Facial Gesture Recognition using Active Appearance Models based on Neural Evolution *

Jorge García Bueno
Robotics Lab
Av. Universidad 30
28911 Madrid, Spain
jgbueno@ing.uc3m.es

Miguel González-Fierro
Robotics Lab
Av. Universidad 30
28911 Madrid, Spain
mgpalaci@ing.uc3m.es

Luis Moreno
Robotics Lab
Av. Universidad 30
28911 Madrid, Spain
moreno@ing.uc3m.es

Carlos Balaguer
Robotics Lab
Av. Universidad 30
28911 Madrid, Spain
balaguer@ing.uc3m.es

ABSTRACT

Facial gesture recognition is one of the main topics in HRI. We have developed a novel algorithm who allows to detect emotional states, like happiness, sadness or emotionless. A humanoid robot is able to detect these states with a ratio of success of 83% and interact in consequence. We use Active Appearance Models (AAMs) to determinate face features and classify the emotions using neural evolution, based on neural networks and differential evolution algorithm.

Categories and Subject Descriptors: I.2.6 [Artificial Intelligence]: Learning – *Connectionism and neural nets*

General Terms: Algorithms

Keywords: gesture recognition, HRI, neural evolution, AAM, humanoid

1. INTRODUCTION

Humanoid robot research of last years has produced a remarkable improvement of robot abilities, mobility and interaction capability. An intelligent and skillful robot requires natural interaction and complex behaviors to perform tasks and offer services to humans. It also needs cognitive models to understand human emotions and expressions. Visual recognition of facial gestures can be useful in accomplishing natural and robust human-robot interaction.

Pattern recognition and computer vision techniques have been successfully used for many gesture recognition systems [2] in very different robots. It involves face tracking, shape detection of facial features, clustering, optical flow, optimization and classification. Facial feature extraction techniques rely on the detection and tracking of several face parameters like mouth shape or eyebrows distance. An approach that has been probed very effective is based on Hidden Markov Models (HMMs). Facial muscles variations are modeled and classified and in [3] the facial expression is decomposed in submotions to enhance the performance.

2. ACTIVE APPEARANCE MODELS

Active Appearance Models (AAM), first introduced by [1], is a method to match deformable statistical models of shape and appearance to an image. The algorithm consists in fitting a AAM model to an image minimizing the error between the input image and the closest deformable model.

To build a model, a set of landmarks of enhanced points are selected . With this set of landmarks, a mesh is wrapped to the image using Delaunay triangles. Using a training set of several images and PCA reduction, the model of the shape and the appearance is a linear combination of the shapes and appearance of all training images.

$$\begin{aligned} x &= \bar{x} + Q_s c \\ g &= \bar{g} + Q_g c \end{aligned} \tag{1}$$

Equation (1) represents the shape and appearance models, where $\bar{x}$ is the mean shape, $\bar{g}$ is the mean normalized texture, c is the parameter controlling the shape and appearance and Q_s and Q_p are matrices representing the shape and appearance variations derived from the training image set.

3. NEURAL EVOLUTION

A novel method for pattern recognition is presented based on a genetic algorithm named Differential Evolution (DE) in conjunction with a Neural Network (NN) in charge of the evaluation process. The genetic optimizer minimizes the global error by tuning the weights and biases of the NN.

3.1 Differential Evolution

As in other population based optimization methods, DE generates new points (perturbations) based on previous points. Those deviations are not reflections as other solutions such as CRS or Nelder-Mead. Those new points are randomly selected from three individuals. Two elements x_{r1}, x_{r2} are subtracted and multiply by a weight (weight and mutation)

*A video of the experiments can be found at `http://goo.gl/PSD9Q` When the human smiles the robot reacts with a movement of happiness, if the human is sad, the robot performs a movement of sadness, if the human is neutral the robot waits. The image in the right corner of the video is what the robot sees. This image is sent to a external server where our algorithm is computed.

HRI'12, March 5–8, 2012, Boston, Massachusetts, USA.
ACM 978-1-4503-1063-5/12/03.

F and a third point x_{r3} is added giving the trial vector

$$u_0 = x_{r3} + F \cdot (x_{r1} - x_{r2}) \tag{2}$$

In the selection step, the trial vector u_0 is compared with the rest of the vectors with the same index. This representation is called the *selection and storage* where the lowest cost vector is taken as the member for the next generation. The process is repeated until a population has competed against the trial vector randomly generated.

3.2 Proposed algorithm

NN is trained with DE instead of classical methods such as back-propagation. The optimized values are the weights and biases of input, hidden and output layers, giving as a result the optimal NN. The input parameters are the location of the features for each face obtained using AAM's location algorithm. To initialize the AAM, AdaBoost classifier has been used with Haar-Like features to determine the original position of the face on the frame as it is shown in Figure 3.2.

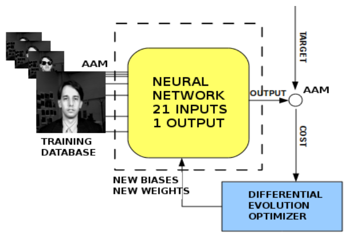

Figure 1: Proposed System mixing AAM with a NN optimized with Differential Evolution. Weights and biases are tuned by the optimizer.

4. EXPERIMENTAL RESULTS

We have constructed a database of 5 faces of several subjects with 20 images per subject, showing different expressions and postures. To compute the AMM[1], 63 characteristic points has been tagged in every image. These points correspond to the important features that define a face, such as eyes, chin, nose, mouth and eyebrows.

The model is trained with all 20 images for every subject, producing a statistical model of shape and appearance. The output is a face tracker, consisting in a mesh of Delaunay triangles which wraps to the subject face (see Figure 4 left). The representation of the deformable face model is computed using splines that connect some of the landmarks (Figure 4 right). Once the AAM model has been obtained, we use that information to characterize facial expressions. The expressions that our algorithm currently identify are happy, sad and neutral (no expression).

The algorithm has been proved in the small humanoid HOAP-3 in real time. The robot detects the face of a human and computes the algorithm. If the human is happy, the robot performs a movement of happiness, moving the arms up and down, if the human is sad, the robot performs a movement expressing sadness, looking down, finally, if the human is neutral, the robot waits. All robot trajectories are preprogrammed and calculated using inverse kinematics. The algorithm is computed in a external server and the mean response is 1.5 sec. In Figure 4 some snapshots of the experiments are shown showing the results of the algorithm.

[1]The AAM algorithm used here is based on the code of Dr. Radhika Vathsan at BITS Pilani Goa Campus, India.

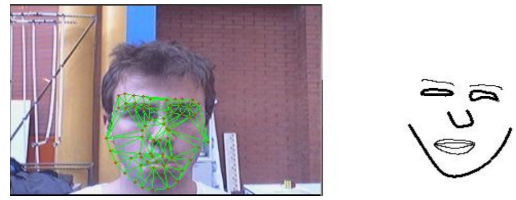

Figure 2: Face tracker mesh (left) and face spline representation (right).

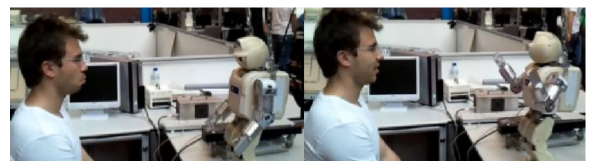

Figure 3: Recognition of happy and sad expressions and interaction of the robot.

In Table 1 a evaluation of the robustness of the algorithm is shown. In a initial phase, the algorithm has been tested using static images. Afterwards, the algorithm has been tested in real time with the humanoid robot. The table shows the only the correct detections.

Table 1: Results of facial gesture recognition

	Static	Real Time	Overall
Happy	96%	90%	96%
Sad	78%	67%	72.5%
Neutral	87%	79%	83%

5. CONCLUSIONS

In this paper a new approach of human-robot interaction has been presented. First, AAM models have been used to obtain a statistical representation of the human face. Extracting some important features of the mesh resulting of the AAM algorithm, like mouth and eyebrows shape, a neural network has been trained in conjunction with a global optimizer to distinguish between three different states, happy, sad, and normal (no emotion).

6. REFERENCES

[1] G. Edwards, C. Taylor, and T. Cootes. Interpreting face images using active appearance models. In *Automatic Face and Gesture Recognition. Proc. Third IEEE International Conference on*, pp 300–305, 1998

[2] Mitra, S., and Acharya, T. Systems, Man, and Cybernetics, Part C: Applications and Reviews. *IEEE Transactions on Gesture recognition*, pp 311–324 vol.37, 2007

[3] Arsic, D., Schenk, J., Schuller, B Submotions for hidden markov model based dynamic facial action recog. In *Image Processing IEEE International Conference on*, pp 673–676, 2006