A Face Attention Technique for a Robot Able to Interpret Facial Expressions

Carlos Simplício, José Prado, Jorge Dias

A Face Attention Technique for a Robot Able to Interpret Facial Expressions

Carlos Simplício[1, 2], José Prado[1] and Jorge Dias[1]

[1] Institute of Systems and Robotics, at Department of
Electrical Engineering and Computers of University of Coimbra, Coimbra, Portugal
{jaugusto, jorge}@isr.uc.pt
[2] School of Technology and Management of
Institute Polytechnic of LeiriaSystems, Leiria, Portugal
{simplicio}@estg.ipleiria.pt

Abstract. Automatic facial expressions recognition using vision is an important subject towards human-robot interaction. Here is proposed a human face focus of attention technique and a facial expressions classifier (a Dynamic Bayesian Network) to incorporate in an autonomous mobile agent whose hardware is composed by a robotic platform and a robotic head. The focus of attention technique is based on the symmetry presented by human faces. By using the output of this module the autonomous agent keeps always targeting the human face frontally. In order to accomplish this, the robot platform performs an arc centered at the human; thus the robotic head, when necessary, moves synchronized. In the proposed probabilistic classifier the information is propagated, from the previous instant, in a lower level of the network, to the current instant. Moreover, to recognize facial expressions are used not only positive evidences but also negative.

Keywords: Facial Symmetry; Focus of Attention; Dynamic Bayesian Network.

1 Introduction

Usually, human beings express their emotional states through paralinguistic cues, e.g., facial expressions, gaze, gestures, body positions or movements. Among all, our main work focuses specifically on human facial expressions.

Automatic facial expression recognition systems have many potential applications. They can be used in medicine / psychology, surveillance or in intelligent human-machine interaction. In this paper is proposed a human face focus of attention technique and a Dynamic Bayesian Network to classify human beings facial expressions. They will be incorporated in a companion robot, an autonomous mobile agent, whose hardware is composed by a robotic platform and a robotic head. The autonomous agent must observe and react according to the facial expressions of a person. This agent will be used in the context of assisted ambiance. The global project addresses the emergent tendencies of developing new devices to the elderly community.

2 Related Work and Contribution to Technological Innovation

A great part of the research done in facial expressions uses the tool FACS (Facial Action Coding System) proposed by Ekman [1]. FACS was developed to describe the visible "distortions" on the face produced by muscular activity.

Some studies about techniques to find automatically the vertical symmetry axis of human faces were published. Hiremath and Danti [2] proposed a bottom-up approach to model human faces. In this method a face is explained by various lines-of-separability, being one of them the axis of symmetry. Chen [3] proposed a method to automatically extract the face vertical middle line. In their method the histogram of gray level differences (between both sides of the face) is build. The symmetry axis corresponds to the line with maximal Y value (obtained as the relation between the number of events equals to the mean and the variance of the histogram). Nakao [4] applied the generalized Hough transform to find the vertical symmetry axis in frontal faces.

Various attempts have been done to develop a system classifying human facial expressions, but only a few through Bayesian networks. One of these classifiers was presented by Datcu and Rothkrantz [5]: it is a Bayesian belief network that handles behaviors along the time. A Bayesian network with a dynamic fusion strategy was developed by Zhang and Ji to classify six facial expressions [6].

Facial expressions recognition becomes easier if done in frontal face images. Normally, the image acquisition device is fixed and does not follow the head movements. To solve this problem, 3D models of the human face, projections of textures in the model and geometric transformations are used to obtain a image of a frontal face. In this paper a different method to get always a frontal face image is proposed: a robotic system is used to follow the human being movements. In the proposed technique of human face focus of attention, the robotic platform navigate, running an arc centered in the human, to keep always targeting the face frontally. If necessary, the robotic head rotates in synchronization with the platform. The approximated symmetry presented by human faces is the attribute to support the extraction of information to control the tasks of movement.

Facial expressions recognition is performed by a Dynamic Bayesian Network (DBN). With this classifier, one of six emotional states (anger, fear, happiness, sadness, neutral or other) will be assigned to the human being. The evidences provided to the DBN are derived from some Action Units as defined by Ekman [1]. However, unlike other authors, positive and negative evidences are used explicitly to obtain the probability associated with each facial expression (for the neutral facial expression are used just negative evidences, therefore it is an exceptional case). There is another difference from previous works: the information is propagated along time in a low level, near the nodes which collect the evidences from the sensors.

3 Autonomous Mobile Agent (AMA)

The Autonomous Mobile Agent (AMA) architecture is presented in Fig. 1. At hardware level, the AMA consists in a Robotic Head and a Robotic Platform. The

commands (*head movement* and *platform movement*), to put the AMA frontally to the human being face, are sent to the hardware by the Robotic Systems Controller (RSC). This module (it is briefly described in sub-section 3.1) receives an input, *face pose*, coming from the module Face Pose Identification System (FPIS). As is explained in sub-section 3.2, FPIS processes an image to provide the respective output, *face pose*. With the Robotic Head positioned "face-to-face" with the human being, the module Automatic Facial Expression Recognition System (AFERS) takes an image and provides the respective Emotional State in probabilistic terms. This output is associated with a discrete random variable with six events (*anger*, *fear*, *happy*, *sad*, *neutral* and *other*). Specifically, the system output is composed by six probabilities: *p(anger)*, *p(fear)*, *p(happy)*, *p(sad)*, *p(neutral)* and *p(other)*. In sub-section 3.3 is presented the Dynamic Bayesian Network which provides this module output.

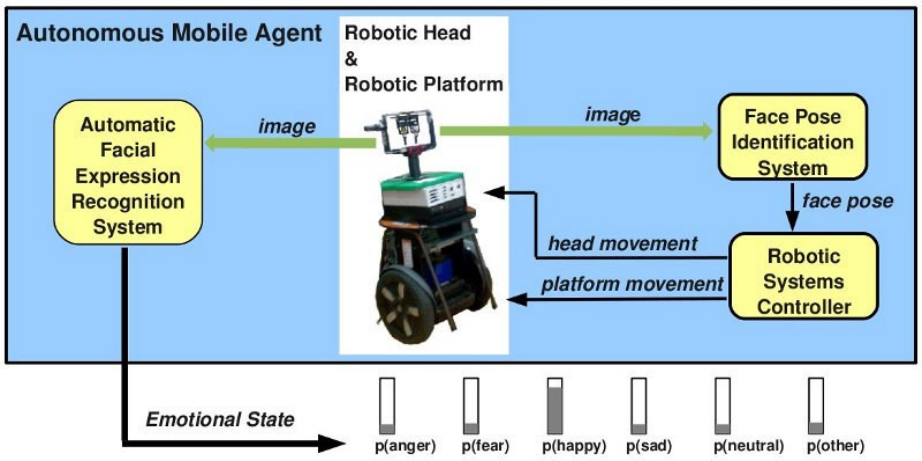

Fig. 1. Autonomous Mobile Agent (AMA) - principal modules.

3.1 Robotic Systems Controller (RSC)

The Robotic Platform does three types of movements (longitudinal or transversal translations and rotations) following the commands provided by the RSC.

Longitudinal translations are performed to approach or move away the AMA from the human being. The objective is keeping the image of the human being face with the same number of pixels. Transversal translations are performed to keep always the face in the centre of the image.

Rotations correspond to an arc of circle centered in the human being. These movements are performed to follow the rotation movements done by the human being, getting always an image of a frontal face.

The Robotic Head can move in synchronization with the platform when it is necessary.

3.2 Face Pose Identification System (FPIS)

In a perfect bilateral symmetric image, the difference between a pixel value and the respective symmetric counterpart is zero. By nature, human faces do not present a perfect bilateral symmetry and it is reflected in the acquired images. Moreover, the "imperfections" in the images are worsened by the noise associated to the acquisition process or due to lights distribution. Despite this, gray-level differences can be used to detect the bilateral symmetry axis of a human face.

The theoretical method used to identify the axis of symmetry is based on rather simple principles but is very effective. A vertical axis is defined to divide the face image region in two parts with equal number of pixels and a Normalized Gray-level Difference Histogram (NGDH) is built. When the face is frontal, this vertical axis bisects it and the information collected in the NGDH is strongly concentrated near the mean. If the camera is fixed, when the face rotates the defined axis is not a symmetry axis and the information is scattered along the NGDH. In practice, instead the mean, is used a narrow region (+/- 10 units) around it: the *pseudomean*.

Our method requires a frontal face just in the initialization, but it assumes in all phases an upright face. The algorithm begins performing the face detection using Haar-like features [7]. In this way, a region of interest is found and a vertical axis is established in the middle of this region. To find the real face orientation, the region of interest is successively rotated about that axis using a 3D transformation. Five rotation angles, taken from interval [-30; +30] with 15º steps, are used to generate five synthetic images. For every synthesized image, the NGDH is built and the probability of *pseudomean* is computed. The five probabilities are compared. The real face orientation, corresponding to the great probability, is sent inside the command *face pose* to the module RSC.

3.3 Automatic Facial Expressions Recognition System (AFERS)

Psychologists identified seven transcultural emotions. In this paper we will only consider five of them (anger, fear, happiness, sadness and the neutral state). Normally, every emotional state has a characteristic facial expressions associated to it. In Fig. 2 are presented examples of facial expressions.

(a) (b) (c) (d) (e)

Fig. 2. Facial expressions associated to some transcultural emotional states: (a) anger, (b) fear, (c) happiness, (d) neutral and (e) sadness.

The AFERS try to recognize local facial actions as defined in FACS [1]. FACS defines a total of 52 Action Units (Aus). Eight of them are related with the head pose.

The remainder 44 concern to facial expressions. Each of these AUs is anatomically related to the activity of a specific set of muscles which produces changes in the facial appearance (in our work, only a sub-set of these AUs is used). A facial expression can be interpreted as a set of specific AUs, which cause "distortions" in facial features (i.e., mouth, eyes, eyebrows or noise). Identified these "distortions", facial expressions can be recognized.

Inside the AFERS, there is a specific module to classify the facial expressions: a Dynamic Bayesian Network (DBN). Fig. 3 presents part of the structure of this classifier of 6 levels.

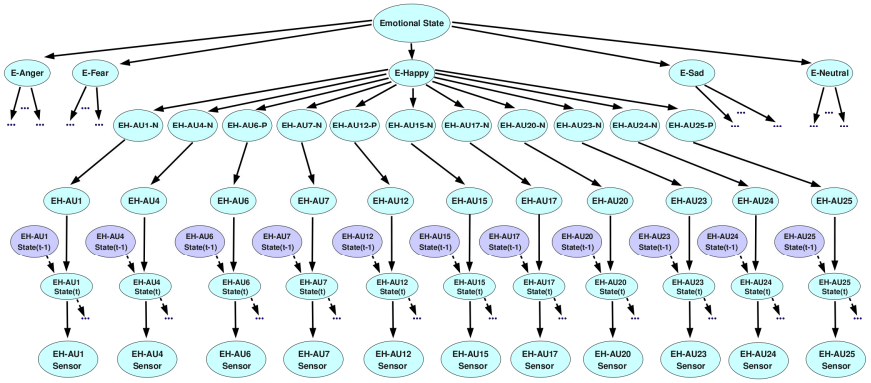

Fig. 3. Dynamic Bayesian Network (DBN) - partial structure. Light nodes are of the present time slice, while dark nodes belong to the previous one (*t-1*). The dashed arrows connect nodes of different time slices: from past to present and from here to future (these nodes are not represented).

At the DBN's first level there is only one node. The overall classification result is provided by the random variable associated with this node: *Emotional State* ∈ *{anger, fear, happy, sad, neutral, other}*. First five events correspond to emotional states that actually are intended to classify. The sixth (*other*) is used for completeness when the facial expression does not fit the five preceding categories.

In the second level there are five nodes, one for every facial expression associated to an emotional state to recognize. To each node is associated one of the following variables: *E-Anger* ∈ *{no, yes}*, *E-Fear* ∈ *{no, yes}*, *E-Happy* ∈ *{no, yes}*, *E-Sad* ∈ *{no, yes}* or *E-Neutral* ∈ *{no, yes}*; where two events are associated with each variable.

After the second level, the structure presented in Fig. 3 is not complete. The figure only shows, from top to bottom, all the nodes and arcs used to classify the facial expression corresponding to happiness (since a similar structure is used to the other emotional states, we avoid to overcharge the figure). In the following, description will only be made to the happiness emotional state.

To the happiness, in the third level of the DBN, there are eleven nodes which are associated with the following variables: *EH-AU1-N* ∈ *{no, yes}*, *EH-AU4-N* ∈ *{no, yes}*, *EH-AU6-P* ∈ *{no, yes}*, *EH-AU7-N* ∈ *{no, yes}*, *EH-AU12-P* ∈ *{no, yes}*, *EH-AU15-N* ∈ *{no, yes}*, *EH-AU17-N* ∈ *{no, yes}*, *EH-AU20-N* ∈ *{no, yes}*, *EH-AU23-N*

∈ *{no, yes}*, *EH-AU24-N* ∈ *{no, yes}* and *EH-AU25-P* ∈ *{no, yes}*. The events associated with these variables are related to the absence or presence of "distortions" (respectively on eyebrows, eyes and mouth) that are relevant to the facial expression. The probabilities assigned to events of these third level variables depend on evidences (positives and negatives) strength.

In table 1 are discriminated, for every facial expression, the AUs that lead to positive or negative evidences. For example, to happiness, the AU6 is positive evidence, but AU7 is negative evidence. From this table it is possible to reconstitute completely the structure of the DBN.

Table 1. Discrimination of the AUs that are considered as Negative evidences (N) or Positive evidences (P) for the facial expressions associated to some emotional states.

	Brows		Eyes		Mouth						
	AU1	AU4	AU6	AU7	AU12	AU15	AU17	AU20	AU23	AU24	AU25
Anger	N	P	N	P	N	N	P	N	P	P	N
Fear	P	P	N	N	N	N	N	P	N	N	P
Happiness	N	N	P	N	P	N	N	N	N	N	P
Sadness	P	P	N	N	N	P	P	N	N	N	N
Neutral	N	N	N	N	N	N	N	N	N	N	N

At the DBN's fourth level are the nodes associated with variables that probabilistically reflect the strength of the evidences (here is not relevant whether they are positive or negative evidences). The variables associated to these nodes are, respectively, *EH-AU1* ∈ *{small, moderate, big}*, *EH-AU4* ∈ *{small, moderate, big}*, *EH-AU6* ∈ *{small, moderate, big}*, *EH-AU7* ∈ *{small, moderate, big}* , *EH-AU12* ∈ *{small, moderate, big}* , *EH-AU15* ∈ *{small, moderate, big}* , *EH-AU17* ∈ *{small, moderate, big}* , *EH-AU20* ∈ *{small, moderate, big}* , *EH-AU23* ∈ *{small, moderate, big}* , *EH-AU24* ∈ *{small, moderate, big}* and *EH-AU25* ∈ *{small, moderate, big}*.

It is through the nodes of the fifth level that information is propagated between time slices. In Fig. 3 this propagation, from past or to future, is represented by dashed arrows, and dark nodes are those of a previous time slice (*t-1*). As an example, *EH-AU1 State(t)*∈ *{minimum, low, medium, high, maximum}* is a variable associated to one node of this fifth level in present. The functionality of the fifth level nodes is to combine / fuse probabilistically, through inertia, information coming from the low level in present time slice with that from the previous instant. After the fusion, the information is used in the present by the high level nodes and is made available to the next time slice to be fused therein.

In the sixth level of the DBN are the nodes associated with variables collecting the evidences provided by the sensors. For example, *EH-AU1 Sensor* ∈ *{minimum, low, medium, high, maximum}*, is one of those variables.

4 Discussion of Results and Critical View

Fig. 4 presents some results obtained by the proposed technique to identify the head orientation when the human is facing the AMA directly. In image (a) is showed a

scene in our laboratory with the human face segmented. From (c) to (g) are the synthesized face images after the application of the 3D transformation and, from (j) to (n), the correspondent NGDHs. The *pseudomean* probability increases as the synthesized image becomes more similar to a frontal face, as it is observable at the histograms. In (h) is illustrated graphically the output as a line segment drawn superimposed to the segmented face. The angle of this segment, referenced to the image vertical axis, corresponds to the real human head orientation. In this case a vertical segment means a frontal face.

Even when the human is not facing the AMA, always one of the synthesized images presents a near frontal face and the real orientation can be found.

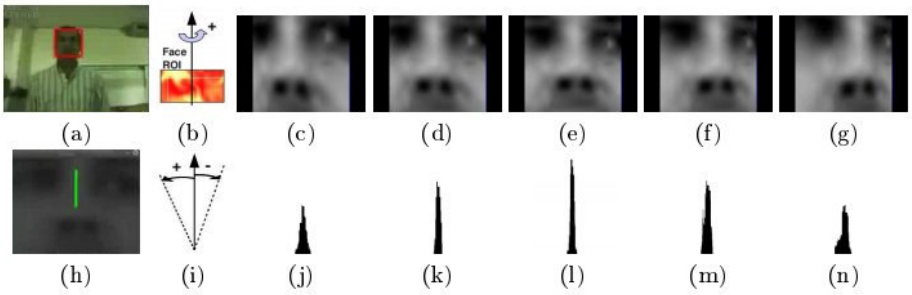

Fig. 4. Working principles of the technique to find the real head orientation. In (a), image of a scene with a frontal face segmented. In (b) is defined the positive direction of the rotations used to synthesize images. Synthesized images after applying a 3D rotation with an angle of (c) *-30º*, (d) *-15º*, (e) *0º*, (f) *+15º* and (g) *+30º*. NGDHs of the synthetic images generated for an angle of (j) *-30º*, (k) *-15º*, (l) *0º*, (m) *+15º* and (n) *+30º*. In (h) is presented the output: a vertical line segment means a *0º* head orientation. In (i) is defined the vertical axis reference of the output image.

The implementation of the presented DBN architecture is actually under test. Reaching conclusions based on negative evidences is something that should be done with special care because it can lead to mistakes. However, unlike other types of classifiers, Bayesian networks can treat negative evidences naturally and correctly through *priors* values near zero. Anyway, in the proposed method, the classification is done using mainly the positive evidences. The two main exceptions are *neutral* and *other* emotional states.

5 Conclusions and Further Work

In this paper is proposed a human face focus of attention technique and a Dynamic Bayesian Network structure to classify facial expressions. Both are to incorporate in an autonomous mobile agent whose high level architecture is also proposed here.

The focus of attention technique has a good performance, is fast and easy to implement. In the future we intend to use it as a valuable help to automatically extract facial feature points without the use of a complex 3D model. Before that, the

technique will be validated more rigorously attaching a inertial sensor to the human head and comparing its outputs with that of the proposed method.

The proposed DBN uses explicitly two types of evidences (negative and positive) and the temporal information is propagated at a low level. The preliminary tests done with the DBN showed that it has great potentialities. We intend to build a simpler version of the proposed DBN (using only positive evidences) and, comparing the two versions, prove the real advantages incoming by the use of negative evidences.

Acknowledgments. The authors gratefully acknowledge support from EC-contract number BACS FP6-IST-027140, the contribution of the Institute of Systems and Robotics at Coimbra University and reviewers' comments.

References

1. Ekman, P., Friesen, W.V., Hager, J.C.: Facial Action Coding System - The Manual. A Human Face. Salt Lake City, Utah, USA (2002)
2. Hiremath, P.S., Danti, A.: Detection of multiple faces in an image using skin color information and lines-of-separability face model. International Journal of Pattern Recognition and Artificial Intelligence, vol. 20 (1), pp. 39--69. February (2006)
3. Chen, X., Flynn, P.J., Bowyer, K.W.: Fully automated facial symmetry axis detection in frontal color images. In: 4-th IEEE Workshop on Automatic Identification Advanced Technologies, pp. 106--111 (2005)
4. Nakao, N., Ohyama, W., Wakabayashi, T., Kimura, F.: Automatic detection of facial midline and its contributions to facial feature extraction. Electronic Letters on Computer Vision and Image Analysis, vol. 6 (3), pp. 55--66 (2008)
5. Datcu, D., Rothkrantz, L.J.M.: Automatic recognition of facial expressions using bayesian belief networks. In: 2004 IEEE International Conference on Systems, Man & Cybernetics, pp. 10--13. October (2004)
6. Zhang, Y., Ji, Q.: Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27 (5), pp. 699--714 (2005)
7. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision, vol. 57 (2), pp. 137--154 (2004)