Recognition of emotions by the emotional feedback through behavioral human poses

Javier G. Rázuri[1], Aron Larsson[1], Rahim Rahmani[1], David Sundgren[1], Isis Bonet[2], and Antonio Moran[3]

[1] Dept. of Computer and Systems Sciences, Stockholm University, DSV,
Stockholm, SE-164 40, Sweden
[2] Antioquia School of Engineering (EIA),
Envigado, Colombia
[3] Pontifical Catholic University of Perú PUCP,
Lima, Lima 32, Perú

Abstract

The sensory perceptions from humans are intertwined channels, which assemble diverse data in order to decrypt emotional information. Just by associations, humans can mix emotional information, i.e. emotion detection through facial expressions criteria, emotional speech, and the challenging field of emotional body language over the body poses and motion. In this work, we present an approach that can predict six basic universal emotions collected by responses linked to human body poses, from a computational perspective. The emotional outputs could be fed as inputs to a synthetic socially skilled agent capable of interaction, in the context of socially intelligent systems. The methodology uses a classification technique of information from six images extracted from a video, entirely developed using the motion sensing input device of Xbox 360 by Microsoft. We are taking into account that the emotional body language contains advantageous information about the emotional state of humans, especially when bodily reaction brings about conscious emotional experiences. The body parts are windows that show emotions and they would be particularly suitable to decoding affective states. The group of extracted images is merged in one image with all the relevant information. The recovered image will serve as input to the classifiers. The analysis of images from human body poses makes it possible to obtain relevant information through the combination of proper data in the same image. It is shown by experimental results that the SVM can detect emotion with good accuracy compared to other classifiers.

Keywords: *Detection of Emotional Information, Affective Computing, Body Gesture Analysis, Robotics, Classification, Machine Learning.*

1. Introduction

Human emotions are strongly connected with bodily states; they are not only connected with mental states, if not also show connections at a physical level [1]. Body signals allow us to communicate through non-verbal cues emotional indicators. The bodily indicators are translated into an informative experience, that we could handle according to the situation in which we find ourselves. The understanding of complex emotions and its physical indicators, have been carefully studied and analyzed in order to understand and track, about how we behave in social interactions [2], [3], [4], [5]. But the challenge is the understanding by a machine of all this emotional information, which will influence positively the closed loop human-robot interaction framework. Cognitive agents are shown to be capable of adapting to emotional information from humans. Taking into account emotional information could be critical for how to handle data captured by a cognitive agent, with direct and precise information of our behavior [6], [7], [8], [9], [10].

Machines which are able to express, recognize and communicate their own feelings and those of others, are capable of enhancing human computer interaction and aiding related research in surprising ways [11], this advance will also allow us to self-scrutinize the variety and complexity of our emotions and manage emotions well in ourselves and in our relationships [12]. Providing emotional abilities to machines could enhance productivity and quality of life, considering the possible interaction that the machine could generate as partner and collaborator. They would be able to manage the criteria of their actions, allowing them through judgments of value to find ways of improving. Machines are getting smarter at understanding human's actions and the environment in which they are developed, the human body motion involved in this interrelation contains a high degree of flexibility, and therefore, how to figure out the emotions implied in the behavior analysis of bodily poses is an interesting challenge to investigate. The link between the emotional differences of the interpretation from the body poses and the dependencies of the poses from persons is very strong. The emotional body language could be a different means

IJCSI International Journal of Computer Science Issues, Volume 12, Issue 1, No 1, January 2015
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

8

of expressing the same set of basic universal emotions as facial expressions and speech [13], [14], [15]. Some questions can be raised over the study of bodily expressions and the meaning of their movements through various physical behaviors. Several members of the body and the dynamics of them could be means of transmission, in unison, of emotion categories. According to [16] and [17], the three spatial dimensions of body poses could convey the intensity of emotions.

One of the motivations for the future interrelation with machines would be the happiness of human beings. This motivation would stimulate the collaboration between machines and humans by exchanging information. The future human-machine interaction could be influenced by an emotional feedback that machines can capture from the environment of humans and evolve with them. We shall also expect improvements in the communicative behavior in machines, which is a very urgent task, and then people could more easily be able to accept and integrate them. Machines have to be spontaneous, polite and must learn how to feel. In this sense, we believe that such an approach can be beneficial, both for machine-emotional feedback, which is still far from reaching the capabilities of biological vision, as well as for neuroscience [18], where these computational approaches can contribute new insights.

In this work, we address the problem of emotion recognition through bodily expression with a technique that will not imply huge memory consumption in a care giver robot. Humans have a strong support of various emotional channels; the combinations of data from different senses will provide to humans the emotional prediction which leads them to socialization with others. We can get towards making similarities between human posture and emotions that the machine could perceive. The input channel of bodily expression may be more powerful than other channels of nonverbal communication and as such presents special challenges.

In this paper, several classification methods were tested to detect six basic bodily expressions in images provided by a group of videos. This research states the relationship between dynamic postures and attributions of emotion in an attempt to describe how emotion may be communicated through the body and captured by an intelligent agent. The output of this system is a recognized emotion related with a body human posture, which will be used in a future model that supports the decision making process of a decision agent, see [19], in a scenario that occurs during the emotional behavior from human bodies, when they are taking part in the closed loop human-robot interaction.

This paper is organized as follows. Section 2 describes the related research background in bodily emotion recognition and it is divided in subsections 2.1, 2.2 and 2.3. Subsection 2.1 describes the problem to being addressed related to capture of emotional bodily information. Subsection 2.2 shows the literature based on emotion recognition using the human body poses. Subsection 2.3 covers the literature based on the bodily emotion detection based in machine learning. Section 3 presents the data and methods used in the controlled study and it is divided in subsections 3.1, 3.2 and 3.3. Subsection 3.1 describes the data set and Subsections 3.2 and 3.3 the classifiers and the performance measures used to the study, respectively. Section 4 describes the experimental results and the discussion about some particularities of the study. Conclusions are presented in Section 4.

2. Related research

Emotion recognition in humans is researched in various scientific disciplines such as neuroscience, psychology, and linguistics. Development of automated emotional recognition systems depends significantly on the progress in the aforementioned sciences, and recently multiplied from the integration of all these disciplines in computational systems for emotion prediction, see [20], and [21]. Focusing on the human body as main source of emotional information we start our analysis by exploring the problem, the background in bodily emotion recognition, and automated procedures.

2.1 Problem

Machines are entering into our life and are capable to support us in our daily tasks. They save time for us at home, freeing our hands and eyes from tablets or mobile phones, to interact, learn, play and collaborate with us and our environment in an autonomous, natural and personalized way [6]. But like humans the mysteries of emotions play an essential role in machines, which includes the ability of them to associate the others feelings with its own emotional internal state and make decisions [7]. In light of this remark, machines will be able to recover and interpret human emotions conveyed through different sources from the human body [11]. For many years, the problem of decrypting emotional information from humans in an automatic way has been studied with diverse methods and theories, many of them are indeed less efficient in terms of computation cost in personal robotics. If we think in the future of an intelligent machine like a robot, and also part of the family, it is required that they become affordable, say at the price of a PC, a tablet or

a smartphone with efficiency according to the computational characteristics.

Emotional bodily information is generally less understood compared to other type of modalities, notably in the case of face [13] and voice [15]. In some cases the option to use the multimodal approach and merge multiple sources to recognize emotions can lead to a high degree of accuracy [23]. But we cannot always hope that our companion robots are capable to react in a timely and sensible manner, especially if they haven't be able to recover all the through their sensors. Not always are the emotional features that the robot must capture provided by different sources from the human body at the same time. Maybe, all the information collected lacks robustness or, purely and simply because the robot lacks the specific sensor to extract the emotional feature. We try to solve the problem of emotion recognition in humans based only on their more common bodily expressions [3]. The method for the analysis of emotional behavior is based on direct classification from the sum of pixels in two-dimensional space images previously processed. Finally we show and interpret the recognition rates of our proposal using different classification algorithms in Weka [22].

2.2 Emotion Recognition using Body Pose

Bodily expressions convey important affective information, although this modality is relatively neglected in the literature as compared to facial expressions and speech and has been a major challenge over several years. During this quest, several misclassifications over interpretations have been found, often due to structure of the human body. In order to obtain an accurate emotion the use of a multimodal framework between the speech and body could replace the wrong data [23], the case of degrees of freedom of human body are higher than the face alone, and its overall shape varies strongly during articulated motion. In machine learning research, recent results about object recognition have shown that even for highly variable visual stimuli, quite reliable categorical decisions can be made from dense low-level visual cues [24]. Many researchers have collected lists of stereotypical features, see [25], in order to decrypt the emotional bodily expressions, whereas others have argued for diverse patterns along a number of more abstract dimensions, terms like force, speed, energy, directness, etc., [26], [27]. Despite the usefulness of these features, based on generalities, they tend to focus on dynamic properties of bodies, and usually, they fail to make clear predictions regarding the bodily poses which may be associated with different emotional states, e.g., likely configurations of the distribution head, trunk, arms, shoulders and legs in an image could be of noticeably grainy appearance and

difficult to predict. Nonetheless, there exists a variety of sources which offer more or less detailed descriptions of emotional bodily poses [28], [29], [30].

Facial gestures, kinetic of human body, bodily poses can be underlined like clear indicators of emotional states [31], some of these indicators are critical in emotional recognition from affective states [13]. The body movement of humans differs from other emotional indicators like speech and face gestures, since it is the only visual stimulus that we can perceive and produce with several degrees of freedom, with several combinations from all members of the human body [32]. The expressive body movements are strongly influenced by emotions and movement qualities highlighted in [33], [25] and [26]. Several experiments with actors that express emotions through body posture were analyzed in [34] by using photos without tridimensional information, the set of photos were decoded using low-level visual data. Some experiments have focused on body posture representation; a system can capture different body positions and generate a set of features, i.e. distance and angles between shoulder and head, etc., [35]. In addition to furthering basic understanding of human behavior, work on biological movement can reveal how stimuli act as triggers, leading to better design of computational models of emotion, focusing only of visual analysis of affective body language [36]. The human brain can generate empathic connections at a social level, for this to be possible, the motion-visual neurons are affected biological by bodily motion in many visual areas. This may allow a deeper understanding about the comprehension of others humans and how emotions are keys for empathy [37].

2.3 Automatic bodily emotion detection based in machine learning

The first research in automated emotional bodily recognition was reported in [38], this work focused on posture analysis through Tekscan's Body Pressure Measurement System (BPMS). The system works in a learning environment sensing the temporal transitions of posture of children. A neural network provided real-time classification of nine static postures with an overall accuracy of 87.6 percent. Machine learning algorithms were used in order to detect boredom, engagement/flow, confusion, frustration, and delight, by kinetics of movements of students during a learning task [39]. Two sets of features were selected from the pressure maps that were automatically computed with the BPMS. The emotion detection showed accuracy in emotions like boredom, confusion, delight, flow, and frustration from neutral with 73, 72, 70, 83, and 74 percent, respectively. In [40], [41] it was held a real-time analysis of expressive gesture in

IJCSI International Journal of Computer Science Issues, Volume 12, Issue 1, No 1, January 2015
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

10

full-body human movement based on computer vision algorithms, were the quantity of motion and contraction index of the upper body as well as velocity, acceleration and fluidity of limbs and head were measured. The Bayesian Network-based classifier achieves a correct recognition rate of 61% from four emotions; anger, joy, pleasure and sadness. The affective states using information of facial expressions and upper-body gestures were decrypted in [42]. Expressions like anger, anxiety, disgust, happiness and uncertainty are recognized by a performance of 90% by using body expressions only and with a Bayes network. The kinetics of four emotional states was recorded using a 3D motion capture system and with these data, several classifiers were tested and trained. Neural networks and support vector machines were able to achieve a correct recognition rate of 84%, [43]. Similar experiments in [44], were developed using the data from records of a standard DV camera. The expressive motions were distinguishable by a Bayesian Network-based classifier at a rate of 90%. The streams of 3D measurements were the input data to binary SVM classifiers, [45]; the experiments covered the multi-class classification of six emotions based on a combination of classifiers that used gesture segmentation by kinetic energy. The reason to use SVM classifiers was that the descriptors were easily separable. A pattern recognition problem was held in [46], the classification used suitable features from gestures provided by Kinect sensor. The joints from the upper body were coded using the angles and positions. The features spaces obtained are classified using a number of different classifiers. The average classification obtained for binary decision tree, ensemble tree, k-NN, SVM with radial basis function kernel and neural network classifier with back-propagation learning were 76.63%, 90.83%, 86.77%, 87.74% and 89.26% respectively. Emotions like disgust, fear, happiness, surprise, sadness and anger, anxiety, boredom, puzzlement and uncertainty were decrypted from FABO database [47], the problem of multi-class classification was solved using a SVM with a RBF kernel as multi-class classifier. The average accuracy raised by the three-fold cross validation was 83.1%. In [48], a Random Forests classifier faced the problem of multivariable time series classification of extracted features from psychological experiments [25]. Features like low-level postural, high-level kinematic and geometric were calculated as well as statistical cues. The experiments over the real-time expressive gesture recognition system achieved an overall recognition rate of 75.41% (138 correctly classified out of 183).

3. Data and Methods

3.1 Data set

We developed a controlled study to evaluate whether our methodology recognizes the emotions in human poses, for this purpose, we used the Xbox 360 Kinect Sensor [49]. In order to capture the body postures, we used the Software Development Kit (SDK) and the HP Laptop Webcam from a HP Envy 17-3077NR. The data was provided by the SDK, that's why, all the data captured in videos are focused on the body postures from the virtualized skeleton, irrespective of the color of the skin or the individual's dress. We performed six sessions with a group of 44 individuals in order to capturing a specific emotion; this gave way to capture the emotions in six different videos. The length was 440 seconds per each emotion, approximately, the time per individual in order to express the emotion took up 10 seconds, as illustrated in Fig.1.

The six universal emotions disgust, sadness, happiness, fear, anger and surprise were captured in six videos with duration of 7 minutes and 20 seconds. The individuals were randomly chosen, with different ethnic groups, sex (20 male and 24 female). During the capture of emotional pose, a diverse set of affective pictures were shown to the individuals in order to manipulate the stimulus (intensity of pleasantness and unpleasantness), [50]. The database was constructed following the background used in experiments related to whole body expressions, based on movements and postures accompanying specific emotions; emotional link between posture and emotions [3], several body poses have been stimulus materials in experiments based on emotions decrypted through emotional-kinetic responses from human body [51], [52], the link between the emotional expression of whole human body and the neural basis [53], autism and normal body perception in patients [54], studies of whole body expression using a stimulus set [55], the classification of seven basic emotional states using a hierarchy of neural detectors to evaluate static views of body poses [56]. The reason for capturing data from individuals was to acquire knowledge about the kinetics of movement related to bodily emotion, the collections of images from different poses does give us an idea of the associate emotion at the time. For example, sadness emotion, depicted in Fig.1, is a group of six images captured in different time spaces, it was represented with head curves to the left or to the right, the shoulders pointing downwards and neck hanging to the left or right, supporting the head, waist unbalanced to the left and bended on itself, motionless, passive, the head hangs on the contracted chest. Anger is depicted with the head of the individual bowed down in a left-backward position, the legs protruding with the knees forward. Arms folded in

IJCSI International Journal of Computer Science Issues, Volume 12, Issue 1, No 1, January 2015
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

11

anguish with shoulders tilted to the left. Surprise is showed with hands pointing forwards, the arms look alert for new things and legs flexibly relaxed. Happiness showed the lower part of the body pruning downwards with the support of the knee pointing forward in an angle. The hands flying upwards, while the shoulders gutting higher to release relief into the face. Disgust showed postures as if to push away or to guard oneself, arms positioned and pressed close to the sides, shoulders raised as when horror is experienced. In fear, the arms are thrown wildly over the head. All the body shrinks, the arms are protruded as if to push away something, raising both shoulders with the bent arms pressed closely against sides or chest. The bodily emotion detection loop starts with the input of a matrix addition, from six different stages of emotion captured in the video. The images pass through a series of phases, turning it into a new matrix addition of new image with the new size of 40 x 30 pixels (that could be named matrix-knowledge). The matrix-knowledge is prepared to the analysis in a classifier, in which the sum of the images represents the knowledge of kinetics behavior of an emotional pose, as illustrated in Fig.2. The set of images are matrices converted to the same dimensions that pass through a series of phases. In order to resize the set of images, we used the Nearest Neighbor Interpolation method per image, because it is very simple and requires less computation, using the nearest neighbor's pixel to fill interpolated point. For each individual image a new value is calculated from a neighborhood of samples and replaces these values in the minimized image; this technique is applied to all the images in the dataset. The matrix-knowledge is the addition of six matrices per individual; all of these matrices are composed by pixels of images captured from the video. The matrix-knowledge are composed of nxm real elements such that matrix-knowledge $\epsilon\ R^{\,nxm}$, in this case the real values for n and m are 40 x 30 pixels respectively as stated earlier. Each position of the matrix-knowledge will become a feature. We clarify that the matrix-knowledge is converted to a row vector of features, whereby each position is a feature, in which the total amount of features will be 1200. The matrix-knowledge and the vector-knowledge are given by Eq.1 and Eq.2.

$$[matrix-knowledge] = \begin{bmatrix} Im1_{11} & \cdots & Im1_{1n} \\ \vdots & \cdots & \vdots \\ Im1_{m1} & \cdots & Im1_{mn} \end{bmatrix} + \begin{bmatrix} Im2_{11} & \cdots & Im2_{1n} \\ \vdots & \cdots & \vdots \\ Im2_{m1} & \cdots & Im2_{mn} \end{bmatrix} +$$
$$+ \begin{bmatrix} Im3_{11} & \cdots & Im3_{1n} \\ \vdots & \cdots & \vdots \\ Im3_{m1} & \cdots & Im3_{mn} \end{bmatrix} + \begin{bmatrix} Im4_{11} & \cdots & Im4_{1n} \\ \vdots & \cdots & \vdots \\ Im4_{m1} & \cdots & Im4_{mn} \end{bmatrix} + (1)$$
$$+ \begin{bmatrix} Im5_{11} & \cdots & Im5_{1n} \\ \vdots & \cdots & \vdots \\ Im5_{m1} & \cdots & Im5_{mn} \end{bmatrix} + \begin{bmatrix} Im6_{11} & \cdots & Im6_{1n} \\ \vdots & \cdots & \vdots \\ Im6_{m1} & \cdots & Im6_{mn} \end{bmatrix}$$

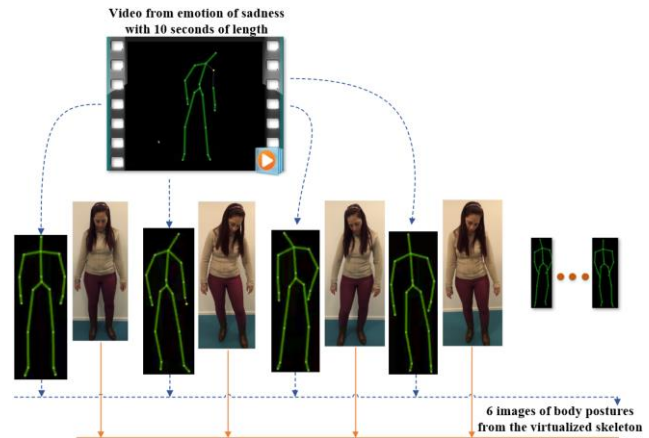$$[vector-knowledge] = [Vec(matrix-knowledge)]^{T} \quad (2)$$
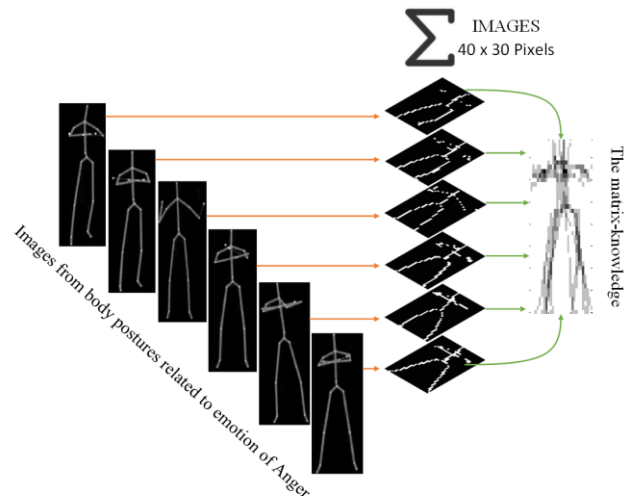

Fig. 1 Sensing of data

Fig. 2 Matrix-knowledge: matrix addition of a new image

3.2 Classifiers

Support Vector Machine (SVM) is a classification algorithm that offers a robust classification to very large number of variables and small samples. Its origins were in statistical learning theory [57]. The SVM can learn both simple and highly complex classification models, applying sophisticated mathematical principles to avoid overfitting. The SVM can be divided into linear and nonlinear the latter being obtained by the introduction of kernel.

IJCSI International Journal of Computer Science Issues, Volume 12, Issue 1, No 1, January 2015
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

12

The most widely used being the polynomial and Radial Basis Function (RBF), as shown in Eq. 3 and Eq. 4 respectively.

$$k(x.x') = \langle x.x' \rangle^d \tag{3}$$

$$k(x.x') = exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \tag{4}$$

Artificial Neural Network (ANN) is a mathematical model classified within connectionist techniques. The ANN tries to replicate the neural processing functions of true neural networks, where neurons set in layers process information. It can be described as a graph of an interconnected group of artificial neurons, with information in the weighs of the arcs that connect the neurons. The topology of an ANN divided the algorithm in two groups: feed-forward and recurrent neural network. The FF networks are supported over a directed acyclic graph, while RR networks have cycles. The Multilayer Perceptron is one of the feedforward algorithms most used, which has a supervised learning algorithm called backpropagation [58]. The learning process involves two steps, the first being a forward processing of input data by the neurons that produces a forecasted output, the other imply the adjustment of weights within the neuron layers, in order to minimize the errors of the forecasted solution compared with the correct output.

Decision tree is another classification model relatively fast compared to other models; it sometimes obtains similar or better accuracy facing others. The algorithm makes simple classification rules that are easy to understand, which represents the information in a tree based in a set of features. The classic decision tree is named ID3 based on growing and pruning [59], although C45 is other top–down decision trees inducers for continuous values [60].

k Nearest Neighbors (kNN) is one of the simplest of classification algorithms available for supervised learning. The algorithm classifies unlabeled examples based on their similarity with examples in the training set. It is a lazy learning method that searches the closest match of the test data in feature space. The most widely used is based on Euclidean metric [61].

The naive Bayes classifier is a supervised learning method as well as a statistical method for classification [62]. The probabilistic classifier is based on the well-known Bayes theorem with strong assumptions; it allows us to capture uncertainty about the model in a principled way by determining probabilities of the outputs. One of the advantages is the robustness to noise in input data. The classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable.

A Bayesian Network (bayesNet) [63] is a graphical model (GMs) for probabilistic relationships among a set of variables, they are used to represent knowledge the uncertainty. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. The random variables are represented by nodes in the graph, and for each node there is a probability table specifying the conditional distribution of the variable given (any possible combination of) the values of its predecessors in the graph. These conditional dependencies in the graph are generally calculated by using known statistical and computational methods.

3.3 Performance measures

Machine learning techniques have several measures in order to evaluate the performance of classifiers, which are principally focused in handling two-class problems. Are more common the classification tasks with more than two classes such as the problem in this research, in which the common measures are around six classes formed by six universal emotions.

Most of the measures to evaluate binary problems could also apply to multi-class problem. In a problem with m classes, the performance of classifiers can be assessed based on an *mxm* confusion matrix. The groups of rows that describe the matrix represent the actual classes, while the columns are the predicted classes. For example, the accuracy is the percentage of correctly classified cases of the dataset. Based on the confusion matrix, the accuracy can be computed as a sum of the main values from the diagonal of the matrix, which represents the correctly classified cases divided by the total number of instances in the dataset (Eq. 5).

$$Accuracy = \frac{\sum_{i=1}^{m} CM_{ii}}{\sum_{i=1}^{m}\sum_{j=1}^{m} CM_{ij}} \tag{5}$$

where CM_{ij} represents the elements in the row i and column j of the confusion matrix.

Some measures like accuracy do not represent the reality of the number of cases correctly classified per each class. In order to make a deeper analysis, the measure of recall

IJCSI International Journal of Computer Science Issues, Volume 12, Issue 1, No 1, January 2015
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

13

has been calculated for each class. Recall provides the percentage of correctness of classification into each class. Eq. 1 represents the recall for class (Eq. 6) [64].

$$Recall_i = \frac{CM_{ii}}{\sum_{j=1}^{m} CM_{ij}} \qquad (6)$$

For the validation stage, a k-fold cross-validation (k = 10) was employed. The goal of this technique is to assess how the model could generalize to an unknown dataset. The dataset is randomly divided into k equal sizes parts or folds; one fold is used as validation set and the remaining k-1 folds as the training set. The process is repeated k times using a different fold as a validation set, this process continues until each fold can used once as validation test. Then, the k results obtained by folds can be averaged to a single result. The advantage of 10-fold cross-validation is that all examples of the database are used for both, training and testing stages [65].

4. Discussion and Results

With the purpose of choosing the best classification for bodily emotional states, six classifiers were tested: Decision tree (J48), Bayes Net (BN), Naive Bayes (NB), Multilayer Perceptron (MLP) and Support Vector Machine (SVM), the last of these with three kernels: linear, polynomial of degree 2 and RBF. First and foremost, the performances of the classifiers selected were validated using 10-fold cross-validation. The accuracy and the recall results were compared.

As you can see in Fig. 3, the accuracy results raised per each classifier show us that the highest performance corresponds to SVM with a linear kernel, which achieves the accuracy of 91.6%. The percentage of most relevant results per emotion positively classified (recall) is depicted in Fig. 4. Emotions like "anger" and "surprise" have achieved the best results using a SVM with linear kernel. Classifiers like BayesNet and Naive Bayes achieved 100% of accuracy in the emotion "sad". The emotion "happiness" using the BayesNet has achieved the maximum percentage in emotion recognition. In case of the multilayer perceptron (MLP), the emotion "disgust" gathered the best results; similarly, the emotion "fearful" has collected the best results with all the group of SVM and MLP previously trained. The average of the results provided by the SVM with linear kernel, SVM with polynomial kernel and MLP have reached a successful of 91.6%, 90.9% and 90.5% respectively. The performances of these methods have not shown significantly differences, but at computational cost level, SVM with linear kernel has less training time. We

can highlight that one of the main current challenges in robotics is to develop feasible models in inexpensive platforms, with its entailed computational limitations.
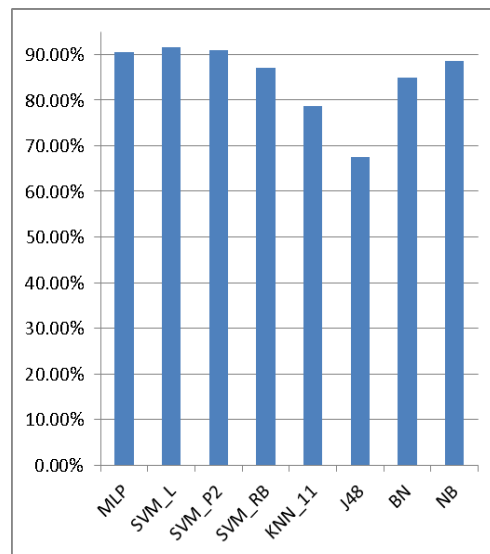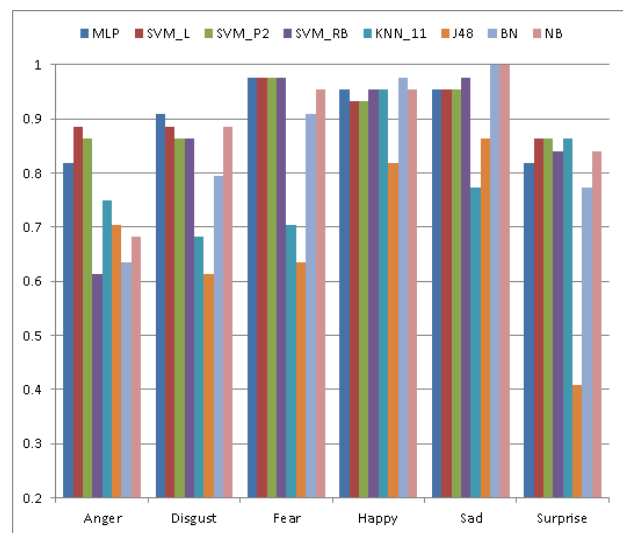

Fig. 3 Accuracy

Fig. 4 Recall

The second analysis is related to the number of images used to construct the matrix-knowledge, in order to describe each case in the database. As previously described, six images were used to build the matrix-knowledge. We decided to use six images after several experiments with trial and error, the percentage of correctly classified instances of more than six images was invariant, just a slight difference over decimals. This difference is partially explained if the classifier made one or two misclassifications, but this is not a big difference. But from the computational standpoint, it would imply less data to

IJCSI International Journal of Computer Science Issues, Volume 12, Issue 1, No 1, January 2015
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

14

process for robot memory. Fig. 5 shows the accuracy of images in the range of 1 to 8. Fig. 6 shows the recall per each emotion, using a different numbers of images in the range of 1 to 6.
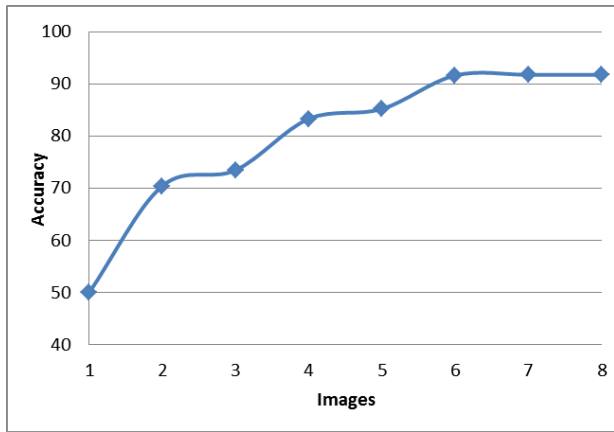

Fig. 5 Accuracy

Grade increments in accuracy and recall due to more images are expected, in this case, the number of images could provide more information to make a prediction. For all the six cases of any emotion it can be shown that the same increase of yield has been the best performance with six images.
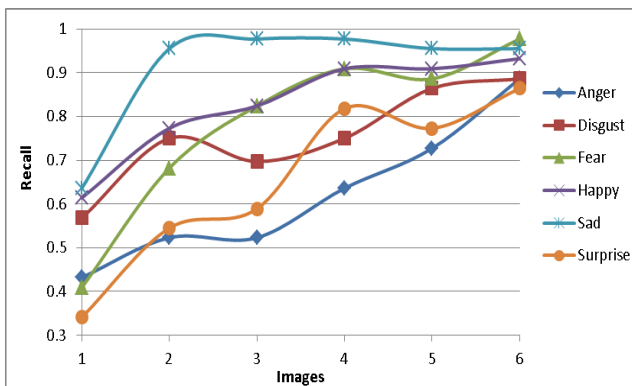

Fig. 6 Recall

We can point out that the emotion "sadness" has raised a good percentage of classification since two images, ahead of the rest of emotions, which have more misclassifications. By increasing the number of images the percentage of the emotion "sadness" is sacrificed to increase the classification of the other emotions. The percentage of

recall is balanced to increasing the number of images as illustrated in Fig. 6.

Table 1: confusion matrix

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	39	4	0	0	0	1
Disgust	5	39	0	0	0	0
Fear	0	1	43	0	0	0
Happy	0	0	0	41	0	3
Sad	0	2	0	0	42	0
Surprise	1	2	0	3	0	38

As mentioned before, the total rate of correct recognitions over all emotions was 91.6% related with the confusion matrix given in Table 1. Some emotions are perfectly classified, while others are simply confused with others. The body language of disgust and anger are most likely to be confused. Some individuals often express both anger and disgust facing the same situation; these two emotions seem to be mixed up with fear. If the individuals express any emotion mixed with sadness it is most probable that the result is disgust, anger, or fear [66]. The body expression in anger and disgust shows similarities [67], e.g., the arms are protruded near to the chest, in case of anger the position is related to a tentative of fight with closed fists, whereas in disgust, the arms are protruded showing a pushing action, these similarities make a misclassification between disgust and anger. The same case is observed between happiness and surprise [68], they have demonstrated three cases of overlap, due to some similarities in upright position from the body and the arms are raised to the sides with forearms straight.

4. Conclusions

The results of this research have clearly showed link between nonverbal bodily behavior and the emotional content. We showed that the emotional triggers systematically affect the patterning of human body poses. The information conveyed by the body modality contains large amounts of emotional data, compared to what has been assumed until now. The results suggest that there may be few emotion specific prototypical patterns of body postures in humans clearly visible and identifiable. We compared the performance of different classifiers in order to select the best result to predict the emotion, based on images that represent bodily poses. The results show that SVM with linear kernel outperforms all the remaining classifiers achieving 91.6% of the accuracy and a range between 86.4 and 97.7 of recall. The features used to describe each case are based on a matrix as the result of

IJCSI International Journal of Computer Science Issues, Volume 12, Issue 1, No 1, January 2015
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

15

combine the pixels matrices of different images. We show that six images can be enough to predict the emotion achieving a high performance and saving memory, thinking about an inexpensive robotic platform. Taking into account that the model could be applied to a low-cost robotic platform that could make decisions with a sufficient computation power and sensing [19], we suggest the use of SVM because of their flexibility, computational efficiency and capacity to handle high dimensional data. The costs directly affects the technology acceptability, thus innovation by using cheaper computer systems, sensors and compute capabilities, are relevant and should be taken into account in the implementation of robotic systems.

Human emotions could be adaptive responses in certain situations or maladaptive in others, would be important define the characteristics of emotionally intelligent of machines with respect to an individual and a social environment. The individual component could be how machines rely on their own emotional reactions as a source of information for the task at hand. The emotional component regarding how machines may handle information detected from individuals to adjust their social behavior. These two characteristics are tightly bound because they are depending directly of the data provided by diverse emotional human sources, and we must clarify that not always machines could acquire the emotional information from all sources.

One potential use of the proposed system is in a context of multi-modal communication to assist the societal participation of persons deprived of conventional modes of communication, e.g., in order to enhance the interaction with deaf people, bearing in mind that in this scenario, the robot cannot handle the audio source.

Acknowledgments

References

[1] L. Nummenmaa, E. Glerean, R. Hari, and J. K. Hietanen, Bodily maps of emotions Proceedings of the National Academy of Sciences, vol. 111, pp. 646-651, January 14, 2014.

[2] S. S. Tomkins, Affect, imagery, consciousness, Springer Verlag, 1962.

[3] C. Darwin, The Expression of the Emotions in Man and Animals. Oxford University Press, USA, 1872.

[4] N. H. Frijda, The Emotions, Cambridge University Press, 1986.

[5] P. Ekman, Should we call it expression or communication? Innovations in Social Science Research, vol. 10, pp. 333-344, 1997.

[6] C. Breazeal, Emotion and sociable humanoid robots, in International Journal of Human-Computer Studies, vol. 59, no. 1-2, pp. 119-155, 2003.

[7] R. W. Picard, Toward computers that recognize and respond to user emotion, in IBM Systems Journal, vol. 39, no. 3-4, pp. 705-719, 2000.

[8] S. Walter, C. Wendt, J. Bohnke, S. Crawcour, J. W. Tan, A. Chan, K. Limbrecht, S. Gruss and H. C. Traue, Similarities and differences of emotions in human–machine and human–human interactions: what kind of emotions are relevant for future companion systems?, Ergonomics, 2013.

[9] K. Church, E. Hoggan and N. Oliver, A study of mobile mood awareness and communication through MobiMood. ACM, pp.128-137, 2010.

[10] R. Likamwa, Y. Liu, N. D. Lane and L. Zhong, Can your smartphone infer your mood?, in Proc. ACM Workshop on Sensing Applications on Mobile Phones (PhoneSense), ACM Press, 2011.

[11] R. Picard, Affective computing: challenges, in International Journal of Human-Computer Studies, vol. 59, no. 1, pp. 55-64, 2003.

[12] D. Goleman, Emotional Intelligence. Bantam Books, New York, 1995.

[13] H. K. M. Meeren, C. van Heijnsbergen and B. de Gelder, Rapid perceptual integration of facial expression and emotional body language, Proc. National Academy of Sciences of the USA, vol. 102, no. 45, pp. 16518-16523, 2005.

[14] B. de Gelder, Towards the neurobiology of emotional body language, Nature Reviews Neuroscience, vol. 7, no. 3, pp. 242-249, 2006.

[15] A. Metallinou, A. Katsamanis and S. Narayanan, Tracking changes in continuous emotion states using body language and prosodic cues, in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2288-2291, 2011.

[16] P. Ekman, Differential Communication of Affect by Head and Body Cues, Journal of Personality and Social Psychology, vol. 2, no. 5, pp. 726, 1965.

[17] B. de Gelder, From body perception to action preparation: a distributed neural system for viewing bodily expressions of emotion, in People Watching: Social, Perceptual, and Neurophysiological Studies of Body Perception, eds Shiffrar M., Johnson K., editors. New York, NY: Oxford University Press, pp. 350-368, 2013

[18] A. Gibaldi, M. Chessa, A. Canessa, S. P. Sabatini, and F. Solari, A neural model for binocular vergence control without explicit calculation of disparity, in ESANN, 2009.

[19] J. G. Rázuri, P. G. Esteban, D. R. Insua, An adversarial risk analysis model for an autonomous imperfect decision agent, in T. V. Guy, M. Kárný and D. H. Wolpert, Eds. Decision Making and Imperfection. SCI, vol. 474, pp. 165-190. Springer, Heidelberg, 2013.

[20] E. Mueller, M. Dyer, Day dreaming in humans and computers, in Proceedings of the Ninth International Joint Conference on Artificial Intelligence, CA: Los Angeles, 1985.

[21] R. Picard, Affective Computing, United States: the MIT Press, 1998

[22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, The WEKA data mining software: an update, ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10-18, 2009.

[23] L. Kessous, G. Castellano and G. Caridakis, Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. Journal on Multimodal User Interfaces vol. 3, Issue. 1-2, pp. 33-48, 2010.

[24] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber and T. Poggio, Robust object recognition with cortex-like mechanisms, IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 29, no. 3, pp. 411-426, 2006.

[25] H. G. Wallbott, Bodily expression of emotion. European Journal of Social Psychology, vol. 28, no. 6, pp. 879-896, 1998.

[26] M. Meijer, The contribution of general features of body movement to the attribution of emotions. Journal of Nonverbal Behavior, vol. 13, no. 4, pp. 247-268, 1989.

[27] T. R. Boone and J. G. Cunningham, Children's expression of emotional meaning in music through expressive body movement. Journal of Nonverbal Behavior, vol. 25, no. 1, pp. 21-41, 2001.

[28] A. Furnham, Body language at work. Channel Islands: The Guernsey Press, 1999.

[29] A. Mehrabian, Silent messages: Implicit communication of emotions and attitudes (2nd ed.). Belmont, CA: Wadsworth, Inc., 1981.

[30] J. Montepare, E. Koff, D. Zaitchik, and M. Albert, The use of body movements and gestures as cues to emotions in younger and older adults. Journal of Nonverbal Behavior, 1999.

[31] A. Kleinsmith, P. R. de Silva, and N. Bianchi-Berthouze, Crosscultural differences in recognizing affect from body posture, Interacting with Computers, vol. 18, no. 6, pp. 1371-1389, 2006.

[32] M. Shiffrar and J. Pinto, The visual analysis of bodily motion, in Prinz, W., Hommel, B. (eds.) Common mechanisms in perception and action: Attention and Performance, pp. 381-399. Oxford University Press, Oxford, 2002.

[33] T. R. Boone and J. G. Cunningham, Children's decoding of emotion in expressive body movement: the development of cue attunement. Developmental psychology vol. 34, pp. 1007-1016, 1998.

[34] K. Schindler, L. van Gool, and B. de Gelder, Recognizing emotions expressed by body pose: a biologically inspired neural model, Neural Networks, vol. 21, no. 9, pp. 1238-1246, 2008.

[35] A. Kleinsmith and N. Bianchi-Berthouze, Recognizing affective dimensions from body posture, in Affective Computing and Intelligent, Lecture Notes in Computer Science, pp. 48–58, Springer, Berlin, Germany, 2007.

[36] M. A. Giese and T. Poggio, Neural mechanisms for the recognition of biological movements. Nature Reviews Neuroscience, vol. 4, pp. 179-192, 2003.

[37] G. Rizzolatti, L. Fogassi and V. Gallese, Mirrors in the mind. Scientific American , vol. 295, pp. 54-61, 2006.

[38] S. Mota and R. Picard, Automated Posture Analysis for Detecting Learner's Interest Level, Proc. Computer Vision and Pattern Recognition Workshop, vol. 5, p. 49, 2003.

[39] S. D'Mello and A. Graesser, Automatic Detection of Learner's Affect from Gross Body Language, Applied Artificial Intelligence, vol. 23, pp. 123-150, 2009.

[40] G. Castellano, S. Villalba, and A. Camurri. Recognising human emotions from body movement and gesture dynamics. In ACII'07: Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction, pp. 71-82, Berlin, Heidelberg, Springer-Verlag, 2007.

[41] G. Castellano, M. Mortillaro, A. Camurri, G. Volpe, and K. Scherer, Automated analysis of body movement in emotionally expressive piano performances. Music Perception, vol. 26, no. 2, pp. 103-119, 2008.

[42] H. Gunes and M. Piccardi, Bi-modal emotion recognition from expressive face and body gestures. Journal of Network and Computer Applications , vol. 30, pp. 1334-1345 , 2007.

[43] A. Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis, and P. Driessen, Gesture-based affective computing on motion capture data. In ACII'05: Proceedings of the First International Conference on Affective Computing and Intelligent Interaction, pp. 1-7, Springer, 2005.

[44] H. Gunes and M. Piccardi, Affect recognition from face and body: early fusion vs. late fusion. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, vol. 4, pp. 3437-3443, 2005.

[45] S. Piana, A. Stagliano, A. Camurri and F. Odone, A set of full-body movement features for emotion recognition to help children affected by autism spectrum condition. In IDGEI International Workshop, 2013.

[46] S. Saha, S. Datta, A. Konar and R. Janarthanan, A study on emotion recognition from body gestures using Kinect sensor, International Conference on Communications and Signal Processing (ICCSP), pp. 056-060, Jadavpur Univ., Kolkata, India, 2014.

[47] H. Gunes, C. Shan, S. Chen and Y. Tian, Bodily Expression For Automatic Affect Recognition, Advances in Emotion Recognition. Wiley-Blackwell, pp. 1-31, 2012.

[48] W. Wang, V. Enescu and H. Sahli, Towards Real-time Continuous Emotion Recognition from Body Movements. Human Behavior Understanding, in Human Behavior Understanding, vol. 8212, pp. 235-245, 2013.

[49] T. Leyvand, C. Meekhof, Y. C. Wei, J. Sun, and B. Guo, Kinect identity: Technology and experience, Computer, vol. 44, no. 4, pp. 94-96, 2011.

[50] U. Schimmack, Attentional interference effects of emotional pictures: Threat, negativity, or arousal? Emotion, vol. 5, pp. 55-66, 2005.

[51] R. L. Bannerman, M. Milders B. de Gelder and A. Sahraie A, Orienting to threat: Faster localization of fearful facial expressions and body postures revealed by saccadic eye movements. Proceedings of the Royal Society of London B: Biological Science, vol. 276, pp. 1635-1641, 2009.

IJCSI International Journal of Computer Science Issues, Volume 12, Issue 1, No 1, January 2015
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

17

[52] R. L. Bannerman, M. Milders and A. Sahraie, Attentional cueing: Fearful body postures capture attention with saccades, Journal of Vision, vol. 10, no. 5, pp. 1-14, 2010.

[53] M. E. Kret and B. de Gelder, Social context influences recognition of bodily expressions, Exp Brain Res vol. 203, no. 1, pp. 169-180, 2010.

[54] N. Hadjikhani and B. de Gelder, Seeing fearful body expressions activates the fusiform cortex and amygdala, Current Biology, vol. 13, pp. 2201-2205, 2003.

[55] M. V. Peelen, A. P. Atkinson, F. Andersson, and P. Vuilleumier, Emotional modulation of body-selective visual areas, Social Cognitive and Affective Neuroscience, vol. 2, pp. 274-283, 2007.

[56] K. Schindler. L. Van Gool and B. de Gelder, Recognizing emotions expressed by body pose: a biologically inspired neural model, Neural Netw vol. 21, no. 9, pp. 1238-1246, (2008).

[57] V. Vapnik, The Nature of Statistical Learning Theory ed.; Springer-Verlag, New York, 1995.

[58] D. E. Rumelhart, G. E. Hinton and R. J. Williams, in Parallel distributed processing: explorations in the microstructure of cognition; D. E. Rumelhart and J.L. McClelland, eds, MIT Press: Cambridge, MA, USA, vol. 1, pp. 318-362, 1986.

[59] J. R. Quinlan, Induction of decision trees. Mach. Learn, vol. 1, no. 1, pp. 81-106, 1986.

[60] J. R. Quinlan, C4.5: Programs for Machine Learning, 1st ed.; Morgan Kaufmann Publishers: San Francisco, CA, USA, 1993.

[61] T. M. Mitchell, Machine Learning; McGraw-Hill: New York, NY, p. 432, 1997.

[62] H. Zhang, The Optimality of Naive Bayes, Proc. the 17th International FLAIRS conference, Florida, USA, pp. 17-19, 2004

[63] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.

[64] P. Flach, Machine Learning: The Art and Science of Algorithms that Make Sense of Data, Cambridge University Press, 2012.

[65] B. Efron and R. J. Tibshirani, An introduction to the Bootstrap. Chapman & Hall: New York, USA, 1993.

[66] S. E. Duclos, J. D. Laird, E. Schneider, M. Sexter, L. Stern, and O. Van Lighten, "Emotion-specific effects of facial expressions and postures on emotional experience," Journal of Personality and Social Psychology, vol. 57, p. 100, 1989.

[67] P. Ekman, Basic emotions, in T. Dalgeish & M. Power (Eds.), Handbook of cognition and emotion, pp. 45-60, New York: John Wiley & Sons, 1999.

[68] M. Coulson, Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence, Journal of Nonverbal Behavior, vol. 28, no. 2, pp. 117-139, 2004.

Javier G. Rázuri is currently a researcher at the Department of Computer and System Sciences, Stockholm University. He was a researcher and Phd student of Department of Statistics and Operations Research at Universidad Rey Juan Carlos and Acting developer of the Affective computing department in AISoy Robotics. In 1995, he received with honors the Bachelor of electronic engineering degree at Peruvian University of Applied Sciences (UPC) in Perú. In 2007, he received a Master's Degree in Business Management at IEDE - Business School of European University of Madrid. In 2008, he received a Master's Degree in Decision Systems Engineering at Universidad Rey Juan Carlos. From 2008, he was a researcher in several projects financed for the university and the European Union focused on Robotics and decision-making related to EU higher education. His research interests are related with, Affective computing, emotional decision making, Human-machine interaction, Robotics, Human Robotics, Robotics and Autonomous Systems, Neuroscience, Sentiment analysis, in order to reproducing behavior patterns similar to human and provide to the agents a type of emotional intelligence and improve the interaction experience making more close the loop human-robot emotional interaction.

Aron Larsson has his PhD degree in Computer and Systems Sciences and is currently a researcher at the Department of Information and Media, Mid Sweden University as well as the Department of Computer and System Sciences, Stockholm University. His main research interest is the use of computer-based decision analysis and process models in complex decision making in which risks, uncertainties and trade-offs exist. Aron is the coordinator for the DECIDE Research Group at Stockholm University and leads the research RDALAB at Mid Sweden University. He has developed and applied risk and decision analytical methods for the paper industry, municipal decision-making problems, international mine clearance efforts and disaster preparedness. Aron is also active in the spin-off company Preference AB, which maintains the decision analysis software DecideIT.

Rahim Rahmani received his MSc in Electrical Engineering, wireless communication from Mid Sweden University in 1997. He worked as junior lecturer at Mid Sweden University and from 2007 he received 50% time for his study toward to PhD, he earned a technical doctorate in computer science from Mid Sweden University in March 2010. He is currently a researcher at the Department of Computer and System Sciences, Stockholm University.

David Sundgren received his Doctoral degree in Computer and Systems Sciences from Stockholm University in 2011. He worked as lecturer at University of Gavle. He defended the thesis on "Apparent Arbitrariness of Second-Order Probability Distributions", dealing with uncertain probabilities. He is currently a researcher at the Department of Computer and System Sciences, Stockholm University.

Isis Bonet Cruz received the B.Sc. degree in Computer Science from the Universidad Central "Marta Abreu" de Las Villas (UCLV), Santa Clara, Cuba, in 2001, her M.Sc. degree in Computer Science at UCLV in 2005 and her Ph.D. in Technical Sciences at UCLV in 2009. She is currently a researcher at Antioquia School of Engineering (EIA), Envigado, Colombia. She has authored/coauthored for some 42 papers in conference proceedings and scientific journals and earned several awards including the Cuban Academy of Sciences Award in 2011. Her research interests include Artificial Intelligence, Neural Networks, Classification Problems, Bioinformatics and Business Intelligence.

Antonio Moran Cardenas has his PhD degree in Engineering from Tokyo University of Agric. and Technology, Japan. His research interests include intelligent systems design based on the integration of neural networks, fuzzy logic and genetic algorithms applied to autonomous control engineering, bio-engineering, robotics, systems modeling, optimization and other applications.