2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

A Real Time and Robust Facial Expression Recognition and Imitation approach for Affective Human-Robot Interaction Using Gabor filtering

Felipe Cid, José Augusto Prado*, Pablo Bustos and Pedro Núñez

Abstract— **Facial expressions are a rich source of communicative information about human behavior and emotion. This paper presents a real-time system for recognition and imitation of facial expressions in the context of affective Human Robot Interaction. The proposed method achieves a fast and robust facial feature extraction based on consecutively applying filters to the gradient image. An efficient Gabor filter is used, along with a set of morphological and convolutional filters to reduce the noise and the light dependence of the image acquired by the robot. Then, a set of invariant edge-based features are extracted and used as input to a Dynamic Bayesian Network classifier in order to estimate a human emotion. The output of this classifier updates a geometric robotic head model, which is used as a bridge between the human expressiveness and the robotic head. Experimental results demonstrate the accuracy and robustness of the proposed approach compared to similar systems.**

I. INTRODUCTION

In the near future, a variety of different robots will inhabit human environments, such as homes or offices. Current robots' abilities for an efficient communication with people in these real environments (e.g., facial expressiveness, body language or gestures) are very limited; nevertheless, these communication skills are the main key in the acceptance of robots by next human generations. Nowadays, the design of social robots takes into account how to enhance the empathy and the attention [1], and also, how to include affective components in Human-Robot Interaction (HRI). Some design alternatives, such as human shaped robots, decreases the gap between robot and human during social interactions. Such user-friendly designs were extensive explored on the development of platforms for affective HRI (*e.g.*, Kismet [2], Saya [3] or WE-4RII [4] robotics heads). In this respect, facial expression recognition and mimicry play an important role in human interaction and non-verbal communication.

In order to have an efficient affective interaction, not only the robot shape is determinant, but also the robot ability to acquire knowledge about the context or the emotional state of the user. Faces are rich and powerful sources of communicative information about human behavior and emotion. Thus, a large number of facial expression recognition systems are used in the literature as a first stage in affective HRI [5]. These systems provide support to the emotional responses of a robot inside a social dialogue through audio signals or visual aids, creating a feedback for the content of the dialogue [6].

Human behavior imitation has been used for learning tasks and for enhancing human-robot communication [7]. From a communication theoretical perspective, mimicry systems have been interpreted as revealing information to define and reinforce the relationship between individuals [9]. For the robot to express a full range of emotions and to establish a meaningful communication with a human being, facial expressions are vital.

The proposed approach presents a real time emotion recognition and imitation system based on facial expression analysis. On one hand, the facial expression recognition system consists of a robust feature extraction algorithm, which consecutively applies morphological and convolutional filters to reduce the noise and the dependence against changes of luminosity. After, an efficient Gabor filter is used for efficient edge detection. The output edge image of this bank of filters is used for detecting and extracting scale-invariant facial features. Contrary to other approaches, these facial features are a combination of independent and antagonistic distortions of the face and constitute the input of a Dynamic Bayesian Network (DBN) used as classifier [8]. Four different emotions as well as a non-emotional state are detected using this Bayesian approach (happiness, sadness, anger, fear and neutral). On the other hand, an imitation system is development and presented in this work, where a robotic head model is used as a bridge to directly map from the detected emotion to the robot's actuators in safety conditions. This model is part of a cognitive module that is able to build selective representations of the self, the environment and the agents in it [10]. Finally, a set of experiments using the 12 Degrees Of Freedom (DOF) robotic head *Muecas* has been achieved in order to present and comment the results of the recognition and imitation systems.

This paper is organized as follows: after discussing known approaches to facial expression and imitation techniques in Section II, Section III presents an overview of the proposed recognition and imitation system, which are described with more detail in Section IV and Section V, respectively. In Section VI, the experimental results are pointed out, and finally, Section VII describes the conclusions and future work of the presented approach.

II. RELATED WORKS

Automatic recognition of emotions has been studied in the last years by several authors. Current literature describes a complete system as multi-modal, that is, a recognition system

F. Cid, P. Bustos and P. Núñez are members of Robotics and Artificial Vision Lab. *Robolab* Group, University of Extremadura, Spain (e-mail: fcid,pbustos,pnuntru@unex.es)

*J.A. Prado is with the Institute of Systems and Robotics, Dept. Electrical and Computer Engineering, University of Coimbra, Portugal (e-mail: jaugusto@isr.uc.pt)

978-1-4673-6357-0/13/$31.00 ©2013 IEEE

that uses different information sources: face, gesture, body language, speech or physiological signals, among others [11], [12]. However, the majority of the approaches are based on the analysis of facial expressions using visual information. An interesting and updated review is shown in [13]. Common frameworks are based on the Facial Action Coding System (FACS) proposed by Ekman et al's [14]. In these works, facial expressions are assigned to a small set of six prototypical expressions conveying the basic emotions.

Once a relevant face in the image is detected, two main problems arises in classical facial expression recognition systems: i) extracting the facial features; and ii) classifying the feature-based facial expressions into different categories. Detection and classification of facial features is a very diversified field in its classification ranges from the use of Active Appearance Models (AAM) [15], Support Vector Machines (SVM) [17] or Gabor filter banks [16]. Gabor filters have been commonly used for directly extracting features for recognition. However, it is computationally expensive. In [8] a method for detecting and classifying facial expressions was proposed. It used color information and analysis of edges in order to extract facial features and use them as input in a Dynamic Bayesian Network. Facial features in this work were very unstable and dependent of light conditions. However, authors claim that the algorithm detects and recognizes four different facial expression in real time. The proposed approach is inspired by this work, with the results improving not only in robustness and accuracy, but also in processing time.

Several authors use robots in domestic environments with untrained users or people with disabilities [18], [19], [20]. In these works, their authors achieve natural HRI by making the robot generate facial expressions with the goal of maintaining a level of empathy and emotional attachment [1]. 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 constraints [4]. In robotic heads with human-like characteristics such as the head used in this paper, different works provide solutions for emotion generation depending on their physical constraints [21], [2]. This paper includes a model of the robotics head, as a module of a cognitive architecture that among other actions, to prevent risk situations.

III. System Overview

This paper presents a real time and robust facial expression recognition and imitation system. An overview of the proposed approach is illustrated in Fig. 1, which flows from left to right. Given an input video sequence S, the algorithm detects a relevant face in the robot's surroundings using Viola and Jones' method [22]. Once the region of the face R has been estimated, the next step divides it into two different regions of interest, R_{top} and R_{bottom}. The proposed method incorporates a series of steps chosen to counter the effects of illumination variations, local shadowing and highlights. Next, morphological and convolutional filters, followed by a Gabor filter, are also used to reduce image noise. The facial feature extraction step detects and extracts invariant features from the edge images $f_i \in F$. These features and their time evolution, are used as input to a Dynamic Bayesian Network, which classifies them into an emotion (happiness, sadness, anger, fear and neutral). Finally, the detected facial expression is mapped into the robot face to generate an imitation behavior *Muecas*.

IV. Facial Expression Recognition System

This section describes the facial expression recognition system presented in this paper in more detail. The algorithm uses the RGB sequence acquired by the robot and processes each image to detect a set of robust and invariant features in the user's face. These features are used as input to the Bayesian classifier proposed in [8]. Next, each stage illustrated in Fig. 1 is described.

A. Face detection

Let S being a sequence of RGB images acquired by the robot in a real interaction. A frame $I(t)$ of this sequence is then processed in order to detect a relevant face in the image. Viola and Jones' method is used in the proposed approach. This method uses Haar-like features and a cascade of simple classifiers [22] to detect the face in the image in real-time. The relevant region R is then normalized to a fixed size and this image constitutes the input of the next stage.

B. Region of interest definition and pre-processing

Once the face has been detected, this image is processed in order to remove noise and improve its light dependence. Besides, computational complexity is reduced by dividing the region R in two sub-regions of interest, R_{top} and R_{bottom} and converting the image to gray scale. These two regions are used to extract invariant facial features in the human face. In the proposed approach, the nose is irrelevant for facial expression recognition and it is removed from R. Thus, R_{top} and R_{bottom} are associated to the upper and lower region of the face, respectively. Let R being the face image of size N x M, and let $p = (u, v)$ being the central pixel of this image, which is the approximated position of the nose in the image. Then R_{top} and R_{bottom} are defined as selective copies of R as follows: R_{top} of size N x $(v - U_{Th})$ and R_{bottom} of size N x $(v + U_{Th})$), where U_{Th} is an user-fixed threshold.

Next, the new image is converted to gray scale and processed. First, the proposed method incorporates a series of steps chosen to counter the effects of illumination variations while still preserving the main elements for use in recognition. The method is based on the approach described in [23]. The processing sequence follows a set of consecutive stages: i) gamma correction; ii) Difference of Gaussian (DoG) Filtering; iii) Masking; and iv) Contrast equalization.

Once the light dependence has been reduced, a set of morphological and convolutional filters is used. Noise in this image is associated with facial hair, wounds or similar elements, and it is mitigated applying Median, Blur and Gaussian filters, consecutively. The final image is processed using a Gabor filter, which improves the edge-based facial feature detection.

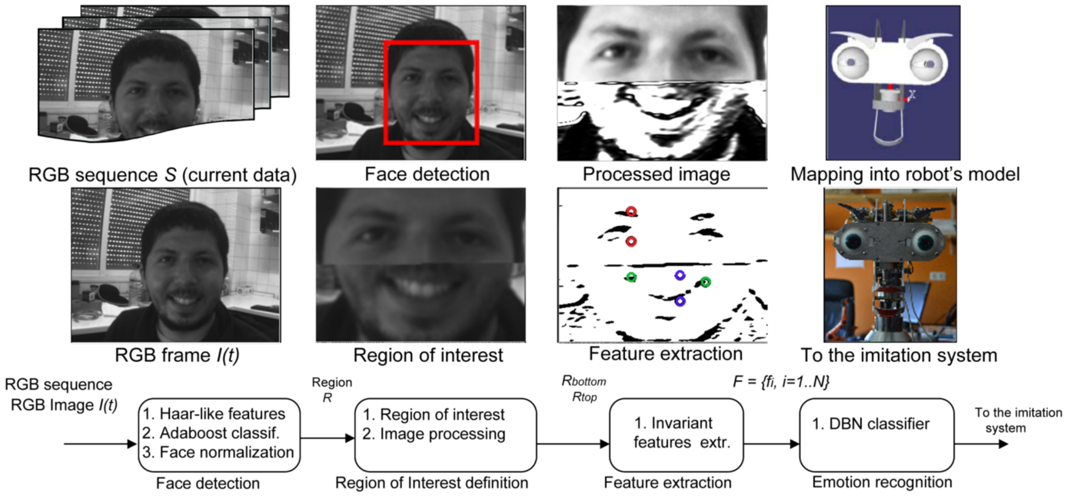
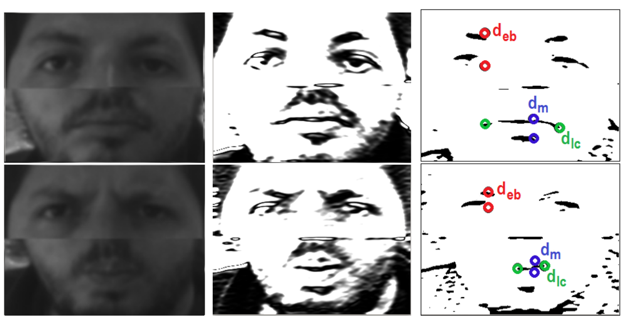


Fig. 1. Overview of the proposed methodology. The approach flows from left to right. See the text for more details.



(a) (b) (c)

Fig. 2. Edge-based facial feature extraction: a) Region of interest in the face image (R_{top} and R_{bottom}); b) image after the pre-processing stage and Gabor filtering; and c) Features extracted in the image.

C. Gabor Filtering

Gabor filter is a linear filter usually very effective and fast in the detection of edges with different orientations. This filter is used here as a previous stage in the detection of facial features, which are extracted using the contours of the eyes, mouth or eyebrows in the face. Gabor impulse response in the spatial domain consists of a sinusoidal plane wave of some orientation and frequency, modulated by a two-dimensional Gaussian envelope. Let $I(u, v)$ be the input image, then the output of the Gabor filter, $G(u, v)$, is given by:

$$G(u, v) = exp(-\frac{1}{2}(\frac{u_\theta^2 + v_\theta^2}{\sigma^2}))cos(2\pi\frac{u_\theta}{\lambda} + \psi), \quad (1)$$

where θ, λ and ψ are associated to the sinusoidal plane wave (orientation, wavelength and phase, respectively), and u_θ and v_θ being described as:

$$\begin{aligned} u_\theta &= ucos\theta + vsin\theta \\ v_\theta &= -usin\theta + vcos\theta \end{aligned} \quad (2)$$

Fig. 2(a) illustrates two different facial images. These images are processed according to the method described in this section. Results after applying light normalization and noise removal methods are shown in Fig. 2(b).

D. Invariant feature extraction

Feature extraction is a crucial step in facial expression recognition systems. The method described in this paper looks for a set of invariant edge-based features in the image F=$\{f_i \mid i = 1..N\}$, that is, features that are independent from the image scale or the distance between the user and the robot. Action Units (AU) are the basis of the proposed feature extraction algorithm. Each of these Action Units is a distortion on the face induced by small muscular activity, as it was described by the Facial Action Coding System (FACS). In contrast with other approaches, in this paper only a set of independent and antagonistic AUs are used (*e.g.*, AU12 and AU15 in Fig. 3 are related to distortions in the lip corners, and they are antagonistic and independent). Only three features are defined in the edge face image, labelled as d_{eb} (red), d_{lc} (green) and d_m (blue) in Fig. 2(c), associated to the Euclidean distances between the upper contour of the eyebrows and the lower edge of the eyes, lip corners and upper and lower contour of the mouth, respectively. These values are easily detected analysing the output of the Gabor filter, as illustrated in Fig. 2(c). In order to become independent to different scales or distances to the user these edge-based features are always normalized using the values extracted in a neutral state.

E. Dynamic Bayesian Network for Facial Expression Recognition

The DBN takes advantage of the existing antagonism in some AUs to reduce the size of the dynamic Bayesian network. Thus, instead of using the 11 AUs as leafs for the DBN (Dynamic Bayesian Network), 7 variables are proposed as combinations of d_{eb}, d_{lc} and d_m. These variables group together the related antagonistic and exclusive Action Units. For the correct detection of these variables, the AUs associated with each variable must have a minimum intensity

Fig. 3. Action Units (AUs) used in this paper.

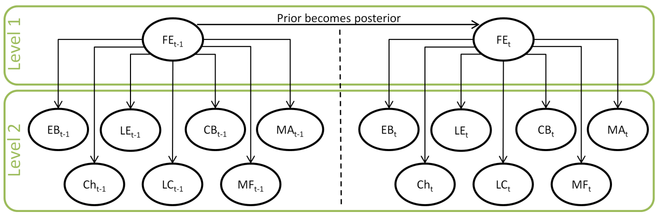


Fig. 4. Facial Expression Dynamic *Bayesian network, two time intervals are shown.*

level B ($i.e.$, slight evidence), within the intensity scale of the AUs (Intensity range $A - E$; where A is commonly referred has a trace, and E has the maximum evidence) [14]. The two-level network structure and the time influence that characterizes this network as a DBN is also represented in Fig. 4. In order to classify the Facial Expression (FE) produced by the user, the overall classification result achieved is the one foreseen by the belief variable FE, in the scope $(FE_{neutral}, FE_{happines}, FE_{sadness}, FE_{fear}, FE_{anger})$.

Bayesian networks need to be supplied with learning data. The most common approach is to use a threshold to find the matches when computing the probability of a new sample. In this approach, to avoid the extant gaps, a pre-processing stage is done before the learning stage, fitting a Gaussian distribution to the data. The learning data was collected via supervised learning; seven random variables (leaf variables of the proposed DBN) were gathered and manually related to the correct classification result.

The leaf random variables of our model, and it's respective virtual-scopes are:

- EB: $\{AU1,\ AU4,\ none\}$; stands for *Eye-Brows*
- Ch: $\{AU6,\ none\}$; stands for *Cheeks*
- LE: $\{AU7,\ none\}$; stands for *Lower Eyelids*
- LC: $\{AU12,\ AU15,\ none\}$ stands for *Lips Corners*
- CB: $\{AU17,\ none\}$ stands for *Chin Boss*
- MF: $\{AU20,\ AU23,\ none\}$ stands for *Mouth's Form*
- MA: $\{AU24,\ AU25,\ none\}$ stands for *Mouth's Aperture*

In this approach, these 7 leaf variables are assumed to be independent given the facial expression (FE). Although some muscular movements from one area of the face may slightly affect other areas, this small influence could not be detected by the cameras of the robot.

Thus, after the feature extraction process, the data (D) is obtained according to the following set up:

$$D = ((x_1, y_1)...(x_n, y_n)), x_i \in \mathbb{R}^d, y_i \in \mathbb{R} \tag{3}$$

Consider that y_1 to y_5 are the five possible emotional states $(FE_{neutral}, FE_{happines}, FE_{sadness}, FE_{fear}, FE_{anger})$; and each dimension of x, corresponds to one of the previously described random variables, namely: EB, Ch, LE, LC, CB, MF and MA. Since the learning data may have gaps between its samples, a model is built assuming that $(X_1, ..., X_n)$ are independent given FE, and

$$X_i \sim N(prior^T x_i, \sigma^2) \tag{4}$$

At first, $prior \sim U(1/n)$, however throughout the iterations, the posterior of $t-1$ becomes the prior on t.

Finally, by using Bayes's rule, we have the posterior equation:

$$P(FE|x_m) = \frac{\prod_1^n P(x_i|FE) * P(FE)}{P(x_m)}, \tag{5}$$

where x_m is the most recent sensory data acquired.
The last dividend can be computed using the Bayesian marginalization rule:

$$P(x_m) = \sum_{FE} \prod_1^k P(x_i|FE) * P(FE), \tag{6}$$

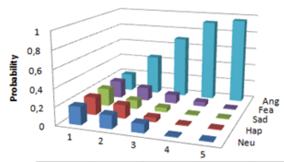
with $k = 7$, the number of random variables of the system.

This model includes a dynamic convergence property over time. The resulting histogram from the previous frame is used as prior knowledge for the current frame. The maximum number of frames for convergence has been limited to 5. If the convergence reaches a 80% threshold before 5 frames, the classification is considered complete (Fig 5). If not, it keeps converging up to the fifth frame. If the fifth frame is reached and no value is higher than the threshold, the classifier selects the highest probability value (usually referred to as the *Maximum a posteriori decision* in Bayesian theory) as the classification result. The threshold is used as a control measure for the classification errors generated in the detection of the Action Units (AUs). Thus, for high intensity values in each AUs submitted by the user, the system presents probability values that are more accurate and robust in every emotional state, reducing the rate of detected errors in the classification.

V. Facial Expression Imitation System

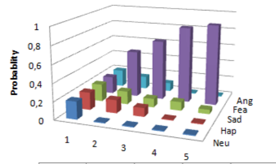
In this section, the proposed facial expression imitation system is described. The outline of the proposal is illustrated in Fig 6. Once the DBN recognizes the facial emotion, it is imitated by the robotics head *Muecas*, which has been designed and built in collaboration with the Spanish start-up Iadex. *Muecas* has 12 DOF's and it has been designed following the main articulated parts of the human anatomy to facilitate the generation of facial expressions[1]. Before mapping the facial emotion to the set of robot's actuators, the imitation system uses an internal simulator over the

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



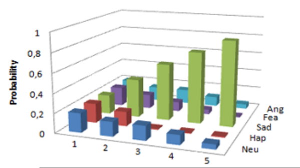
	1	2	3	4	5
Neu	0,2	0,15	0,1	0,01	0,01
Hap	0,2	0,15	0,05	0,01	0,01
Sad	0,2	0,1	0,05	0,01	0,01
Fea	0,2	0,15	0,1	0,05	0,01
Ang	0,2	0,45	0,7	0,92	0,96

(a)



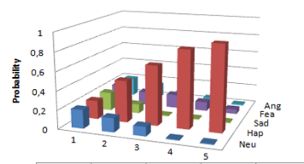
	1	2	3	4	5
Neu	0,2	0,01	0,01	0,01	0,01
Hap	0,2	0,15	0,1	0,01	0,01
Sad	0,2	0,15	0,1	0,1	0,05
Fea	0,2	0,54	0,69	0,87	0,92
Ang	0,2	0,15	0,1	0,01	0,01

(b)



	1	2	3	4	5
Neu	0,2	0,15	0,15	0,1	0,05
Hap	0,2	0,15	0,01	0,01	0,01
Sad	0,2	0,4	0,59	0,74	0,88
Fea	0,2	0,15	0,1	0,05	0,01
Ang	0,2	0,15	0,15	0,1	0,05

(c)



	1	2	3	4	5
Neu	0,2	0,15	0,1	0,01	0,01
Hap	0,2	0,45	0,64	0,83	0,91
Sad	0,2	0,1	0,01	0,01	0,01
Fea	0,2	0,15	0,15	0,1	0,05
Ang	0,2	0,15	0,1	0,05	0,01

(d)

Fig. 5. Results from the classifier for different facial expressions.

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

TABLE I

MAPPING OF FACIAL EXPRESSIONS (AUS) TO THE ROBOT'S MECHANICAL COMPONENTS.

mesh model of the robotics head. Thus, assuming that all the characteristics of each mechanical element are modelled, the system looks for possible collisions among them. In this respect, the system generates the whole kinematic chain associated to the facial expression, and if there is a collision, a retargeting of each mobile component is done to regenerate the kinematic chain.

Table I describes the relation between the AUs and the mobile elements of the mesh model of the robotic head *Muecas* (*i.e.*, mapping) associated to each facial expression. Fig. 7(a) depicts the facial expression recognized by the system (e.g. Fear). This facial expression is previously mimicked by the robotic head model (Fig. 7(b)) and by *Muecas*, after determining a collision-free configuration (Fig. 7(c)).

VI. EXPERIMENTAL RESULTS

In this section, a set of tests has been achieved in order to evaluate the performance of the proposed facial expression recognition and imitation system. Robustness and effectiveness are evaluated. Besides, a comparative study with the the method proposed in [10] has been done, where edge-based features using color analysis were proposed. Our method has been evaluated using a real video sequence. The algorithms were developed in C++ and the benchmarks were

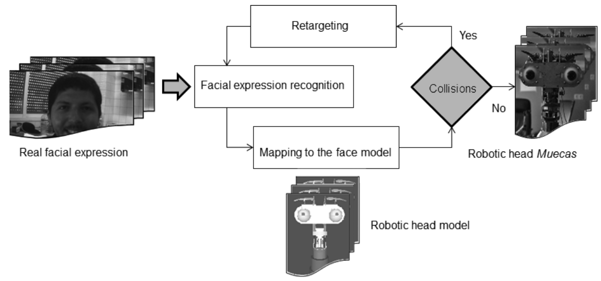
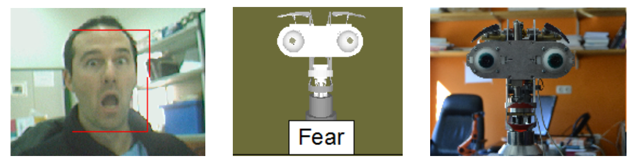


Fig. 6. Description of the imitation system proposed in this paper.



(a) (b) (c)

Fig. 7. a) Facial expression estimated by the system; b) Mapping of the facial expression into $Muecas$; and c) Imitation of the facial expression.

performed on a computer with a 2.8 GHz Intel(R) Core(TM) i7 CPU and 4Gb RAM running using GNU/Linux Ubuntu 10.10. Real data was acquired using a Firewire camera at 25 fps. The software to control the system is built on top of the RoboComp [24] robotics framework. By building on the set of components provided and on its communication middleware, an efficient and easy to understand architecture has been developed.

The proposed system is running on-line, acquiring and estimating the facial expressions in real-time. Thus, the system updates the emotional state at 12 or 24 milliseconds (processing time of the recognition system), and imitates the facial expression using the robotic head *Muecas* in real-time. The experiments consist of two tests, which were performed with untrained users in uncontrolled environments for quantifying the effectiveness of the proposed system. For each test, the user performed a series of continuous and random facial expressions. Besides, to verify the robustness of the system, the tests were performed with different lighting conditions and users of different facial features.

The first test measured the robustness of the facial expression recognition system. A set of 20 interlocutors with different gender and facial features was used, mainly students and researchers with an estimated age between 20 and 50 years. Each interlocutor generated 5 random sequence of emotions. Ground truth (*i.e.,* real facial expression) was selected by an expert. Meanwhile, a third person monitored the correct detection of emotional states by the system through the images and data saved in each experiment. The percentages of correctly detected facial expressions (r) are shown in Table II.

The second test was designed to compare the proposed method with the system described in [10]. Identical conditions were used for both methods, including the same database composed of 18 interlocutors of different gender,

Test	Percentage of correctly detected facial expressions (r)
Sadness	91%
Happiness	94%
Fear	96%
Anger	95%
Neutral	96%

TABLE II

ROBUSTNESS OF THE FACIAL EXPRESSION RECOGNITION SYSTEM

Test	Percentage of improvement in the detection of FE (p_c)
Sadness	18.4%
Happiness	4.2%
Fear	15.1%
Anger	11.1%
Neutral	32.4%

TABLE III

PERCENTAGE OF IMPROVEMENT OF THE FACIAL EXPRESSION RECOGNITION SYSTEM

age and facial features. The rate of correct detections by each system were compared, using an evaluation methodology identical to that of the previous test. Results of this comparison are shown in Table III, where the percentages of improvement in the detection of each facial expression p are illustrated. These results demonstrate an improvement in each detected emotion, especially in the neutral state. Furthermore, the proposed system improves the previous method in several aspects such as: i) better performance in the detection of emotion; ii) less error correction among states; iii) recognition of the neutral state expressions; iv) greater accuracy in the recognition of facial features and v) the use of a smaller amount of data during the training.

VII. CONCLUSIONS

In this paper, a real-time and robust facial expression recognition and imitation system for robotics head has been proposed. The facial expression recognition stage is based on the use of invariant features of the edge image, which are robustly extracted by way of a set of consecutive filters. Gabor filtering is used for an effective edge detection in the face image. Next, a Dynamic Bayesian Network is used for classifying these invariant features into an emotional state. The output of the Bayesian classifier is imitated by the robotic head *Muecas*. The system described in this paper was tested in order to verify the robustness, accuracy and improvement in respect to other approaches for facial expression recognition, using different users, environments and lighting conditions.

Future work will be focused on multi-modal interaction, where auditory information (*e.g.*, speech) will be used to estimate the emotional state. Besides, the proposed method will integrate RGBD information acquired by a low-cost range sensor to estimate the facial expressions more accurately.

ACKNOWLEDGMENT

This work has been partially supported by the MICINN Project TIN2012-38079-C03-01.

REFERENCES

[1] 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, 2007.

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

[3] 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, 2006.

[4] M. Zecca et al., ”Emotional Expression Humanoid Robot WE-4RII-Evaluation of the perception of facial emotional expressions by using fMRI”, In *Robotics and Mechatronics Conference ROBOMEC2007*, Japan, pp. 2A1-O10, 2007.

[5] 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.

[6] T. Chen, “Audio-Visual Integration in multimodal Communication”. In *IEEE Proceedings*, 1998.

[7] R. W. Picard, ”Affective Computing”. MIT Press, pp. 88-91, 2000.

[8] J. Prado, C. Simplício, N. F. Lori, J. Dias. ”Visuo-auditory Multimodal Emotional Structure to Improve Human-Robot-Interaction”, in International Journal of Social Robotics Vol. 4, Issue 1, pp 29-51, 2012.

[9] J. B. Bavelas, A. Black, C. Lemery, and J. Mullett, ”I show how you feel: motor mimicry as a communicative act,” In Journal of Personality and Social Psychology, vol. 50, pp. 322-329, 1986.

[10] F. Cid, J.A. Prado, P. Bustos and P. Núñez. ”Development of a Facial Expression Recognition and Imitation Method for Affective HRI”, in Workshop of Physical Agents, pp. 119-124, 2012.

[11] A. Jaimes and N. Sebe, ”Multimodal Human Computer Interaction: A survey”, In *IEEE International Workshop on Human Computer Interaction in conjuntion whith ICCV 2005*, beijing, China, 2005.

[12] N. Sebe, I. Cohen, T. Gevers, T. S. Huang, ”Multimodal Approaches for Emotion Recognition: A Survey”, In *Internet Imaging VI, SPIE’05*, USA, 2005.

[13] V. Bettadapura, ”Face Expression Recognition and Analysis: The State of the Art”, Tech Report, College of Computing, Georgia Institute of Technology , 2012.

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

[15] 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.

[16] 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.

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

[18] T. Mukai et al., ”Development of a Nursing-Care Assistant Robot RIBA That Can Lift a Human in Its Arms”, In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taiwan, pp. 5996-6001, 2010.

[19] 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*, Vol. 4, No.5, 2007.

[20] C. Jayawardena et al. ”Deployment of a Service Robot to Help Older People”. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems* Taiwan, pp. 5990-5995, 2010.

[21] T. Fukuda , M. J. Jung, M. Nakashima, F. Arai and Y. Hasegawa, ”Facial Expressive Robotic Head System for Human-Robot Comunication and Its Application in Home Environment”, In *Proc. of the IEEE*, Vol. 92, No. 11, pp. 1851-1865, 2004.

[22] P. Viola and M.J. Jones, ”Robust real-time face detection”, In International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.

[23] X. Tan and B. Triggs. ”Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions”, IEEE Transactions on Image Processing, Vol. 19, pp. 1635-1650, 2007.

[24] L. Manso, P. Bachiller, P. Bustos, P. Núñez, R. Cintas and L.Calderita ,”RoboComp: a Tool-based Robotics Framework”, In *Proc. of SIMPAR Second International Conference on Simulation, Modeling and Programming for Autonomous Robots*. pp 251-262. 2010.