Multimodal Emotion Recognition

Marco Paleari & Benoit Huet

{Paleari, Huet} @ eurecom.fr
Outline

- Motivations
- Possible HCI scenarios and objectives
- Automatic emotion recognition
  - Introduction
  - Approach
    - Audio: speech prosody
    - Video: facial expressions
    - Multimodal emotion recognition
- Conclusions
Motivations

Why Emotions?
Motivations

- **Emotions are important communication means**
  
  - For example **how** one thing is said is sometimes more important than **what** is said.

Search results for “good morning”
Motivations

- Emotions are important communication means
  
  ➢ "Art is a vehicle for the expression or communication of emotions and ideas" (1).

Edward Munch, The Scream

Francisco Hayez, The Kiss

Motivations

- Human memory tends to group together events that present similar emotive meaning
Possible HCI Scenarios

How to use Emotions?
Possible HCI Scenarios

- Emotions could be used for indexing and retrieving media
Possible HCI Scenarios

- Summarization

![Diagram showing emotion recognition with labels: surprise, sadness, happiness.]

Comedy
Tailored
Summary
Possible HCI Scenarios

- Telemedicine
- Gaming
- E-learning
- Ubiquitous / pervasive computing
Automatic Emotion Recognition

Introduction
Automatic Emotion Recognition

- Different possibilities
  - Detect people emotions:
    - Facial Expressions
    - Vocal Prosody
    - Autonomous Nervous System Signals
      (i.e. heart rate, skin conductivity, etc.)
    - Gestures / Posture
    - Others
  - Detect object emotions (music, images, etc.)
    - Colors
    - Tempo / Timber
    - Others
Emotion Recognition From Facial Expression

- Duchenne 1862
  - First studies of facial expressions

- Ekman & Friesel 1971
  - Definition of 6 universal emotions:
    - Anger
    - Disgust
    - Fear
    - Happiness
    - Sadness
    - Surprise
Emotion Recognition From Facial Expression

- Different techniques are possible
  - 3D mesh mapping
    [Essa 1995, Cohen et al. 2000]
  - Dense motion flow
    [Lien 1998, Busso 2004]
  - Feature point movement
  - Facial movement energy estimation
    [Essa 1995]
Emotion Recognition From Vocal Prosody

- Different techniques are possible
  - Pitch and Intensity
  - Spectral Envelop (Formants)
  - Harmonicity analysis
  - Linear Predictive Coefficient (LPC)
  - Mel-Frequency Cepstral Coefficients (MFCC)
  - Speech rate
  - Combination of the above

  ~23/46% ~71%
Inter and Intra Individual Differences

**AVERAGE**

**Inter-individual variability**

**Intra-individual variability**

- Psychological meaning
  - e.g. arousal

- Emotional cue
  - e.g. Heart Rate, pitch, eyebrow shape

- Personality or Mood Dependence
  - [Fiorito and Simons 1994]

- Day Dependence
  - [Picard et al. 2001]
Automatic Emotion Recognition

Approach
Experimental Setup

- eNTERFACE database [Martin et al. 2005]
  http://www.enterface.net/enterface05/

- 44 characters (male and female)
- Each “acting” 6 emotions: Anger, Disgust, Fear, Happiness, Sadness and Surprised
- Each emotions is “played” 5 times by each character
- Resulting in 1420 videos
- Emotion learned on 30 characters acting 4 times each emotion
- Recognition is evaluated on the remainder of the database

[Image: Faces of a person with various expressions]
Experimental Setup
Audio: Speech Prosody
Emotion Recognition From Speech Prosody

Speech Feature Extraction
- Pitch
- Formants
- Energy
- Harmonicity (Harmonics/"Noise" Energy Ratio)
- Linear Prediction Coefficients (LPC)
- Mel-Frequency Cepstral Coefficients (MFCC)

Feature Vector Generation
- Mean
- Standard Deviation
- Variance
- Min (& its position)
- Max (& its position)
- [0.05, 0.25, 0.5, 0.75, 0.95] Quantiles Coefficients
- Polynomial Regression

Classification
- Support Vector Machines
- Neural Networks
- Hidden Markov Models
Audio Result

Result obtained with 1 NN using Polynomial & Statistical analysis and averaging results on 0.6 seconds
Video: Facial Expressions
Emotion Recognition from Facial Expressions
Lukas Kanade Technique

- This method assumes that gray values (n x n feature window) do not change between two consecutive frames, but only shift from one position to another.
Emotion Recognition from Facial Expressions

- Follow points and analyze the movement signals to extract meaningful features

![Image of facial expressions with marked points]
Emotion Recognition from Facial Expressions

- Mean
- Standard Deviation
- Variance
- Min (& its position)
- Max (& its position)
- [0.05, 0.25, 0.5, 0.75, 0.95] Quantiles
- Polynomial Regression

SVM
NN
HMM
Video Result

Result obtained with 6 SVM (one for each emotion)
Example
Example
Example
Multimodal Emotion Recognition
Multimodal Emotion Recognition
Multimodal Result

![Bar chart showing emotion recognition results](image)

Result obtained with NN and data filtering (average on 0.6 sec). Decision Fusion results are obtained by averaging on 0.4 sec.
Conclusion

- Similar results (on average over all emotions) are obtained based on audio and video.

- Simple decision fusion provides more improvement than simple feature fusion.

- More modalities should lead to improved emotion recognition accuracy.
Future Work

- **Video Analysis**
  - Facial Feature Point Detection

- **Classification**
  - Hidden Markov Models

- **Multimodal Fusion**
  - Temporally controlled Multimodal Fusion
Multimodal Fusion
Thank you for your attention