Development of a Facial Expression Recognition and Imitation Method for Affective HRI

F. Cid, J.A. Prado, P. Bustos and P. Nuñez

***Abstract*—This paper presents a system for recognizing and imitating facial expressions using visual information acquired by a robot. The proposed approach is capable of estimating the emotion state of a human interlocutor through the recognition of facial expressions (i.e. happiness, sadness, anger, fear, neutral) using a Bayesian approach, which is achieved in real time. This information updates the knowledge of the robot about the people in its field of view, and thus, allows the robot to use it for future actions and interactions. In this paper, the human facial expression is imitated by Muecas, a 12 degree of freedom (DOF) robotics face. This paper also introduces the concept of the human and robot facial expressions models, which are included inside of a new cognitive module that builds and updates selective representations of the robot and the agents in its environment for enhancing future HRI. Experimental results demonstrate the quality of the detection and imitation using different scenarios with Muecas avatar.**

***Index Terms*—Facial Expression Recognition, Imitation, Human Robot Interaction.**

I. INTRODUCTION

Human Robot Interaction (HRI) is one of the most important task in social robotics. In the last decades, HRI has become an interesting researching topic where different non-trainer users interact with robots in real scenarios. Most of the HRI methodologies use the non-invasive techniques based on natural language (NL), in a similar way that people interact in their daily life. In this regards, verbal communication (speech, among others) or non-verbal communication (corporal language, gestures or facial expressiveness) have been successfully used for enhancing the empathy, the attention or the understanding of the social skills in a human-machine interaction [1], [2].

Social robots are usually designed in order to enhance the empathy and the attention of the communication [3]. Thus, human shaped robots are typically used for improving the communication, decreasing the gap between the machine and the human being behaviors. Besides, it allows the robot to adapt itself to the emotional state of the human interlocutor, which could be used for different purposes in a social affective communication. In order to have an efficient HRI, not only the robot shape is important, but also the knowledge of different elements of the human interlocutor state: pose in the environment, number of interlocutors in the scenario or the emotional state, among others. To acquire this information, several techniques and methodologies have been studied and applied, such as facial expression recognition [13], skeletal modeling [15], use of corporal language [16] or speech recognition [17].

F. Cid and P. Bustos and P. Nuñez are with University of Extremadura.
E-mail: fcidb@alumnos.unex.es
J.A. Prado is with University of Coimbra, Portugal
E-mail: jaugusto@isr.uc.pt

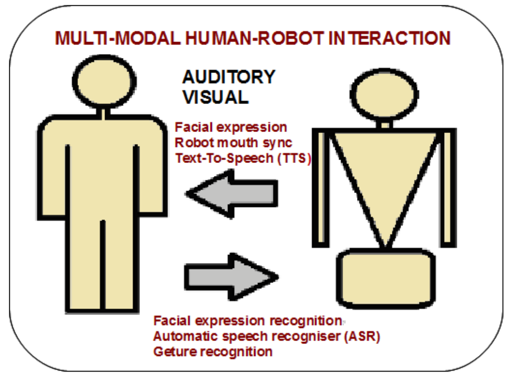

Fig. 1. Multi-modal HRI is usually based on visual and auditory information.

Therefore, in order to interact with people, robot have to be able to perceive and share information with them using visual and auditory messages. Natural language, in conjunction with visual information is a very efficient method for an interaction paradigm with robots (see Fig. 1). On one hand, facial expression recognition provides an estimate of the interlocutor's emotional state through the understanding of the visual information, providing support to the emotional responses of a robot inside a social dialog through audio media or visual aids creating a feedback for the content of the dialog [8]. In fact, interactive NL-based communication provides a fast feedback that is successfully used for handling errors and uncertainties. On the other hand, human behavior imitation has been used for learning tasks and for enhancing the human-robot communication. Thus, imitation of emotions plays an important role in cognitive development and has been studied in the last year in social robotics [7], [6]. Both, visual and auditory information, are used for mimicking human expressions as a mean of developing social and communication skills. Into social robots, the robotic heads (e.g. Kismet[9], Saya [10] or WE-4RII [18]) mainly imitates facial expressions, through the modification of the positions of the mobile elements as the eyes or mouth.

The proposed approach presents a facial expression recognition system that allows to detect and recognize four different emotions (happiness, sadness, anger and fear) besides of the neutral state. This system is based on a real-time Bayesian classifier where visual signal is analyzed in order to detect the expressiveness of the interlocutor. Besides, an imitation system is development and presented in this approach, where a robotic expressiveness model is used as a bridge between the human expressiveness and the final robotic head. This model is part of a new cognitive module that is able to build selective representations of the self, the environment and the agents in it. Finally, a set of experiments using a robotic head Mueca's avatar has been achieved in order to present and comment the results of the recognition and imitation systems.

This paper is organized as follows: In Section II, previous works in facial expression recognition and imitation are briefly described. Next, Section III presents the emotional state models associated to both interlocutors, robot and human, which are integrated inside the cognitive architecture for the proposed social robot. In Section IV, an overview of the proposed recognition and imitation system is presented. In Section V, experimental results are pointed out, and finally, Section VI describes the conclusions and future work of the presented approach.

II. Previous Works

To achieve affective Human-Robot Interaction, this paper primarily focuses on presenting different methodologies commonly used for facial expression recognition and imitation. The automatic recognition of emotions is necessary multimodal, that is, it requires of verbal and non-verbal channels (face, gesture, body language), physiological signals or midterm activity modeling, among others [13], [20], [19]. One of the most significant works used by the scientific community in facial expression recognition using visual information is based on Paul Ekman's study [11], [12]. This author identifies and classifies the facial expressions through the study of different facial muscles in each expression, giving rise to so-called Facial Action Coding System(FACS). The recognition of facial expressions is a very diversified field in its classification or detection methods, ranging from the use of active Appearance Models (AAM)[21], Support Vector Machines(SVM)[6], Gabor filter bank [22] and Dynamic Bayesian Network (DBN).

On the other side, several authors use robots in domestic environments with untrained users or people with disabilities [4], [5]. In these works, authors achieve a natural HRI through the generation of facial expressions by the robot with the goal of maintaining a level of empathy and emotional attachment to robots [3]. These facial expression and emotion generation methods differs in the amount of facial expression that are possible to generate by the robot due to physical constraint [18]. In robotic heads with human-like characteristics such as the robotic head used in this paper, different works provide solutions for emotion generation depending on their physical constraints [9].

Finally, robot's capability of imitating facial expressions determines the design of the heads used in social robotics. Usually, imitation of facial expressions is achieved through mobile elements of the head (e.g. eyelids, eyebrows, eyes or mouth) [6], [18].

III. Emotional state modelling

The proposed approach is part of a new robotics cognitive architecture that builds selective representations (i.e. models) of the robot, the environment and the agents in it.

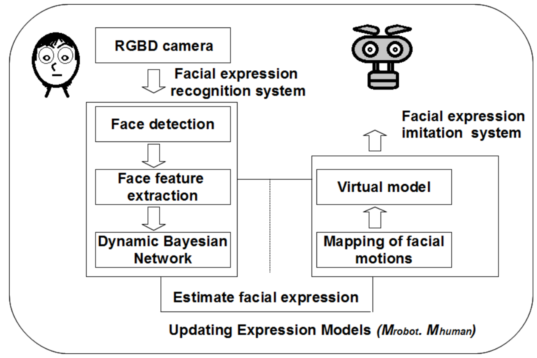

Fig. 2. Overview of the facial expression recognition and imitation system proposed in this paper.

This cognitive architecture performs internal simulations over these models to anticipate the outcome of future actions and interactions (e.g. navigation, path-planning, grasping of objects or more complex interactions). Model-based representations of reality to help social robots achieve their tasks has been used in the last years with interesting results [15]. In order to achieve a affective HRI, non-contact interaction is modelled in the cognitive architecture, including gesture and facial expression recognition, and detection of human emotional state. This last model is presented in this paper. Thus, human and robot emotional state models are similar and defined as: $M_{\{robot,human\}} = \{(m_0,p_0),(m_1,p_1),...(m_i,p_i)\}$, where m_i represents an emotional state $m_i = \{happy, sad, anger, fear, neutral\}$ and p_i the probability of this emotional state $0 \leq p_i \leq 1$. Both models M_{robot} and M_{human} will be updated once the facial expression and emotional state have been estimated by the proposed system.

IV. Facial Expression Recognition and Imitation System

In this paper, a real-time facial expressions recognition and imitation system is presented. This system will be integrated inside a cognitive architecture as a new module that provides a representation of the agent and robot's emotional states. The proposed approach is described in Fig. 2. The robot acquires the information using a RGBD (RGB and distance) Kinect sensor. This measurement is preprocessed in order to estimate the pose of the face in the robot's surrounding. Then, once the region of interest (i.e human face) is detected, the system extracts facial features for the later classification task. This is achieved using a Dynamic Bayesian Network (DBN), allowing to recognize the emotional state associated to the facial expression. In the next stage, the system updates the emotional state of the agents in the communication (self and interlocutor emotional state models), and finally, the facial expression is played by the robotic head Muecas' avatar.

A. Facial expression recognition system

When designing the classifier, we made our own interpretation of FACS (Facial Action Coding System) [11], [12], which drives a set of random variables different from those defined t

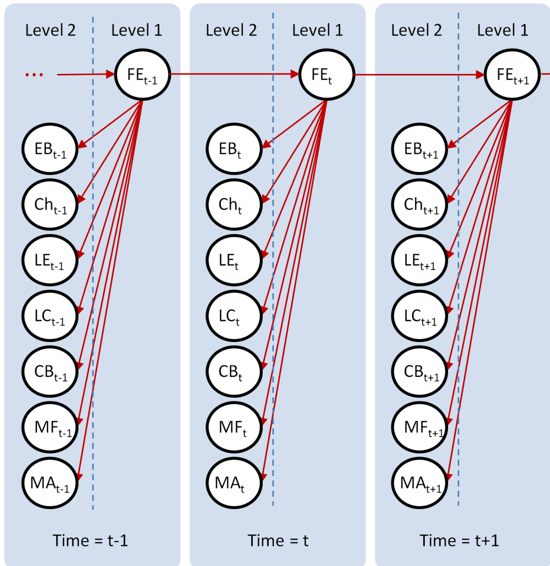

Fig. 3. Facial Expression Dynamic *Bayesian network,*three time intervals are shown $(t-1,\ t,\ t+1)$.

by other researchers. Each facial expression is composed by a specific set of Action Units. Each of these Action Units is a distortion on the face induced by small muscular activity. Normally, a well determined set of face muscles is associated to a specific Action Unit, which can give the idea that all these basic distortions are independent. Nevertheless, some of these Action Units are antagonistic. One clear and understandable example is the case of two Action Units related with the lips corners: AU12 and AU15. When performing AU12, lips corners are pulled up. Oppositely, when performing the AU15, the lip corners are pulled down. Therefore, if by one way the movements of the lip corners can be considered independent because they are performed by distinct muscle sets, by another, when analyzed visually they are antagonistic, and exclusive.

We assumed that the state space is discrete, and in this case, hidden Markov models (HMM) can be applied. A hidden Markov model can be considered to be an instantiation of a dynamic Bayesian network and though exact inference is feasible. Based in these principles, belief variables were defined and a dynamic Bayesian classifier of facial expressions was developed.

Facial Expression dynamic Bayesian network: We took advantage of the antagonism extant in some AUs to reduce the size of the dynamic Bayesian network. Though, instead of using the 11 AUs as leafs for our DBN (Dynamic Bayesian Network), we propose 7 (seven) variables. These variables groups the related antagonist and exclusive Action Units. The structure of the network of two levels is illustrated in figure [fig:DBN-facial-expressions], also in this figure, the time influence that characterizes this network as a dynamic Bayesian network is represented.

In the *dynamic Bayesian network*'s first level there is only one node. The global classification result obtained is provided by the belief variable associated to this node: $FE \in \{Anger,\ Fear,\ Sad,\ Happy,\ Neutral\}$, where the variable name stands from Facial Expression. Considering the structure of the dynamic *Bayesian network*, the variables in their second level have as parent this one in the first level: FE.

In the second level there are seven belief variables:

- $EB \in \{AU1,\ AU4,\ none\}$ is a belief variable related with the *Eye-Brows* movements. The events are directly related to the existence of AU1, and AU4.
- $Ch \in \{AU6,\ none\}$ is a belief variable which is related with *Cheeks* movements; more specifically, the events indicate if the cheeks are raised (AU6 is performed).
- $LE \in \{AU7,\ none\}$ is a belief variable which is related with the *Lower Eyelids* movements; AU7 is associated to lower eyelids set to up.
- $LC \in \{AU12,\ AU15,\ none\}$ is the belief variable associated with the movements of the *Lips Corners*. When the corners did not perform any movement then the event *none* has a high probability. The event *AU12* has a big probability when the corners of the lips are pulled up. If the lip corners moves down the event *AU15* must have a big probability.
- $CB \in \{AU17,\ none\}$ is the belief variable collecting the probabilities related with the *Chin Boss* movements. The event *none* is related with the absence of any movement, while the event *AU17* had a great probability when the chin boss is pushed upwards.
- $MF \in \{AU20,\ AU23,\ none\}$ is the belief variable associated with the Mouth's Form. The events *AU20* and *AU23* indicated, respectively, if the mouth is horizontally stretched or tightened.
- $MA \in \{AU24,\ AU25,\ none\}$ is the belief variable associated with the Mouth's Aperture. The events *AU24* and *AU25* are related, respectively, with lips pressed together or with lips relaxed and parted.

The movements performed by the human in one area of the face can slightly affect muscles on other area, however, this influence is very small and cannot be detected by the cameras of the robot. Thus, conditional independence among the 7 proposed variables was assumed.

The following equations illustrate the joint distribution associated to the Bayesian Facial Expressions Classifier.

$$\begin{aligned}&P(FE,EB,Ch,LE,LC,CB,MF,MA)=\\&P(EB,Ch,LE,LC,CB,MF,MA|FE)*P(FE)=\\&P(EB|FE)*P(Ch|FE)*P(LE|FE)*P(LC|FE)*\\&P(CB|FE)*P(MF|FE)*P(MA|FE)*P(FE)\end{aligned} \tag{1}$$

The last equality is written assuming that the belief variables in the second level of the dynamic *Bayesian network* are independent.

From the joint distribution, the *posterior* can be obtained by the application of the Bayes rule as follows:

$$\begin{aligned}&P(FE|EB,Ch,LE,LC,CB,MF,MA)=\\&P(EB|FE)*P(Ch|FE)*P(LE|FE)*P(LC|FE)*\\&\quad P(CB|FE)*P(MF|FE)*P(MA|FE)*P(FE)/\\&\quad P(EB,Ch,LE,LC,CB,MF,MA)\end{aligned} \tag{2}$$

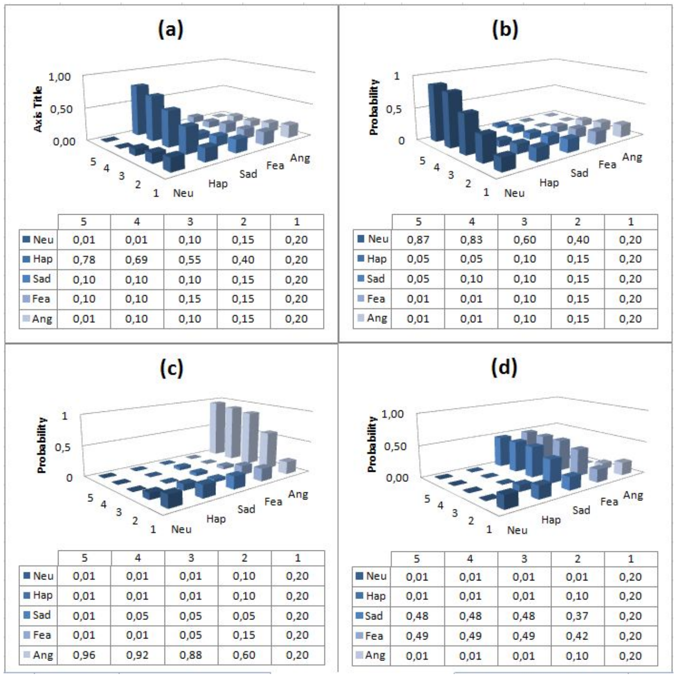

Fig. 4. Results from facial expression classifier.

From the Bayesian marginalization rule we can calculate:

$$P(EB, Ch, LE, LC, CB, MF, MA) = \sum_{FE} P(EB|FE) * P(Ch|FE) * P(LE|FE) * P(LC|FE) * P(CB|FE) * P(MF|FE) * P(MA|FE) * P(FE) \quad (3)$$

As a consequence of network to be dynamic, convergence happens along the time, the resultant histogram from the previous frame is passed as prior knowledge for the current frame. We limited the maximum number of frames for convergence as 5. If the convergence reach to 80% before 5 frames, the classification is considered complete (Fig 4). If not it keeps converging up to the fifth frame. If the fifth frame is reached and no value is higher than 80%, the classifier selects the highest probability value (usually refered as the maximum a posteriori decision in Bayesian theory) as a classified result.

In figure 4, camera grabbing was set to 5 fps, therefore, the iteration axis represents the 5 utterances that happens inside one second. The expression axis is the selected scope of possible expressions. Notice that the sum of probability at each iteration among the five possible expressions is always 1. In examples (a), (b) and (c), respectively, inputs were given for happy, neutral and fear; the dynamic Bayesian network was capable of classifying the expected expression with a fast convergence. In (d), an example of ambiguity and misclassification is shown, where the expected result was sad but the result of classification was fear.

B. Facial expression imitation system

Imitation is a key process in several social robotics applications as a mean of developing social and communication skills (e.g learning or gesture imitation in the context of HRI). In the most of facial expression mimicking approaches, visual and auditory information are used for achieving a multi-modal imitation system. In addition, it has been demonstrated that a more realistic communication derive from a robotic head with similar characteristic than human face [9]. Thus, a facial expression imitation system is described in this paper, where visual information is used to perform non-verbal communication in a more friendly and intuitive way using the robotic head Muecas (Fig 5a). with a graphical representation using a ”Muecas” virtual model (avatar) (Fig 5b). Besides, an mesh model is able to be used for allowing the robot to be pro-active, by interpreting sensory information to predict the immediately relevant future inside the cognitive architecture.

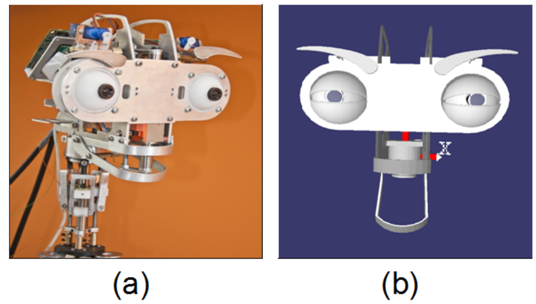

Fig. 5. b) 12DOF Robotics head Muecas; and a) Robotics Head Mueca’s avatar.

1) Robotic head Muecas: The robotic head Muecas consist of 12 DOF and it has been designed by Iadex S.L in cooperation with RoboLab as a mean to transmit expressions for social robots [25][1]. One of the main goals in the design of Muecas was to imitate human emotional states according to the generation of facial expressions. Thus, the movement of the elements present in the recognition of facial expressions (e.g. eyes, eyebrows or mouth, among others) is similar to those of the human face, resulting in simplest and most natural imitations.

Muecas has also its own virtual model, which consists of 16 DOF, with four degrees more that the real robotic head (Eyelids). besides, the mesh model of the robotic head ”Muecas” is used as a bridge between the facial expression estimated by the system and the emotion reproduced by the robotic head, performing the necessary retargeting. That is, before generating facial expression in the real robotic head, first the system tries to generate all the cinematic chain of the mechanical motions and graphic representation of each expression imitated by the ”Muecas” avatar.

2) Facial Expression Generation: Facial expressions are detected and recognized using the recognition system described in Section IV-A. Four different emotional states are estimate (i.e. Happiness, sadness, fear and anger), also a neutral state (i.e. no expression associated with an emotion). Fig. 6a illustrates the facial expressions estimate by the recognition system for different examples. These facial expressions are then mapped over mesh model, modeling each one of the

[1] For more information, you can visit www.robolab.unex.es

Emotion	AUs	Muecas' component
Neutral	-	-
Happy	AU6-AU12-AU25	Eyebrows-Eyelids-Eyes-Mouth
Sad	AU1-AU4-AU15-AU17	Eyebrows-Eyelids-Eyes
Fear	AU1-AU4-AU20-AU25	Eyebrows-Eyelids-Mouth
Anger	AU4-AU7-AU17-AU23-AU24	Eyebrows-Eyelids

TABLE I
MOVEMENTS FOR THE ROBOTIC HEAD MUECAS' COMPONENTS ASSOCIATED TO THE EMOTION RECOGNITION

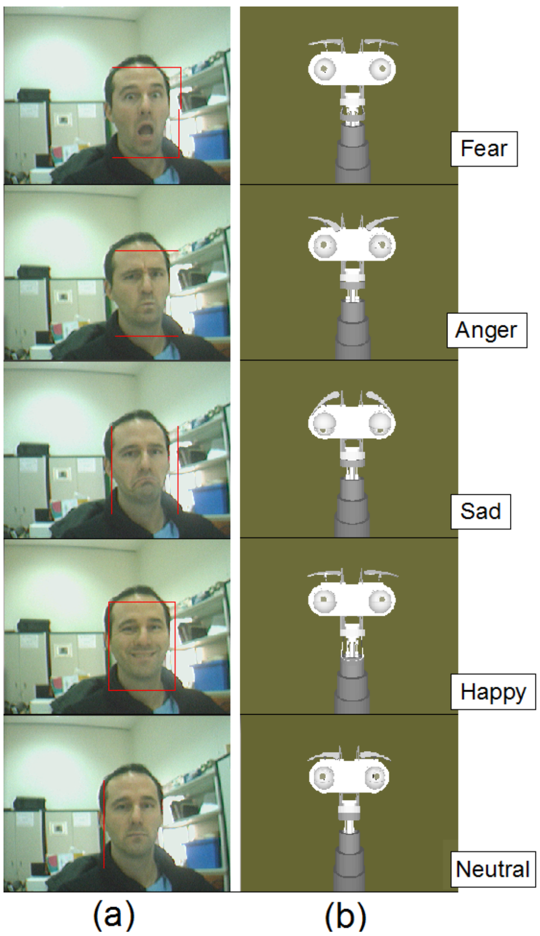

Fig. 6. a) Facial expression estimated by the recognition system; and b) Facial expressions imitated on the Muecas' avatar.

movements needs to generate the emotional state. Table I describes the set of mobile elements of the robotic head and the AUs for each emotion. In Fig. 6b the facial expressions generated by the mimicking system are illustrated using the virtual model of the robotic head.

V. EXPERIMENTAL RESULTS

In this section, a set of test has been achieved in order to evaluate the effectiveness of the facial expression recognition and imitation system described in this paper. The software to control the system is built on top of the robotics framework *RoboComp* [23]. Making use of the components and tools it provides and its communication middleware, an easy to understand and efficient architecture has been developed.

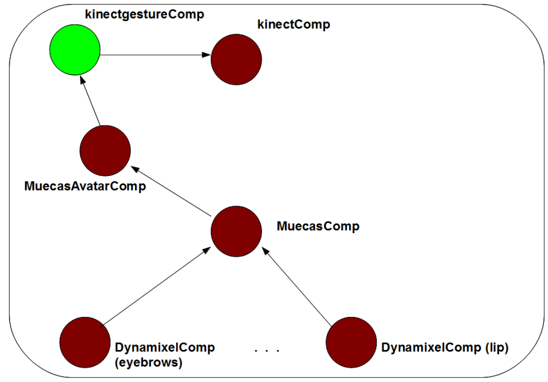

Fig. 7. Dependence Relationships between the different components used in the proposed approach.

The relationships between the different components used for the experimental setup have been drawn in Fig. 7. The main component of the proposed system is *kinectgestureComp*. It is connected, directly or indirectly, to the rest of the software components, such as: *kinect* or *robotic head* among others (In the figure, not all components of the robotic head Muecas have been drawn to provide a simple explanation). The RGBD image is provided by the *KinectComp* component, which sends it to the *kinectgesturecomp* component that estimate the interlocutor's facial expressions. This component also updates the robot and human emotional state models and assigns the motion of each mobile element of the robotic head in order to generate a realistic facial expression. Then, *MuecasavatarComp* is used as a bridge between the facial expression recognition and the final imitation over the robotic head Muecas (*MuecasComp*). Once the robotic head receives the positions of each mobile element, each motor commands are received and executed by its associated *dynamixelComp*.

Since the system was designed an implemented using component oriented design/programming, these components can be easily used for other purposes, which is a very important feature in robotics development.

In order to evaluate the recognition and imitation system, a set of experimental tests has been achieved in a real HRI scenario. A human interlocutor is located opposite to the robot, achieving different facial expressions (from sadness to happiness) in a continue mode. The proposed system is running on-line, acquiring and estimating the facial expressions in real-time. Then, the system updates the emotional state models (M_{robot}, M_{human}) and imitates the facial expression using Muecas in real-time too. These experiments are run 20 times with different interlocutors and generating different facial expressions. An example of the results are shown in the Table. II, where the evolution of each p_i is given. Robustness of the approach is given in Table II for the set of experiments achieved in the described scenario. As is shown in Table II, the most of the facial expressions are correctly estimate.

VI. CONCLUSION

In this paper, an approach to facial emotion state recognition and imitation in image sequences has been presented. First, a

TABLE II
ROBUSTNESS OF THE FACIAL EXPRESSION RECOGNITION SYSTEM

Test	Percent of correctly detected facial expression (p_i)
sad	74%
happy	89%
fear	95%
Anger	79%

Dynamic Bayesian Network (DBN) structure has been used to classify facial expressions (happiness, sadness, anger, fear and neutral). This paper demonstrate the robustness of the solution for a common HRI scenario with different users and environmental conditions. Next, this facial expression is imitated in the robotic head Muecas, which has been designed for generating emotions. The full system has been incorporated in a social robot whose cognitive architecture has been pointed out in this paper. Thus, the robot and human emotional states are updated and tracking by the architecture in order to plan future actions and interactions.

Future works are focused on a multi-modal interaction, where auditory information (e.g. speech or intensity) will be used in order to estimate the interlocutor's emotional state. This new module will be integrated in the architecture, taking into account the probabilities associated to each one of these emotional states. Besides, to achieve an affective HRI it will be interesting to study the empathy level of the presented solution in real scenarios with non-trainer scenarios.

ACKNOWLEDGMENT

This work has been partially supported by the Spanish Ministerio de Ciencia e Innovación TIN2011-27512-C05-04 and AIB2010PT-00149, and by the Junta de Extremadura projects IB10062. The authors also gratefully acknowledge support from Institute of Systems and Robotics at University of Coimbra (ISR-UC), Portuguese Foundation for Science and Technology (FCT) [SFRH/BD/60954/2009].

REFERENCES

[1] A. Paiva, J. Dias, D. Sobral, R. Aylett, P. Sobreperez, S. Woods, C. Zoll and L. Hall, "Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents", In *Third International Joint Conference on Autonomous Agents and Multiagents Systems*, Vol. 1, pp. 194-201, New York, USA, 2004.

[2] M. Siegel, C. Breazeal, and M. I. Norton, "Persuasive Robotics: The influence of robot gender on human behavior". In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2563-2568, October 2009.

[3] A. Tapus and M.j. Mataric, "Emulating Empathy in Socially Assistive Robotics", In *AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics*, Stanford, USA, March 2007.

[4] M.J. Matarić , J. Eriksson , D.J. Feil-Seifer1 and C.J. Winstein. "Socially assistive robotics for post-stroke rehabilitation". In *Journal of NeuroEngineering and Rehabilitation*. 4:5. 2007.

[5] C. Jayawardena, I. H. Kuo, U. Unger, A. Igic, R. Wong, C. I. Watson, R. Q. Stafford, E. Broadbent, P. Tiwari, J. Warren, J. Sohn and B. A. MacDonald. "Deployment of a Service Robot to Help Older People". In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems* Taiwan, pp. 5990-5995, October 2010.

[6] S. S. Ge, C. Wang, C.C. Hang, "Facial Expression Imitation in Human Robot Interaction", In *Proc. of the 17 IEEE International Symposium on Robot and Human Interactive Communication*, Germany, pp. 213-218, 2008.

[7] S. DiPaola, A. Arya, J. Chan, "Simulating Face to Face Collaboration for Interactive Learning Systems", In *In Proc: E-Learn 2005*, Vancouver, 2005.

[8] T. Chen, "Audio-Visual Integration in multimodal Communication". In *IEEE Proceedings*, May, 1998.

[9] Kismet, Available at: *http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html*.

[10] T. Hashimoto, S. Hitramatsu, T. Tsuji, and H. Kobayashi, "Development of the Face Robot SAYA for Rich Facial Expressions". In *Proc. of 2006 SICE-ICASE International Joint Conference* Korea, pp. 5423-5428, October 2006.

[11] P.Ekman, WV Friesen, JC Hager, "Facial Action Coding System FACS", the manual, 2002.

[12] P.Ekman, E. Rosenberg, "What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS)", 2nd edn. Oxford Press, London.

[13] J.A. Prado, C. Simplcio, N.F. Lori and J. Dias, "Visuo-auditory Multimodal Emotional Structure to Improve Human-Robot-Interaction", In *International Journal of Social Robotics*, vol. 4, no. 1, pp. 29-51, December 2011.

[14] W. Zhiliang, L. Yaofeng, J. Xiao, "The research of the humanoid robot with facial expressions for emotional interaction". In Proc. *First International Conference on Intelligent Networks and Intelligent Systems*, pp. 416-420. 2008.

[15] J.P. Bandera, "Vision-Based Gesture Recognition in a Robot Learning by Imitation Framework", Ph.D. Thesis, University of Malaga, 2009.

[16] A. Aly and A. Tapus, "Speech to Head Gesture Mapping in Multimodal Human-Robot Interaction", In *Proc. of the 5th European Conference on Mobile Robots ECMR 2011* Sweden, pp. 101-108, September 2011.

[17] C. Breazeal and L. Aryananda, "Recognition of Affective Communicative Intent in Robot-Directed Speech", Artificial Intelligence, pp. 83-104, 2002.

[18] M. Zecca, T. Chaminade, M. A. Umilta, K. Itoh, M. Saito, N. Endo, "Emotional Expression Humanoid Robot WE-4RII- Evaluation of the perception of facial emotional expressions by using fMRI", In *Robotics and Mechatronics Conference ROBOMEC2007*, Akita, Japan, pp. 2A1-O10, 2007.

[19] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann and S. Narayanan, "Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information", In *Proc. of ACM 6th International Conference on Multimodal Interfaces (ICMI 2004)*, 2004.

[20] Z. Zeng, M. Pantic, G. I. Roisman and T. Huang, "A Survey of Affect Recognition Methods: Audio, Visual and Spontaneous Expressions", In textit IEEE Transactions on Pattern Analysis and Machine Intelligence,Vol. 31, pp. 39 - 58, 2008.

[21] K.-E. Ko, K.-B. Sim, "Development of a Facial Emotion Recognition Method based on combining AAM with DBN", in *International Conference on Cyberworlds 2010*, pp. 87-91 ,2010.

[22] H.-B. Deng, L.-W. Jin, L.-X. Zhen and J.-C. Huang. "A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA", International Journal of Information Technology. Vol.11, no.11. pp. 86-96, 2005.

[23] L.J. Manso, P. Bachiller, P. Bustos, P. Nuñez, R. Cintas and L. Calderita. "RoboComp: a Tool-based Robotics Framework". In *Simulation, Modeling and Programming for Autonomous Robots* (SIMPAR). Pages 251-262. 2010.

[24] Intel open source computer vision library, Available at: *http://www.intel.com/technology/computing/opencv*.

[25] Available at: *http://iadex.es*.