4M20 - Introduction to Robotics

Coursework - Part D

Human-Robot Interaction

Chapa Sirithunge, Fulvio Forni

Michaelmas 2025

Objective: Facial expressions play a central role in conveying human intentions during interaction. In this coursework, we examine how these expressions can be quantitatively analyzed using information-theoretic measures to inform robots in human–robot interaction (HRI). Specifically, micro-expressions captured as Facial Action Units (AUs) through OpenFace are used to construct probability distributions, from which entropy and information rate are calculated. This task highlights how tools such as OpenFace can be combined with fundamental concepts of information theory to measure the richness of human–robot communicative signals.

Task 1: Dataset creation

A dataset containing emotion data, suitable for analysis with information-theoretic methods, will be generated.

(i) Install OpenFace 2.0. This is an open-source facial behavior analysis toolkit. This will be used to extract facial features relevant for examining different emotions from video data through information-theoretic equations.

(ii) Download publicly available datasets for recorded facial expressions that corresponds to different emotions: Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS), CMU-MOSEI.

(iii) Select four video clips, each representing a standard emotion for evaluation (e.g., anger, happiness, disgust, surprise, sadness, fear, neutral). In your report, include representative frames from each video clip to illustrate the nature of the emotion.

(iv) Use OpenFace 2.0 to extract features from each video clip, such as eye gaze, head pose and action units (AUs) for right and left sides of the face.

(v) Extract all available AUs for your analysis. Save the extracted data to a CSV file or a data structure. Use this file for all subsequent analyses.

Task 2: Dataset preparation

Write a Python code that opens the CSV file and normalizes AUs per frame, using the following equations. Let $a_{t,j}$ denote the raw intensity of AU_j at frame t, with $j = 1, \ldots, k$ (where k is the number of AU features). First compute:

$$s_t = \sum_{j=1}^{k} a_{t,j}. \tag{1}$$

Then define the normalized distribution:

$$p_{t,j} = \begin{cases} \dfrac{a_{t,j}}{s_t}, & s_t > 0 \\ 0, & s_t = 0. \end{cases} \tag{2}$$

Store these quantities for the next steps.

Task 3: Quantitative analysis of the dataset

For each video clip, develop the following analysis (keep results separated).

(i) Based on the normalized data, compute the Shannon entropy

$$H_t = -\sum_{j=1}^{k} p_{t,j} \, \log_2(p_{t,j}) \tag{3}$$

with the convention $0 \log 0 = 0$.

(ii) Compute the mutual information

$$I(t; t + \Delta t) \approx I_t := H(t + \Delta t) - H(t) \tag{4}$$

(iii) Compute information transfer rate R_t. Two cases exist:

- If timestamps are present:

$$R_t = \frac{H_{t+1} - H_t}{\Delta t_t}, \quad \Delta t_t = \text{timestamp}_{t+1} - \text{timestamp}_t \tag{5}$$

- If timestamps are not available (assume 30 fps):

$$R_t = 30\big(H_{t+1} - H_t\big) \tag{6}$$

(iv) Calculate the channel bandwidth (B) across both H(t) and R(t). For each case, you may assume H(t) and R(t) at fs = 30 Hz, remove the mean and compute the FFT-based power spectrum using only positive frequencies (DC excluded). The spectrum can be normalized to unit area. Bandwidth is defined as the half-power width—the frequency interval where power is at least 50% of its peak—and reported as the width of that interval. Series that were too short or had negligible power can be returned as NA/0. Plot bandwidth for H(t) and R(t) for the selected four cases.

In the report, for each video clip, include plots of H_t and R_t. These will reveal where expressions become more complex or dynamic, linking observable behavior with information-theoretic descriptors.

Task 4: Discussion

In your report, answer/discuss in detail the following questions. Support your argument with plots, images, diagrams, and mathematical formulas. State any assumptions you made.

(i) Temporal Dynamics: How did the information flow H_t and R_t vary over time? Correlate specific events in your videos (e.g., a punchline in a joke, a moment of disagreement) with peaks or dips in your plots.

(ii) Emotion Structure: What does a high Mutual Information value suggest about the relationship between two emotions, e.g: anger vs happiness?

(iii) Communication Channel: What does a high or low average Bandwidth tells about the chosen scenario?

(iv) Communicative Effectiveness: Based on your analysis, which of selected emotions seemed to have a higher information content?

(v) Encoding for Robotics: Propose a simple method for a robot to perceive emotions using chosen feature such as AUs, gaze etc. in real-time.

(vi) Application to Human-Robot Interaction: How could a robot change its behavior based on the quantitative analysis of real-time visual data. Discuss two scenarios. Support each scenario with one-two references from literature.

Useful links and/or bibliography

1. Ash, Robert B. *Information theory.* Courier Corporation, 2012.
2. Jack, Rachael E., Oliver GB Garrod, and Philippe G. Schyns. "Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time." Current Biology 24.2 (2014): 187-192.
3. Breazeal, C. (2003). Emotion and sociable humanoid robots. International journal of human-computer studies, 59(1-2), 119-155.
4. Coupé, C., Oh, Y. M., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances*, 5(9), eaaw2594.

Your Report

- Please refer to the Moodle website for the exact submission deadline for this report. Be aware that penalties will be applied to all late submissions

  Undergraduates must use only their Coursework Candidate Number (CCN) and must not include their name in the report or file name. MPhil and postgraduate students must include their name and CRSID on their reports.

- Please submit your report in the preferred four-page, double-column format. Templates for both LaTeX and Microsoft Word are available on the Moodle website. Your report should be a concise exposition that focuses on the salient features of your findings. Ensure that each figure is properly annotated and directly contributes to your explanations. While the evaluation will be based solely on the four pages of the report, you are welcome to include an appendix for any additional materials.

- Please upload an archive file with your code and/or simulation files. Do NOT upload data!

- This assignment is worth 25% of your final grade, with an expected time commitment of approximately 10 hours. For any questions regarding this assignment, please contact Chapa Sirithunge (csh66) and Fulvio Forni (ff286).