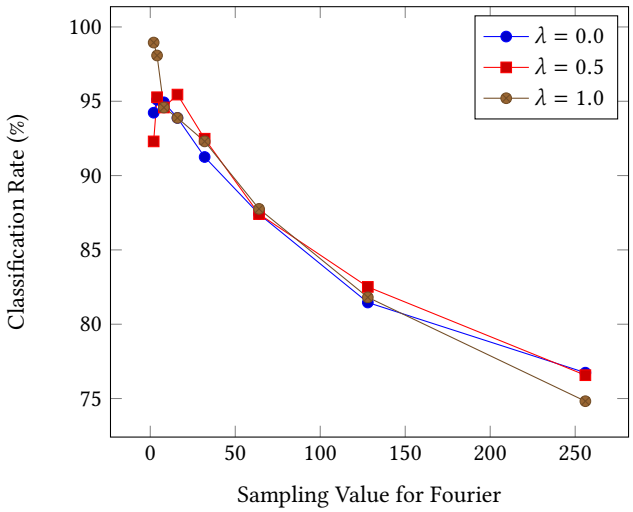

Figure 6: Comparison of the classification rate with different values of λ and sampling on the SIGGRAPH Database.

body movement. We argue that a spectral representation of a motion is well adapted to separate the expression from the gesture. This assumption is supported by the work of Yumer and Mitra [33] that manages to transfer a style between animations in the spectral domain. We calculate the spectral representation of the input motion and the synthesized neutral motion. Then, we subtract both magnitudes for each body joints and obtain values that include mainly the expression of animation.

We recall the formulation of the **Discrete Fourier Transform (DFT)**. Let x_n be a discrete time domain signal of one of the degrees of freedom (DOF) of a human motion data. The Discrete Fourier Transform X_k of x_n is given by:

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i2\pi kn/N} \tag{4}$$

where N is the length of the signal and $i^2 = -1$. The single-sided spectrum X_ω is given by:

$$X_\omega = \frac{2}{N} X_k \quad k = 0, ..., N/2 \tag{5}$$

where $\omega = (x_s/N)k$ is the frequency transformed from the samples k in the spectral space. We only use the single-sided spectrum in the positive frequency range ($\omega = 0 : x_s/2$). Here, x_s is the sampling frequency of the original time domain signal x_n. From this spectral representation, we can extract the magnitude and phase of the spectra. The magnitude defines the existence and intensity of a motion whereas the phase describes the relative timing.

The residue between the neutral animation and the expressive animation is calculated for each degree of freedom (DOF) of each joint in the skeleton independently from the others. It consists as a subtraction between the neutral spectral magnitude and the expressive spectral magnitude. In formal manner, we describe it by the Equation 6 where $M_s(\omega, j, l)$ (respectively $M_o(\omega, j, l)$) is the spectral magnitude for the joint j and for the DOF l at the frequency ω during an action computed on the synthesized neutral movement (respectively on the original movement).

$$\begin{aligned} \textbf{Residue} = & (|M_o(\omega, j, l) - M_s(\omega, j, l)|) \\ & j \in \theta \\ & l \in DOF \\ & \omega \in 1..N \quad \text{where N is the signal length} \end{aligned} \tag{6}$$

The magnitude contains the information about the motion and the expression of an animation, it provides enough information. The residue forms the feature vector we use as input data of the classifiers in order to get the expression type. The classification rate given in the Section 4 is the ratio between the number of good classification divided by the number of tested animation. In our approach, we have to define the number of sampling of the input signal N. The Figure 6 illustrates the variation of the classification rate (in %) with different sampling value. The classification rate decreases when the sampling value is increased, this is due to the fact that it increases the size of the feature vector which generate noise for the classifier. The size of our feature vector is given by the number of DOF in the skeleton multiplied by this sampling value of the input signal. We have set the sampling value to 8 for the Section 4.

4 RESULTS AND ANALYSIS

We have tested our method on four databases where characteristics are detailed in the Table 1. Three of them are acted databases, while the last one consists of synthetic animations generated by the method of Xia et al. [32]. We will refer the last database as the SIGGRAPH database in the rest of this paper.

(1) Biological Motion [20]. This database contains 4080 motions (walking, knocking, lifting and throwing) with 4 expressions (angry, neutral, happy, sad). Unfortunately, we used only a subset of 1356 motions, especially knocking as it is the set of motions available on the web. The state of the art presented in Table 2 uses the same set of motions we used. This database contains 15 male and 15 female amateur actors with a mean age of 22 years. Motions have been recorded using a motion capture system leading to a set of 35 body joints.
(2) UCLIC Affective Body Posture and Motion [18]. This acted database contains 183 animations with 4 expressions (fear, sad, happy, angry). This database contains 13 human subjects from different cultural regions. Actors were directed to perform the emotion postures in their own way. They used a motion capture system to collect 3D affective postures leading to a set of 32 body joints.
(3) MPI Emotional Body Expressions Database for Narrative Scenarios [29]. This database consists of 1447 motions of amateur actors narrating coherent stories. This database contains 8 actors, 4 females and 4 males with a mean age of 25 years. Actors were asked to imagine that they were narrating several stories to children. It contains 11 emotions (amusement, anger, disgust, fear, joy, neutral, pride, relief, sadness, shame, surprise). They used a motion capture system to collect 3D postures leading to set of 22 joints. This database is highly imbalanced is terms of expressions. Joy

is the most represented expression with 227 instances and shame is the less represented expression with 58 instances.

(4) SIGGRAPH database [32]. The SIGGRAPH database is a database used in synthesis of animation. They recorded 11 minutes of motion capture data. It contains 572 animations with 8 expressions or style (angry, childlike, depressed, neutral, old, proud, sexy, strutting). Notices that the SIGGRAPH database includes the largest range of movements: jump, run, kick, walk, punching and transitions between these motions.

Table 1: Description of the databases used for the test of our method.

DataBase	Number of movements	Number of expressions
UCLIC [18]	183	4
Biological [20]	1356	4
SIGGRAPH [32]	572	4 and 4 styles
MPI [29]	1443	11

Figure 7 shows the influence of the size of the training set on the performance of the three classifiers used in our method. We compared the performance of our method with Support Vector Machine (SVM) with χ^2 kernel, Random Forest with 100 trees and 2-Nearest neighbor based on Euclidean distance. Figure 7 has been produced on the SIGGRAPH database. For all the classifiers, we have computed the classification rate using different number of folds (k's) for the k-fold cross validation technique. The performance of our framework was evaluated using a conventional Random Forest with 10 folds and 100 trees as its achieved highest recognition rate. Results obtain with the Random Forest show how well our feature space was clustered. Table 2 shows the comparison of the achieved recognition rate of the proposed framework with the state-of-the-art methods using the same database. Table 2 illustrated that our framework exceeds state of the art methods in terms of expression recognition accuracy. Indeed, we are comparing our method with specific methods developed for one database containing often only on type of movement whereas our method is intended to be generic. We have evaluated our approach on a PC i7-4710MQ with 8GB of Ram. The neutral animation synthesis take 180ms for a sequence of 2 seconds with 120 frames, after running the optimization to find the parameters that satisfy the cost function. The optimization process takes 8 seconds for the same sequence of 2 seconds in C#. We believe that an optimized code can be real time.

Table 2: Comparison of our methods using all features mentioned in this paper to the state of the art methods. Notice that the specific method on the Biological database uses a dedicated algorithm to one kind of movements by computing an average motion and can not be generalized.

Database	Results from state-of-the-art	Our results
UCLIC	78% [31]	**83%**
Biological	50% (general) to 80% (specific) [6]	**57%**
SIGGRAPH	93% [9]	**98%**
MPI	–	**50%**

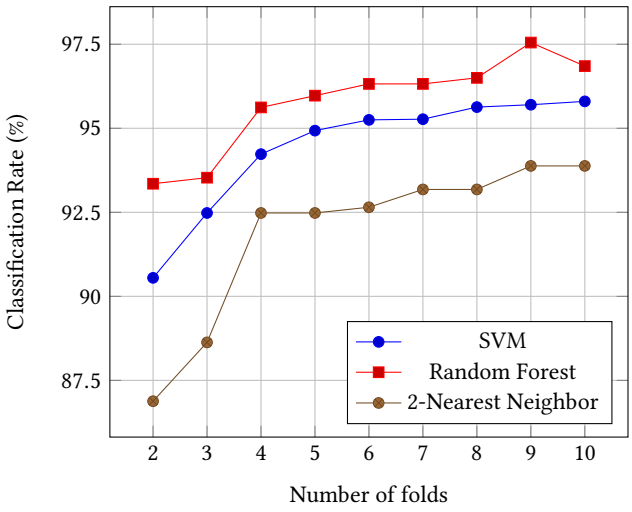

Figure 7: Evolution of the classification rate for the SIGGRAPH Database with the increasing number of folds for the k-fold cross validation method

Table 2 shows that the proposed framework is comparable to any other state-of-the-art method in terms of expression recognition accuracy. We obtain better recognition rate than the state of the art on the SIGGRAPH database and the UCLIC database. In the Biological Motion Database, movements are mainly knocking at a door ($\approx$ 1200 animations out of 1356 animations). The state of the art approach [6] uses this particularity to compute the average movement of knocking at a door and then subtracting this movement before running the recognition in order to emphasize the expression. Their recognition rate for this unbiased method is 81%. Nevertheless, this trick is possible for very specific movements, when you assume that all movements are similar. Also, they proposed a segmentation method to decompose a knocking action into motion primitives which are analyzed in terms of dynamic features. Since the purpose of our proposed framework is to be robust against heterogeneous movements, we cannot apply this assumption. We believe that to compare the approach [6] and ours, their biased recognition rate is 50% whereas our approach obtains recognition rate of 57%. Finally, to the best of our knowledge, no method in the literature on body expression recognition has tested the MPI database, our method obtains a recognition rate of 50%. This database is difficult because of the number of expressions 11 combined with a highly imbalanced database. By using a common re-sampling filter to deal with imbalanced database, our method obtains a recognition rate of 67%.

5 DISCUSSIONS AND CONCLUSION

We have presented a novel approach for automatic recognition of body expressions through 3D skeleton provided by motion capture data. We argue that body expressions can be robustly recognized by analyzing the difference between neutral and expressive animations in the frequency domain. One of the contribution of the proposed method is the synthesis of plausible neutral motion from an expressive motion, problem that was never tackle to our knowledge. Proposed method is able to generate a neutral motion from

an expressive animation even in complex cases: jump, run, kick, etc. From the synthesized neutral motion, proposed method classifies expressions of the input motion by computing the spectral difference in the Fourier domain between the neutral and expression motion. We have evaluated our approach on four databases that contain heterogeneous movements and expressions and obtained results that exceeds state of the art. Thus, our approach opens up many possibilities for human-computer interaction applications. One such application that can benefit from proposed approach is the generation of video games using the Kinect-like device. Computer games can benefit from real time body expression analysis to adapt its content at run time i.e. dynamic game play based on expressions exhibited by player's body .

In future we aim to continue working on method that can generate more realistic neutral animations by adding a technique to change the timing of the neutral animation. This can be achieved by analyzing the phase of the neutral animation produced by proposed method. Another direction that we are looking forward to explore is the usage of the Quaternion Fourier Transform in order to have a more accurate signal information since the joint channels are processed in a single unit. Finally, an improvement of the validation will be to test our approach on multi-simultaneous actions during a real live demonstration with a Kinect for instance.

ACKNOWLEDGEMENTS

This work was funded by the Region Auvergne-Rhône-Alpes http://www.auvergnerhonealpes.fr/

REFERENCES

[1] Kenji Amaya, Armin Bruderlin, and Tom Calvert. 1996. Emotion from motion. In *Graphics interface*, Vol. 96. Toronto, Canada, 222–229. http://www.graphicsinterface.org/wp-content/uploads/gi1996-26.pdf

[2] Andreas Aristidou, Yiorgos Chrysanthou, and Joan Lasenby. 2016. Extending FABRIK with Model Constraints. *Comput. Animat. Virtual Worlds* 27, 1 (Jan. 2016), 35–57. https://doi.org/10.1002/cav.1630

[3] Andreas Aristidou and Joan Lasenby. 2011. FABRIK: A fast, iterative solver for the Inverse Kinematics problem. *Graphical Models* 73, 5 (Sept. 2011), 243–260. https://doi.org/10.1016/j.gmod.2011.05.003

[4] Stephen W. Bailey and Bobby Bodenheimer. 2012. A Comparison of Motion Capture Data Recorded from a Vicon System and a Microsoft Kinect Sensor. In *Proceedings of the ACM Symposium on Applied Perception (SAP '12)*. ACM, New York, NY, USA, 121–121.

[5] Avi Barliya, Lars Omlor, Martin A. Giese, Alain Berthoz, and Tamar Flash. 2013. Expression of emotion in the kinematics of locomotion. *Experimental brain research* 225, 2 (2013), 159–176. http://link.springer.com/article/10.1007/s00221-012-3357-4

[6] Daniel Bernhardt and Peter Robinson. 2007. Detecting affect from non-stylised body motions. In *Affective Computing and Intelligent Interaction*. Springer, 59–70. http://link.springer.com/chapter/10.1007/978-3-540-74889-2_6

[7] Vinay Bettadapura. 2012. Face expression recognition and analysis: the state of the art. *arXiv preprint arXiv:1203.6722* (2012). http://arxiv.org/abs/1203.6722

[8] Armin Bruderlin and Lance Williams. 1995. Motion signal processing. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. ACM, 97–104. http://dl.acm.org/citation.cfm?id=218421

[9] Arthur Crenn, Rizwan Ahmed Khan, Alexandre Meyer, and Saida Bouakaz. 2016. Body expression recognition from animated 3D skeleton. IEEE, 1–7. https://doi.org/10.1109/IC3D.2016.7823448

[10] Simon Fothergill, Helena Mentis, Pushmeet Kohli, and Sebastian Nowozin. 2012. Instructing People for Training Gestural Interactive Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1737–1746.

[11] M. Melissa Gross, Elizabeth A. Crane, and Barbara L. Fredrickson. 2010. Methodology for Assessing Bodily Expression of Emotion. *Journal of Nonverbal Behavior* 34, 4 (Dec. 2010), 223–248. https://doi.org/10.1007/s10919-010-0094-x

[12] Eugene Hsu, Kari Pulli, and Jovan Popović. 2005. Style translation for human motion. *ACM Transactions on Graphics (TOG)* 24, 3 (2005), 1082–1089. http://dl.acm.org/citation.cfm?id=1073315

[13] Heechul Jung, Sihaeng Lee, Junho Yim, Sunjeong Park, and Junmo Kim. 2015. Joint Fine-Tuning in Deep Neural Networks for Facial Expression Recognition. In *The IEEE International Conference on Computer Vision (ICCV)*.

[14] Michelle Karg, Kolja Kühnlenz, and Martin Buss. 2010. Recognition of Affect Based on Gait Patterns. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 40, 4 (Aug. 2010), 1050–1061. https://doi.org/10.1109/TSMCB.2010.2044040

[15] R. A. Khan, A. Meyer, H. Konik, and S. Bouakaz. 2012. Human vision inspired framework for facial expressions recognition. In *2012 19th IEEE International Conference on Image Processing*. 2593–2596. https://doi.org/10.1109/ICIP.2012.6467429

[16] Rizwan Ahmed Khan, Alexandre Meyer, Hubert Konik, and Saida Bouakaz. 2013. Framework for reliable, real-time facial expression recognition for low resolution images. *Pattern Recognition Letters* 34, 10 (2013), 1159 – 1168.

[17] Andrea Kleinsmith and Nadia Bianchi-Berthouze. 2013. Affective body expression perception and recognition: A survey. *Affective Computing, IEEE Transactions on* 4, 1 (2013), 15–33. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6212434

[18] Andrea Kleinsmith, P. Ravindra De Silva, and Nadia Bianchi-Berthouze. 2006. Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers* 18, 6 (2006), 1371–1389. http://www.sciencedirect.com/science/article/pii/S0953543806000634

[19] Andrea Kleinsmith, Tsuyoshi Fushimi, and Nadia Bianchi-Berthouze. 2005. An incremental and interactive affective posture recognition system. In *International Workshop on Adapting the Interaction Style to Affective Factors*. 378–387. http://www0.cs.ucl.ac.uk/staff/n.berthouze/paper/KleinsmithFushimi.pdf

[20] Yingliang Ma, Helena M. Paterson, and Frank E. Pollick. 2006. A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior research methods* 38, 1 (2006), 134–141. http://link.springer.com/article/10.3758/BF03192758

[21] Albert Mehrabian and John T Friar. 1969. Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology* 33, 3 (1969), 330.

[22] Microsoft. 2017. Kinect. (2017). https://developer.microsoft.com/en-us/windows/kinect

[23] Lars Omlor and Martin A. Giese. 2007. Extraction of spatio-temporal primitives of emotional body expressions. *Neurocomputing* 70, 10 (2007), 1938–1942. http://www.sciencedirect.com/science/article/pii/S0925231206004309

[24] C.L. Roether, Lars Omlor, Andrea Christensen, and Martin A. Giese. 2009. Critical features for the perception of emotion from gait. *Journal of Vision* 9, 6 (06 2009), 1–32. https://doi.org/10.1167/9.6.15 reviewed.

[25] Simon Senecal, Louis Cuel, Andreas Aristidou, and Nadia Magnenat-Thalmann. 2016. Continuous body emotion recognition system during theater performances: Continuous body emotion recognition. *Computer Animation and Virtual Worlds* 27, 3-4 (May 2016), 311–320. https://doi.org/10.1002/cav.1714

[26] Jochen Tautges, Arno Zinke, Björn Krüger, Jan Baumann, Andreas Weber, Thomas Helten, Meinard Müller, Hans-Peter Seidel, and Bernd Eberhardt. 2011. Motion Reconstruction Using Sparse Accelerometer Data. *ACM Trans. Graph.* 30, 3, Article 18 (May 2011), 12 pages.

[27] Arthur Truong, Hugo Boujut, and Titus Zaharia. 2016. Laban descriptors for gesture recognition and emotional analysis. *The Visual Computer* 32, 1 (Jan. 2016), 83–98. https://doi.org/10.1007/s00371-014-1057-8

[28] Munetoshi Unuma, Ken Anjyo, and Ryozo Takeuchi. 1995. Fourier principles for emotion-based human figure animation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. ACM, 91–96. http://dl.acm.org/citation.cfm?id=218419

[29] Ekaterina Volkova, Stephan de la Rosa, Heinrich H. Bülthoff, and Betty Mohler. 2014. The MPI Emotional Body Expressions Database for Narrative Scenarios. *PLoS ONE* 9, 12 (Dec. 2014), e113647. https://doi.org/10.1371/journal.pone.0113647

[30] R. von Laban and L. Ullmann. 1971. *The mastery of movement*. Number vol. 1971,ptie. 1 in The Mastery of Movement. Macdonald & Evans. https://books.google.fr/books?id=-RYIAQAAMAAJ

[31] Weiyi Wang, Valentin Enescu, and Hichem Sahli. 2015. Adaptive Real-Time Emotion Recognition from Body Movements. *ACM Transactions on Interactive Intelligent Systems* 5, 4 (Dec. 2015), 1–21. https://doi.org/10.1145/2738221

[32] Shihong Xia, Congyi Wang, Jinxiang Chai, and Jessica Hodgins. 2015. Realtime style transfer for unlabeled heterogeneous human motion. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 119.

[33] M. Ersin Yumer and Niloy J. Mitra. 2016. Spectral style transfer for human motion between independent actions. *ACM Transactions on Graphics* 35, 4 (July 2016), 1–8. https://doi.org/10.1145/2897824.2925955