

# Comparing videos

---

*M2R Internship proposal, September 2015*

Contact : Rémi Ronfard, IMAGINE team, INRIA/LJK Grenoble  
<http://team.inria.fr/ imagine/remi-ronfard/> [remi.ronfard@inria.fr](mailto:remi.ronfard@inria.fr)

## Context

The goal of the internship will be to propose novel and original methods for comparing two videos with similar contents, such as multiple performances of a theatre play, or multiple rehearsals of a performance, or multiple takes of a scene during the shooting of a movie. Conceptually, the goal of the internship will be to provide a “diff” function [1,2] for comparing different “versions” of a video, as in a software versioning system (Git, Mercurial, SVN). This includes methods to compute frame-by-frame similarities between the two videos; methods to compute optimal alignments between the two videos; and methods to display the aligned videos to properly convey the differences and similarities between them.

## Objectives

In dialogue scenes, alignments can be computed automatically from the audio track using speech processing [3]. The general problem of video alignment is much harder because it requires multimodal dynamic time warping [4]. Recent work [5] has shown excellent performance in computing video alignments in cases when the camera is moving, and the background motion provides useful information.

In the case of dramatic performances, as in theatre or film, a different set of image features will be needed to better capture the similarities and differences between actor performances. Actor-specific features such as [6] can be used to track the movements of actors reliably. The internship will investigate such features in the context of video alignment. A high-quality dataset of theatre actors recorded during rehearsals and live performances will be made available for training and testing.

## References

1. Eugene W. Myers. An  $O(ND)$  Difference Algorithm and Its Variations. *Algorithmica*, 1986.
2. John W. Ratcliff and D. E. Metzener. Pattern Matching: The Gestalt Approach. *Dr. Dobb's Journal*, 1988.
3. Sakoe and Chiba. *Dynamic programming algorithm optimization for spoken word recognition*. *IEEE Trans. on Acoust., Speech, and Signal Process.*, ASSP **26**, 43-49 (1978).
4. Feng Zhou and F. Della Torre. Generalized time warping for multi-modal alignment of human motion. *CVPR* 2012.
5. Oliver Wang, Christopher Schroers, Henning Zimmer, Markus Gross, and Alexander Sorkine-Hornung. 2014. VideoSnapping: interactive synchronization of multiple videos. *ACM Trans. Graph.* 33, 4, Article 77 (July 2014).
6. Vineet Gandhi and Remi Ronfard. 2013. Detecting and Naming Actors in Movies Using Generative Appearance Models. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*.