COMPARING BAYESIAN NETWORKS TO CLASSIFY FACIAL EXPRESSIONS

Carlos Simplício
Institute of Systems and Robotics
University of Coimbra, Portugal
Institute Polythechnic of Leiria, Portugal
carlos.simplicio@ipleiria.pt

José Prado
Institute of Systems and Robotics
University of Coimbra, Portugal
jaugusto@isr.uc.pt

Jorge Dias
Institute of Systems and Robotics
University of Coimbra, Portugal
jorge@isr.uc.pt

ABSTRACT
In this paper are presented two distinct Bayesian networks to analyse human beings' facial expressions. Both classifiers are completely defined: structure of the networks, belief variables and respective events, likelihoods, initial priors and procedure to change dynamically priors. The performance (relatively to the convergence) of the two approaches is compared. For both networks, the classification is done associating the facial expression to the probabilities of five emotional states: anger, fear, happy, sad and neutral. A justification for the usage of this set is presented: it is based in emotional states presented by human beings during social relationships. Classifiers as these described here can be used in Human Robot Interation. We believe that this interaction shall be done in a similar way of that used by human beings to communicate between them and, after all, facial expressions is one of the main non-verbal means of communication used by human.

KEY WORDS
Facial Expressions – Bayesian Networks – HRI.

1 Introduction

In last years, greats efforts was been dedicated to reduce the estrangement between humans and machines. Nevertheless, the problem is not yet solved. We believe that Human Robot Interaction shall be done in a similar (but, not exactly equal) way of that used by human beings to communicate between them.

In they daily relationships, human beings communicate using a extensive number of channels. By example, when performing a dialogue, the communication is done not only through speech. Normally, voice, body position, gestures, gaze and facial expressions are used, as well as speech.

Facial expressions can be described as distortions or movements of some features in the face (e.g., eyebrows, eyes, nose, mouth) which arise as result of the muscular activity. Human beings can perform volutarily these muscular activity when making, for example, a grimace. However, in their daily activities, human beings performed systematically involutary facial expressions. These are a form of non-verbal communication used in social contacts as means to express emotions. Normally, to every emotional state is associated a characteristic facial expressions. In figure 1 are presented examples of some facial expressions.

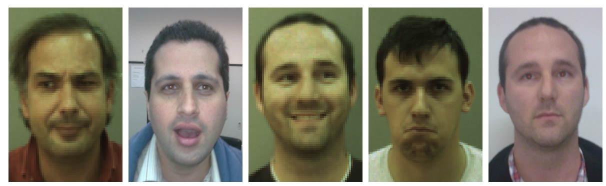

Figure 1. Facial expressions associated to emotional states: anger, fear, happy, sad and neutral.

Spinoza [11], in the seventeenth century, studied how the humans beings behave in social relationships when in distinct emotional states. Recently, Damasio reformulated and extended Spinoza point of view. In him approach, Damasio [2] [3] proposed a joint behavior of four groups composed by emotional states and respective emotional competent stimulus. Following Spinoza's point of view, three of the groups present a negative charge and are related with the lost of some capabilities to communicate; that is: *de-alternative*, *de-enabled* and *de-message*. A fourth group, associated to *success*, is also considered by Damasio. The association between the principal emotional state in each group and respective communicative (negative or positive) capabilities are the following: *anger* is linked to *de-alternative*, *fear* is linked to *de-enable*, *sad* to *de-message* and *happy* to *success*. Damasio considers that, normally, human beings are in one of these emotional states: the absolute neutral emotional state is a rare exception. We understand the position of Damasio but, when in the context of Human Robot Interaction, we prefer considering the neutral emotional state to replace another if it is weak. By this reason, we are considering the five emotional states which facial expression is illustrated in figure 1.

In human beings, the association between emotional states and facial expressions are so extensive that, for certain emotions, it can be very difficult to avoid performing the characteristic facial expressions, even when it is wanted to hide the real emotional state.

Through the development of facial expression's classifiers we are giving a contribution to reduce the estrangement between humans and robots. In section 2 are presented some concepts taken from the psychology literature. These concepts, which are related with human beings' facial expressions, are important to the development of the classifiers. In section 3 two approaches of a facial expressions' classifier (Bayesian networks) are presented. In section 4 learning results (histograms tables used by the Bayesian networks) and classifications results are presented. The classifiers are compared relatively to their convergence. Finally, in section 5 conclusions and perspectives for future work are presented.

2 Facial Expressions Classification

Darwin [4] studied the various means used, by animals in general and humans in particular, to express emotional states. However, a special focus was put to describe how human beings express their emotional states through movements and distortions of facial features (i.e., eyebrows, eyes, cheeks, or mouth) and changes on face's skin color. More recently, Paul Ekman devoted to the specific subject of emotional states and facial expressions [6] [7] [8].

Various attempts have been done to develop a automatic system to recognize the facial expressions associated to emotional states. The great majority of then makes the classification with basis in Facial Action Coding System [6]; however, only a few of them use dynamic Bayesian networks to make the classification. A three-layer Temporal Exemplar-based Bayesian Network for facial expression recognition which improves the accuracy of probability estimation without no assumption on the prior distribution was proposed by [10]. A Bayesian network with a dynamic fusion strategy to classify facial expressions is proposed in [12]. A Bayesian belief network that handles behaviors along the time (the authors do not use the term but, in fact is a dynamic Bayesian network) is described in [5]. When compared to these approaches, the Bayesian networks described here present a structure much more simple without compromising the classification results. A method to search for the correct Bayesian network structure applied to facial expressions classification was proposed in [1]. However, our work is distinct from this one because the structures of proposed classifiers are very simple and we do not want to learn the structure of Bayesian networks.

Facial Action Coding System defines a total of 52 Action Units (AUs) where 8 of them are related with the head pose. The remainder 44 concern to small distortions over the face, which could be used to characterize the facial expressions. Each of these Action Units is anatomically related to the activity of a specific set of muscles which produces changes in face's appearance.

In our work, only a small sub-set of the Action Units introduced in Facial Action Coding System is used (some examples are presented in figure 2). In fact, the sub-set used is chosen to permit the classification of the facial expressions associated to the five emotional states illustrated in figure 1. In table 1 are discriminated the Action Units normally associated to each one of these facial expressions.

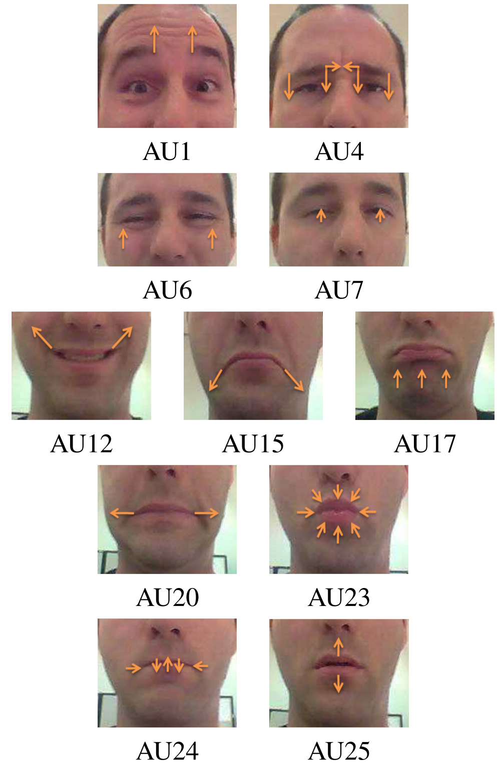

Figure 2. Examples of Action Units (AUs): AU1 - inner portion of the brows raised; AU4 - brows lowered and draw together; AU6 - cheeks raised; AU7 - lower eyelids raised; AU12 - lip corners pulled obliquely; AU15 - lip corners pulled down; AU17 - chin boss pushed upwards; AU20 - mouth stretched horizontally; AU23 - lips tightened; AU24 - lips pressed together; AU25 - lips relaxed and parted.

	Upper Face				
	Anger	Fear	Happy	Sad	Neutral
EyeBrows	AU4	AU1+4	—	AU1+4	—
Cheeks	—	—	AU6	—	—
Lower Eyelids	AU7	—	—	-	—
Lip Corners	—	—	AU12	AU15	—
Chin Boss	AU17	—	—	AU17	—
Mouth Form	AU23	AU20	—	—	—
Mouth Aperture	AU24	AU25	AU25	—	—

Table 1. Discrimination of the AUs that are present in the facial expressions associated to some emotional states.

A facial expression is composed by a specific set of Action Units. Each one of these Action Units is a distortion of a facial feature induced by muscular activity. Normally, a well determined set of muscles is associated to a specific Action Unit, what can give the idea that all these basic distortions are independent. Nevertheless, some of these Action Units are antagonistic. One concrete, and understandable, example is the case of two Action Units related with the movements of the corners of the mouth, that is AU12 and AU15. When performing the first one of these, the lip corners are pulled obliquely in the direction of the ears and eyes; that is the corners move up and back. On the other end, when performing just the AU15 the lip corners are pulled down. Therefore, if by one way the movements of the lip corners can be considered independent because are performed by distinct muscles sets by another, when analyzed visually, they are, almost by some extend, antagonistic, exclusive and no-independent.

Nevertheless, some Action Units seeming antagonistic and mutually exclusive can occurs simultaneously: in this case the used term to describe this situation is "non-additive combination". One example of this situation occur some times when a human being is showing a facial expression of fear or of sadness. In these cases, the set of muscles responsible by the AU1 are activated together with another set responsible by the AU4. In terms of appearance, when AU1 occurs alone the inner eyebrows are pulled upwards and, when AU4 occurs alone the eyebrows are pulled together and downwards. Therefore AU1 and AU4 are antagonistic. In reality, it is possible the activation of the two sets of muscles and in this situation we are in the presence of a "non-additive combination". In this case the notation used is AU1+4 (it is different of the notation AU1 + AU4, which would be used if these AUs could appends in an "additive combination").

3 Bayesian Networks to Classify Facial Expressions

To classify the facial expressions two approaches were developed. In the first, described in sub-section 3.1, the belief variables where the evidences are collected are directly related with Action Units relevant to classify the distinct facial expressions presented in figure 1. In this case, associated to every variable, there are only two events (*no* and *yes*) which indicate if the Action Unit is absent or present.

In the second approach, which is described in subsection 3.2, the belief variables where the evidences are collected are directly related with the type of movements associated to a facial feature. In this second approach, the number of events associated with distinct belief variables is not necessarily equal. This number depends of how much Action Units, associated to a facial feature, are relevant to classify the distinct facial expressions presented in figure 1.

3.1 Facial Expressions Classification – Approach 1

The structure of the classifier, of this first approach, is presented in figure 3: it is a two levels Bayesian network.

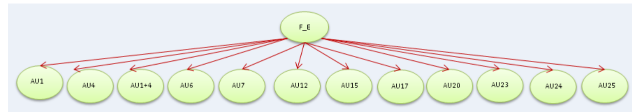

Figure 3. Bayesian Network to classify facial expression – approach 1.

In the Bayesian network's first level there is only one node. The global classification result obtained is provided by the belief variable associated with this node: $F\_E \in \{anger, fear, happy, sad, neutral\}$, where the variable name stands from Facial_Expression.

Considering the structure of the Bayesian network, the variables in the second level have as parent F_E. In the second level there are twelve belief variables:

- $AU\_1 \in \{no, yes\}$ is a belief variable related with the eyebrows movements; that is, if the inner portion of the eyebrows are raised or no. The two events (*no* and *yes*) are directly related with the absence, or existence, of this type of movement.

- $AU\_4 \in \{no, yes\}$ is a belief variable which is related with another type of eyebrows movements; more specifically, if the eyebrows are lowered and draw together or no. As in previous the case, there are two events (*no* and *yes*). They are directly related with the absence, or existence, of this type of movement.

- $AU\_1+4 \in \{no, yes\}$, it is a belief variable related with a third type of eyebrows movements. In this case, we are in presence of a "non-additive combination" (it is the result of the joint action of the set of muscles responsible by the movements corresponding to AU1 and AU4 alone).

- $AU\_6 \in \{no, yes\}$ is a belief variable related with cheeks movements. In this case, the two events indicate if the cheeks are in an usual position or if they are raised.

- $AU\_7 \in \{no, yes\}$ is a belief variable indicating if the lower eyelids are in normal position or if they are raised.

- $AU\_12 \in \{no, yes\}$ is a belief variable associated with lip corners' movements. The two events are related with the absence or presence of movements pulling the corners obliquely, up and backwards.

- $AU\_15 \in \{no, yes\}$ is another belief variable associated with the movements of the lip corners. In this case, the two events are related with the absence, or presence, of downwards movements.

- $AU\_17 \in \{no, \, yes\}$ is the belief variable collecting the probabilities related with absence or presence of movements performed by the chin boss (in this case, the movements push upwards the chin boss).

- $AU\_20 \in \{no, yes\}$, it is a belief variable associated with mouths form. More specifically, the two events are related with muscles' actions directly related with the absence, or presence, of a mouth stretched horizontally.

- $AU\_23 \in \{no, yes\}$ it is another belief variable associated with the form of the mouth; more concretely, with tightened lips. As previsouly, the events indicate the probability of the absence, or presence, of this type of movements.

- $AU\_24 \in \{no, \, yes\}$ is a belief variable associated with aperture of the mouth. The two events are related with the presence of a normal closed mouth or if the lips are pressed.

- $AU\_25 \in \{no, \, yes\}$ it is another belief variable associated with the mouth's aperture, more specificaly with the fact of the lips are relaxed and parted. The two events indicate the probability of this situation be absent or present.

The following equations illustrate the joint distribution associated to this approach 1 of the Bayesian network.

$$
\begin{aligned}
&P(F\_E, AU\_1, AU\_4, AU\_1{+}4, AU\_6,\\
&AU\_7, AU\_12, AU\_15, AU\_17, AU\_20,\\
&AU\_23, AU\_24, AU\_25) = \\
&= P(AU\_1, AU\_4, AU\_1{+}4, AU\_6, AU\_7,\\
&\quad AU\_12, AU\_15, AU\_17, AU\_20, AU\_23,\\
&\quad AU\_24, AU\_25|F\_E) \,.\, P(F\_E) = \\
&= P(AU\_1|F\_E).P(AU\_4|F\_E).P(AU\_1{+}4|F\_E).\\
&\quad .P(AU\_6|F\_E).P(AU\_7|F\_E).P(AU\_12|F\_E).\\
&\quad .P(AU\_15|F\_E).P(AU\_17|F\_E).\\
&\quad .P(AU\_20|F\_E).P(AU\_23|F\_E).\\
&\quad .P(AU\_24|F\_E).P(AU\_25|F\_E).P(F\_E)
\end{aligned}
$$

The last equality is written assuming that the belief variables in the second level of the Bayesian network are independent. From the joint distribution, the *posterior* can be obtained by the application of the Bayes Formula as follows.

$$
\begin{aligned}
&P(F\_E|AU\_1, AU\_4, AU\_1{+}4, AU\_6, AU\_7,\\
&AU\_12, AU\_15, AU\_17, AU\_20, AU\_23, AU\_24, AU\_25) = \\
&= \frac{P(AU\_1|F\_E). \cdots .P(AU\_25|F\_E).P(F\_E)}{P(AU\_1, \cdots, AU\_25)}
\end{aligned}
$$

3.2 Facial Expressions Classification – Approach 2

In figure 4 is presented the structure of this second approach of a facial expressions' classifier. Like the classifier in the first approach, it is a two levels Bayesian network.

Also like in the first approach, in the first level of the Bayesian network there is only one node. It is in belief variable associated to this node ($F\_E \in \{anger, \, fear, \, happy, \, sad, \, neutral\}$) where the global classification result is obtained. The name of this variable stands from Facial_Expression and the five events are directly related with emotional states presented in figure 1.

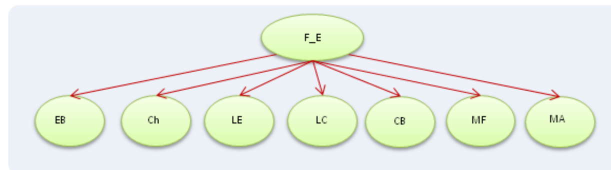

Figure 4. Bayesian Network to classify facial expression – approach 2.

The variables in the second level of the Bayesian network have as parent F_E. The seven belief variables in the second level are the following:

- $EB \in \{none, \, au1, \, au4, \, au1+4\}$ is a belief variable related with *Eye-Brows* movements. The four events are directly related with absence or with the existence of AU1, AU4 or of their "non-additive combination" (AU1+4).

- $Ch \in \{none, \, au6\}$ is a belief variable which is related with *Cheeks* movements; more specifically, the events indicates if AU6 is absent or if the cheeks are raised.

- $LE \in \{none, \, au7\}$ is a belief variable which is related with the *Lower Eyelids* movements; AU7 is the action unit associated with the raising of the lower eyelids.

- $LC \in \{none, \, au12, \, au15\}$ is the belief variable associated with the movements of the *Lips Corners*. The event *none* must have a high probability when the corners do not perform any movement. The event *au12* must have a great probability when the lip corners are pulled obliquely up and backwards. If the lip corners moves downwards the event *au15* must have a great probability.

- $CB \in \{none, \, au17\}$ 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* has a great probability when the chin boss is pushed upwards.

- $MF \in \{none, \, au20, \, au23\}$ is the belief variable associated with the *Mouth's Form*. The events *au20* and *au23* indicates, respectively, if the mouth is stretched horizontally or, inversely, if the lips are tightened.

- $MA \in \{none, au24, au25\}$ is the belief variable associated with the *Mouth's Aperture*. The events *au24* and *au25* are related, respectively, with the act of the lips are pressed together or lips are relaxed and parted.

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

$$P(F\_E, EB, Ch, LE, LC, CB, MF, MA) =$$

$$= P(EB, Ch, LE, LC, CB, MF, MA | F\_E).P(F\_E) =$$

$$= P(EB|F\_E).P(Ch|F\_E).P(LE|F\_E).$$
$$.P(LC|F\_E).P(CB|F\_E).P(MF|F\_E).$$
$$.P(MA|F\_E).P(F\_E)$$

The last equality is written assuming that the belief variables in the second level of the Bayesian network are independent. From the joint distribution, the *posterior* can be obtained by the application of the Bayes Formula as follows.

$$P(F\_E|EB, Ch, LE, LC, CB, MF, MA) =$$

$$= \frac{P(EB|F\_E).P(Ch|F\_E). \cdots .P(MA|F\_E).P(F\_E)}{P(EB, Ch, LE, LC, CB, MF, MA)}$$

4 Results Presentation and Discussion

After defining the structure of the Bayesian networks it is necessary to provide the probabilities: priors and likelihoods. For both classifiers (approach 1 and 2) these likelihoods are provided as histograms tables. To build these distribution tables the Cohn-Kanade database was used [9].

The initial priors are defined through a uniform distribution. In both approaches, as there are five events associated to the belief variable in the Bayesian networks' first level, the priors are $P(F\_E = anger) = \cdots = P(F\_E = neutral) = 0.2$. These priors are changed dynamically: that is, systematically, after each classification, the posterior is transformed in the new prior of the Bayesian network.

As opposite to what appends with priors, the likelihoods are not changed dynamically and remains the same over the time. But, in the case of these probabilities, as the variables in the second level of both Bayesian networks are different, two histograms tables, one for each approach, is needed. In tables 2 and 3 are presented the histograms of likelihoods obtained as result of the learning. In both tables there are a considerable number of probabilities whose value is near zero but not null. It appends because we "force" it during the learning phase, when the histograms are built. The justification for this procedure is the following: "*it is considered that, if a occurrence is not observed in the learning phase it is because it has a low probability, not because it is impossible*".

		F_E				
		anger	fear	happy	sad	neutral
AU1	no	0.99	0.99	0.99	0.99	0.99
	yes	0.01	0.01	0.01	0.01	0.01
AU4	no	0.01	0.99	0.99	0.99	0.99
	yes	0.99	0.01	0.01	0.01	0.01
AU1+4	no	0.99	0.01	0.99	0.01	0.99
	yes	0.01	0.99	0.01	0.99	0.01
AU6	no	0.99	0.99	0.01	0.99	0.99
	yes	0.01	0.01	0.99	0.01	0.01
AU7	no	0.01	0.99	0.99	0.99	0.99
	yes	0.99	0.01	0.01	0.01	0.01
AU12	no	0.99	0.99	0.01	0.99	0.99
	yes	0.01	0.01	0.99	0.01	0.01
AU15	no	0.99	0.99	0.99	0.01	0.99
	yes	0.01	0.01	0.01	0.99	0.01
AU17	no	0.01	0.99	0.99	0.01	0.99
	yes	0.99	0.01	0.01	0.99	0.01
AU20	no	0.99	0.01	0.99	0.99	0.99
	yes	0.01	0.99	0.01	0.01	0.01
AU23	no	0.01	0.99	0.99	0.99	0.99
	yes	0.99	0.01	0.01	0.01	0.01
AU24	no	0.01	0.99	0.99	0.99	0.99
	yes	0.99	0.01	0.01	0.01	0.01
AU25	no	0.99	0.01	0.01	0.99	0.99
	yes	0.01	0.99	0.99	0.01	0.01

Table 2. Learned histogram – likelihoods to approach 1.

		F_E				
		anger	fear	happy	sad	neutral
EB	none	0.01	0.01	0.97	0.01	0.97
	au1	0.01	0.01	0.01	0.01	0.01
	au4	0.97	0.01	0.01	0.01	0.01
	au1+4	0.01	0.97	0.01	0.97	0.01
CH	none	0.99	0.99	0.01	0.99	0.99
	au6	0.01	0.01	0.99	0.01	0.01
LE	none	0.01	0.99	0.99	0.99	0.99
	au7	0.99	0.01	0.01	0.01	0.01
LC	none	0.98	0.98	0.01	0.01	0.98
	au12	0.01	0.01	0.98	0.01	0.01
	au15	0.01	0.01	0.01	0.98	0.01
CB	none	0.01	0.99	0.99	0.01	0.99
	au17	0.99	0.01	0.01	0.99	0.01
MF	none	0.01	0.01	0.98	0.98	0.98
	au20	0.01	0.98	0.01	0.01	0.01
	au23	0.98	0.01	0.01	0.01	0.01
MA	none	0.01	0.01	0.01	0.98	0.98
	au24	0.98	0.01	0.01	0.01	0.01
	au25	0.01	0.98	0.98	0.01	0.01

Table 3. Learned histogram – likelihoods to approach 2.

In figures 5 and 6 are illustrated two specific cases of classification performed by each approach presented in section 3. In both cases, only the Action Units AU23, AU24 and AU 25 are present in the human's face. As can be see through these graphics the approach 2, defined in sub-section 3.2 converges more quickly. For approach 2, at Time = 2 the final probabilities are obtained; whereas for the approach 1 at Time = 4 the probabilities are yet converging.

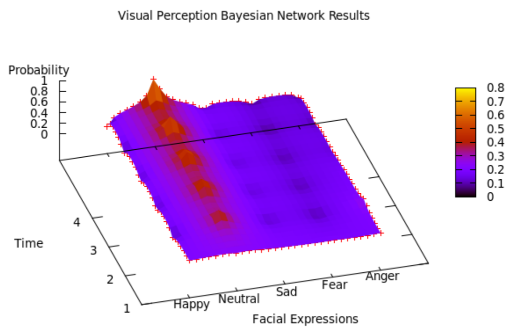

Figure 5. Classifying facial expressions: approach 1 results.

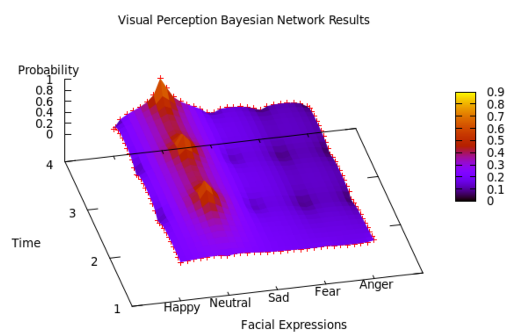

Figure 6. Classifying facial expressions: approach 2 results.

5 Conclusions and Futher Work

Two distinct Bayesian networks to classify human facial expressions were presented. The definition of these two approaches was done completely: structure of the networks, events for every belief variable, likelihoods, initial priors and procedure to change dynamically priors. The performance (relatively to the convergence) of the two approaches was compared.

The classification done through each of the Bayesian networks determines the probabilities of the human be anger, fear, happy, sad and neutral. The choice of these five emotional states is not arbitrary and a justification, based in philosophical works of others authors, is presented.

As future work we intend develop a robot with some humorist characteristics. To obtain true interaction, the robot must owns a face to show distinct expressions. The process of synthesis will be develop inside the Bayesian framework. This means that a Bayesian network (similar to that used in the analysis of human being's face) will be used to synthesize robot's facial expressions. It this case the evidences provided to the network are the emotional states and results are probabilities associated to the Action Units.

Acknowledgments

The authors gratefully acknowledge support from Institute of Systems and Robotics at Coimbra University (ISR-UC), Institute Polytechnic of Leiria (IPL) and from Portuguese Foundation for Science and Technology (FCT).

References

[1] I. Cohen, N. Sebe, F. G. Gozman, M. C. Cirelo, and T. S. Huang. Learning bayesian network classifiers for facial expression recognition using both labeled and unlabeled data. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages I–595 – I–601 vol.1, 2003.

[2] Antonio Damasio. *The Feeling of what happens*. Harcourt, Inc - ISBN 978-0-15-601075-7, 2000.

[3] Antonio Damasio. *Looking for Spinoza*. Harcourt, Inc - ISBN 978-0-15-100557-4, 2003.

[4] C. R. Darwin. *The expression of the emotions in man and animals*. London: John Murray, 1872.

[5] D. Datcu and L. J. M. Rothkrantz. Automatic recognition of facial expressions using bayesian belief networks. In *2004 IEEE International Conference on Systems, Man & Cybernetics*, pages 10–13, October 2004.

[6] Paul Ekman, Wallace V. Friesen, and Joseph C. Hager. *Facial Action Coding System - The Manual*. A Human Face, 2002.

[7] Paul Ekman and W.V Friesen. *Unmasking the face: A guide to recognizing emotions from facial clues*. Ma: Malor Books, 2003.

[8] Paul Ekman and E.L. Rosenberg. *What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS)*. Oxford University Press. Second expanded edition, 2004.

[9] T. Kanade, Y. Tian, and J. F. Cohn. Comprehensive database for facial expression analysis. In *4th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 46–53, 2000.

[10] L. Shang and K.-P. Chan. Temporal exemplar-based bayesian networks for facial expression recognition. In *7th International Conference on Machine Learning and Applications*, pages 16 – 22, 2008.

[11] Spinoza. *Ethics*. 1677.

[12] Y Zhang and Q. Ji. Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):699–714, 2005.