

Movie and Script Alignment without Dialogues



Supervisor: Rémi Ronfard, IMAGINE team, remi.ronfard@inria.fr Context

The alignment of a movie and its script provides insight into the movie making process and can be used to build annotated data sets useful in many areas, including actor and action recognition. A common approach, first described by Ronfard and Thuong [1] consists in separating dialogues from action descriptions, and aligning the dialogues with the subtitles in the movie. Cour et al have taken this approach further by using the alignments as a starting point for also aligning scene breaks and for naming actors in the movie [2]. One problem with such approaches is that they are limited to dialogue scenes. In the case of actions scenes with little or no dialogue, the alignment needs to be performed based on visual cues such as object and actor appearances, actor movements and actions.

Objectives

In this master's thesis, we will explore deep learning methods used in sequence-to-sequence machine translation [3,4,5] and adapt them to the case of movie/script alignment by discovering frequently occurring visual cues and motion patterns from the video [6] that can be mapped to actor names and motion verbs in the script.

The candidate will have to answer the following questions : (1) which elements of the script and the movie can be matched to each other manually ? (2) which subset of those matches can be obtained automatically, using state of the art computer vision ? and (3) assuming that an optimal alignment between the text and the movie is one that maximizes the number of correct matches, how can such an optimal alignment be obtained in finite time using dynamic programming techniques ?

For evaluation, scenes from the silent movie "the artist" will be annotated and aligned with its (English) script, both manually and automatically.

The Master thesis is expected to lead to a PHD thesis on a related topic in 2018-2020.

For this thesis, we are looking for candidates with a strong background in machine learning and an interest in computer vision and natural language processing.

References

- [1] Remi Ronfard, Tien Tran Thuong. A framework for aligning and indexing movies with their script. Multimedia and Expo, 2003. ICME'03.
- [2] Timothee Cour, Chris Jordan, Eleni Miltsakaki, Ben Taskar. Movie/Script: Alignment and Parsing of Video and Text Transcription. ECCV 2008.
- [3] Ilya Sutskever, Oriol Vinyals, Quoc V. Le. Sequence to Sequence Learning with Neural Networks. NIPS, 2014.
- [4] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. International Conference on Learning Representations, 2015.
- [5] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, Yann N. Dauphin. Convolutional Sequence to Sequence Learning. arXiv, 2017.
- [6] Daniel Weinland, Remi Ronfard, Edmond Boyer. Automatic Discovery of Action Taxonomies from Multiple Views. CVPR 2006.