Multimodal Caricatural Mirror

eNTERFACE
The Similar NoE Summer Workshop on Multimodal Interfaces
July 18th – August 12th 2005
Faculté Polytechnique de Mons, Belgium

Abstract :

This project aims at creating a caricatural mirror where users could see their own emotions amplified (image+speech) by an avatar (mainly facial animation), on a wide screen facing them. It includes multimodal face tracking, multimodal (facial features + prosody of speech) emotion recognition, and multimodal emotion synthesis.

1. Project Objective

The main goal of this project is to build a system that recognizes the user's emotionnal state and displays an avatar expressing the recognised emotion(s) in an exaggerated way. The technical challenges of such a project may be decomposed in several components:

- Multimodal face tracking (using image AND speech to detect the user's face)
- Facial features' extraction (mouth, eyes, eyebrows,…)
- Vocal features' extraction (pitch, energy, speaking rate, …)
- Emotions modelling (in terms of facial/vocal features)
- Multimodal emotion recognition
- Multimodal emotions synthesis

2. Background Information

Researches on automatic emotion recognition and synthesis is currently focusing the attention of an ever-growing community of researchers from various fields (signal processing, artificial intelligence, psychologists, human-computer interactions, …). Many different prototypes of emotion recognition systems have been already developed but it remains very difficult to compare the performances of such systems, due to the lack of common databases and experimenting protocols.

In this project, we will then focus on the development of a system that will involve the user both interactively and emotionnaly, giving the opportunity to users to assess the usability of such a system while at the same time generating real emotionnal experiences on the user's side. This affective response from the user will be used to further train the system, the emotions generated being expressed in a more natural

way than when expressed ‘on demand’, as it is the case for most of the existing databases. We could then view the system as a way of building a multimodal database, which could later be used by researchers of the SIMILAR NoE to compare the performances of their recognition algorithms.

3. Detailed technical description

The project can be divided in several modules :

- Multimodal face tracking
  Face detection and tracking, while being an extremely easy task for humans, remains an important problem for automatic emotion recognition. In this project, we will explore different approaches using both intramodal fusion and multimodal fusion. The techniques proposed so far are using one or several of the following criteria :

  - Skin chrominance information (detecting ‘pink’ patterns)
  - Ellipsoid-shaped properties of head
  - Luminance/chrominance gradient
  - User’s segmentation prior to face tracking
  - An array of microphones and locate the user’s mouth
  - Masks of face and powerful classifiers (Support vector Machines,…)
  - Masks of facial features and infering the face localization
  - …

  In the first part of the project, we will select the most effective techniques and build an efficient multimodal face tracker.

- Facial features extraction
  Another crucial step in emotion recognition is the design of both a reliable and fast facial features extraction algorithm, whose goals is both the localisation of the different facial features (mouth, eyes, eyebrows, nose, …) and the analysis of the emotional information that they contain (through the shape, the displacements,…)

  Several approaches will be compared (Active contours, trace transform, statistical analysis, …) and tested.

- Vocal features extraction
  An abundant litterature relates the influence of emotions on the prosody. In this work-package, we will focus on the techniques used to extract affective non-verbal information from the speech signal and see how this information can be handled to amplify the prosodic aspects of user’s expression of emotions.

- Emotion modeling
  The problem addressed here is to find how emotions are related to both facial and vocal features. *A priori* knowledge may be used to initiate learning

algorithms whose goals would be to learn the relationships between the emotions and their expressions in terms of facial and vocal features.

- Multimodal emotion recognition
  This core module will be responsible for classifying the emotional state of the user, based on the values of the vocal and facial features extracted. The goal is twofold : measure the recognition rate of both modalities (and see which modality is best to detect each of the emotions) and design efficient multimodal fusion algorithms to take into account effectively the influence of both modalities on the decision. An interesting aspect of this work will be to quantify the gain obtained by the joint use of both modalities compared to the performances of the two respective monomodal recognition systems.

- Multimodal emotion synthesis
  The challenge of this group will be to amplify both the displacements of facial features and the prosodic expressions of emotions. The results should be displayed to the user, thereby generating an emotional response. The displayed result can both be seen as a way to measure the efficienty of the recognition module, and as a way to measure the ability of the user to express specific emotions (such a system could be used for instance to train actors to express emotions).

  The ideal system should be real-time for the facial modality to maximize the interactivity, but a real-time system would not be desired for the speech modality (listening to what you are saying while you are saying it is both unpleasant and very disturbing). Therefore, we will introduce a delay between the user's expression and the system reaction (ideally, the system should react as soon as the user stops talking).

4. Organisation

- Equipment
  - One fast computer per participant
  - Firewire Digital camera + tripod
  - Projector + wide screen (for display)
  - Loudspeakers
  - Array of microphones
  - One large room for the experiment

- Software
  - Rapid prototyping languages such as Python or Matlab are ideal tools.
  - C/C++ modules can of course also be handled.
  - OpenInterface will be used to interface the different modules

- Team needed
  Experts in the following fields are welcome to participate in this project :

  - Face detection and tracking
  - Face analysis

- Prosody analysis
- Multimodal systems
- Facial expression recognition
- Emotion synthesis
- Multimodal fusion and fission
- Vocal features extraction
- Human-computer interaction
- Image processing
- Speech processing
- Intelligent data analysis
- … any researcher who may contribute is welcome

- Project leaders : Prof. Benoit Macq (UCL), Olivier Martin (UCL)

  Benoit Macq was born in 1961. He is currently Professor at Université catholique de Louvain (UCL), in the Telecommunication Laboratory. Benoit Macq is teaching and doing is research work in image processing for visual communications. His main research interests are image compression, image watermarking and image analysis for medical and immersive communications.

  Olivier Martin is a research assistant at Communications and Remote Sensing Lab. (UCL) since March 2003. He is involved in the AIDA project ("Autonomy and Intelligence for Dynamic Interactive Applications"). Co-author of several papers about "Mixed Reality Interactive Storytelling", he is now focusing his research on his PhD thesis whose subject is "Multimodal Emotion Recognition for Interactive Applications".

5. References

- Di Fiore F., Van Reeth F. : **Mimicing 3D transformations of emotional stylised animation with minimal 2D input**, *Proceedings of the 1st international conference on Computer graphics and interactive techniques in Australasia and South East Asia, 2003.*

- Chen L.S., Huang T.S., Miyasato T., Nakatsu R. : **Multimodal Human Emotion/Expression Recognition,** *3rd. International Conference on Face & Gesture Recognition, Nara, Japan, 1998*

- Lisetti C., Nasoz F., LeRouge C., Ozyer O., Alvarez K. : **Developing multimodal intelligent affective interfaces for tele-home health care,** *International Journal of Human Computer Studies, 2003*

- Schuller B., Rigoll G., Lang M. : **Hidden Markov model-based speech emotion recognition**, *IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP'03), 2003*