Deep visual semantics for multimodal film analysis

Keywords

Computer vision, data representation, film analysis

Location

Laboratoire I3S, Université Côte d'Azur, CNRS Sophia Antipolis

Supervisors

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

Contact

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

Duration

3 to 6 months

Target

Level: M1

Salary

French internship gratification (ca. 600€/month)

To apply

Send CV, master and bachelor transcripts to above email addresses

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 [1,5].

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. Extraction of scene elements: survey and test of existing models for character tracking and identification [2], object detection and tracking [3], human pose estimation [4].
2. Extraction of cinematographic elements [5,8]: survey and test of existing models for 3D camera motion type and camera pose [1,6,8], shot-type detection [7], frame layering estimation [8].

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

References

[1] J. Suchan and M. Bhatt, “The geometry of a scene: On deep semantics for visual perception driven cognitive film, studies,” in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Mar. 2016, pp. 1–9. doi: 10.1109/WACV.2016.7477712.
[2] Omkar M. Parkhi, Esa Rahtu, Qiong Cao, and Andrew Zisserman. 2020. Automated Video Face Labelling for Films and TV Material. *IEEE Trans. Pattern Anal. Mach.*

Intell. 42, 4 (April 2020), 780–792. DOI:https://doi.org/10.1109/TPAMI.2018.2889831
[3] 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).
[4] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, Jul. 2017, pp. 1302–1310. doi: 10.1109/CVPR.2017.143.
[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.
[6] T. Gu, Z. Wang, D. Li, H. Yang, W. Du and Y. Zhou, "OnionNet: Single-View Depth Prediction and Camera Pose Estimation for Unlabeled Video," in IEEE Transactions on Cognitive and Developmental Systems, vol. 13, no. 4, pp. 995-1009, Dec. 2021, doi: 10.1109/TCDS.2020.3042521.
[7] M. Svanera, S. Benini, N. Adami, R. Leonardi and A. B. Kovács, "Over-the-shoulder shot detection in art films," 2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI), 2015, pp. 1-6, doi: 10.1109/CBMI.2015.7153627.
[8] Robin Courant, Christophe Lino, Marc Christie, Vicky Kalogeiton. High-Level Features for Movie Style Understanding. ICCV 2021 - Workshop on AI for Creative Video Editing and Understanding, Oct 2021, online, France. pp.1-5.