A set of Full-Body Movement Features for Emotion Recognition to Help Children affected by Autism Spectrum Condition

Stefano Piana
DIBRIS*
University of Genova
steto84@infomus.org

Alessandra Staglianò
DIBRIS*
University of Genova
alessandra.stagliano@unige.it

Antonio Camurri
DIBRIS*
University of Genova
antonio.camurri@unige.it

Francesca Odone
DIBRIS*
University of Genova
francesca.odone@unige.it

ABSTRACT
The EU FP7 ICT Project ASC-INCLUSION aims at developing ICT solutions to assist children affected by Autism Spectrum Conditions (ASC), in order to monitor their behaviour while playing with an interactive serious game that will help them in understanding and reproducing emotions. The complete framework will analyse the voice, the face and the body of the player to have a general description of the emotion the child is feeling/trying to play. In this work the focus is on the body movement analysis. A set of 2D and 3D features is introduced and Dictionary Learning is described as a possible method to help classifying emotions. Preliminary results are shown to assess the importance of the studied features in solving the problem.

Keywords
body gesture analysis, motion features, motion capure systems, RGB-D cameras, dictionary learning

1. INTRODUCTION
This work is part of the EU FP7 ICT 3-years Project ASC-INCLUSION, that aims at developing ICT solutions to assist children affected by Autism Spectrum Conditions (ASC). In particular, it focuses on the development of serious games to support children to understand and express emotions. The general ASC-Inclusion framework will be able to process facial expressions, voice, and full-body movement and gesture. Automatic monitoring of children behaviour in ecological environments (e.g., their home) is an important component to detect their emotional state and stimulate them to interact socially. The proposed framework will monitor the ASC children while they are interacting with other people or playing serious games and will evaluate their ability to express and understand emotions. Then the system will try to help the children improve their knowledge of emotions using interactive multimodal feedback. The whole system is described in Figure 1 [20]; the module marked in red is the part that will be described in this paper.

State of the art research in emotion recognition mainly focuses on facial expression or voice analysis. The movements that involve the entire body also contribute to the inference of distinct emotions [8, 27, 7, 9]. Research in experimental psychology demonstrated how some qualities of movement are related to specific emotions: for example, the fear brings to contract the body as an attempt to be as small as possible, surprise brings to turn towards the object capturing our attention, joy may bring to movements of openness and acceleration of forearms toward the high [5]. Body turning away is typical of fear and sadness; body turning towards is typical of happiness, anger, surprise; we usually spread out when we are happy, angry or surprised; we can either move fast (fear, happiness, anger, surprise) or slow (sadness).

In [8] de Meijer presents a detailed study of how the body movements are related to the emotions. He individuates the following dimensions and qualities: Trunk movement: *stretching - bowing*; Arm movement: *opening - closing*; Vertical direction: *upward - downward*; Sagittal direction: *forward - backward*; Force: *strong - light*; Velocity: *fast - slow*; Directness: *direct - indirect*.

Those dimensions and qualities can be found in different combinations in the different emotions. For instance a joyful feeling could be characterized by a strong Force, a fast Velocity and a direct trajectory but it could have a Light force as well, or be an indirect movement. For a more detailed description of the study, please see [8].

This paper focuses on the automated emotions recognition

*This work has been developed inside the InfoMus Lab in Casa Paganini, Piazza di Santa Maria in Passione 34, Genoa, Italy

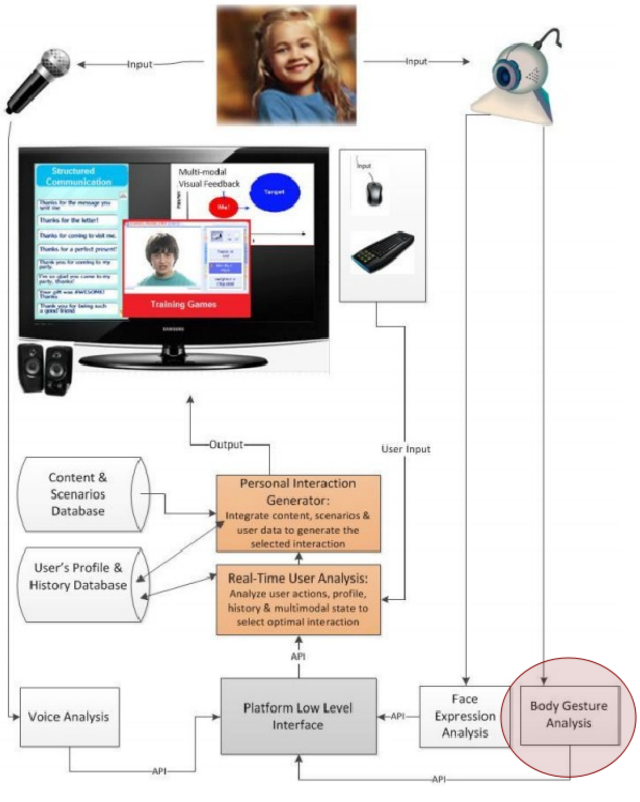

Figure 1: A scheme of the ASC-Inclusion Project, with a focus on the part that will be analysed in this paper.

starting from entire body movement. We base our research on the previously mentioned studies from experimental psychology (de Meijer[8], Wallbott [27], Boone & Cunningham [5]) and from humanistic theories (Laban effort [15, 14]). We start from a detailed analysis of full-body movement, by means of movement data recordings using professional grade optical motion capture systems (e.g., Qualysis[2]) and video cameras. Once refined, the developed emotion analysis algorithms are applied on low-cost (and less precise) RGB-D sensors (e.g., Kinect[1]). The adoption of low-cost measuring devices will enable us to integrate our analysis techniques in serious games supporting autistic children to learn to recognize and to express emotions, a main goal of the ASC-INCLUSION EU ICT Project. In the following sections we will introduce the set of movement features that we propose for emotion recognition (Section 2); then, we propose a method to represent the extracted features in order to exploit the information they carry to perform emotion recognition (Section 3); finally we present preliminary results on feature extraction performed on a dataset of 3D user tracking data created filming actors expressing emotions (the ASC-Inclusion Emotion Video Repository) (Section 3.3).

2. MOVEMENT FEATURES FOR EMOTION RECOGNITION

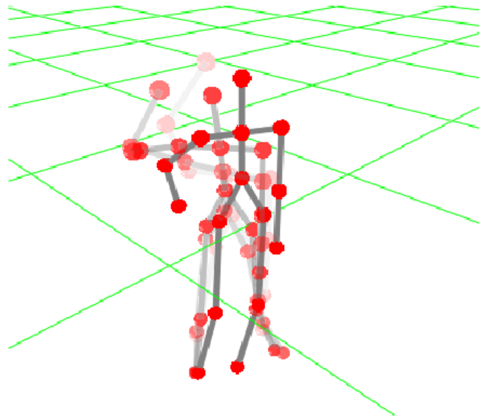

Figure 2: A short sequence showing a 3D skeleton movement.

We will now introduce a set of movement features that are deemed important in the process of recognizing emotions. The set we consider includes both features computed from 3D user tracking information, coming either from professional grade optical motion capture systems, composed by several high resolution, high speed, infra-red cameras or low-cost RGB-D cameras, and features computed from (2D) image segmentation of the shape of the users.

2.1 3D Features

In this section we describe the features computed from three-dimensional user tracking information captured by optical motion capture systems. The following set of features includes indexes computed on the user's movement and its qualities, on the body posture and on the trajectories drawn by the user's joints. An example of input data can be seen in Figure 2.

2.1.1 *Kinetic Energy*

The Kinetic Energy (KE) is the overall energy spent by the user during movement, estimated as the total amount of displacement in all of the tracked points. The amount of movement activity is important for differentiating emotions. The highest values of energy are related to anger, joy and terror emotions, the lowest values correspond to sadness and boredom. Camurri and colleagues [7] showed that the movement activity is a relevant feature in recognizing emotion from the full-body movement. For these reasons, we include in the set of expressive features an approximated measure of the overall motion energy at time frame f. Given 3D user tracking information, let $v_i(f) = \sqrt{\dot{x_i}^2(f) + \dot{y_i}^2(f) + \dot{z_i}^2(f)}$ denote the magnitude of velocity of the i-th tracked point at time frame f. We then define $KE(f)$, the Kinetic Energy index at the frame f, as an approximation of the body kinematic energy, the weighted sum of each joints' kinetic energy:

$$KE(f) = \frac{1}{2}\sum_{i=1}^{n} m_i v_i^2(f) \quad (1)$$

where m_i is the approximation of the mass of the i-th joint. The m values are computed starting from anthropometric tables [19].

2.1.2 Spatial extent: Contraction Index and Density

The Contraction Index ($3DCI$), is a measure, ranging from 0 to 1, of how the user's body occupies the space surrounding it. It is related to Laban's "personal space" [15, 14]. The *Contraction Index* is the normalized volume of the *Bounding Volume* (BV) that is the volume of the minimum parallelepiped surrounding the user's body.

A different index of spatial extent is the density (DEI). The density of a set of coordinates can be computed as follows:

$$DEI = \frac{1}{n}\sum_{i=1}^{n} d_i \tag{2}$$

where d_i is the Euclidean distance of the i-th point of the set from the centroid C that is computed as the centre of mass of the set.

2.1.3 Smoothness and Fluidity

In general, *smoothness* (SMI) is synonymous for "having small values of high-order derivatives." Wallbott [27], in his analysis of qualitative aspects of psychiatric patients' hand movements, noticed that movements judged as smooth are "characterized by large circumference, long wavelength, high mean velocity, but not abrupt changes in velocity or acceleration (standard deviations of velocity and acceleration). Thus, smooth movements seem to be large in terms of space and exhibit even velocity". We have therefore adapted Wallbott's statements on the qualitative dimensions of underconstrained arm movements and we have computed hands trajectories curvature to identify trajectories' smoothness. Curvature (k) measures the rate at which a tangent vector turns as a trajectory bends. It is defined as the reciprocal of the radius (R) of the curve described by the trajectory:

$$k = \frac{1}{R} \tag{3}$$

A joint's trajectory following the contour of a small circle will bend sharply, and hence will have higher curvature that means low fluidity; by contrast, a point trajectory following a straight line will have zero curvature so high smoothness. The curvature is computed for a single point's trajectory at time frame f as follows:

$$k_i = \frac{\dot{r_i} \times \ddot{r_i}}{\dot{r_i}^3} \tag{4}$$

where $\dot{r_i}$ is the velocity of the trajectory of the i-th point and $\ddot{r_i}$ is its acceleration.

The principle of the curvature can be applied to the movement's velocity and its variation in time to calculate the movement's fluidity (FI): high curvature of the speed's trajectory in time means low fluidity, while low curvature means high fluidity. To calculate the fluidity we compute the curvature of the tangential velocity of the desired joint as described in equation (4).

2.1.4 Symmetry

Lateral symmetry (SI) of emotion expression has long been studied in face expressions, resulting in valuable insights about a general hemisphere dominance in the control of emotional expression. An established example is the expressive advantage of the left hemiface that has been demonstrated with chimeric face stimuli, static pictures of emotional expressions with one side of the face replaced by the mirror image of the other. A study by Roether et al. on human gait demonstrated pronounced lateral asymmetries also in human emotional full-body movement [22]. Twenty-four actors (with an equal number of right and left-handed subjects) were recorded by using a motion capture system during neutral walking and emotionally expressive walking (anger, happiness, sadness). For all the three emotions, the left body side moved with significantly higher amplitude and energy. Perceptual validation of the results were conducted through the creation of chimeric walkers using the joint-angle trajectories of a half body to animate completely symmetric puppets. Considering that literature pointed out the relevance of symmetry as behavioural and affective features, we address the symmetry of gestures and its relation with emotional expression. It is measured evaluating limbs spatial symmetry with respect to the body computed symmetry on each of the available dimensions. Each partial symmetry (SIx, SIy, SIz) is computed from the position of the centre of mass and the left and right joints (e.g., hands shoulders, foots, knees) as described below:

$$SI_{Xi} = \frac{(x_{Li} - x_B) - (x_{Ri} - x_B)}{|x_{Li} - x_B| + |x_{Ri} - x_B|} \quad i = 0, 1, ..., n \tag{5}$$

$$SI_{Yi} = \frac{(y_{Li} - y_B) + (y_{Ri} - y_B)}{|y_{Li} - y_B| + |y_{Ri} - y_B|} \quad i = 0, 1, ..., n \tag{6}$$

$$SI_{Zi} = \frac{(z_{Li} - z_B) + (z_{Ri} - z_B)}{|z_{Li} - z_B| + |z_{Ri} - z_B|} \quad i = 0, 1, ..., n \tag{7}$$

where x_B, y_B, z_B are the coordinates of the centre of mass, x_{Li}, y_{Li}, z_{Li} are the coordinates of a left joint (e.g., left hand, left shoulder, left foot, etc.) and x_{Ri}, y_{Ri}, z_{Ri} are the coordinates of a right joint (e.g., right hand, right shoulder, right foot, etc.). The three partial indexes are then combined in a normalized index that expresses the overall estimated symmetry.

2.1.5 Forward-backward leaning of the upper body and relative positions

Head and body movements and positions are relied on as an important feature for distinguishing between various emotional expressions [23]. The amount of forward and backward leaning JL_i of the joint i at the time frame f is measured by the velocity of the joint's displacement along its z component (depth) respective to the body position and orientation.

$$JL_i = \frac{(z_B - z_{Li}) - (z_B - z_{Ri})}{z_{Ri} - z_{Li}} \tag{8}$$

2.1.6 Directness

Directness (DI) is one of the features that are deemed important in the process of recognizing emotions [8]. A direct movement is characterized by almost rectilinear trajectories. Movement Directness Index is computed from a trajectory drawn in the space by a joint as the ratio between the euclidean distance, calculated between the starting and the ending point of the trajectory, and the trajectory's actual length. The directness index tends to assume values close to 1 if a movement is direct and low values (close to 0) otherwise. In the case of three dimensional trajectories the index is computed as follows:

$$DI = \frac{\sqrt{(x_E - x_S)^2 + (y_E - y_S)^2 + (z_E - z_S)^2}}{\sum_{k=1}^{N-1} \sqrt{(x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2 + (z_{k+1} - z_k)^2}} \tag{9}$$

where x_S, y_S, z_S are the coordinates of the trajectory's starting point, x_E, y_E, z_E are the coordinates of the ending point and N is the length of the trajectory.

2.1.7 *Periodicity*

Periodicity (PI) can be calculated using periodicity transform [24], The periodicity transform decomposes sequences into a sum of periodic sequences by projecting onto a set of "periodic subspaces". The Periodicity Transform looks for the best periodic characterization of the length N sequence x. The underlying technique is to project x onto some periodic subspace P_p. This periodicity is then removed from x leaving the residual r stripped of its *p-periodicities*. A sequence of real numbers is called *p-periodic* if there is an integer p with $x(k+p) = x$ for all k integers.

2.1.8 *Impulsiveness*

In human motion analysis, Impulsiveness (II) can be defined as a temporal perturbation of a regime motion (Heiser et al. [13]]). Impulsiveness refers to the physical concept of impulse as a variation of the momentum. This contributes to define and reach a reference measure for impulsiveness. From psychological studies (Evenden [11]; Nagoshi et al. [21]), an impulsive gesture lacks of premeditation, that is, it is performed without a significant preparation phase. We developed an algorithm for impulsiveness detection, derived by (Mancini and Mazzarino [18]), where a gesture is considered an impulse if it is characterized by a short duration and high magnitude. In addition to the observations made by Mazzarino and Mancini we included the condition that to be recognized as impulsive a gesture has to be performed without preparation. This includes sudden change of the movement's direction or intensity. The algorithm for the computation of the impulsiveness index is the following:

let δt=0.45sec;
let energyThreshold=0.02;
if (KE >= energyThreshold) **then**
  if (0 <= dt<=δt) **then**
    evaluate energy peaks and movement direction;
    if (the energy peak is solitaire in the given direction) **then**
      $II = \frac{\overline{KE}}{dt}$;
    end if
  end if
end if

where dt is the gesture duration, KE is the kinetic energy and $\overline{KE}$ is the mean of the kinetic energy computed over the gesture duration. The values of the proposed thresholds have been empirically evaluated through perception tests on videos portraying people who performed highly and lowly impulsive gestures. Figure 3 shows an example of how the *Impulsiveness Index* is computed.

2.2 2D+T features

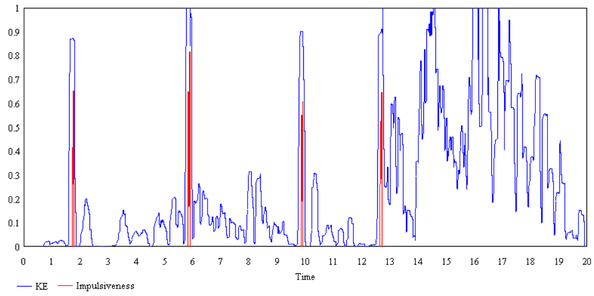

Figure 3: Example of *Impulsiveness Index* computation, a gesture is recognised as impulsive only if executed with high energy, quick time and there is a clear change in the movement qualities (change of movement's direction, isolated peak of energy)

In this section we summarize the 2D features extracted from the image sequences. As a pre-processing step all the images have been segmented with a change detection algorithm to discard the background portion, thus the measurements are computed on the silhouette of the person moving in the scene. These features are affected by acquisition and preprocessing noise ad by numerical noise accumulated by the image processing algorithms. Thus they are noisier than the 3D features. At the same time, they can be computed easily starting from image sequences acquired by off-the-shelf low cost cameras and do not require any particular acquisition set up. All these considerations make them appealing for an ecological set up.

2.2.1 *Quantity of Motion and MHI/MHG*

An important clue of an emotion is how much a person is moving during the action, and the direction of the movement itself. To compute the *Quantity of Motion* (QoM), in [4] the authors propose to represent patterns of motions with successive layering of image silhouettes. The Motion History Image (MHI) is the result of this layering, where every time a new frame arrives the older silhouettes are decreased in value and the new silhouette is added with the maximum brightness (see Figure 4 for some examples). The Quantity of Motion is computed as the weighted sum of the area of the layered silhouette: a wide movement will have a higher QoM than a small one. In [6], this technique is extended in order to take into account the direction of the motion as well, computing the gradients of the MHIs resulting from a bunch of frames. In this case the authors talk about Motion History Gradients ($MHGs$). These features can be used to cluster the different videos in groups with high or low quantity of motion and/or up or down movements.

2.2.2 *Barycentre Tracking and 2D Contraction Index*

The motion of the barycentre (BMI) of a silhouette over time gives a good idea of the way a person is moving. Figure 5 shows the trends of the BMI of the same actor performing two different emotions (excitement and sadness). In the first case, the barycentre spans a bigger area and its movements are wider, while in the second case its movements are more quiet.

(a) (b) (c)

Figure 4: 10 frames layered silhouettes of a person standing (a), waving arms (b) and jumping (c).

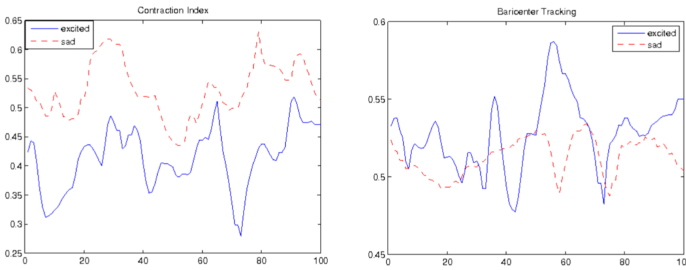

Figure 5: Comparison of the Contraction Index (left) and the Barycentre motion (right) of an excited person and a sad person

The 2D Contraction Index ($2DCI$) is a measure, ranging from 0 to 1, related to the bounding region, i.e., the minimum rectangle surrounding the person's body. A low $2DCI$ will correspond to an open shape of the silhouette, with the limbs stretching away from the body. Vice versa, a high CI will carry the information of a close shape of the silhouette. Another important contribution is the variation of the $2DCI$ over time. Intuitively, if we are observing an action of an excited person, we expect him or her to open and close quickly the limbs, leading to a quickly varying $2DCI$. The variances, rather than the means, are more expressive measures to discriminate between emotions.

3. ENHANCING THE FEATURES DESCRIPTIVE POWER

The features described so far will be the starting point for the discrimination of the emotions. They will have to be combined in order to obtain a descriptor that will allow the system to understand automatically what the person in the video is feeling. Assuming to have a set of labelled examples, we aim at training the system so that it will be able to give a correct answer when a new, unlabelled datum arrives. Since it is not easy to get labelled data, both in terms of time and accuracy (subjectivity) of the labelling, we want to maximise the amount of information that can be extracted from the data, so we need a data representation that has got a good descriptive power with respect to the problem. In the next subsection we describe a technique, called Dictionary Learning, that in an unsupervised fashion is able to extract the intrinsic characteristics of the processed data. Its goal is to learn a set of prototype signals from the data so that they can be represented by means of a linear combination of such basis. The final reconstruction will be a de-noised version of the original data, maintaining only the actual, important characteristics and discarding noise. Once the features are decomposed with respect to the dictionary, the sparse vectors containing the coefficients of the linear combination will be used as input for classification algorithms.

3.1 Sparse Coding and Dictionary Learning

Let $\mathbf{x}$ be a signal in $\mathbb{R}^d$. We assume we can decompose it with respect to a dictionary $\mathbf{D}$ in $\mathbb{R}^{d\times K}$, that is a collection of K basis signals, or *atoms*. In other words, there exists a vector $\mathbf{u}$ in $\mathbb{R}^K$ such that $\mathbf{x} \approx \mathbf{Du}$. In general $K > d$ and we want only a few atoms to be active in the linear combination, thus we look for a sparse vector of coefficients. The solution to our problem can be found minimizing with respect to $\mathbf{u}$ the functional in (10), where the first term is the reconstruction error, that ensures that our data are not distant from the linear combination of atoms, and the ℓ_1-norm term is added to enforce sparsity [25].

$$\|\mathbf{x} - \mathbf{Du}\|_2^2 + \lambda\|\mathbf{u}\|_1 \tag{10}$$

Having a batch of n data, the problem can be written in the following matrix form

$$\|\mathbf{X} - \mathbf{DU}\|_2^2 + \lambda\sum_{i=1}^{n}\|\mathbf{u}_i\|_1 \tag{11}$$

The dictionary can be fixed analytically a priori, using for example Wavelets or Discrete Cosine transform [17], or it can be learned directly from data as for example in [16, 10]. To this purpose, the functional in (11) can be minimized with respect to both $\mathbf{D}$ and $\mathbf{U}$. The problem is still convex fixing one variable at a time and minimizing with respect to the other, in an alternate scheme. The solution can be achieved using for example the PADDLE algorithm (see [3]).

As stated before, we can now use the vectors $\mathbf{u}$ as data representations and decompose new data with respect to $\mathbf{D}$ (using only one step of the optimization).

3.2 Learning a Dictionary of Emotions

Once we have extracted all the features described in Section 2, our data will be represented by a set of time series describing the type of motion in 2D or 3D evolving over time. These time series may be split in sub-sequences of a fixed length which will be fed into a Dictionary Learning algorithm. For example, suppose to have a feature vector $\mathbf{x}$ describing N frame. We divide it into n sub-vectors describing $j = \frac{N}{n}$ frames each, resulting in an input matrix $\mathbf{X} \in \mathbb{R}^{\mathbf{j}\times\mathbf{n}}$. We then learn $\mathbf{D} \in \mathbb{R}^{\mathbf{j}\times\mathbf{K}}$ and $\mathbf{U} \in \mathbb{R}^{\mathbf{K}\times\mathbf{n}}$, so for each vector we get a description matrix composed by the $\mathbf{u}$ vectors associated to each sub-vector. If we learn a dictionary for each feature, we can then concatenate every descriptor to get a global one.

Dictionary Learning and Sparse Coding techniques have been successfully used in [12] for gesture recognition. In this work the authors aim at recognizing an action extracted from a set of predefined actions, performed by a human. To do so, they extract features from both the motion and the appearance of the performer and then they learn a dictionary for each feature. The final descriptor used for the proper action

Table 1: Summary of the extracted features for the six basic emotions: anger, disgust, fear, happiness, sadness and surprise

	Anger		Disgust		Fear		Happiness		Sadness		Surprise	
	Mean	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**
KE	1.7078	5.3978	0.7591	0.4834	1.5567	2.7342	1.5705	6.6397	0.7576	3.2839	1.1101	6.0395
3DCI	0.4733	0.0315	0.6055	0.0191	0.6787	0.3536	0.4388	0.06032	0.6868	0.0167	0.4843	0.0351
DEI	0.2904	0.0015	0.2743	0.0018	0.5731	0.3381	0.3026	0.0010	0.2676	0.0014	0.2724	0.0012
SI	0.5751	0.0168	0.6178	0.0314	0.7500	0.3475	0.6274	0.2284	0.6872	0.0248	0.6892	0.0254
SMI	0.1621	0.0243	0.1301	0.0145	0.4594	0.3505	0.5801	0.0119	0.9774	0.0118	0.0698	0.0267
FI	0.6784	0.0617	0.7911	0.0364	0.7295	0.3942	0.7864	0.0315	0.7631	0.0660	0.6213	0.1122
JL	0.0175	0.8710	-0.0075	0.8701	0.3720	0.3435	0.0069	0.9067	0.0155	0.7869	-0.0098	0.8081
DI	0.8532	0.0754	0.5272	0.0681	0.3684	0.1698	0.8249	0.2789	0.4529	0.0347	0.7246	0.0781
PI	0.2823	0.0019	0.2595	0.0012	0.5495	0.3383	0.2639	0.0017	0.2708	0.0010	0.2518	0.0010
II	0.1438	0.0262	0.0203	0.0135	0.4010	0.3554	0.0372	0.0242	0.0123	0.0074	0.0970	0.0025
QoM	1.5210	0.1011	1.3791	0.0579	1.517	0.146	1.4611	0.0772	1.3663	0.0562	1.5964	0.2604
2DCI	0.4621	0.0057	0.4921	0.0057	0.4645	0.0047	0.4030	0.0122	0.4802	0.0050	0.4324	0.0101
BMI	0.5128	0.0007	0.5179	0.0004	0.5631	0.0007	0.5105	0.0006	0.5341	0.0009	0.5355	0.0009

classification is the vector obtained by concatenating each sparse code computed. The classification task is carried out using Support Vector Machines (SVMs) [26] because the descriptors are easily separable.

Our setting is slightly different from the one described above, because we don't have to recognize an action, but the different shades of the motion/appearance allowing us to understand the feelings of the person we are observing. Despite this, we think that this model can help us building a space where the dimensions of the different emotions lie and having a set of easily separable representations.

3.3 Preliminary Results

We will now review some preliminary results on feature extraction performed on videos and MoCap data coming from recordings of actors expressing emotions through full body movements. The dataset is composed by audiovisual and 3D MoCap recordings (captured by an RGB-D camera) of seven actors (4 women and 3 men), each actor was asked to interpret the six basic emotions (anger, disgust, fear, happiness, sadness and surprise) for a short period of time then we extracted movement features using the algorithms described in Section 2. Table 1 shows some measurements about the extracted features, namely the mean and the average over time obtained for each of the described features measured on different instances of each emotions and then averaged. Despite the fact that mean and average produce a big loss of information, the numbers obtained allow us to perform a raw classification over the videos. For instance, Anger have a high value for the Kinetic Energy, combined with a huge variance, while Disgust does not. Happiness is another emotion with hight mean/variance for the Kinetic Energy, but these two emotions can be differentiated for a different Impulsiveness, for instance. The table shows that the combination of all the features described in Section 2 gives a very rich descriptor that emphasizes the differences and the similarities among the different emotions.

4. CONCLUSIONS AND FUTURE WORK

In this paper we have reviewed a set of body movement's features that can be extracted from video/3D sequences to understand emotions automatically. These features are deeply inspired by the psychological literature that have found different dimensions that allow a human being to interpret what another subject is feeling.

We have also introduced a Machine Learning technique, named Dictionary Learning, that we aim at exploiting in order to have good responses on emotion classification.

The dataset used to test the feature extraction algorithms isn't suitable for the final objective of helping ASC children to improve their ability of emotion expression because it is based on adult actors recordings. At the moment, we are creating a dataset composed by recordings of children (aged from 5 to 10) that express natural emotions. This dataset will be used for training the system for automatic emotion recognition.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 289021 (ASC-Inclusion).

5. REFERENCES

[1] Kinect for windows, 2013. http://www.microsoft.com/en-us/kinectforwindows/.

[2] Qualisys motion capture systems, 2013. http://www.Qualisys.com.

[3] C. Basso, M. Santoro, A. Verri, and S. Villa. Paddle: proximal algorithm for dual dictionaries learning. *Artificial Neural Networks and Machine Learning–ICANN 2011*, pages 379–386, 2011.

[4] Aaron F. Bobick and James W. Davis. The recognition of human movement using temporal templates. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(3):257–267, 2001.

[5] R Thomas Boone and Joseph G Cunningham. Children's decoding of emotion in expressive body movement: The development of cue attunement. *Developmental psychology*, 34:1007–1016, 1998.

[6] Gary R Bradski and James W Davis. Motion segmentation and pose recognition with motion history gradients. *Machine Vision and Applications*, 13(3):174–184, 2002.

[7] A. Camurri, I. LagerlÃűf, and G. Volpe. Recognizing emotion from dance movement: comparison of

spectator recognition and automated techniques. 59(1-2):213–225, 2003.
[8] M. de Meijer. The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, 13(4):247–268, 1989.
[9] D.Glowinski, N.Dael, A.Camurri, G.Volpe, M.Mortillaro, and K.Scherer. Towards a minimal representation of affective gestures. 2(2):106–118, 2011.
[10] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15:3736–3745, December 2006.
[11] John L Evenden. Varieties of impulsivity. *Psychopharmacology*, 146(4):348–361, 1999.
[12] Ilaria Gori, Sean Ryan Fanello, Giorgio Metta, and Francesca Odone. All gestures you can: A memory game against a humanoid robot.
[13] P Heiser, J Frey, J Smidt, C Sommerlad, PM Wehmeier, J Hebebrand, and H Remschmidt. Objective measurement of hyperactivity, impulsivity, and inattention in children with hyperkinetic disorders before and after treatment with methylphenidate. *European child & adolescent psychiatry*, 13(2):100–104, 2004.
[14] R. Laban. *Modern Educational Dance,*. Macdonald & Evans, London, 1963.
[15] R. Laban and F. C. Lawrence. *Effort*. Macdonald & Evans, London, 1947.
[16] H. Lee, A. Battle, R. Raina, and A.Y. Ng. Efficient sparse coding algorithms. In *Advances in Neural Information Processing Systems 19 (NIPS 2006)*, pages 801–808, 2006.
[17] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, New York, 2nd edition, September 1999.
[18] B. Mazzarino and M. Mancini. The need for impulsivity & smoothness: improving hci by qualitatively measuring new high-level human motion features. In *Proceedings of the International Conference on Signal Processing and Multimedia Applications (IEEE sponsored), SIGMAP is part of ICETE -The International Joint Conference on e-Business and Telecommunications*. INSTICCPress, 2009.
[19] John T McConville, Charles E Clauser, Thomas D Churchill, Jaime Cuzzi, and Ints Kaleps. Anthropometric relationships of body and body segment moments of inertia. Technical report, DTIC Document, 1980.
[20] Noga Meir-Goren. First platform basic prototype. Technical report, October, 2012.
[21] Craig T Nagoshi, James R Wilson, and Lawrence A Rodriguez. Impulsivity, sensation seeking, and behavioral and emotional responses to alcohol. *Alcoholism: Clinical and Experimental Research*, 15(4):661–667, 2006.
[22] C.L. Roether, L. Omlor, and M.A. Giese. âĂIJlateral asymmetry of bodily emotion expression. 18(8), 2008.
[23] J Schouwstra Sanneke and Johan Hoogstraten. Head position and spinal position as determinants of perceived emotional state. *Perceptual and motor skills*, 81(2):673–674, 1995.
[24] William A Sethares and Thomas W Staley. Periodicity transforms. *Signal Processing, IEEE Transactions on*, 47(11):2953–2964, 1999.
[25] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
[26] Vladimir Vapnik. Statistical learning theory, 1998.
[27] H. G. Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28:879–896, 1998.