Automated Video Object Removal System

Degree programme : MAS | Specialisation : MAS Data Science

Vehicle simulation systems which use actual route footage often do not have full control over what appears in the video. In this project, a prototype automated video editing system is demonstrated which can remove common objects observed in street traffic situations. The end result is an inpainted video which can be used for further simulation purposes.

Background

Locomotive simulators generally utilise a combination of real-world hardware and in-cab video footage obtained for the particular simulation route. Through later post-processing, the video stream is augmented with various image overlays to simulate traffic signal controls along the rail route. The video footage obtained insofar, is based solely on the particular instance when it was taken. It may contain scenes with vehicle or pedestrian traffic which may or may not be desirable. Ideally, the simulation engineer should have the ability to remove unnecessary objects (i.e. people or vehicles) in order to focus the simulation towards the locomotive aspect.

In this project, a prototype video editing system is demonstrated. This system allows for the automated post-processing of video files in order to remove common objects such as people and vehicles. The end result is an edited video in the original format where the objects have been replaced with the background.

Project Requirements:

For street traffic situations, the video editing platform must address the removal of commonly observed items (and their accompaniments):

- 'person', 'handbag', 'backpack', 'suitcase'
- 'car', 'truck', 'bus', 'motorcycle', 'bicycle'

For the platform, the simulation engineers must focus their attention towards the aspects of hardware and signal simulation. The introduction of a video editing system for object removal should aid their process, and not require extensive overhead in terms of maintenance or interaction. Taking this into consideration, the requirements of the system are:

- Allow a high degree of automation for video editing
- Reduced complexity of the interface
- Accessibility from various platforms

Product Development:

The system utilises a three-step approach for editing: (1) detection, (2) mask grouping, and (3) inpainting. The detection step involves the identification of objects within the video based upon a given object class definition, and the production of pixel level masks. Subsequently, the object masks are grouped and tracked through the frame sequence to determine persistence and allow correction of classified results. Finally, the grouped masks are used to target specific objects instances in the video for inpainting removal.

The system utilises separate detection and inpainting models -- each based on Pytorch and CUDA -- but with different versions of these libraries. In order to accommodate each model's platform requirements, a Linux-based NVIDIA Docker container system was used. Each model operates from within its own container, and communicates with the other via SSH protocols and a common file system. To facilitate interaction with the user, a web interface has been implemented. This allows remote use of the system, with full platform independence at the user side.

Results:

The end result is a demonstration of the video editing platform in the context of locomotive route simulation. The detection, grouping, and inpainting results are accessible in separate steps to allow pre-emptive user interaction. The final video output demonstrates the system's ability to automatically remove moving pedestrians in a video sequence, which commonly occur in most street tram simulations. This work also addresses the limitations of the system, in particular, the inability to remove quasi-stationary objects. The overall outcome of the project is video editing system with automation capabilities equal to or surpassing most commercial inpainting software.

Daniel Cameron

dan.cameron.2007@gmail.com