Learning Emotional Affordances based on Affective Elements in Human-Robot Interaction Scenarios

F. Cid and P. Núñez

***Abstract*—The proposed work presents the concept of emotional affordances as an extension of the classical perceptual affordances for human-robot interaction. In this paper, emotional affordances represent the relation between affective elements, such as the objects in the scenario or the emotional states, and the effects and oportunities for the robot's reactions. Thus, the system proposed in this work describes the use of the affordances in affective scenarios using robotic agents that can recognize and predict the emotional information of the user through elements of the environment and the human natural language ($i.e.$ facial expressions). The major purpose of this paper is to evaluate the proposed conceptual idea through simple experiments to demonstrate the validity and impact of this work.**

***Index Terms*—Social Robot, Facial Expression Recognition, Human-Robot-Interaction, Affordances.**

I. Introduction

In the field of Human-Robot Interaction (HRI), the development of systems capable of learning through communication and interaction with the world presents one of the immediate challenges for the design of autonomous social agents. In this topic, biologically inspired concepts such as *Affordances* [1] represent novel approaches to the perception of the environment. The definition of affordances "the affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill" describes what possible actions associated to an object in the environment are restricted by their physical abilities of the observer. In the last decades, this theory is related to several researching in the area of robotics, such as the prediction of the affordances in objects for grasping tasks [2] or learning by imitation based on interaction with objects [3], among others.

Based on the idea of learning methods where the imitation presents the main solution to the transfer of knowledge from the user to the robot, this paper extends the classic concept of perceptual affordances including the learning of affective behaviors. These new affordances, introduced as affective affordances, were presented in the previous authors's work [4]. In the emotional affordances, the affective elements ($stimulus$) create relationships with potential effects and reactions associated with the emotional state of the user. Therefore, in this approach the robotic agent can manipulate the emotional state of the user by predicting the user's reaction and the effect on each element (see Fig. 1).

In this work, the affective elements are divided into two groups of variables: i) emotional states of the user; and ii) objects in the environment. On one hand, the process of estimating human emotional states are based on the real-time facial expression recognition system presented in [5], which uses the Candide-3 reconstruction model. This system extracts a set of independent

F. Cid and P. Núñez are member of Robotics and Artificial Vision Lab. *Robolab* Group, University of Extremadura, Spain.
E-mail: fcid@unex.es; pnuntru@unex.es

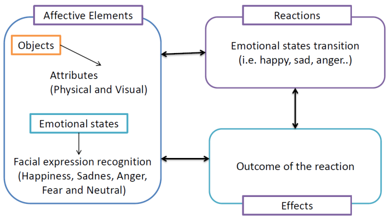


Fig. 1. Emotional affordances: a relationship between stimulus (affective elements, observer's reactions and effects)

and antagonistic facial features, which constitutes the input of a Dynamic Bayesian Network (DBN) used as classifier. The output of this network estimates five possible emotional states, such as: happiness, sadness, anger, fear and neutral (non-associated with an emotion). On the other hand, the object recognition consists of a Marked-based system despite the classical systems based on acquisition of visual features [2], [6]. the use of marks avoids the limitations associated with the number of attributes and objects recognized by the robot, which allows to train a large number of objects at real-time through learning by imitation. The training process is based on the physical and visual attributes, which describe a level of representation of the objects. These attributes are the input variables of a Bayesian Network (BN), which classifies and trains some objects with the five emotional states of the user mentioned in the previous system.

This paper is organized as follows: After discussing previous works in the literature related to facial expression recognition and perceptual affordances for learning in Section II, Section III presents an overview of the proposed system. In Section IV and V presents a detailed description of the acquisition of the features of the affective elements. Next, Section VI described a learning system by imitation proposed in this paper. In Section VII, the experimental results are outlined, and finally, Section VIII describes the conclusions and future works of the presented approach.

II. Related Works

To achieve an affective Human-Robot Interaction, this paper primarily focuses on the concept of emotional affordances as an extension of the perceptual affordances defined by Gibson in [1]. This concept is not new, and was already included by other authors as in Morie *et al.*'s work [7] for the creation of more emotionally affective Virtual Enviroments. In robotics, emotional affordances were defined in a previous work as part of a cognitive system for social robots [4]

The problem of learning perceptual affordances in robotics has been attacked of different ways in the literature. Most of the approaches are based on the imitation and analysis of visual cues. Other methods describe a general model based on Bayesian networks coupling actions, effects and objects features [8]. This formulation is also used in other interesting works, such as [9] or [10], where the authors extend affordances by taking the environmental context. In the basis of learning perceptual affordances system, this paper presents a new method for learning emotional affordances that analyzes the information of the affective elements during the interaction.

The development of affordances within cognitive systems presents a new field of interest in the last decades. These cognitive systems should consider elements such as the sensory motor and the learning aspects, as was described in the review [11]. Among the most relevant works, Ivaldi *et al.*[12] describes an architecture based on affordances of the objects. In other approach, Scheutz *et al.* [13] presented an architecture for complex affective HRI, where the *affect* plays an important role in the integration of the subsystems. Currently, the two major aspects that should be simultaneously present in affective cognitive architectures was presented in [14]: *intentions* and *emotions*. This recommendation has been used for the development of the emotional learning system addressed in this paper, where affective affordances are the basis of the method.

III. Emotional Affordances Modelling

In order to model the emotional affordances, the robotic agent must perform specific actions that allow to predict an effect or expected result. These actions are related to the basic skills of the robotic agent and depend on the type of interaction with each affective element. In the case of the affordances of the objects (*i.e.*, perceptual affordances), the robot interacts with the object expecting results for each actions [3], [2], [15]. Contrary to these perceptual affordances, the proposed method describes an experiment based on affective elements that affect the emotional state of the user through the interaction with a robot. In both two cases, perceptual and emotional affordances, these basic skills are the key of the learning systems by imitation, allowing to acquire information from the user and the objects in the environment. In this proposal two basic skills are provided: First, a facial expression recognition system to estimate the emotional state of the user. Second, an recognition system based on marks can be used to acquire information about the physical attributes of the objects in the HRI scenario. An overview of the proposed method is shown in Fig. 2.

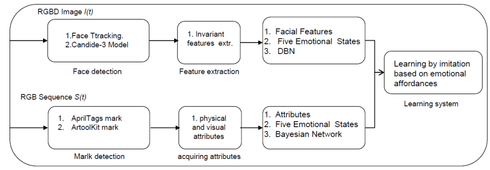


Fig. 2. Overview of the proposed system

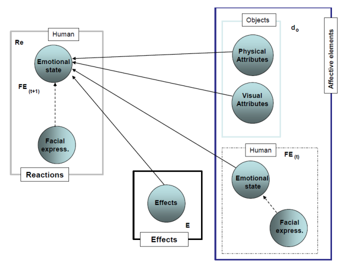


Fig. 3. Model to represent the emotional affordances. See the text for more details.

In this study, the set of reactions performed by the user at an instant of time t are represented by the discrete random variables Re = $\{re_t\}$. Affective elements (objects and the human emotional state) and effects during the HRI are also modeled using discrete random variables. In the case of the emotional elements, let FE be the user's emotional state estimates by the facial expression recognition system, and let d_o = $\{d_o(1), ..., d_o(n_o)\}$ define the set of attributes of the objects detected by the object recognition method (this set of attributes is composed of visual and physical features, such as:color, size, simetry, or shape). Finally, let E = $\{E(1), ..., E(n_e)\}$ represent the effects provided by the robot in the human emotional state. The set of nodes G is formed by the discrete variables Re, FE, d_o and E, Z = (Re, FE, d_o, E). In Fig. 3 is illustrated the representation of the model based on emotional affordances (see [4] for more details).

The learning process uses a Bayesian Network BN based on a set of attributes of each object d_o and the emotional states of the user at an instant of time t, $FE_{(t)}$. This network trains each object through their relevant attributes, and associates each object to an emotional state. This learning approach is inspire in several works in the literature [3], [8]. The application of this learning by imitation system allows to quantify the effects E of each object on the emotional state of the user through the facial expression recognition system $FE_{(t)}$. Therefore, the emotional affordances predict the reaction of the user at $t+1$ (Re = $FE_{(t+1)}$), and allows to modify the reaction estimated through the manipulation of the affective elements.

IV. Facial expression recognition system

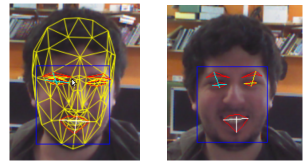
In this section, a facial expression recognition system is presented. This system consist of an algorithm that estimates the user's emotional state based on *Candide-3* mesh Model described in [5]. The *Candide-3* model is a face deformable model controlled by global and local Action Units (AUs) [16]. First, this algorithm uses the *Microsoft Kinect SDK* [17] for detecting and extracting facial features based on the distance between the 3D points of the mesh model. Thus, these distances and their associated features are used as input to a Dynamic Bayesic Network (DBN), wich estimates the user's emotional state ($i.e.$, happiness, Sadness, Fear, Anger and Neutral).

A. Invariant Facial Feature Extraction

The feature extraction process is based on the information obtained from the $Candide$-3 mesh model. This model provides a list of p points belonging to regions and relevant features of the face, R^I=$\{r_i^I | i = 1...p\}$, which are directly related to the Action Units (AUs) described in the Facial Action Coding System (FACS) [18].

These features r^I are calculated using the Euclidean distances between different points in the mesh. The main distances are the following:

- $\mathbf{d_{eb}}$: represents the distance between the upper contour of the eyebrows and lower edge of the eyes. Besides, it is associated directly with the variable EB as shown in Fig. 4(b) (d_{eb} - yellow).
- $\mathbf{d_{lc}}$: represents the distance between the lip corners. Besides, it is associated directly with the variable LC as shown Fig 4(b) (d_{lc} - white).
- $\mathbf{d_{ma}}$: represents the distance between the upper contour and the lower edge of the lips (mouth's aperture). Besides, it is associated directly with the variable MA as shown Fig 4(b) (d_{ma} - white).



(a) (b)

Fig. 4. Facial feature extraction based on Candide-3 model: a) Mesh model of the face; and b) Features extracted in the mesh model of the image.

- $\mathbf{d_{ch}}$: represents the distance in the cheek. Besides, it is associated directly with the variable CH as shown Fig 4(b) (d_{ch} - yellow).

In order to become independent of the limitations as the scale or distance from the user to the sensor, all these features are normalized by using the 3D information of the mesh model within the range of the *SDK face tracking* algorithm. In Fig. 4(a) shows the face tracking and the *Candide-3* model displayed over the face. Finally, Fig. 4(b) illustrates the facial features obtained by this algorithm.

B. Dynamic Bayesian Network

The Bayesian approach implemented in this study, uses independent and antagonistic properties in some AUs to take advantage, increase the performance and reduce the number of variables to be considered in the dynamic bayesian network. The proposed system uses only 11 Action Units to reduce the information processing. These 11 AUs will be grouped according to antagonistic and exclusive properties into only 7 variables.

In this work, the use of a dynamic bayesian network with a two-level structure and the time influence are illustred in Fig. 5. The first level of the DBN contains the belief variable *FE*, that represent the emotional state resulting from the classifier. Each facial expression is associated with a possible emotional states of the user such as: $FE_{[Neutral]}$, $FE_{[Anger]}$, $FE_{[Sad]}$,

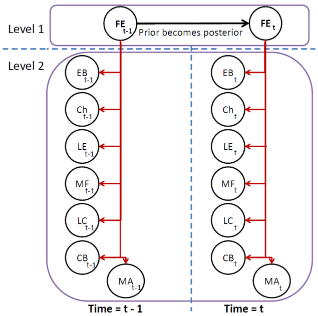


Fig. 5. Dynamic Bayesian Network, two time intervals are shown.

$FE_{[Happy]}$ and $FE_{[Fear]}$. The variables in the second level have as parent this one in the first level: *FE*. The 7 belief variables of the second level in this model, are described in the Table I.

Variable	Action Units AUs	Element of the face
EB	AU1, AU4	Eye-Brows
Ch	AU6	Cheeks
LE	AU7	Lower Eyelids
LC	AU12, AU15	Lips Corners
CB	AU17	Chin Boss
CB	AU20, AU23	Mouth's Form
MA	AU24, AU25	Mouth's Aperture

TABLE I
LEAF VARIABLES OF THE DBN

Usually, in each new probability a threshold is used to find the matches. But, in this case to avoid the extant gaps, a pre-processing stage is done before the learning stage, fitting a Gaussian distribution to the data. In the process of detection of the AUs, the mesh model performs a tracking of the elements of the face, allowing these 7 variables to be independent given the facial expression *FE*.

The data *D* obtained in the facial feature extraction algorithm proposed in this paper, have the following set up:

$$D = ((x_1, y_1)...(x_n, y_n)), x_i \in \mathbb{R}^d, y_i \in \mathbb{R} \quad (1)$$

Consider that y_1 to y_5 are the five possible emotional states ($FE_{[neutral]}$, $FE_{[happy]}$, $FE_{[sad]}$, $FE_{[fear]}$, $FE_{[anger]}$); 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

$$X_i \sim N(prior^T x_i, \sigma^2) \quad (2)$$

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

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

$$P(FE|x_m) = \frac{\prod_1^n P(x_i|FE) * P(FE)}{P(x_m)}, \quad (3)$$

where x_m is the most recent sensory data acquired.

Finally, the last dividend can be calculated using the Bayesian marginalization rule:

$$P(x_m) = \sum_{FE} \prod_1^j P(x_i|FE) * P(FE), \quad (4)$$

being $j = 7$, the number of random variables of the system.

In this work, the model of the bayesian network has a dynamic property that cause a convergence over time. The effect of time is based on the resulting histogram from the previous frame is used as prior knowledge for the current frame. This convergence process is considered complete, when a threshold is exceeded 85% after 5 frames. On the contrary, if the threshold is not exceeded in the 5 frame, the classifier selects the higuest probability value (usually called as $Maximum\ a\ posteriori\ decision$ in Bayesian theory) as the classification result.

V. OBJECTS

In this work, the attributes of each object are associated to the affective elements in the interaction. Each element consists of different attributes related to physical (shape) and visual (color) properties. The approach proposed in this paper determines that these attributes are associated with different emotional states (emotional affordances) and are learned by the robot by imitation of the user's behavior. First, the process implements a representation and categorization of objects based on a system of marks. The main features of these objects are related to the perception and manipulation capabilities in the environment by the robot. The objects used in the development and evaluation of this process, can be easily found in scenarios of the daily routine (cups, toys or pets, among others).

A. *Marks*

The process of acquiring information for each object is achieved through the use of different types of marks commonly used in robotics. Therefore, each mark is associated with a specific information based on the attributes of each object. Fig. 6 illustrates the real object (Fig. 6(a)) and the two mark-based system used in the approach (Fig. 6(b) and Fig. 6(c)). The architecture of the system uses two different types of mark-based libraries for the development and evaluation of the system:

1) ARToolKit: presents an interesting solution for vision-based marker tracking, which has been tippically used for augmented reality applications [19]. Fig. 6(b) shows the specific mark used in the system to represent a real mug.

2) AprilTags: represents a system of codes uses for localizing features, wich contains just several bits of information. Usually, the use of AprilTag presents better results in comparison to similar methods as ARToolKit due to its line detection system that can recognize and identify the marks easily, even if only a small part of the image is visible or is tilted. In the developed system, the AprilTags is based on default marks, mainly the families $tag36h11$. In Fig 6(c), the AprilTag mark associated to the real mug is shown (mark with identifier - $id:1$).

(a) (b) (c)

Fig. 6. Representation of the object (mug) based on marks: a) Real mug; b) ARToolKit mark; and c) AprilTags mark.

B. *Attributes*

The elements or objects in interactions based on affordances must have specific attributes that enable it to be easily distinguishable from other objects. For this reason, the acquisition of features or attributes for each object is performed through the use of marks that provide information to the system, contrary to other methods that consist of the acquisition of visual clues acquired through some vision algorithms based on RGB-Depth information [6], [22]. Thus, the number and types of attributes is not restricted to only visual characteristics but also physical. The choice of each attribute is based on several studies related to affordances for manipulation [2], [3].

Finally, the attributes of different objects are defined in 6 sets of variables v: Shape, Color, Material, Height, Size and Action. These sets of attributes were quantified in scales with numeric values, where the number of values in each scale for each variable is shown in the Table II.

Variable v	Attributes	values	No Values
AS	Shape	ball, cylinder, triangule...	12
AC	color	white, red, black, blue...	19
AM	Material	plastic, metal, ceramic...	9
AW	Height	light, medium, heavy...	6
ASi	Size	little, medium, big...	6
$A1$	Action	pushable, rollable...	7

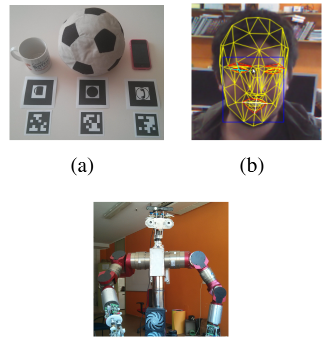
TABLE II
DESCRIPTION OF ATTRIBUTES IN OBJECTS

VI. Learning emotional affordances

In this work, the objects are trained through learning by imitation to be associated with different emotional states of the user. On one hand, the learning process uses the emotional information acquired through facial expressions to classify the objects in the interaction. Therefore, an experiment of learning based on Bayesian network is performed at real time, where the user showes an object and generates a recognizable facial expression by the robot. This Bayesian approach acquires the information associated with the unknown object in the environment through marks, due to the fact that each mark provides information related to the attributes that characterize each object. Thus, let $d_o = \{AS, AC, AM, AW, ASi, A1\}$ represent the set of attributes used as input to the Bayesian network. Finally, the output of the classifier estimates the emotional state associated with each object within 5 possible results, such as: ($ES_{[Fear]}$, $ES_{[Happy]}$, $ES_{[Anger]}$, $ES_{[Sad]}$, $ES_{[Neutral]}$). On the other hand, the emotional affordances use this learning to recognize emotional relationships between the objects and the user, which is necessary to obtain an estimate of the effect of each element on the user's reaction. The learning data corresponds to the perception of some users to train the system with basic elements and a limited number of emotions. Even so, this information may vary in relation to the user.

A. Agents

In the HRI, the agents are responsible for interacting with the user through physical and psychological elements. Therefore, the social robot called *Loki* is presented as a novel platform that combine interaction systems through natural lenguage and manipulation modules. *Loki* has 37 DoF and a architecture composed mainly of two key elements as the expressive robotic head *Muecas* and two arms with manipulators.In this case, the main skills of this platform in the current study are related to the manipulation and the generation of facial expressions. On the one hand, the modules and the sensors on the head estimate and imitate the facial expressions of the user. On the other hand, the manipulator interacts with the user through the manipulation of the elements of the environment, and thus generate reactions in the user. In Fig. 7(c) shows the robot *Loki*.



(a) (b)

(c)

Fig. 7. Elements of the interaction based on objects: a) objects and marks; b) users; and c) Agent.

B. HRI Scenarios

In this study, the affective scenario is located inside RoboHome, a living laboratory located at University of Extremadura. This scenario is adapted to the agent's ability to interact with the objects in the environment in an affective communication. Due to the training and testing of the overall system, the use of this controlled environment allows the robot to obtain information about the affective elements at real time. Thus, the effects of each affective element and the changes caused by the user reactions, are quantified by robust and supervised methods. Fig. 8 illustrates a three-dimensional representation of RoboHome and the elements used in system described in this study.

VII. Experimental Results

In this section, a set of test has been achieved in order to evaluate the perfor-

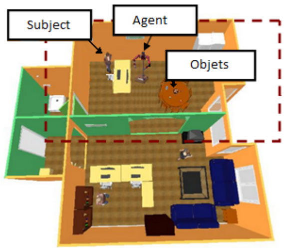


Fig. 8. HRI scenario

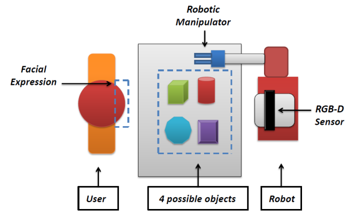


Fig. 9. Learning Representation in the tests

mance of a learning system based on emotional affordances. This evaluation is divided into two different modalities: the initial tests were performed to the facial expressions recognition system through a heterogeneous group of users, which allows evaluating the performance, robustness and effectiveness of the proposed algorithm. The second experiment is performed to evaluate the system of learning by imitation, assessing the global system that uses the algorithms and elements described above.

A. Evaluation of facial expression recognition system

In this first experiment, the evaluation of the robustness of the facial expression recognition system is presented. For this evaluation, the RGBD information and the $Candide$-3 mesh model is acquired through a server with an Intel(R) Core(TM) i3 CPU running Windows and the *Microsoft Kinect SDK*, to another computer on the local network, using the ICE middleware. This client computer on which the tests were performed has a 2.8 Ghz Intel(R) Core(TM) i7 CPU and 4Gb RAM running Linux. This client receives data from the server, using the framework Robocomp [23] that controls the facial expression recognition system.

To test the robustness and homogeneity of the results, a group of 10 users with different gender and facial features has been used. For each test the user must perform 10 random sequences of facial expressions. In table III, the percentages of correctly detected facial expression (r) are shown.

Test	Sadness	Happiness	Fear	Anger	Neutral
P_{fe}	92%	98%	96%	95%	94%
Errors	4%	2%	1%	3%	4%

TABLE III
ROBUSTNESS OF THE FACIAL EXPRESSIONS RECOGNITION SYSTEM

B. Learning by imitation

The second test is achieved in order to implement and evaluate the imitation learning system mentioned above. In this second test a group of 10 users with different gender and facial features must perform facial expressions to interact with each object in the environment. The total number of objects in this experiment is limited to 21, where only three-fifths of the objects will be trained. The algorithm for the detection of objects uses the RGB data from the cameras in the eyes of the robot, which allows to acquire information of the marks associated with the objects in real time. In order to correctly evaluate this method, a affective scenario is implemented with the following requirements: a robotic agent with grasping capacity (in this test, the robot points only to the object), a maximum of 2 humans per session (user and advisor), estable lighting conditions, a normalized distance from the user to the RGB-Depth sensor (60 - 150 cm.) and the equipment described in the first test.

This test is summarized in an affective scenario, where a user will train various elements associated with a emotional state through learning by imitation. These elements are mixed with other non-trained, where the robot must choose an object between a maximum of 4 objects in their environment that cause a specific reaction in the user (Fig. 9). Users began the test with a random emotional state within the five possible emotional states described above. In this case, it is expected a reaction of the user to maintain a constant emotional state or contrary through the choice of objects with the same emotional state or a neutral state, in the absence of more options in the elements of the environment by the robot. In this test, the limited actions of the robot such as the imitation of the facial expressions (only in the absence of a suitable object in the environment) and the movement of the arm to point to the object, avoid feelings of rejection (uncanny Valley[24]).

The results of this tests are shown in Table IV, where the percentages of correctly learning in each learning p_l are illustrated.

Test	Percentage of correct learning ($P_l a$)
Sad	83%
Happy	78%
Fear	71%
Anger	67%
Neutral	42%

TABLE IV
PERCENTAGE OF CORRECT LEARNING BASED ON IMITATION

VIII. CONCLUSION AND FUTURE WORKS

In this paper, an unsupervised learning by imitation has been presented. The proposed system extends the classic concept of affordances, which has been defined as emotional affordances. These emotional affordances represent the relation between affective elements, effects ad oportunities for the robot's reactions. In this work, a learning experiment using emotional affordances has been implemented, where the relationship between the objects in the environment and the emotional states of the user has been defined. This allows the robot to alter the emotional state of the user through the handling of certain objects in a human-robot interaction. This study is divided into several parts described below. On one hand, the method of facial expression recognition based on the use of the Candide-3 face model, which extracts a set of features features associated to the AUs of the user. The high performance and robustness of this method to estimate the emotional state of the user in real time, regardless of environmental conditions and a heterogeneous population, is the key to the dynamics of interaction. On the other hand, the overall system based on learning by imitation presents good results with a set of elements with different attributes. Even so, the disparity in training cause errors associated with overtraining and ambiguity in the attributes. Therefore, there is scope for improving this system through new types of affective elements and applications based on affordances. The purpose of this work allows to evaluate a learning system based on affordances for its implementation in the design of robotic social agents. The implementation of these systems in different agents allows to acquire information through imitation based on the natural language of humans.

Future work will focus on the integrating a multimodal system that allows analyzing not only facial expressions, but also the user's body language in estimating the user's emotional states. Besides, the development of an algorithm of interaction based on behavior predictive models to modify the user's emotional state through a efficient and clear control of the personality of the robot in a specific context of interaction.

ACKNOWLEDGMENT

This work has been supported by the MICINN Projects TIN2012-38079-C03-01 and TIN2012-38079-C03-03 and by the Extremadura Project GR10144.

REFERENCES

[1] J. J. Gibson, "The ecological approach to visual perception". Boston: Houghton Mifin. 1979.

[2] T. Hermans, J. M. Rehg, A. Bobick. "Affordance Prediction via Learned Object Attributes". In *Proc. of IEEE International Conference on Robotics and Automation (ICRA): Workshop on Semantic Perception, Mapping, and Exploration* , 2011.

[3] M. Lopes, F. S. Melo and L. Montesano, "Affordance-based imitation learning in robots". In *Proceedings of the IEEE/RSJ International Conference on Robots and Systems*, pp. 1015-1021, 2007.

[4] F. Cid, A.J. Palomino and P. Núñez. "A New Paradigm for Learning Affective Behavior: Emotional Affordances in Human Robot Interaction". In *Proc. of the XIV Workshop on Physical Agents - WAF2013*, ISBN 978-84-695-8319-7, pp. 47-52, 2013.

[5] P. Romero, F. Cid and P. Núñez. "A Novel Real Time Facial Expression Recognition system based on Candide-3 Reconstruction Model". In *Proc. of the XIV Workshop on Physical Agents - WAF2013*, ISBN 978-84-695-8319-7, pp. 41-46, 2013.

[6] J. Sun, J. L. Moore, A. Bobick and J. M. Rehg. "Learning visual object categories for robot affordance prediction. In *International Journal of Robotics Research*, Vol. 9, pp. 174-197, 2009.

[7] J. Morie, J. Williams, A. Dozois, and D. Luigi, "The Fidelity of "Feel": Emotional Affordance in Virtual Environments". In *Proceedings of the 11th International Conference on Human-Computer Interaction*, 2005.

[8] S. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor. "Modeling Affordances using Bayesian networks". In *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007.

[9] P. Osório, A. Bernardino, R. Martinez-Cantin, and J. Santos-Victor, "Gaussian mixture models for affordance learning using Bayesian Networks". In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4432-4437, 2010.

[10] M. Kammer, M. Tscherepanow, T. Schack, and Y. Nagai, "From Affordances to Situated Affordances in Robotics - Why Context is Important". Frontiers in Computational Neuroscience, 2011.

[11] Toward a Comparative Repository of Cognitive Architectures, Models, Tasks and Data in *http://bicasociety.org/cogarch/*.

[12] S. Ivaldi, et al. "Perception and human interaction for developmental learning of objects and affordances", In *Proc. of the 12th IEEE-RAS International Conference on Humanoid Robots - HUMANOIDS*, pp. 1-8, 2012.

[13] M. Scheutz, P. Schermerhorn, C. Middendorff, J. Kramer, D. Anderson and A. Dingle. "Toward Affective Cognitive Robots for Human-Robot Interaction". In *Proceeding of 17th Innovative Applications of Artificial Intelligence Conference*, pp.1737-1738, 2005.

[14] I. Infantino, "Affective Human-Humanoid Interaction Through Cognitive Architecture". In *The Future of Humanoid Robots - Research and Applications*, In-Tech, Edited by Dr. Riadh Zaier, pp. 147-164. 2012.

[15] S. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Learning object affordances: From sensorymotor coordination to imitation". In *IEEE Transactions on Robotics*, Vol. 24, pp. 15-26, 2008.

[16] Jrgen Ahlberg, "CANDIDE-3 - an updated parameterized face", Report No. LiTH-ISY-R-2326, Dept. of Electrical Engineering, Linkping University, Sweden, 2001.

[17] Face Tracking, http://msdn.microsoft.com, last visited: 30.03.2014

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

[19] Library for building Augmented Reality (AR) applications, ARToolKit, http://sourceforge.net/projects/artoolkit/, last visited: 24/03/2014.

[20] Library ARToolKit for OpenSceneGraph- OSGART, http://www.artoolworks.com/community/osgart/, last visited: 24/03/2014.

[21] E. Olson, "AprilTag: A robust and flexible visual fiducial system", In *Proceedings of the IEEE International Conference on Robotics and Automation* (ICRA), 2011.

[22] H. S. Koppula, R. Gupta and A. Saxena."Learning human activities and object affordances from RGB-D videos". In *International Journal of Robotics Research*, Vol. 32, pp. 951-970, 2013.

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

[24] M. Mori. Bukimi no tani - *The uncanny valley*. In *Energy*. Vol. 7(4). pp. 33-35. 1970.