Digital Systems for Humans Graduate School

**2024-2025 PhD Subject Proposition**

Proposition de Sujet de Thèse 2024-2025

# Fine-grained analysis of emotions and human behavior based on multi-sensor fusion

**Doctoral School:** Doctoral School in Law, Political Science, Economics and Management (ED DESPEG)

**Thesis supervisor: Eric GUERCI et François BREMOND**

**Host laboratory: GREDEG-CNRS (UMR 7321)**

**Subject description:**

In the last decade, scientific research has made significant advancements in investigating the role of emotions in social behavior by observing physiological, brain signals, facial and language expressions. Cheap, powerful and unobtrusive devices or solutions to collect data about human behavior provide large data-sets to learn from with cutting-edge machine/deep learning algorithms. We are already developing this original research project in the unique laboratory CoCoLab at University Côte d'Azur in Nice. The candidate is expected to learn to set-up ad-hoc experiments with humans and to develop efficient machine/deep learning algorithms to detect fine-grained human emotions capable of predicting strategic human behavior. The candidate will be part of a dynamic and engaging context with other PhD and Post-Docs.

**The experimental challenge**

A social interaction between two participants[1] (bargaining, negotiation, discussion, trade, …) can be reproduced in a laboratory while participants are observed (video and voice signals) through cameras (normal and infrared) and microphones, unobtrusive devices to measure physiological activity (EDA, ECG) and neural activity (dry helmet EEG). Three different scenarios can be configured in the laboratory to investigate different modalities of social interaction:

1. total anonymity: participants don't see each other, but they interact through a computer (screen-keyboard-microphone)
2. physical distance: participants see each other through a webcam, but they interact through a computer (screen-keyboard-microphone-camera)
3. proximity: participants see each other in presence, and they can interact by normal communication ((screen-keyboard-microphone-camera-body gestures)

In all scenarios, participants can execute the same tasks, but the engagement and contextual information are different thus altering the outcomes of the decision-making. For instance, in scenarios 2 and 3 the presence of visual cues about the opponent's face can have a relevant impact in the social dynamic.

The experimental strategy can be summarized as:

[1] The experiment could be extended to involve a larger group of participants.

1) Can we predict choices/behavior by observing the participants bodily/physiological activity and expressions?

We aim to adopt multi-modal deep-learning architectures to learn this task by integrating a large feature space comprising physiological signals, facial expressions, EEG activity, bodily movements ...

2) Can we learn to reduce significantly the dimensionality of the feature space, to enable a prediction on a subset of relevant signals, in particular facial expressions?

**The machine learning challenge**

This work consists in the improvement of Emotion Recognition/Human Behavior Coding algorithms using RGB video cameras at test time, but using multi-modalities at training time [6, 8].

The objective is to develop and test a model on multiple datasets with various modalities to characterize human behavior during interactions [2] and to identify specific emotions, such as stress, anxiety, joy, empathy. The approach will consist of advanced Deep Learning methods for combining multimodal inputs, comparing various strategies such as multi-task learning, Knowledge Elicitation (infusion) using Student-Teacher paradigm, contrastive learning and co-training or Transformer. Several levels of ground truth (GT) supervision (e.g. weak-supervision) will be used to trained the model.

Typical pipeline can combine transformers for 3D pose, eye-gaze and facial expression estimation, depending on the emotions to detect [1, 4, 5, 7]. Short temporal aspects of the actions can be handled through TCN or 3DCNN. The objective of this first step is to extract meaningful mid-level features that can be further processed thanks to more long-term reasoning based on TCN or Transformers or even ontology-based reasoning. A challenge will be to propose an approach to leverage the knowledge acquisition process and the long-term reasoning with a weakly supervised setting.

This work aims at reducing the supervision to conceive a general and robust algorithm enabling the detection of the emotions of an individual (together with his/her facial expressions) living in an unconstrained environment and observed through a limited number of sensors (restricting to a single video camera).

Given the scarcity of the real-world emotion data, another challenge is to design learning algorithms pre-trained on large datasets [3], which will allow to transfer learned behavior-patterns onto unseen subjects, such as potential customers or patients. In this direction, we intend to explore domain adaptation, transfer learning, as well as metric learning. Domain adaptation and transfer learning have been able to mitigate dataset noise and to increase cross-dataset accuracy.

The impact of this research project is broad for social sciences; indeed, it is relevant for economics (game theory, negotiation, bargaining, …), marketing science (consumer behavior, …) as well as for computer science. The supervisors are involved in these research domains.

**References:**

[1] Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

[2] Damasio A: Descartes’ Error: Emotion, Reason, and the Human Brain. Avon; 1994.

[3] Tasha Poppa, Antoine Bechara, The somatic marker hypothesis: revisiting the role of the ‘body-loop’ in decision-making, Current Opinion in Behavioral Sciences, Volume 19, 2018, Pages 61-66, ISSN 2352 1546,https://doi.org/10.1016/j.cobeha.2017.10.007. https://www.sciencedirect.com/science/article/pii/S2352154617300736

[4] Barrett, L. F. (2017). How emotions are made: The secret life of the brain.

[5] Hugo D Critchley, Sarah N Garfinkel, Interoception and emotion, Current Opinion in Psychology, Volume 17, 2017, Pages 7-14, ISSN 2352-250X,

[6] Kandasamy, N., Garfinkel, S., Page, L. et al. Interoceptive Ability Predicts Survival on a London Trading Floor. Sci Rep 6, 32986 (2016). https://doi.org/10.1038/srep32986

[7] D, Yang, Y. Wang, A. Dantcheva, L. Garattoni, G. Francesca, and F. Bremond. View-invariant Skeleton Action Representation Learning via Motion Retargeting. International Journal of Computer Vision, VISI, ISSN: 0162-8828, VISI-D-23-00139R1, Jan. 2024.

[8] M. Balazia, P. Muller, A. Levente Tánczos, A. von Liechtenstein and F. Bremond. Bodily Behaviors in Social Interaction: Novel Annotations and State-of-the-Art Evaluation. In Proceedings of the 30th ACM International Conference on Multimedia, ACM-Multimedia 2022, Lisbon, 10-14 October, 2022.

[9] T. Agrawal, M. Balazia, P. Muller and F. Bremond. Multimodal Vision Transformers with Forced Attention for Behavior Analysis. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, video, WACV 2023, Hawai, USA, January 3-8, 2023.

[10] D. Yang, Y. Wang, A. Dantcheva, L. Garattoni, G. Francesca and F. Bremond. Self-supervised Video Representation Learning via Latent Time Navigation. In Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI-23, Washington DC, February 7-14, 2023.

[11] D. Yang, Y. Wang, A. Dantcheva, Q. Kong, L. Garattoni, G. Francesca and F. Bremond. LAC - Latent Action Composition for Skeleton-based Action Segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, October 2-6, 2023.

[12] P. Müller, M. Balazia, T. Baur, M. Dietz, A. Heimerl, D. Schiller, M. Guermal, D. Thomas, F. Brémond, J. Alexandersson, E. André and A. Bulling MultiMediate'23: Engagement Estimation and Bodily Behaviour Recognition in Social Interactions. In Proceedings of the MultiMediate Challenge: Multi-modal Group Behaviour Analysis for Artificial Mediation, part of the 30th ACM International Conference on Multimedia, ACM-Multimedia 2023, Ottawa, 29 Oct - 2 Nov, 2023.

[13] R. Dai, S. Das, M. Ryoo and F. Bremond. Attributes-Aware Network for Temporal Action Detection. In Proceedings of the 34th British Machine Vision Conference, BMVC 2023, Aberdeen, UK, 20th - 24th November 2023.

[14] H. Chaptoukaev, V. Strizhkova, M. Panariello, B. Dalpaos, A. Reka, V. Manera, S. Thummler, E. Ismailova, N. Evans, F. Bremond, M. Todisco, M.A. Zuluaga, and L. Ferrari. StressID: a Multimodal Dataset for Stress Identification. In Proceedings of the NeurIPS 2023 Datasets and Benchmarks Track, part of the Thirty-seventh Conference on Neural Information Processing Systems, NeurIPS 2023, New Orleans, 11th - 14th December 2023.

Candidater / Apply