Introduction to Media Computing

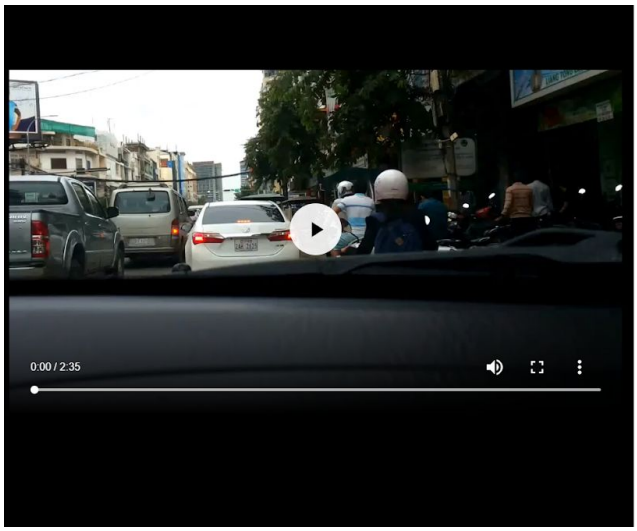
Trimming the Fat

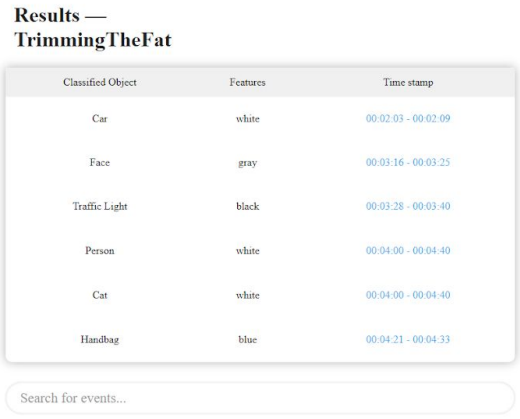
Deep learning based searching in videos

Palash Bansal (2014072), Mridul Gupta (2015061)

Introduction

Finding clips of interest in videos such as that from CCTV footage is an extremely important problem for many companies and security agencies. Many techniques are being used to do it more efficiently with the least amount of human effort. Techniques include merging multiple time snippets into one so that they can be analysed faster by human analysts, other ways are to tag multiple segments to create an index which can then be searched to improve speeds. But with the recent advancements in deep learning, this problem is more easily solvable by automatic detection and tagging. According to our research, till this day, no proper consumer based solution exists to tackle this problem in the right way.





About

Our application is a software to analyze a video file or feed and allow the user to search for objects in it. It is also extended to recognize object color and recognize specific people in the video.

Technologies:

The application uses the YOLO model trained with the COCO dataset to detect various objects in the image. After the objects are detected, they are cropped and further detection is performed using more models. The complete framework that we developed is extendible and as a proof of concept we have implemented color detection and face recognition. One is based on K-Means clustering and the other on Deep learning, hence showing that model of any kind can be applied to the object to get the features.

The application hosts a python flask backend for interaction with the user. On the website, the users can upload a video file for analysis, further view and interact with the results using a clean and intuitive user interface.

In the application, the user will analyse the video, get basic description and can query strings which would search and return a list of all the matched video segments.

User Interface

The application has a very intuitive user interface that can be used to interact with the video and the data. On the left, the video is played and can be controlled by normal video controls. On the right, a table shows the list of all the detected objects in a timeframe, with their features like color or detected faces and the timestamp when they occur. Users can click on the timestamp link to seek to the appropriate time and view the video from there.

With the search bar, users can type and search for various tags and get a list of all the objects matching that search term.

Color Detection

To detect the color of an object we are using K-Means clustering algorithm to find the most dominant color in the cropped part of the whole frame where the object is present. K Means clusters the complete image into 3 color values and the most dominant one is taken.

The color value is then converted to a corresponding english name by first matching the color to the closest one by Euclidean distance to a set of known colors. The nearest color is then taken and the function returns the color name.

Face recognition

The application does face recognition on all the frames to find the known faces. The user can put all the known faces in the known_faces folder. These images are automatically queried and matched with the faces in the video.
Face recognition is performed by precomputing the face encodings and storing those in respective files. These encodings are then matched with the detected face encodings from the video.

Evaluation and State of the Art:

We couldn't find any consumer based product that uses deep learning to search in videos with efficiency.

Companies which are involved in a similar kind of work include:

- Ella: Web-based AI but needs a ICR Box products on the camera network
- Camio: Too hard to able to be used by the general media, hence not suitable for many non-tech-savvy markets like India.

Seeing the existing products in the market, we feel that what we have developed is unique. Most models run to determine a finite set of classes, but with model chaining, we can detect a vast number of objects with proper searching functionality. Plus with the addition of a proof of concept User Interface, the analysis becomes fast.

Business Model

There are many companies that provide video analytics solutions, but since Deep Learning has broken all benchmarks with accurate tagging and classification, most commercial solutions perform poorly in comparison. With this project, our aim is to make a full fledged video searching and tagging software package that can be used by companies and government organizations that operate in the non-tech domain.

The software is developed with model chaining, such that new machine learning or general image analysis models can be chained to detect further features of specific objects. This is demonstrated in our project with face recognition and color extractor features.

In our business model, we plan to make 2 products with different payment models.

1. Server-based analytics: This is for small enterprises and individuals who would like to test the product or perform analysis on a small set of videos. Here the user will be able to upload a video file onto the website and be able to view the results in a web-based environment. A search bar on the webpage will enable the user to search for objects by tags and some properties like color and human face recognition.
2. Local software suite: This is for big companies and government organizations where the data is huge and cannot be uploaded to our servers for analysis. This software will be completely offline and the users will see a similar UI to the website. With further extensions, live camera and CCTV feeds can be fed into the system for real-time analysis and tagging.

Sample business use-cases for companies include:

- Finding and recognizing suspects in extremely long video feeds.

- Tracking suspected objects like handbags, clothes, etc and the people carrying them.
- Counting and recognizing people in crowds.
- Taking attendance in classes/offices using face-recognition
- Automated tagging of color and other features on e-commerce websites for fashion items.

Link to code/ dataset/ presentation:
https://drive.google.com/open?id=19hCif9eGduZp473mgzmbcc6H78L3V-pq