



**HAL**  
open science

# Human Motions and Emotions Recognition Inspired by LMA Qualities

Insaf Ajili, Malik Mallem, Jean-Yves Didier

► **To cite this version:**

Insaf Ajili, Malik Mallem, Jean-Yves Didier. Human Motions and Emotions Recognition Inspired by LMA Qualities. *The Visual Computer*, 2019, 35, pp.1411–1426. 10.1007/s00371-018-01619-w . hal-01971026

**HAL Id: hal-01971026**

**<https://hal.science/hal-01971026>**

Submitted on 6 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Human Motions and Emotions Recognition Inspired by LMA Qualities

Insaf Ajili · Malik Mallem · Jean-Yves Didier

the date of receipt and acceptance should be inserted later

**Abstract** The purpose of this paper is to describe human motions and emotions that appear on real video images with compact and informative representations. We aimed to recognize expressive motions and analyze the relationship between human body features and emotions. We propose a new descriptor vector for expressive human motions inspired from the Laban Movement Analysis method (LMA), a descriptive language with an underlying semantics that allows to qualify human motion in its different aspects. The proposed descriptor is fed into a machine learning framework including, Random Decision Forest, Multi-Layer Perceptron and two multiclass Support Vector Machines methods. We evaluated our descriptor first for motion recognition and second for emotion recognition from the analysis of expressive body movements. Preliminary experiments with three public datasets, MSRC-12, MSR Action 3D and UTkinect showed that our model performs better than many existing motion recognition methods. We also built a dataset composed of 10 control motions (move, turn left, turn right, stop, sit down, wave, dance, introduce yourself, increase velocity, decrease velocity). We tested our descriptor vector and

achieved high recognition performance. In the second experimental part, we evaluated our descriptor with a dataset composed of expressive gestures performed with four basic emotions selected from Russel's Circumplex model of affect (happy, angry, sad and calm). The same machine learning methods were used for human emotions recognition based on expressive motions. A 3D virtual avatar was introduced to reproduce human body motions, and three aspects were analyzed 1) how expressed emotions are classified by humans, 2) how motion descriptor is evaluated by humans, 3) what is the relationship between human emotions and motion features.

**Keywords** Motion recognition · Emotion recognition · Laban movement analysis · Features importance · Machine learning · Human perception

## 1 Introduction

Humans communicate with each other and with their environment through different modalities: facial expressions, speech, touch, gestures, etc. The body moves as we talk, think, play, work, etc. It plays an important role in human social interaction. Moreover, technology today integrates the human body for different applications in order to naturally interact with it. For example, the Kinect for Xbox 360 project enables users to control and naturally interact with a video game console using gestures, without the need to physically touch a game controller. The human body is an important medium for emotional analysis and recognition [1]. In some cases, body language is a way of expressing feelings even better than spoken language. For example, when we want to express our emotions to deaf peo-

---

I. Ajili  
IBISC, Univ Evry, Université Paris-Saclay, 91025, Evry,  
France  
Site Pelvoux, UFR-ST  
36, Rue du Pelvoux  
CE1455 Courcouronnes 91020 Evry Cédex  
Tel: 01 69 47 75 51  
Fax: 01.69.47.06.14  
E-mail: Insaf.Ajili@ufrst.univ-evry.fr

M. Mallem  
E-mail: Malik.Mallem@ufrst.univ-evry.fr

J.Y. Didier  
E-mail: Jean-Yves.Didier@ufrst.univ-evry.fr

ple, when we are in a long distance conversation or in a noisy environment, we are unable to convey our emotions through spoken language. Then, we interpret the emotional state from the sight of the body. Body language involves gestures, body posture and facial expressions. The bulk of research into emotion perception has usually focused on facial expressions as the primary modality through which emotion is conveyed [2,3]. However, during social interactions, emotional face perception can be influenced when integrating body movement. Some researchers [4,5] have demonstrated that body contexts can dramatically influence the recognition of emotions from facial expressions. Therefore, the role of body movement has become of great importance in the field of emotion, and many motion recognition systems have been used in several application fields, such as human-computer interaction, education, affective computing, etc. However, when it comes to whole body expressions, affective movements can be highly variable in different circumstances (energy, intensity, emotion, expression, etc.), even when the same person repeats the same movement multiple times. Such variations are not easy to measure and it is very difficult to find universal algorithms to completely describe the human motion. To cope with such challenges, body movement representation should be considered as the most important step for human motion recognition. It is than essential to find reliable and discriminative features to represent expressive human motions. In this paper, which extends our preliminary work [6,7], we propose a suitable representation of motions that captures both the quantitative and qualitative aspects of the movement. Two experimental parts are performed: in the first one we evaluate our descriptor vector for motion recognition using three public datasets and our own dataset of control motions. In the second part, we test our descriptor in recognizing human emotions based on expressive gestures. A second study of emotion perception and descriptor evaluation is carried out by humans in order to evaluate the performance of our recognition system.

The main contributions of this work are:

- A motion representation inspired by LMA qualities for motion and emotion recognition. We also evaluate our system against human perception.
- An accessible expressive motion database composed of 5 gestures performed with 4 emotions selected from the arousal-valence model [8] (happy, sad, angry and calm).

This paper is organized as follows: section 2 discusses previous works in this research area. Section 3 describes our proposed system. In section 4, two experimental

parts are presented where the performance of the proposed strategy is analyzed and discussed for the motion and emotion recognition processes, respectively. In the second experiment, we test our descriptor ability to characterize expressive motion with two approaches (learning method and statistical method). The paper ends with conclusions and perspectives for future work.

## 2 Related works

Human motion recognition is an active field in computer vision with a wide variety of potential applications such as video surveillance, video content analysis, human-computer interaction, robotics, etc. However, the high variability of human appearance and motion is problematic. It is therefore critical to extract robust representations of these variations. Several descriptors have been proposed in the literature based on the type of gesture. We will review two approaches. The first one considers a gesture as a simple movement without integrating the emotional aspect. The second one integrates the expressive dimension to analyze gestures and recognize emotions.

### 2.1 Body motion description

Gong et al. [9] represented human motion sequences by multivariate time series in the joint-trajectories space. In each frame, the joint-position of several key human body points was formulated as a point in the multi-dimensional space. Junejo et al. [10] proposed 2D representations by extracting silhouettes from actors. Each extracted silhouette was converted to a 1D time series. Jiang et al. [11] introduced a hierarchical model for action recognition, which consists in assigning each action to a group based on the motion states of each body part. For each group, joints motion and their relative positions were adopted as input to the KNN model for actions classification. Wang et al. [12] introduced a descriptor based on motion boundary histograms. They extracted dense trajectories obtained by tracking densely sampled points using optical flow fields. Xia et al. [13] extracted STIPs from depth videos (DSTIPs) and proposed the Depth Cuboid Similarity Feature (DCSF) to boost the performance of their system. Oreifej et al. [14] proposed a novel descriptor, a histogram of oriented 4D normals, in which they encoded the distribution of the surface normal orientation in the 4D space of depth, time and spatial coordinates.

## 2.2 Expressive motion description

Previous researchers mainly concentrated on the quantitative aspect of motions and have neglected the qualitative side of a movement which is important to recognize high-level gestures and to highlight the expressive qualities of the movements. For example, the difference between punching and reaching for an object is very slight in terms of body structure, both rely on the extension of the arm. But as the dynamics of the movement differs resulting in different intentions behind motions. Also, when performing the same motion with different moods, we need to characterize the quality of the movement to differentiate them. So another field appeared in the literature dedicated to expressive motion analysis aimed at different purposes.

*Motion retrieval and synthesis* There is a wide array of research devoted to motion retrieval and synthesis, such as Chi et al. [15] which developed a 3D character animation system named the EMOTE model. They employed Effort-Shape LMA components to produce expressive limb motions using empirical mappings between LMA components and motion parameters. This system enables the modification of a predefined motion in terms of its style in order to produce more expressive and natural simulated motions. Likewise, Kapadia et al. [16] encoded motions using keys which represent motion properties based on LMA factors for indexing motions. These keys were then combined to search for complex motions in large motion databases. Müller et al. [17] introduced a set of geometric features describing geometric relations between certain body points of a pose to identify logically similar motions. For the same purpose, Durupinar et al. [18] proposed a motion representation inspired by LMA qualities. They performed a mapping between the human motion parameters and different personality traits in an effort to synthesize motions with personality.

*Emotion stylization and style transfer* The integration of motion style editing or modification research has also been addressed in the computer vision literature. This approach consists in transferring style from one motion to another to obtain a newly synthesized motion. Hsu et al. [19] used iterative time warping to map from the input to the output sequence. They transformed a motion sequence into a new motion style while retaining its original content. Xia et al. [20] designed a data-driven animation system to add various styles to the existing animation. They constructed a series of local mixtures of autoregressive models to represent the complex relationships between styles and automatically transform an unlabeled heterogeneous mo-

tion data into different styles. Ymer and Mitra [21] proposed an approach for style transfer based on spectral analysis, which not only handles heterogeneous motion sequences but also transfers style between independent actions. Aristidou et al. [22] employed the LMA model to synthesize human motions from existing motion capture data. They extracted quantitative and qualitative characteristics of the movement based on LMA components. They applied the RBF regression model for mapping motion features to their emotion coordinates on Russell's Circumplex Model (RCM) of affect. The same authors [23] adopted the LMA model to synthesize motion by calculating both posture and style correlations. They first extracted the movement's features inspired from LMA components and found their stylistic correlations.

*Expressive motion analysis and recognition* Expressive motion representation requires high-level features describing both kinematics and semantic characteristics, and LMA answers these requirements. It is a formal and universal language developed by Rudolf Laban [24]. It allows to qualify human motion in its different aspects. This approach was used for many purposes, such as motion representation [29], segmentation [30], style transfer [22], etc. Many authors were inspired by the LMA method to design expressive motion descriptors, such as Glowinski et al. [29] who proposed a minimal representation of expressive movements. They focused only on the upper body. Visual tracking of trajectories of head and hands was carried out from a frontal and a lateral view. In their approach, the PCA indicated four components (roughly indicating motion activity, temporal and spatial excursion of movement, spatial extent and postural symmetry, motion discontinuity and jerkiness) which could characterize the affective behavior expressed by the human in 2D arousal-valence emotion space. Samadi et al. [31] quantified the Effort component and directional movement factor to analyze hand and arm movements. They designed a database composed of 6 motions in order to convey 6 basic emotions (anger, happiness, sadness, fear, disgust, surprise). After, their dataset was annotated by a certified movement analyst (CMA) to study the statistical correlation between the CMA-annotated and the quantified Laban components. Truong et al. [26] proposed a descriptor based on the LMA model to recognize the gesture and analyze the emotional content of orchestra conductors. For emotion analysis, they asked professional orchestra conductors to annotate each motion by choosing emotions among the categories proposed in their dataset. However, only a few authors have addressed the issue of automatic classification of emotions expressed in body motions using the machine learning

method [25,32]. Aristidou et al. [25] presented an LMA-based framework to extract motion features, aiming to classify acted dance movements with regard to the expressed emotion. Professional dancers were invited to perform an emotional state, together with music of their choice. In total, they obtained 72 dances, with approximately 130 minutes of motion. Their framework has also been used by Senecal et al. [32] who created constituent social agents that recognize the emotion of users by mapping motion onto the Russell Circumplex model diagram using neural networks.

In [25] professional dancers were invited to express their emotions with dance movements. Their gestures are explicitly chosen for special emotions, for example happy emotion will be expressed with open arms. So, their algorithm is helped by the type of gesture. In our case, we want to optimize the algorithm so that it is more effective in the case of the same gesture and able to recognize the emotion without being influenced by the type of gesture. So, we constructed a generic descriptor vector for characterizing expressive motions, able to distinguish different emotions expressed through the same motion. We created our database with emotions expressed by non-professional people for our final application referred to Human-Robot interaction. Also we proposed a statistical method based on human in order to evaluate the reliability and the adequacy of our proposed system. Only very few works have addressed the topic of human emotion recognition while performing the same motion, such as Cimen et al. [33]. The authors focused on a single movement task which was "walking". They used the SVM method for classification. In our work, we used a larger dataset composed of 5 gestures acting with 4 emotions, which requires a more robust descriptor able to characterize the movement and its qualities. We constructed a system invariant to initial positions and orientations of people. We evaluated our approach with 4 learning methods. Moreover we studied the importance of our motion descriptor to estimate the pertinence of each motion feature in characterizing each emotion and we compared the results obtained from the automatic learning method against human perception.

### 3 The proposed approach

In this section, the proposed human motion recognition approach is detailed. It includes three important steps: data preprocessing, motion representation and classification.

#### 3.1 View invariant system

We use the Kinect sensor for data acquisition to extract 3D skeleton joints in real time. The first step after data acquisition is the normalization of all skeletons which consists in aligning all the skeletons in the initial frame. So if two users perform the same gesture from different positions, our system should give the same result. This step makes our system independent of the initial position and orientation of the user. Given an action sequence  $S = \{J_{t,j}, t \in 1, \dots, T, j \in 1, \dots, N\}$  with  $T$  frames and  $N$  joints,  $J_{t,j}$  corresponds to the coordinates of the joint  $j$  captured at frame  $t$ .

We define a local coordinate system for the skeleton placed at the hip center joint ( $J_c$ ), represented by three axes ( $X', Y', Z'$ ) and equipped with three unit vectors respectively, the left hip joint vector  $\mathbf{n}_{lhi}$ , the spine vector  $\mathbf{n}_s$  and their cross product  $\mathbf{n}_c = \mathbf{n}_{lhi} \wedge \mathbf{n}_s$ . For each sequence, we first apply a translation to move the skeleton to the center of Kinect, and after a rotation to align both coordinate systems (Figure 1). So the transformed joint yields:

$$[J_{t,j}]_{(X',Y',Z')} = R_{(X,Y,Z) \leftarrow (X',Y',Z')}^{-1} ([J_{t,j}]_{(X,Y,Z)} - J_{1,c}) \quad (1)$$

$$R_{(X,Y,Z) \leftarrow (X',Y',Z')} = \begin{bmatrix} \frac{n_{lhi}}{\|n_{lhi}\|} & \frac{n_s}{\|n_s\|} & \frac{n_c}{\|n_c\|} \end{bmatrix} \quad (2)$$

Where:

$$J_{1,c} = ([J_{1,lhi}]_{(X,Y,Z)} + [J_{1,rhi}]_{(X,Y,Z)})/2 \quad (3)$$

$$n_{lhi} = [J_{1,lhi}]_{(X,Y,Z)} - J_{1,c} \quad (4)$$

$$n_s = [J_{1,s}]_{(X,Y,Z)} - J_{1,c} \quad (5)$$

$[J_{1,lhi}]$  and  $[J_{1,rhi}]$  refer to the positions of left and right hip joints, respectively, captured at initial time ( $t=1$ ). Since the rotation vectors are orthogonal unit vectors, then  $R^{-1} = R^T$ .

$$[J_j]_{(X',Y',Z')} = R_{(X,Y,Z) \leftarrow (X',Y',Z')}^T [J_j]_{(X,Y,Z)} \quad (6)$$

Once we apply transformations to all sequences, our representation becomes invariant to the position and orientation of the subject in the scene, we convert the skeleton joint data to a descriptor vector based on the LMA method for human motion representation.

#### 3.2 Motion representation

Human body motion representation is considered as one of the most important steps in the motion recognition process since each movement must be described with specific features. For example in walking, the motion of

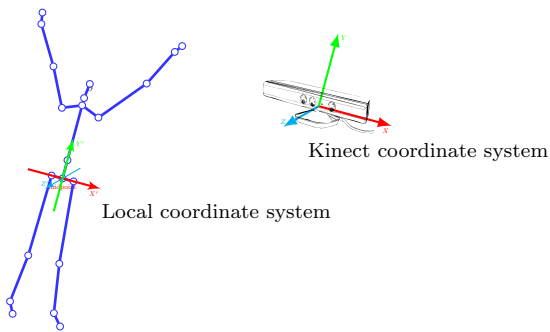


Fig. 1 Kinect and local coordinate systems.

the hip joint is considered an important feature for motion description, which is not the case in sitting, where the bending feature is the most expressive features. Our descriptor vector is derived from the LMA model, used as a tool by dancers, athletes, physical and occupational therapists. This approach can describe and interpret all varieties of human movements. LMA provides a rich description of movement, focusing on the four major components of movement: body, space, shape and effort.

The **Body** describes structural and physical characteristics of the body in movement. It indicates how the whole body is connected and organized and which body parts are influenced by others. In the upper body part, we compute the following angles in left and right parts respectively (Figure 2): between hands and shoulders ( $\theta_1^l, \theta_1^r$ ), between elbows and hips ( $\theta_2^l, \theta_2^r$ ), between elbows and shoulders in the symmetrical part ( $\theta_3^l, \theta_3^r$ ). And we describe outstretched arms by computing the distance between two hands ( $d_{Hs}$ ). To know more about the hands pathway, we add three other features, the distances between the head and left ( $d_{h,lh}$ ) and right hand ( $d_{h,rh}$ ), and the angle between two hands with respect to the shoulder center joint ( $\theta_{Hs}$ ). In the lower part, we describe leg extension by computing the angle between foot and hip joints, respectively in left and right parts ( $\theta_4^l, \theta_4^r$ ). Also, we describe leg spread by computing the opening angle of the knees ( $\theta_{Ls}$ ). To encode body quality, three measures of variability are used for each feature: the mean, the standard deviation and the range.

The **Space** component describes the movement in the space. It defines directions and paths of a movement. We compute the length of the trajectory followed by upper extremities including the head and hands. It is interesting to know the length ( $L$ ) of joint trajectories which adds more information about motion type.

$$L = \sum_{t=1}^{T-1} \|J_{t+1} - J_t\| \quad (7)$$

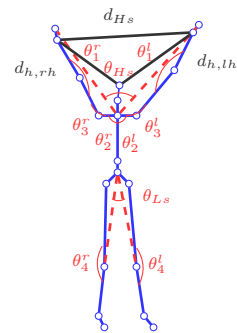


Fig. 2 Body features.

The **Shape** component describes the way the body changes shape during movement. It is characterized by the body's interaction with itself as well as the space around it. The shape component is composed of three subcomponents: shape flow, directional movement and shaping.

- *Shape Flow* refers to how a body can interact and change shape in a self-to-self relationship. We construct the smallest convex envelope of the skeleton based on the Quickhull algorithm [34] and compute its volume to characterize the deformation of the skeleton's shape over time. The convex hull offer a more precise fit than a simple bounding box. Figure 3 illustrates the 3D skeleton convex hull and its shape deformation in the "start music" gesture from the MSRC-12 dataset [35]. In the first frame, the person is in her initial position. After she raises her hands, the volume of the body envelope convex increases. So, by computing the volume of the convex hull we have more idea about body extension. Mean, standard deviation and the range of the 3D convex hull volume feature are computed to quantify the shape flow factor.

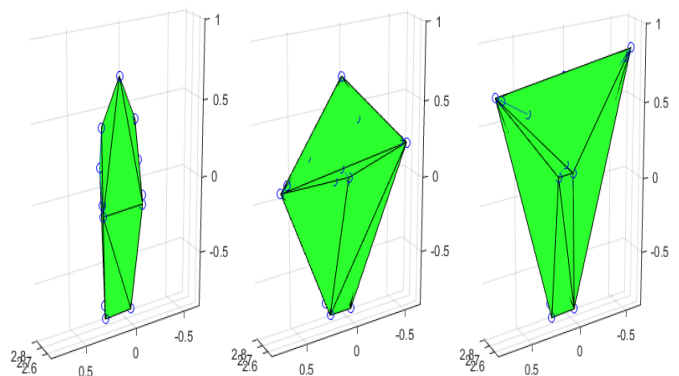


Fig. 3 Variation of convex hull feature in "start music" gesture from the MSRC-12 dataset [35].

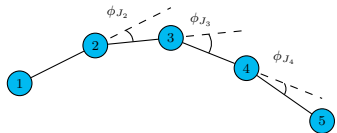
- *Directional movement* defines the pathway of the movement through space. We describe the pathway of the upper extremities, hands and head, by computing their local curvatures. We usually focus on the upper body part, especially hands and head, since they are the most flexible human body parts. We consider the gradual angular change  $\phi$  occurring between two successive frames as shown in Figure 4, defined as follows:

$$\phi_{J_t} = \arccos\left(\frac{\overrightarrow{J_{t-1}J_t}}{\|\overrightarrow{J_{t-1}J_t}\|} \cdot \frac{\overrightarrow{J_tJ_{t+1}}}{\|\overrightarrow{J_tJ_{t+1}}\|}\right) \quad (8)$$

where  $J_t$  is the position of joints (hands and head) at frame  $t$  and  $J_{t+1}$  at the next frame. This equation describes the local curvature of the upper extremities pathway. From this angle, we derive the curvature feature ( $C$ ) defined as:

$$C = \sum_{t=2}^{T-1} \phi_{J_t} \quad (9)$$

where  $T$  is the number of frames. In straight-line trajectories the Curvature index ( $C$ ) is close to 0 and describes a linear movement. In curved paths it changes to a very high value.



**Fig. 4** Gradual angles between successive frames.

- *Shaping* describes the qualitative changes in the shape according to three planes: the **Horizontal** plane describes spreading and enclosing movements; spreading is, for example, a movement of both hands starting in front of the body and opening to the sides; enclosing is the opposite motion, as used in a hug. The **Frontal** plane characterizes rising and sinking movements and the **Sagittal** plane includes forward and backward movements. We describe body extension according to the three planes by computing average distances of all skeleton joints ( $N$  joints) with respect to the spine joint ( $J_{1,s}$ ) at initial frame (Figure 5).

- For Spreading/Enclosing movements we compute the average spread/enclose movement relating to spine joint position at initial frame.

$$D_H = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N \sqrt{([J_{t,i}]_X - [J_{1,s}]_X)^2} \quad (10)$$

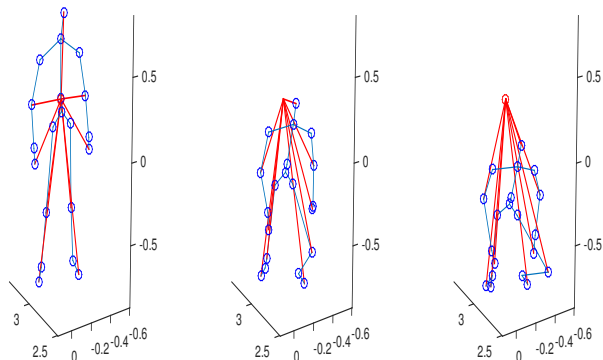
- For Rising/Sinking movements we compute the average rise/sink movement relating to spine joint position at initial frame.

$$D_F = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N \sqrt{([J_{t,i}]_Y - [J_{1,s}]_Y)^2} \quad (11)$$

- For Advancing/Retreating movements we compute the average forward/backward movement relating to spine joint position at initial frame.

$$D_S = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N \sqrt{([J_{t,i}]_Z - [J_{1,s}]_Z)^2} \quad (12)$$

$T$  is the number of frames and  $[J_{t,i}]_{X,Y,Z}$  is the 3D position of joint  $i$  at frame  $t$ .



**Fig. 5** Distance variation relating to the spine joint, in the "crouch" gesture from the MSRC-12 dataset [35].

The **Effort** component encodes the change in the intensity of exertion during motion. It is a form of expression, mostly related to the change of mood or emotions. It describes the qualitative aspects of movement by identifying the type of energy used in the movement. For example, pushing a heavy object and closing the door are very slight in terms of body organization. Both require the extension of the arm but the intention of the movement is different. The effort component is divided into four factors, each factor ranges between two opposite qualities: Weight (Light/Strong), Time (Sustained/Abrupt), Space (Indirect/Direct) and Flow (Free/Bound). It is the most important element in LMA to describe the quality and the dynamics of the movement. This component is essential, especially for the emotion recognition process. We will focus on four joints (head, spine, left and right hands) since, in our emotions corpus, all expressive gestures were performed by the upper body part.

- The Time factor differentiates between abrupt and sustained motion. Abrupt movement involves less

time consumption, so it is a high-speed and Sustained movement is a continuous with constant velocity. It can be represented as the change of acceleration of movement. So, to quantify this factor we compute the mean, the standard deviation and the range of velocity of upper body joints (head, spine and hands).

- The Weight factor is about the level of force or pressure used throughout a motion with the contrasting elements: Strong like a punching action or Light like dabbing paint on a canvas action. Light movement is characterized by an invariant rhythm of motion. We can describe it with zero acceleration. Strong movement requires force and acceleration. We compute the accelerations of head, spine and hands to quantify the weight factor. Like the Time factor, mean, standard deviation and the range of the acceleration are computed to present the Weight Effort factor.
- The Flow factor characterizes the attitude towards bodily tension and control, with the contrasting elements Free and Bound. We compute the yaw and pitch range of head and hands motions. For Free motion, we will obtain a higher range compared to a Bound motion.
- The Space factor characterizes the quality of attention that the person make in a movement and distinguishes a direct movement from an indirect movement. A direct motion is characterized by a single destination where all your attention is focused on a single spatial possibility. For example, when pointing to a particular object we have a Direct gesture. Indirect motion represents a widening of focus like the waving away bugs action. We describe the directness of the upper body parts movements by computing the Straightness index ( $S$ ) of head and hands motions as the ratio between the distance between the positions of the articulation at the first and last frame ( $D$ ) and the sum of the distances between successive frames ( $L$ ) for head, spine and hands joints.

$$S = \frac{D}{L} \quad (13)$$

A Straightness Index close to 1 means there is a Direct motion, and is close to 0 in Indirect motion.

After encoding all the LMA components, we obtain a descriptor vector composed of 85 features.

## 4 Experiments

This section describes our experiment which is divided in two parts. The first one is about the motion

recognition where we evaluated our model with public databases (MSRC-12 [35], Utkinect [36] and MSR Action 3D [37]) and our dataset named CMKinect-10. In the second part, we tested our model for emotion recognition with our dataset composed of expressive motions. We analyzed the importance of our descriptor vector towards each expressed emotion. Both processes have been carried out with learning and statistical methods.

### 4.1 Motion recognition

#### 4.1.1 CMKinect-10

CMKinect-10 is a dataset composed of 10 control motions as shown in Figure 6. Twenty subjects (10 men and 10 women) from the University of Evry Val d'Essonne, ranged in age from 27 to 36 years old ( $M=29.85$  years,  $SD=2.47$ ), took part in this study. Each subject is asked to make a gesture ten times. This dataset has in total 2000 sequences (20 subjects  $\times$  10 motions  $\times$  10 times). For data acquisition, the OpenNI driver under ROS (Robot Operating System) provided a high-level skeleton tracking module. This module requires initial calibration to record the 3D position of skeleton joints at  $640 \times 480$  resolution at 30 fps.

#### 4.1.2 Learning method parameter adjustment

For learning method parameter adjustment, we first used the MSRC-12 dataset [35] and we computed the F-score with a cross-validation test.

**Random Decision Forest** is an algorithm for classification that consists of many decision trees. During tree induction, nodes are subdivided progressively into child nodes and the tree is then grown until the node becomes pure or when a maximal depth is reached. There are various functions to measure the quality of a split, such as the information gain [38] or the Gini impurity [39]. We varied the number of trees from 10 to 200. The best result of 94% was obtained with 100 trees. The other parameters were set in their default values: for the maximum depth the nodes are expanded until all leaves are pure or until all leaves contain less than 2. The number of features to consider when looking for the best split is equal to  $\sqrt{85}$ . The minimum number of samples required to split an internal node is equal to 2, and the minimum number of samples required to be at a leaf node is equal to 1 as recommended in [40].

**Multiclass SVM** was originally designed for binary classification. The often suggested implementations for SVM multiclass classification are One Against



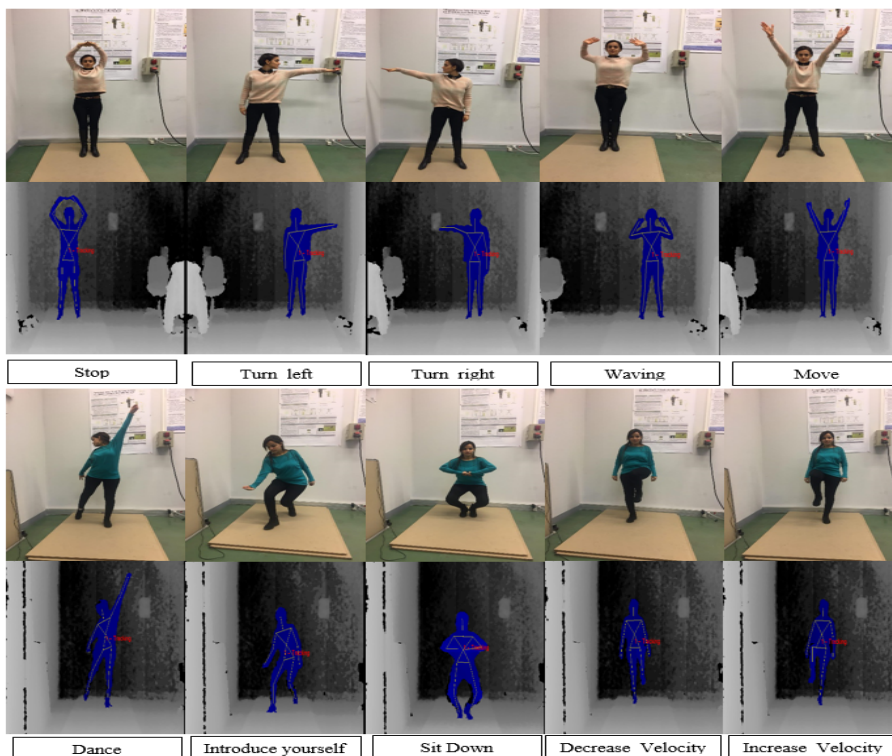


Fig. 6 CMKinect-10 dataset.

One (OAO) and One Against All (OAA). For an  $n$ -class classification problem, OAO constructs  $n(n-1)/2$  classifiers where each one is trained on data from two classes, while OAA involves  $n$  binary SVM classifiers, one for each class. Multi-class SVM provides various parameters such as kernel type including linear, Gaussian radial basis function, sigmoid and polynomial with kernel parameters and regularization parameter  $C$ . We varied the value of  $C$  starting from 0.01 to 10 in order to control the trade-off between achieving a low error on the training data and margin maximization. For RBF and sigmoid kernels, we varied  $\gamma$  from 1 to 10. For the polynomial kernel, experiments have been conducted with  $d$  values ranging from 1 to 5. The best recognition rate of 92% has been obtained with the polynomial kernel function for  $d = 2$ ,  $\gamma = 4$  and  $C = 1$ .

**Multi-Layer Perceptron** is a feed-forward neural network, consisting of a number of neurons organized in several layers (one input and one output layer with one or more hidden layers). The input layer receives an external activation vector and passes it through weighted connections to the neurons in the first hidden layer. These compute their activations and pass them to neurons in the following layer. To find the optimal number of hidden neurons, we varied the number starting from 10 to 100 neurons. For the activation function, we compared between three functions (logistic sigmoid, hyper-

bolic tangent and Rectified linear unit function). The MLP method utilizes a supervised learning technique called backpropagation. Two Backpropagation parameters were adjusted, the learning rate ( $\epsilon$ ) which controls the step-size in updating the weights, it should be between 0 and 1 and the momentum ( $m$ ) to accelerate the convergence network while avoiding instability. Best results were achieved with a single layer of 60 hidden neurons and a logistic sigmoid activation function. For the backpropagation algorithm, optimal parameters were obtained by setting  $m = 0.9$  and  $\epsilon = 0.001$  which yield a recognition rate of 93%.

#### 4.1.3 Experimental results of motion recognition

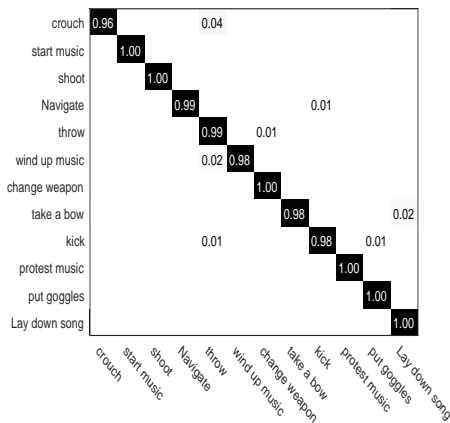
##### MSRC-12 dataset

The performance of the system is measured in terms of precision and recall with the F-score [41]. We performed 5-fold cross-validation: the dataset is randomly divided into 5 equal groups. For each of the 5 experiments, 4 folds are used for training and the remaining one for testing. The cross-validation process is then repeated 5 times, with each of the 5 folds used exactly once as the test set. The performance measure is then obtained by the average of the values computed in the loop. Table 1 illustrates recognition rates obtained in MSRC-12 for each learning method. Best results are

**Table 1** Recognition rates (%) of our method compared to the state of the art methods on the MSRC-12 dataset.

Methods		Iconic	Metaphoric
Lehrmann et al. [45]		90.90	-
Song et al. [46]		79.77	57.21
Truong et al. [47]		88.6	75.2
Our method	RDF	99	93
	OAO	98	90
	OAA	98	91
	MLP	98	89

achieved with the RDF method with recognition rates of 93% in the metaphoric category and 99% in iconic gestures. Figure 7 illustrates the confusion matrix for MSRC-12 dataset when using the RDF method. The diagonal elements represent the number of elements where the predicted values are equal to the expected values, while off-diagonal elements are those that are mislabeled by the classifier. The higher the proportion of values on the diagonal of the matrix in relation to values off of the diagonal, the better the classifier. For example, the "wind up the music" gesture was correctly recognized on 98% of the data, while it was incorrectly recognized as "throw an object" gesture by 2%. From the confusion matrix, we can conclude that our proposed descriptor is very good for discriminating the different gestures. We also compared our result to state of

**Fig. 7** Confusion matrix for MSRC-12 dataset.

the art methods in Table 1. As we can see, our method outperforms their methods, which confirms the robustness of our descriptor in characterizing both iconic and metaphoric gestures.

### MSR Action 3D dataset

To compare the performance of our descriptor with state of the art methods on the MSR Action 3D dataset, we followed the cross-subject test setting where half of

the subjects are considered as training data and the rest as testing data. As shown in Table 2, we obtained recognition rates close to the state of the art methods, 89.5%, 89% and 88.1%, with OAO, OAA and MLP, respectively. The dataset is divided in three subsets AS1, AS2 and AS3. The subsets AS1 and AS2 include actions with similar movements, while AS3 groups complex actions. As a learning method, we used the one that gave the best recognition results, the RDF method. Overall accuracies of 90.3%, 88.7% and 93.1% were achieved for AS1, AS2 and AS3, respectively. Figure 8 shows

**Table 2** Recognition rates (%) of our method compared to the state of the art methods on the MSR Action 3D dataset.

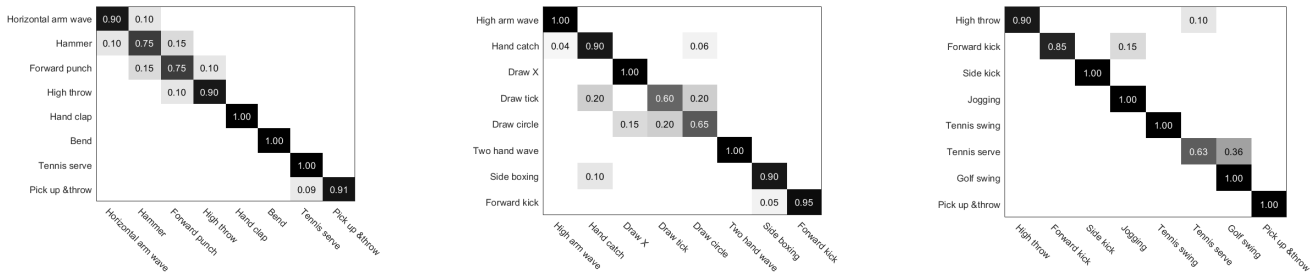
Methods	Recognition rates (%)	
Xia et al. [13]	89.30	
Slama et al. [43]	86.21	
Oreifej et al. [14]	88.89	
Our method	RDF	90.5
	OAO	89.5
	OAA	89
	MLP	88.1

the confusion matrices for AS1, AS2 and AS3, respectively, when using the RDF method. The AS1 subset contains high similar actions, which explains some misclassifications that we found in this subset, for example, the confusion between the following actions: Horizontal arm wave and Hammer, Forward punch and High throw. Also, actions in AS2 are likely to be confused, for example, "draw tick" and "draw circle", since they are quite similar. However, we get perfect recognition results for actions in AS3.

### UTkinect dataset

To compare our results with the state of the art approaches on the UTKinect dataset, we followed the experiment protocol proposed by Xia et al. [36] where they applied Leave-One-Out Cross Validation (LOOCV) [42] and achieved an overall mean accuracy of 90.92% and the best accuracy of 95%. As shown in Table 3, our method outperforms their results by a factor of about 1% when using the RDF classifier and the same when applying MLP and OAA methods, but just 1.8% lower with the OAO method. The resulting confusion matrix with the RDF method is shown in Figure 9. Some confusions occurred between some classes, like the confusion between "throw" and "push" classes. On the other hand, we achieved recognition rates of 100% in the following action classes: stand up, pull, wave hands and clap hands.

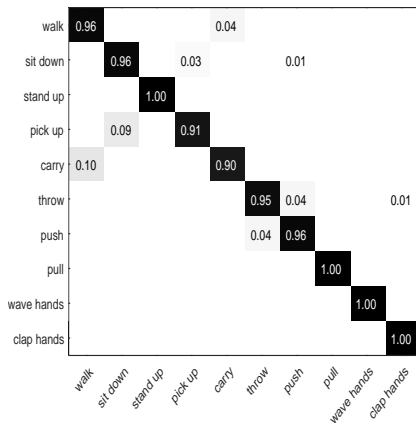
### CMKinect-10 dataset



**Fig. 8** Confusion matrices for MSR Action 3D dataset.

**Table 3** Recognition rates (%) of our method compared to the state of the art methods on the UTkinect dataset.

Methods	Recognition rates (%)	
Xia et al. [36]	95	
Our method	RDF	96
	OAO	93.2
	OAA	95.3
	MLP	95



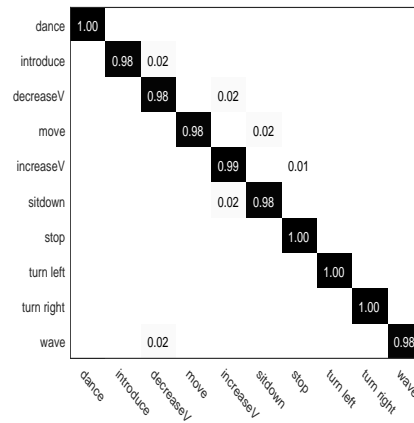
**Fig. 9** Confusion matrix of our proposed approach for UTkinect dataset.

We evaluated our descriptor with our dataset composed of 10 control motions. Using the four learning methods for motions classification. We used the 5-fold cross-validation technique to assess the predictive performance of our model. Table 4 summarizes recognition results obtained for each learning method. As we can see, we have almost the same result when applying the four learning methods, the mean of the F-scores was close to 1. The confusion matrix of our control motion dataset is illustrated in Figure 10. We obtained perfect recognition results for the following motions: dance, stop, turn left and turn right.

In conclusion, we evaluated our descriptor against different public datasets and we can confirm the robustness of our descriptor in human motion recogni-

**Table 4** CMKinect-10 dataset: The mean and standard deviation of F-scores results for each learning methods (RDF, OAO, OAA and MLP).

Methods	RDF	OAO	OAA	MLP
<b>Mean</b>	0.98	0.98	0.98	0.99
<b>F-scores</b>	(+/-0.04)	(+/- 0.03)	(+/-0.01)	(+/- 0.01)



**Fig. 10** Confusion matrix for CMKinect-10 dataset.

tion. Now, we will evaluate how well it recognizes human emotions expressed through body motions.

## 4.2 Emotion recognition and analysis

### 4.2.1 Based on learning method

The purpose of this study is to build an emotional dataset composed of 4 basic emotions (happy, angry, sad and calm) performed with expressive motions. Such a dataset can be used in a robotic application, for example for robot teleoperation in an expressive way which makes the interaction between human and robot more natural. In such a case, the robot will be able to recognize not only the actions of the human but also his mood. In the experimental part, we will first evaluate the pertinence of our descriptor in characterizing human emotions based on learning methods and viewer

**Table 5** Emotions dataset: The mean and standard deviation of F-scores results for each learning methods (RDF, OAO, OAA and MLP).

Methods	RDF	OAO	OAA	MLP
Mean	0.89	0.78	0.71	0.71
F-scores	(+/-0.05)	(+/- 0.06)	(+/-0.13)	(+/- 0.12)

ratings. We will also study the relationship between body features and expressed emotions in order to conclude the importance of each Laban feature to characterize each emotion.

**Emotion classification:** Our emotional dataset is composed of 5 control motions selected from the CMKinect-10 dataset (move, turn left, turn right, wave and stop) described in section 4.1.3. Each motion is acted with 4 different emotions (happy, angry, sad and calm). The same subjects who participated in the previous dataset (CMKinect-10) took part in this study. We proposed scenarios presenting emotional situations. Each participant was asked to read the proposed scenario and take the time to feel the situation. After each participant was asked to stay in front of the kinect and perform each motion while acting the situation. Two examples of scenarios proposed respectively for sad and happy emotions were:

- Someone calls you and informs you that your best friend is dead, so you feel very sad.
- You are selected in a game and you will win a fortune, so you feel very happy.

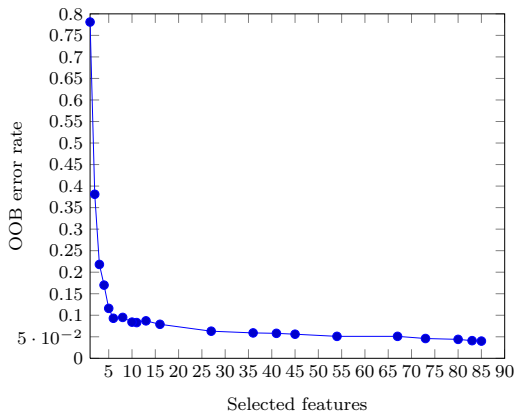
During the recording sessions, the order of scenarios, emotions and motions were randomized from one participant to another. We recorded 400 videos (20 participants  $\times$  5 motions  $\times$  4 emotions). As classification methods, we used the same four methods applied for motion recognition. RDF trained and tested with 5 fold cross-validation achieves the best result compared to OAO, OAA and MLP with recognition rates of 89%, 78%, 71% and 71%, respectively (Table 5).

**Features importance:** As shown in the section above, our motion descriptor has successfully classified the movements in the different public datasets and the emotions in our expressive motions dataset when applying the different learning methods. Now, we want to study the importance of chosen features and see if there are redundant features that contribute less to the performance of our descriptor. In the expressive motions classification step, some authors have also discussed this point and tried to characterize expressive gestures through body movement parameters. The most relevant is the one of Aristidou et al. [25] which is also based on the LMA model to quantify motion features. Compared to them, we work on short gestures (4 seconds of du-

ration) instead of complete dances during between one to two minutes. This is a strong requirement of our application which is in the context of human-robot interaction. As the movement sequence is significantly shorter, it means the amount of data is reduced. To cope with this constraint, we decided to rely on all components of LMA. Whereas it is not the case for Aristidou et al. [25]. That is why we made a faithful association with the LMA model. We quantified all LMA factors. For example, in the Shape component, following the definition of LMA, we quantified the three Shape factors. Unlike Aristidou et al. [25], we considered the directional movement factor as important since it allows us to describe the attention that the person carries in a movement. Also, for the Shape flow factor, we constructed the skeleton convex hull that offers a more precise fit than a simple bounding box and we computed its volum to characterize the body shape deformation during movement. For the Space Effort factor, the authors considered a movement as direct when the character is moving in the same direction as the head orientation. They computed the angle between the head’s orientation and the trajectory of the root joint. However, based on our statistical experiments, if we make an anger pointing gesture with the hand while keeping the head not oriented to the pointing direction means a direct movement. To avoid a such confusion between two extreme qualities of the space factor (direct and indirect), we described the directness of the upper body parts (head, spine, and hands) by computing the Straightness index ( $S$ ) of each articulations. For the Space component, we quantified the trajectory of the root joint as Aristidou et al. [25] but it is also important to know about the trajectories of hands in the space to characterize emotions from body movement. For example, in the pointing gesture a long trajectory of the hand indicates sadness whereas a short trajectory can indicate anger emotion.

In the next part, we will evaluate the importance of our features. Aristidou et al. [25] studied the importance of each feature in their descriptor. In our case, we prefer to study the importance of the set of features to find which combinations of features give the best classification result. We used the RDF method to measure the importance of each feature. During the training phase, each tree is grown using a different bootstrap samples from the original training data, leaving 1/3 as OOB (Out Of Bag) to estimate the prediction error. The importance of the feature  $f_i$  is measured as the difference between the OOB prediction error of each tree before and after permuting  $f_i$ . We started with the whole set of features and we computed the OOB error rate. After, we sorted the features in descending order

of importance, and removed the features with low importance. We recomputed the OOB error rate of the new set of features. We repeated the process until there remained only one feature in the set. Finally, we sorted all OOB error rate results obtained in each subset of features and we selected the one that corresponds to the smallest OOB error rate. As we can see in Figure 11, with the whole set of features (85 features), we obtained a minimal OOB error rate of 0.04, which confirms the importance of the combination of all features to characterize our expressive motion dataset.



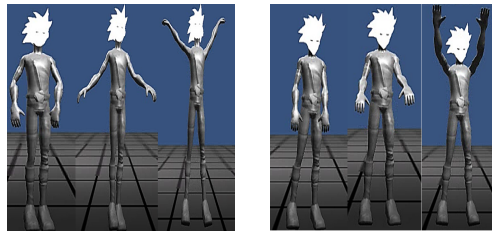
**Fig. 11** Optimal features set giving the minimal value of the OOB error.

#### 4.2.2 Human perception

**Emotions classification:** We tried to make an external evaluation of the emotions expressed in the videos. We recruited 20 viewers (10 men and 10 women) from the University of Evry Val d'Essonne, ranged in age from 28 to 37 years old ( $M=30.9$  years,  $SD=3.16$ ). They watched recorded videos and rated the degree to which they perceived the expressed emotion using a 5-item Likert scale (1=strongly disagree, 3=neutral, 5=strongly agree). For a reliable evaluation, expressive motions in videos were reproduced by a virtual avatar as shown in Figure 12 which represents the moving gesture acted with happy and sadness emotions. This helps viewers to make scores without being influenced by some factors like facial expressions, gender, etc. Table 6 illustrated the mean of perception ratings given by all viewers for each expressed emotion. Each cell represented the score ( $s$ ) computed with the following formula:

$$s = \frac{1}{N} \frac{1}{V} \sum_{i=1}^N \sum_{v=1}^V s_{iv}$$

$N$  is the number of expressed emotions,  $V$  is the number of viewers, and  $s_{iv}$  refers to the perception rating given by viewer  $v$  for the expressed emotion  $i$ , using the 5-item Likert scale. To trust the evaluation of viewers we check whether viewer ratings were due to chance and if there was inter-viewer consistency. So, we computed the Cronbach's alpha, a measure used to assess the reliability of a set of scale, it is expressed as a number between 0 and 1. According to [44], the acceptable range is between 0.70 to 0.95 [44]. In our case, the mean of Cronbach's alpha was greater than 0.8 for all emotions (happy 0.895, angry 0.916, sad 0.895, calm 0.873) which indicates the high consistency between the viewers in rating emotions.



**Fig. 12** Move gesture reproduced by a virtual avatar, expressed with happy and sad emotions.

**Table 6** Confusion matrix of emotion recognition based on viewer ratings.

Expressed emotions	Perceived emotions			
	Happy	Angry	Sad	Calm
Happy	<b>3.08</b>	2.16	1.77	2.04
Angry	2.29	<b>3.27</b>	1.80	1.81
Sad	1.67	1.81	<b>3.41</b>	2.15
Calm	2.14	1.95	2.28	<b>3.17</b>

**Features evaluation:** Once we have finished emotion evaluation we proceeded to feature rating in each observed video in order to study the correlation between emotions and body features. We asked the same 20 viewers to watch the same videos and rate each selected feature using a 7-point Likert scale. Cronbach's alpha has been used to determine the internal consistency between viewers when evaluating features. We obtained values higher than 0.7 (Time 0.958, Weight 0.874, Space 0.823, Flow 0.885, Shaping 0.914, Shape Flow 0.843). So, we can say that we had a very good reliability between viewers in feature rating, except in the Directional Movement quality where we obtained a value of 0.570 inferior to 0.7 which indicates an unsatisfactory internal consistency reliability between viewers. We can explain this by the type of motions selected which makes the distinction between direct and indi-

**Table 7** Relationship between features and left/right extremities of Likert scale.

Components	Factors	1=Left	7=Right
Shape	Shape Flow	shrinking	Growing
	Directional Movement	Curvilinear	Rectilinear
	Shaping	Contracted	Extended
Effort	Time	Sustained	Abrupt
	Weight	Light	Strong
	Space	Indirect	Direct
	Flow	Free	Bound

rect movement sometimes difficult. Some gestures of our dataset are limited (stop gesture) compared to free motions like dance movements, hence the challenge of recognizing emotions through restricted gestures. We used the Pearson correlation to evaluate whether body features are associated with expressed emotions. It ranges from -1 for a perfect negative correlation to +1 for a perfect positive correlation, 0 representing no correlation. We focused on the two factors of LMA, Shape and Effort, which describe the qualitative side of the movement. Table 8 summarizes the relationship between the four emotions and Effort-Shape features: The **Happy** emotion was significantly characterized by Shape factors, there was a moderate positive correlation between Happy emotion and Shape factors. It is then characterized by the growth of the shape and the extension of body members. About Effort factors, there was a moderate positive correlation with Abrupt Time quality and a moderate negative correlation with Bound Flow quality. The **Angry** emotion was characterized by a moderate positive correlation with Shape factors. Between Angry emotion and Effort features, there was a strong positive correlation with Abrupt Time and Strong Weight and a strong negative correlation with Bound Flow quality. The **Sad** emotion was significantly characterized with Contracted and Shrinking Shape, Sustained, Bound and Light movement. The **Calm** emotion there was a low correlation between calm emotion and most LMA features. We found a weak negative correlation with Shape factors. There was a moderate negative correlation with Time and Weight factors and a positive correlation with Bound Flow factor.

Table 8 presented the correlation results obtained between LMA factors and emotions. The scores greater than 0.3 are written in bold (a correlation coefficient of 0.3 is considered as a moderate correlation). From this table, we can notice that the Directional Movement quality has the least importance in characterizing human emotions in expressive control motions.

#### 4.2.3 Discussion

The purpose of this statistical part is to study the recognition of emotions through body movement, and subsequently to study the importance of body features to characterize human emotions. For the first study, according to the results in Table 6, we can notice that viewers confuse Happy and Angry emotions both have high arousal and differ in valence. Also, there is a confusion between Sad and Calm emotions sharing a low arousal and differing in valence. But generally the high scores are presented in the diagonal of the confusion matrix of emotion recognition which indicates the success of viewers in classifying emotions. Same results are obtained with learning methods especially with the RDF method achieving a f-score of 0.89. For the second study, according to results illustrated in Table 8, we can notice that Happy and Angry emotions were associated with similar features but with different importance. The Angry emotion was strongly characterized by abrupt, strong and free movements. However, the Happy emotion was significantly characterized by growing shape and extended body members. On the other hand, Sad and Calm emotions had the same feature qualities but with different importance. The Sad emotion was strongly associated with sustained, light and bound movements with shrinking shape and contracted body limbs. The Calm emotion was strongly characterized by bound movements. According to the scores given by the viewers we can say that most of the LMA characteristics are important for the characterization of emotions with the exception of the Directional movement factor. However, with the RDF method we found that the combination of all LMA features is important. This can be explained by the type of our gestures where in some cases the human is unable to differentiate between rectilinear and curvilinear movement for a same movement acting with two different emotions.

Despite having recruited non-professional actors to express emotions for the construction of the database, the viewers were able to distinguish the Laban features in the movements. This may be an advantage for the application of our framework to be used even by non-experts. On the other hand, we still have the opportunity to enhance the consistency between viewer ratings with some training Laban's theories and techniques.

## 5 Conclusion and perspectives

In this paper, we have proposed an expressive motion representation based on the Laban Movement Analysis method. Our motion descriptor was inspired by LMA

**Table 8** Pearson's correlation coefficients  $r$  between LMA features rating and emotions recognition (\*\*. Correlation is significant at the 0.001 level).

	<b>Happy</b>	<b>Angry</b>	<b>Sad</b>	<b>Calm</b>
Time	<b>0.456**</b> <b>Abrupt</b>	<b>0.734**</b> <b>Abrupt</b>	<b>-0.791**</b> <b>Sustained</b>	<b>-0.383**</b> <b>Sustained</b>
Weight	0.145** Strong	<b>0.765**</b> <b>Strong</b>	<b>-0.579**</b> <b>Light</b>	<b>-0.345**</b> <b>Light</b>
Flow	<b>-0.463**</b> <b>Free</b>	<b>-0.711**</b> <b>Free</b>	<b>0.614**</b> <b>Bound</b>	<b>0.627**</b> <b>Bound</b>
Space	<b>0.340**</b> <b>Direct</b>	<b>0.417**</b> <b>Direct</b>	-0.278** Indirect	<b>-0.455**</b> <b>Indirect</b>
Shape flow	<b>0.573**</b> <b>Growing</b>	<b>0.392**</b> <b>Growing</b>	<b>-0.657**</b> <b>Shrinking</b>	-0.188** Shrinking
Shaping	<b>0.509**</b> <b>Extended</b>	<b>0.475**</b> <b>Extended</b>	<b>-0.724**</b> <b>Contracted</b>	-0.217** Contracted
Directional movement	0.095	0.093	-0.114** Curvilinear	-0.158** Curvilinear

qualities and aiming at characterizing both quantitative and qualitative aspects of human movement. Our descriptor vector was tested for motion and emotion recognition. Four learning methods were used for learning and classification motions (RDF, OAO, OAA and MLP). In the first study, we compared our approach to state of the art methods where we first analyzed the robustness of our descriptor in the recognition of motions with three public datasets and our dataset dedicated to control motions. In a second study, we evaluated the pertinence of our descriptor in recognizing human emotions based on five expressive motions. Each motion was performed with four basic emotions selected from the arousal-valence model (Happy, Angry, Sad and Calm). The same learning methods were applied for emotion recognition, and a study was proposed to analyze the relationship between our descriptor and the four selected emotions. Different scenarios were proposed to participants in order to be able to express their emotions in a natural way, and a 3D virtual avatar was introduced to reproduce human body motions. A statistical study based on viewer ratings was proposed to evaluate our recognition system. An external judgment was made for the evaluation of emotions and the rating of features across all motions in order to analyze the correlation between body features and the selected emotions and then concluded the importance of each feature to express each emotion.

As perspectives, we will expand our database with other emotions to provide more information about the relationship between body features and human emotions. We will apply our framework in a robotic application in order to control NAO robot with expressive motions from our dataset.

**Acknowledgements** We would like to thank the staff of the University of Evry Val d'Essonne for participating in our

datasets. This work was partially supported by the Strategic Research Initiatives project iCODE accredited by University Paris Saclay.

## References

- De Gelder, B.: Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**(1535), 3475-3484 (2009)
- Russell, J.A.: Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*. 115, 102-141 (1994).
- Paul, E.: *Facial Expressions. Handbook of Cognition and Emotion*, Wiley-Blackwell. 16, 301-320 (2005)
- Aviezer, H., Hassin, R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M., Bentin, S.: Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological science*. **19**(7), 724-32 (2008)
- Aviezer, H., Bentin, S., Dudarev, V., Hassin, R.: The automaticity of emotional face-context integration. *Emotion*. **11**(6), 1406-14 (2011)
- Ajili, I., Mallem, M., Didier, J.Y.: Robust human action recognition system using Laban Movement Analysis. *Procedia Computer Science, Knowledge-Based and Intelligent Information & Engineering Systems*, 112, 554-563 (2017)
- Ajili, I., Mallem, M., Didier, J.Y.: Gesture recognition for humanoid robot teleoperation. *26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 1115-1120 (2017)
- Russel, James, A.: A circumplex model of affect. *Journal of personality and social psychology*. **39**(6), 1161 (1980)
- Gong, D., Medioni, G., Zhao, X.: Structured Time Series Analysis for Human Action Segmentation and Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(7), 1414-1427 (2014)
- Junejo, I.N., Junejo, K.N., Al Aghbari, Z.: Silhouette-based human action recognition using SAX-Shapes. *The Visual Computer*. **30**(3), 259-269 (2014)
- Jiang, X., Zhong, F., Peng, Q., Qin, X.: Online robust action recognition based on a hierarchical model. *Vis. Comput.* **30**(9), 1021-1033 (2014)
- Wang, H., Kläser, A., Schmid, C., Liu, C.L.: Dense trajectories and motion boundary descriptors for action recognition. *Int. J. Comput. Vis.* **103**(1), 60-79 (2013)

13. Xia, L., Aggarwal, J.K.: Spatio-temporal Depth Cuboid Similarity Feature for Activity Recognition Using Depth Camera. *IEEE Conference on Computer Vision and Pattern Recognition*, 2834-2841 (2013)
14. Oreifej, O., Liu, Z.: HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences. *IEEE Conference on Computer Vision and Pattern Recognition*, 716-723 (2013)
15. Chi, D., Costa, M., Zhao, L., Badler, N.: The EMOTE Model for Effort and Shape. *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '00*, 173-182 (2000)
16. Kapadia, M., Chiang, I.K., Thomas, T., Badler, N.I., Kider Jr., J.T.: Efficient Motion Retrieval in Large Motion Databases. *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, 19-28 (2013)
17. Müller, M., Röder, T., Clausen, M.: Efficient Content-based Retrieval of Motion Capture Data. *ACM Trans. Graph.* **24**(3), 677-685 (2005)
18. Durupinar, F., Kapadia, M., Deutsch, S., Neff, M., Badler, N.: PERFORM: Perceptual Approach for Adding OCEAN Personality to Human Motion Using Laban Movement Analysis. *ACM Trans. Graph.* **36**(1) (2016)
19. Hsu, E., Pulli, K., Popović, J.: Style Translation for Human Motion. *ACM Trans. Graph.* **24**(3), 1082-1089 (2005)
20. Xia, S., Wang, C., Chai, J., Hodgins, J.: Realtime Style Transfer for Unlabeled Heterogeneous Human Motion. *ACM Trans. Graph.* **34**(4), 119:1-119:10 (2015)
21. Yumer, M.E., Mitra, N.J.: Spectral Style Transfer for Human Motion Between Independent Actions. *ACM Trans. Graph.* **35**(4), 137:1-137:8 (2016)
22. Aristidou, A., Zeng, Q., Stavrakis, E., Yin, K., Cohen-Or, D., Chrysanthou, Y., Chen, B.: Emotion control of unstructured dance movements. *Symposium on Computer Animation*. (2017)
23. Aristidou, A., Stavrakis, E., Papaefthimiou, M., Papa- giannakis, G., Chrysanthou, Y.: Style-based motion analysis for dance composition. *The Visual Computer*. 1432-2315 (2017)
24. Rudolf, V.L., Lisa, U.: *The Mastery of movement*. Mac Donald and Evans, 1971.
25. Aristidou, A., Charalambous, P., Chrysanthou, Y.: Emotion Analysis and Classification: Understanding the Performers' Emotions Using the LMA Entities. *Computer Graphics Forum*. **34**(6), 262-276 (2015)
26. Truong, A., Boujut, H., Zaharia, T.: Laban descriptors for gesture recognition and emotional analysis. *The Visual Computer*. **32**(1), 83-98 (2016)
27. Knight, H., Thielstrom, R., Simmons, R.: Expressive path shape (swagger): Simple features that illustrate a robot's attitude toward its goal in real time. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1475-1482 (2016)
28. Nishimura, K., Kubota, N., Woo, J.: Design support system for emotional expression of robot partners using interactive evolutionary computation. *IEEE International Conference on Fuzzy Systems*, 1-7 (2012)
29. Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., Scherer, K.: Toward a Minimal Representation of Affective Gestures. *IEEE Transactions on Affective Computing*, **2**(2), 106-118 (2011)
30. Bouchar, D., Badler, N.: *Semantic Segmentation of Motion Capture Using Laban Movement Analysis. Intelligent Virtual Agents. Springer Berlin Heidelberg*, 37-44, Berlin, Heidelberg (2007)
31. Samadani, A., Burton, S., Gorbet, R., Kulic, D.: Laban Effort and Shape Analysis of Affective Hand and Arm Movements. *Humaine Association Conference on Affective Computing and Intelligent Interaction*, 343-348 (2013)
32. Senecal, S., Cuel, L., Aristidou, A., Magnenat-Thalman, N.: Continuous body emotion recognition system during theater performances. *Computer Animation and Virtual Worlds*, **27**(3-4), 311-320 (2016)
33. Cimen, G., Ilhan, H., Capin, T., Gurcay, H.: Classification of human motion based on affective state descriptors. *Computer Animation and Virtual Worlds*. **24**(3-4), 355-363 (2013)
34. Barber, C.B., Dobkin, D.P., Huhdanpaa, H.: The Quick-hull Algorithm for Convex Hulls. *ACM Trans. Math. Softw.* **22**(4), 469-483 (1996)
35. Fothergill, S., Mentis, H., Kohli, P., Nowozin, S.: Instructing People for Training Gestural Interactive Systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*, ACM, 1737-1746 (2012)
36. Xia, L., Chen, C.-C., Aggarwal, J.K.: View invariant human action recognition using histograms of 3D joints. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 20-27 (2012)
37. Li, W., Zhang, Z., Liu, Z.: Action recognition based on a bag of 3D points. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 9-14 (2010)
38. Quinlan, J.R.: *Learning With Continuous Classes*. World Scientific, 343-348 (1992)
39. Breiman, L.: *Classification and regression trees*. 358, Wadsworth International Group, Belmont, Calif (1984)
40. Diaz-Uriarte, R., Alvarez de Andrés, S.: Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*. **7**(3) (2006)
41. Hripscak, G., Rothschild, A.S.: Technical Brief: Agreement, the F-Measure, and Reliability in Information Retrieval. *JAMIA*. **12**(3), 296-298 (2005)
42. Arlot, S., Celisse, A.: A survey of cross-validation procedures for model selection. (2009)
43. Slama, R., Wannous, H., Daoudi, M.: Grassmannian Representation of Motion Depth for 3D Human Gesture and Action Recognition. *22nd International Conference on Pattern Recognition*, 3499-3504 (2014)
44. Bland, J.M., Altman, D.G.: Statistics notes: Cronbach's alpha. *BMJ*. **314**(7080), 572 (1997)
45. Lehrmann, A.M., Gehler, P.V., Nowozin, S.: Efficient Nonlinear Markov Models for Human Motion. *IEEE Conference on Computer Vision and Pattern Recognition*, 1314-1321 (2014)
46. Song, Y., Morency, L.P., Davis, R.: Distribution-sensitive learning for imbalanced datasets. *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 1-6 (2013)
47. Truong, A., Zaharia, T.: Dynamic Gesture Recognition with Laban Movement Analysis and Hidden Markov Models, *Proceedings of the 33rd Computer Graphics International, CGI '16*, ACM, Greece, 21-24 (2016)