Affective Computing in Psychology and Computer Science, Challenges and Advancements

Mohammad Bagheri, Negar Namdarian
Technical University of Munich
Department of Informatics

Master Seminar Emotional Awareness in Autonomous driving
Outline

1. Introduction
2. Challenges
3. Proposed Approaches (Emotion Recognition Techniques)
   a. Speech Emotion Recognition
   b. Facial Expression Recognition
   c. Multimodal Systems
4. Conclusion and Summary
Introduction

1. introduction
   a. Affect Definition in Psychology and Computer science
   b. The Necessity of Affective Computing
   c. Levels of Automation in Autonomous Driving
   d. Affective Computing in Autonomous Driving
Affect Definition in Psychology and Computer science:

- **Psychology**: Discrete categories such as fear, sadness, anger, joy, disgust, surprise, love.
- Computer Science: Affect Recognition vs. Sentiment Recognition
- **Affect Recognition**: 
  fine-grained affect recognition, as it aims to classify data to a large set of emotion labels
- **Sentiment Recognition**: 
  coarse-grained affect recognition - binary classification task (positive versus negative)
Affect Taxonomy:
The Necessity of Affective Computing:

- Could improve the systems’ functionality, Artificial intelligence and the ability to respond according to user affective feedback.
- Increase Trustworthiness, Reliability and Security.

# Levels of Automation in Autonomous Driving

<table>
<thead>
<tr>
<th>SAE Level</th>
<th>Name</th>
<th>Narrative Definition</th>
<th>Execution of Steering and Acceleration/Deceleration</th>
<th>Monitoring of Driving Environment</th>
<th>Fallback Performance of Dynamic Driving Task</th>
<th>System Capability (Driving Modes)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0</strong></td>
<td>No Automation</td>
<td>the full-time performance by the human driver of all aspects of the dynamic driving task, even when enhanced by warning or intervention systems</td>
<td>Human driver</td>
<td>Human driver</td>
<td>Human driver</td>
<td>n/a</td>
</tr>
<tr>
<td><strong>1</strong></td>
<td>Driver Assistance</td>
<td>the driving mode–specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the human driver perform all remaining aspects of the dynamic driving task</td>
<td>Human driver and system</td>
<td>Human driver</td>
<td>Human driver</td>
<td>Some driving modes</td>
</tr>
<tr>
<td><strong>2</strong></td>
<td>Partial Automation</td>
<td>the driving mode–specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the human driver perform all remaining aspects of the dynamic driving task</td>
<td>System</td>
<td>Human driver</td>
<td>Human driver</td>
<td>Some driving modes</td>
</tr>
<tr>
<td><strong>3</strong></td>
<td>Conditional Automation</td>
<td>the driving mode–specific performance by an automated driving system of all aspects of the dynamic driving task with the expectation that the human driver will respond appropriately to a request to intervene</td>
<td>System</td>
<td>System</td>
<td>Human driver</td>
<td>Some driving modes</td>
</tr>
<tr>
<td><strong>4</strong></td>
<td>High Automation</td>
<td>the driving mode–specific performance by an automated driving system of all aspects of the dynamic driving task, even if a human driver does not respond appropriately to a request to intervene</td>
<td>System</td>
<td>System</td>
<td>System</td>
<td>Some driving modes</td>
</tr>
<tr>
<td><strong>5</strong></td>
<td>Full Automation</td>
<td>the full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver</td>
<td>System</td>
<td>System</td>
<td>System</td>
<td>All driving modes</td>
</tr>
</tbody>
</table>
Affective Computing in Autonomous Driving

Why do we need Emotion Recognition in Autonomous Driving?

- Increase the acceptance and reliability among users
- Inducing a sense of security and reliability to users
- Detecting drowsiness of the driver while driving [facial analysis]
- Making sure the driver is paying enough attention to the road
- Prevention of following dangerous orders from the driver
- Correcting the misinterpreted command from the driver
- Introducing a more enjoyable and comfortable atmosphere to the cabin for the passengers
Affective Computing in Autonomous Driving

Challenges

Challenges of Emotion Recognition in Autonomous Driving:
- Robustness
- Accuracy
- Reliability
- Security
- Challenges of Affective Computing
  (will be discussed in each Affective Computing methods)
Proposed Approaches
(Emotion Recognition Techniques)

- Speech Emotion Recognition
- Facial Expression Recognition
- Physiological Emotion Recognition Techniques
- Body Gesture and Movement
- Multimodal Systems
Emotion Recognition Techniques

- Emotion Recognition or Affect Recognition
  - Unimodal Features for Affect Recognition
    - Visual Modality
      - Facial Expressions
    - Audio Modality
    - Physiological Emotion Recognition Techniques
  - Multimodal Features for Affect Recognition
    - Multimodal Emotion Analysis (Audio + Video)
    - Multimodal Sentiment Analysis (Audio + Video + Text)
    - Textual Modality (Considered Out of Context for Autonomous Driving)
Proposed Approaches (Emotion Recognition Techniques)

Speech Emotion Recognition:

Recognizing speakers emotional state from speech signal

The main issues:
- selection of suitable Feature set,
- choice of proper classifiers
- preparation of suitable data sets/data bases.
Speech Emotion Recognition:

Applications:
- speech synthesis
- customer service
- education (E-learning)
- Medical purposes (rehabilitation, counseling)
- Entertainment (Music player)
Speech Emotion Recognition:

Speech databases:

1. **Natural speech databases:**
   collected naturally/spontaneously from speakers in real conditions
   example: AIBO database

1. **Acted speech databases:**
   consist of emotional utterances spoken by professional or unprofessional actors
   example: Berlin Database of Emotional Speech
### Speech Emotion Recognition:

#### Speech databases:

<table>
<thead>
<tr>
<th>Corpus Name</th>
<th>language</th>
<th>Type</th>
<th>Emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIBO</td>
<td>German</td>
<td>natural</td>
<td>anger, boredom, empathic, helpless, ironic, joyful, motherese, reprimand, rests, surprise, touchy</td>
</tr>
<tr>
<td>Berlin database of emotional speech</td>
<td>German</td>
<td>Acted</td>
<td>happiness, anger, anxiety, fear, boredom, disgust, neutral</td>
</tr>
<tr>
<td>INTERFACE</td>
<td>English, Spanish, French, Slovenian</td>
<td>Acted</td>
<td>Anger, Sadness, Joy, Fear, Disgust, Surprise, neutral (soft, slow, normal, loud, fast)</td>
</tr>
<tr>
<td>Danish Emotional Speech Database (DES)</td>
<td>Danish</td>
<td>Acted</td>
<td>angry, happy, sad, surprise, and neutral</td>
</tr>
</tbody>
</table>
Speech Emotion Recognition:

Speech emotion recognition System steps:

- Preprocessing: remove noise from signal and obtain high frequency characteristics of signals. Afterwards send signal to Feature extraction and selection steps.

![Diagram of Speech Emotion Recognition System with steps: Emotional speech Input → Feature Extraction → Feature Selection → classification → Emotional speech output]
Speech Emotion Recognition:

speech Feature vectors categorization:

1. short-time(segmental): calculated once for every small time frame(25-50 msec)
2. Long-time(suprasegmental): calculated over the entire utterance duration

1. low-level-descriptions (LLDs): contain prosodic features(suprasegmental) and spectral features(segmental)
2. functionals. LLDs: contain prosodic features(suprasegmental) and spectral features(segmental)
Speech Emotion Recognition:

speech Feature vectors categorization:

1. continuous(pitch, energy,..),
2. qualitative(voice quality such as harsh, tense,..),
3. spectral(LPC, MFCC,..)
4. TEO (Teager energy operator)-based features.
Speech Emotion Recognition:

Feature Extraction Methods:

reduce the dimension of input vector while maintaining the perceptive power of the signal.

- mel-frequency cepstral coefficients (MFCCs)
- linear predictive cepstral coefficients (LPCCs),
- perceptual linear predictive coefficients (PLPs)
Speech Emotion Recognition:

Feature Selection Methods:

- Principal Component Analysis (PCA)
- Canonical Correlation Analysis (CCA)
Speech Emotion Recognition:

![Diagram of speech emotion recognition process]

**Single Classifiers:**

- support vector machine (SVM)
- hidden Markov models (HMM)
- Artificial neural network (ANN)
- K-nearest neighbor
- Gaussian mixture model (GMM)
Speech Emotion Recognition:

Hybrid Classifiers:

- Use of various classifiers in order to improve the performance
Speech Emotion Recognition:

Hybrid Classifiers:
Speech Emotion Recognition:

Challenges:

- selection of best features set
- Different languages, Accents, speaking styles
- Limitations of Speech databases and difficulties of obtaining natural speech database (ethics,..)
Proposed Approaches (Emotion Recognition Techniques)

Facial Expression Recognition:

- one of the most popular and effective approaches in autonomous vehicles
- classifies facial feature deformations and facial motions into emotion categories.
Facial Expression Recognition:

Facial Expression:

could be described by three characteristics:

1. **Onset Phase** (**attack**): the intensity of the muscle activation increases toward the apex phase.

1. **Apex phase** (**sustain**): the plateau when the intensity of the muscle activation stabilizes.

1. **Offset Phase** (**relaxation**): progressive muscular relaxation toward the neutral (no manifestation of muscle activation)
Facial Expression Recognition:

Facial Expression Measurement Methods:

- **Judgment-based approach**: specific facial patterns directly mapped into emotional categories (Anger, Disgust, Fear, Happiness, Sadness and Surprise).

- **Sign-based approach**: facial motion and deformation are coded into visual classes. Facial actions are hereby abstracted and described by their location and intensity.
Facial Expression Recognition:

Facial Expression Analysis steps:

- Face Acquisition
- Facial Feature Extraction
- Facial Expression Classification
- output
Facial Expression Recognition:

Facial Acquisition \[\rightarrow\] Facial Feature Extraction \[\rightarrow\] Facial Expression Classification \[\rightarrow\] output

automatic face detector which can locate faces in any complicated or cluttered background.

Pose and Illumination variations affect the accuracy, Hence some normalizations are recommended before advancing to next steps.
Facial Expression Recognition:

- **Local vs holistic approaches:** In holistic feature expression analysis the whole face is being processed whereas in local feature expression analysis the areas of face which are affected by the facial expressions are being processed.
Facial Expression Recognition:

- **Deformation vs motion based approaches:**

  Motion-based concentrate directly on the facial changes which occur as a result of facial expressions. Deformation based approach requires a neutral image of the face or a face model in order to function. Deformation based approaches extract the facial features which are only relevant to facial actions and not the irrelevant data (wrinkles of old age).
Facial Expression Recognition:

- Face Acquisition
- Facial Feature Extraction
- Facial Expression Classification
- Output

- SVM (Support Vector Machine)
- NN (Nearest neighbor)
- K-NN
- Linear Regression (LR)
- Partial Least Square (PLS)
- Neural Networks
- Rule based classifiers.
Facial Expression Recognition:

Application in Autonomous Driving:
Detecting fatigue in Drivers and hence provide effective methods to improve driving safety.

Fatigue Detection Methods:
- detecting changes and movements of the eyes (driver’s gaze direction, blinking rate or eye closure)
- Assessing mouth shape or head position of the driver

Examples:
- **DAISY (Driver Assisting System)**: monitoring and warning aid for the driver in longitudinal and lateral control on German motorways
- **Copilot**: video based system which measures slow eyelid closure.
Facial Expression Recognition:

Challenges in Autonomous Driving:

- Illumination
- Pose Variations
- Not accurate enough to be used in fully automated vehicles, only warning system
Multimodal Fusion Emotion Recognition

Information Fusion Techniques In Multimodal Affect Recognition

- Feature Level Fusion Or Early Fusion
- Decision Level Fusion Or Late Fusion
- Hybrid Multimodal Fusion
- Model-level Fusion

  - Rule Based Fusion Methods
  - Classification Based Fusion Methods
  - Estimation Based Fusion Methods
1. Feature Level Fusion or Early Fusion:

Features extracted from different channels like audio, text, visual, etc. are fused as a general feature vector and finally the combined features are sent for the analysis.

- **Advantages:**
  
  The correlation between various multimodal features at an early stage can potentially provide better task accomplishment.

- **Disadvantages:**

  The features obtained belong to diverse modalities and can differ widely in many aspects
2. Decision level fusion or late fusion:

Features of each modality are examined and classified separately, and then the results are fused as a decision vector to obtain the final decision.

**Advantages:**

- The fusion of decisions obtained from various modalities becomes easy compared to feature-level fusion, since the decisions resulting from multiple modalities usually have the same form of data.
- Every modality can utilize its best suitable classifier or model to learn its features.

★ Researchers have tended to prefer decision-level fusion over feature-level fusion.
Information Fusion Techniques

3. Hybrid multimodal fusion:

Combination of both feature-level and decision-level fusion methods.

- Advantages:

  Advantages of both early fusion and late fusion techniques!
Information Fusion Techniques

4. Model-level fusion:

It is a technique that uses the correlation between data observed under different modalities, with a relaxed fusion of the data.

- **Rule-based fusion methods:**
  - linear weighted fusion, majority voting, custom-defined rules

- **Classification-based fusion methods:**
  - SVMs, Bayesian inference, Dempster-Shafer theory, dynamic bayesian networks, neural networks and maximum entropy models.

- **Estimation-based fusion methods:**
  - kalman filter, extended kalman filter and particle filter based fusion methods.
Challenges in Multimodal fusion affect computing:

• Continuous data from real noisy sensors may generate incorrect data.

• Identifying whether the extracted audio and utterance refer to the same content.

• Multimodal affect analysis models should be trained on Big Data from diverse contexts, in order to build generalized models.

• Effective modeling of temporal information in the Big Data.

• For real-time analysis of multi-modal Big Data, an appropriately scalable Big Data architecture and platform needs to be designed, to effectively cope with the heterogeneous Big Data challenges of growing space and time complexity.
Conclusion and Summary

**Speech**

*Advantages:* Low Hardware Cost, already existing in cars

*Disadvantages:* not all speech captured by the microphone may be relevant.

Drivers aren’t speaking all the time! → no constant and reliable driver monitoring

**Vision Modality:**

*Advantages:* Omnipresent

*Disadvantages:* it suffers from changing lighting conditions, occlusion and different pose

**Multimodality approach:**

*Advantages:* more robustness to the environment and sensor noise, redundant information improves human performance → computers

*Disadvantages:* Complexity, finding the best approach for fusing multimodal information
Questions?
Thank you for your attention!