Travellers well-being measuring and dynamic facial expression recognition

Thomas Robin          Michel Bierlaire          Javier Cruz

12th July 2010
The context

- Recent interest for emotion recognition in transportation:
  
  Well-being measuring of users
  - Improve public transportation offers
  - Improve car comfort


  Driving assistance
  - Safety
  - Mobility
The context

- Emotion: *mental* and *physiological* state associated with a wide variety of feelings, thoughts and behavior.

- Emotions signs easy to measure with non-intrusive techniques for transportation users:
  - Behavior
    - Facial expression
  - Voice intonation
The context

Well being measuring

Improve level of service

Driving assistance

Adapt car behavior to a danger
Objectives

- Model the facial expression recognition made by a person looking at a facial video sequence
- Model explicitly the causal effects
- No classification
- Estimate the model on *behavioral* data (relax ground truth assumptions)
Outline

● Introduction
● Data: video
● Features extraction
● Data: behavioral data
● Models
● Model predictions
● Conclusion and Perspectives
Introduction

- Model overview:

  Videos $\rightarrow$ Active Appearance Model $\rightarrow$ Features $\rightarrow$ Model $\rightarrow$ Probability distributions among expressions

  Behavioral data
Videoclips → Active Appearance Model → Features → Model → Probability distributions among expressions

Behavioral data
Data: video database

- The Technical University of Munich database (TUM) Facial Expression and Emotion Database (FEED)

Students faced to a video, natural expressions recorded

138 sequences, 18 subjects
Videos → **Active Appearance Model** → Features → **Model** → Probability distributions among expressions

Behavioral data
Features extraction: Active Appearance Model

Shape, \( x = (x_1, y_1, ..., x_m, y_m)^T \)

(facial action coding system) FACS

Texture (C vectors)

AAM
Videos $\rightarrow$ Active Appearance Model $\rightarrow$ Features $\rightarrow$ Model $\rightarrow$ Probability distributions among expressions

Behavioral data
Data: internet survey

- Survey conducted at the address below (English, French, Italian, Spanish):
  
  http://transp-or2.epfl.ch/videosurvey/

- Respondents have to:
  - create an account
  - label some video sequences with expressions
  - observations

- 1 database of video is used:
  - Facial Expression and Emotion Database (FEED)
Videos → Active Appearance Model → Features → Model → Probability distributions among expressions

Model → Behavioral data
Models: introduction

- 3 models based on different assumptions:
  1. **Reduced** model: only last frame is relevant
  2. **Latent** model: only one frame is relevant
  3. **Smoothed** model: a group of frames is relevant

- Example:
Models: Reduced model

- Example:

- Inspired from the static version of the work:

Models: Reduced model

- Discrete choice model (DCM)
- Choice set: 9 expressions (Happiness, Surprise, Fear, Disgust, Sadness, Anger, Neutral, Other, Don’t know)
- Logit model

Utility specification:
- Alternative specific constants (ASC)
- Facial measures for AUs (FACS)
- Elements of C vectors (outputs of AAM)
Models: Latent model

- Example:

- Combination of 2 DCMs:
  - Instantaneous expression perception sub-model
  - Frames weighing sub-model
Models: Latent model

\[ P_{M_2}(i|o, \theta_{M_2}, \alpha) = \sum_{t=1}^{T_0} P_{M_2}(i|t, o, \theta_{M_2,1}, \alpha)P_{M_2}(t|o, \theta_{M_2,2}) \]

- \( P_{M_2}(i|t, o, \theta_{M_2,1}, \alpha) \): Instantaneous expression perception sub-model (DCM).
- \( P_{M_2}(t|o, \theta_{M_2,2}) \): Video frames weighing sub-model (DCM).
- \( P_{M_2}(i|o, \theta_{M_2}, \alpha) \): Model.
Models: Smoothed model

- Example:

  🔄 Group of frames

- Combination of 2 DCMs:
  - Instantaneous expression perception sub-model
  - Sub-model handling with the detection of the first frame of the relevant group of frames
**Models: Smoothed model**

\[
P_{M_3}(i|o, \theta_{M_3}) = \sum_{t=1}^{T_o} P_{M_3}(t|o, \theta_{M_3,2}) \frac{1}{T_o - t + 1} \sum_{l=t}^{T_o} P_{M_3}(i|l, o, \theta_{M_3,1})
\]

- \(P_{M_3}(i|l, o, \theta_{M_3,1})\): Instantaneous expression perception sub-model (DCM).
- \(P_{M_3}(t|o, \theta_{M_3,2})\): Detection of the first frame of the relevant group (DCM).
- \(P_{M_3}(i|o, \theta_{M_3})\): Model.
Models: Estimation results

- Simultaneous estimation of sub-models by likelihood maximization

- Estimation of the models using codes based on BIOGEME

<table>
<thead>
<tr>
<th></th>
<th>Reduced model</th>
<th>Latent model</th>
<th>Smoothed model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nb of observations</td>
<td>369</td>
<td>369</td>
<td>369</td>
</tr>
<tr>
<td>Nb of parameters</td>
<td>32</td>
<td>45</td>
<td>44</td>
</tr>
<tr>
<td>Null log-likelihood</td>
<td>–810.78</td>
<td>–810.78</td>
<td>–810.78</td>
</tr>
<tr>
<td>Final log-likelihood</td>
<td>–475.79</td>
<td>–441.28</td>
<td>–447.67</td>
</tr>
<tr>
<td>$\bar{\rho}^2$</td>
<td>0.374</td>
<td><strong>0.400</strong></td>
<td>0.394</td>
</tr>
</tbody>
</table>

- Parameters are interpretable and have the good signs:

Videose $\rightarrow$ Active Appearance Model $\rightarrow$ Features $\rightarrow$ Model $\rightarrow$ Probability distributions among expressions

Behavioral data
Model predictions: Reduced model

- Expressions order: H, SU, F, D, SA, A, N, O, DK
Model predictions: Latent model

- Expressions order: H, SU, F, D, SA, A, N, O, DK
Model predictions: Smoothed model

- Expressions order: H, SU, F, D, SA, A, N, O, DK
Conclusion and perspectives

- **Conclusions:**
  - Behavioral approach of the facial expression recognition
  - Pre-validated models
  - Models ready to use for applications

- **Perspectives:**
  - Validation of the models on another dataset
  - Couple the model with a tracker of facial characteristics
  - Applications of the models on a case study
Thanks for your attention

http://transp-or2.epfl.ch/videosurvey/