

Image and text processing for multimodal film analysis

Keywords

Computer vision, text processing, data representation, film analysis

Location

Centre Inria d'Université Côte d'Azur

Supervisors

- Hui-Yin Wu, Chargée de recherche (ISFP), Inria, [https://www-sop.inria.fr/members/Hui-Yin.Wu/](https://www.sop.inria.fr/members/Hui-Yin.Wu/)
- Lucile Sassatelli, Maîtresse de conférences HDR, IUF, Université Côte d'Azur, CNRS I3S, <https://www.i3s.univ-cotedazur.fr/~sassatelli/>

Contact

- hui-yin.wu@inria.fr, + 33 4 92 38 79 28
- lucile.sassatelli@univ-cotedazur.fr , +33 4 89 15 43 47

Duration

3 to 6 months

Description

This internship lies in the framework of ANR TRACTIVE, which is a national-funded project that regroups researchers from computer science, media studies, linguistics, and gender studies for the understanding of gender representation in visual media such as film. We integrate AI, linguistics, and qualitative media analysis in an iterative approach that aims to pinpoint the multimodal discourse patterns of gender in film, and quantitatively reveal their prevalence.

The objective of this internship is to provide crucial support at the beginning of the project to establish a workflow for the collection, curation, and analysis of film and video data.

This project offers hands-on experience on a wide variety of multimedia processing tools, experience in developing code for deployable and open frameworks, and the opportunity to work on a project of large scale and strong societal impact. Specifically, it will involve the following steps:

1. Processing of low-level features in a film corpus to extract:
 - image information such as face tracking [1], motion from optical flow, and object detection [2]
 - textual processing such as word indexing, stemming, and sentiment analysis using libraries such as Lemur [3] and NLTK [4]
2. Exploration of a data format for representing multimodal (text and image) film data and generating embeddings for machine learning models
3. Analysis and transformation of low-level features into high-level annotations on film style and storytelling [5]
4. Training preliminary learning models such as decision trees and random forests to establish baselines for film style classification

Pre-requisites

Mandatory:

- Excellent level of Python programming
- A background in machine learning

Highly recommended

- Good knowledge of methods for image segmentation and analysis
- Knowledge on building and maintaining databases
- Knowledge in software engineering, software architecture
- Experience with natural language processing toolkits and libraries

References

- [1] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition.
- [2] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
- [3] The Lemur Project <http://www.lemurproject.org/lemur/background.php>
- [4] Natural language toolkit <https://www.nltk.org/>
- [5] Wu, H. Y., Palù, F., Ranon, R., & Christie, M. (2018). Thinking like a director: Film editing patterns for virtual cinematographic storytelling. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(4), 1-22.