

Automated Vehicle Detection and Tracking using CCTV Surveillance Video

Mustafa Ellam

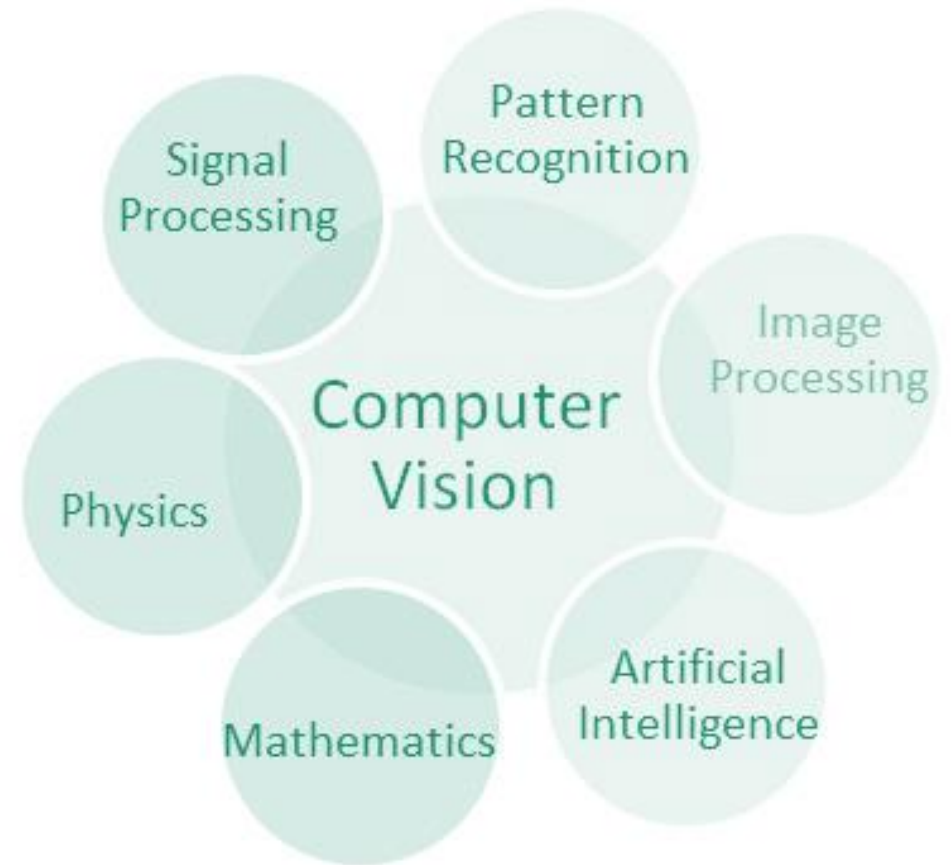
10th January 2018

mustafa.ellam@hmrc.gsi.gov.uk



OVERVIEW

1. Project Objectives
2. Transport for London (TfL) CCTV Data Feed
3. Key Challenges
4. Computer Vision Techniques
 - Background Subtraction
 - Morphological Transformations
 - Contour Detection
 - Vehicle Association
5. Vehicle Count Results
6. Beneficiaries and Implications
7. Recommended Further Work
8. Conclusions

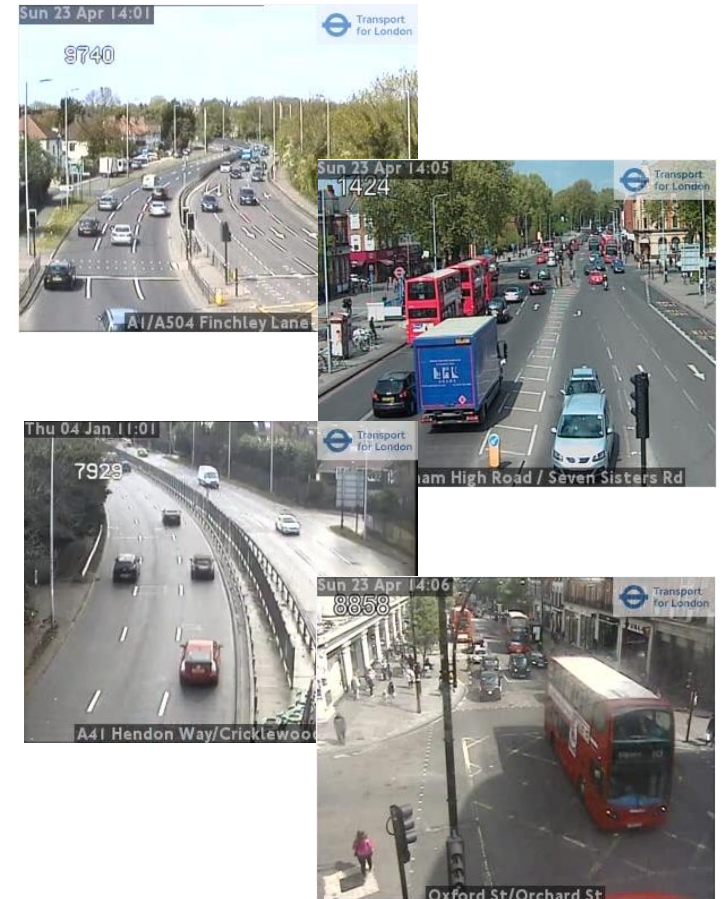


OBJECTIVES

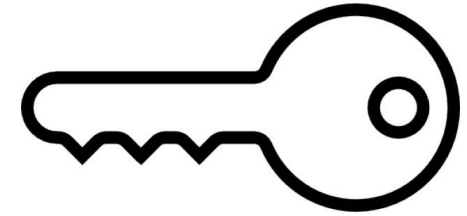
1. Demonstrate the application of Computer Vision Methods to raw CCTV footage from non-calibrated cameras
2. Acquire and Parse CCTV data using the TfL Application Programming Interface (API)
3. Develop Proof of Concept of Computer Vision System using Python-OpenCV
 - Vehicle Detection
 - Vehicle Tracking
 - ~~Vehicle Classification~~ - The classification task requires constructing an adequate labelled dataset from the vehicle detection step, however the scale required was beyond the scope of this project.
4. Compute Traffic Parameters, such as the number of vehicles

TFL UNIFIED API AND CCTV FEED

- Tfl open data platform makes data available across all modes of transport in London; Tube, roads, buses, cycling
- Approximately 600 apps built on this data with around 42% of Londoners using apps powered by Tfl open data (Deloitte, 2017)
- CCTV Feeds are seemingly under used compared to other Tfl data sources
- Tfl live traffic camera feed (aka “JamCams”) provide still images and video feeds from 914 locations across London
- The videos are nine second snippets of the latest traffic scenarios in a given location which is refreshed every three minutes



PERMISSION & ACCESS TO API



Free API Access - the developer should register for access tokens and append the unique “app_id” and “app_key” tokens as query string parameters

Anonymous Access - it is possible to access the API data anonymously, though this type of access is regulated for fair use and it is possible to be kicked off the feed. Free access provides dedicated, higher rate limits than anonymous

There are some restrictions for using this data under the open data licence –

‘Use any automated system, software or process to extract content and/or data, including trawling, data mining and screen scraping in relation to the Oyster, Congestion Charging and Santander Cycles websites, except where expressly permitted under a written licence agreement with TfL.’

The scope of this project was outlined to TfL to ensure there was no breach in the T&C

JSON FILE FORMAT

```
{
  '$type': 'Tfl.Api.Presentation.Entities.Place, Tfl.Api.Presentation.Entities',
  'id': 'JamCams_00001.07450',
  'url': '/Place/JamCams_00001.07450',
  'commonName': 'Piccadilly Circus',
  'placeType': 'JamCam',
  'additionalProperties': [
    {
      '$type': 'Tfl.Api.Presentation.Entities.AdditionalProperties, Tfl.Api.Presentation.Entities',
      'category': 'payload',
      'key': 'imageUrl',
      'sourceSystemKey': 'JamCams',
      'value': 'https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.07450.jpg',
      'modified': '2018-01-04T11:59:07.773Z'},
    {
      '$type': 'Tfl.Api.Presentation.Entities.AdditionalProperties, Tfl.Api.Presentation.Entities',
      'category': 'payload',
      'key': 'videoUrl',
      'sourceSystemKey': 'JamCams',
      'value': 'https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.07450.mp4',
      'modified': '2018-01-04T11:59:07.773Z'},
    ...
  ],
  'lat': 51.5096,
  'lon': -0.1348
}
```

CAMERA LOOKUP TABLE

Common Name	id	Lat	Lon	url	jpgLink	mp4Link
A406 Billet Upass E	JamCams_00002.00865	51.6007	-0.01594	/Place/JamCams_00002.00865		
Romford Rd / Tennyson Rd	JamCams_00001.02151	51.5421	0.00524	/Place/JamCams_00001.02151	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.02151.jpg	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.02151.mp4
Piccadilly Circus	JamCams_00001.07450	51.5096	-0.13484	/Place/JamCams_00001.07450	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.07450.jpg	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.07450.mp4
Blackheath Rd/Greenwich High Rd	JamCams_00001.03675	51.4742	-0.02073	/Place/JamCams_00001.03675	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.03675.jpg	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.03675.mp4
Edgware Way / Broadfields Ave	JamCams_00001.09747	51.6216	-0.27384	/Place/JamCams_00001.09747	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.09747.jpg	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.09747.mp4
Cromwell Rd/Earls Court Rd	JamCams_00001.06600	51.4946	-0.1957	/Place/JamCams_00001.06600	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.06600.jpg	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.06600.mp4
Oxford St/Orchard St	JamCams_00001.08858	51.514	-0.15409	/Place/JamCams_00001.08858		
A205 Tulse Hill/ Thurlow Park Rd	JamCams_00001.03758	51.441	-0.10633	/Place/JamCams_00001.03758	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.03758.jpg	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.03758.mp4
East Cross Route/Tredeger Rd	JamCams_00001.02900	51.5336	-0.02256	/Place/JamCams_00001.02900	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.02900.jpg	https://s3-eu-west-1.amazonaws.com/jamcams.tfl.gov.uk/00001.02900.mp4

KEY CHALLENGES

1. Vehicle occlusion
2. Cannot easily detect stationary vehicles
3. Sensitivity to variable lighting conditions (due to the time of day/weather)
4. Noise and pedestrian movement
5. Non-calibrate cameras (intrinsic parameters) make it difficult to compute vehicle speed

VEHICLE DETECTION: BACKGROUND SUBTRACTION

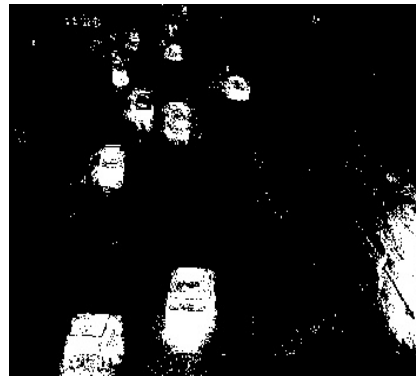
- Background Subtraction performs image segmentation on each frame into foreground and background pixels by:
 1. Subtracting the current frame from a static reference frame that represents the background scene
 2. Thresholding the resulting difference between the two frames to generate a binary object mask
 3. The background is represented in black (pixels = 0), and moving objects in white (pixels = 1)
- Reliant on an accurate representation of the background scene
 - Use a Dynamic Background Reference by using a set number of historical frames to predict the most probable background pixels. This therefore adapts to changes in lighting conditions

BACKGROUND SUBTRACTION

Background Reference: History Parameter



Actual



100 frames



500 (default) frames



800 frames

Threshold Value



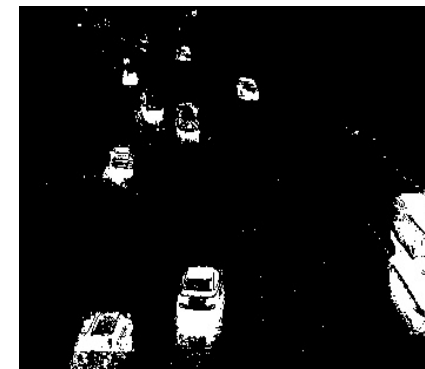
Threshold = 16 (default)



Threshold = 25



Threshold = 50



Threshold = 150

MORPHOLOGICAL TRANSFORMATIONS

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

- Clean the binary image of any residual noise to separate the object from the background and each other as much as possible in order to provide the optimal conditions for the object detection
- Uses a Kernel by overlaying it onto each pixel in the frame (convolution), and resets the pixel value to 1 or 0 depending on the exact specification of the kernel (size, shape)

Mask Filters:

- **Erosion** – Discards pixels near the object boundaries that will help more clearly define the shape of a vehicle, for instance separating to connected objects
- **Dilation** – Increases the size of the white foreground region. The erosion operation followed by dilation is called ‘Opening’
- **Closing** – This closes small holes in an object to produce a solid object by performing the reverse of ‘Opening’ (dilation followed by erosion)

MASK FILTERING

Binary Frame output
from Background
Subtraction



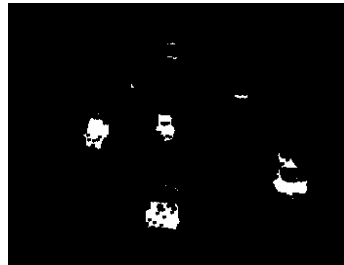
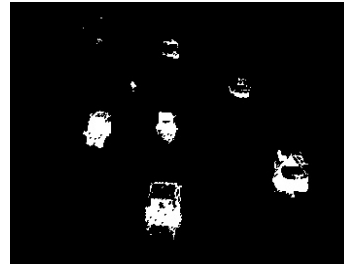
History: 100 frame
Threshold Value: 300

(2,2) Kernel

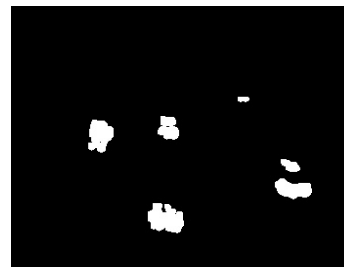
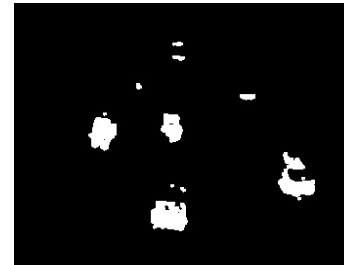
(4,4) Kernel

(6,6) Kernel

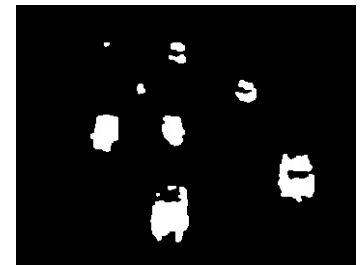
Erosion



Dilation



Closing/Opening



VEHICLE DETECTION: CONTOUR DETECTION

- Identifies the convex hull of the vehicles and traces the contours dynamically from frame to frame
- Provides the appropriate coordinates for the contours of each vehicle
- These points are used to display a bounding box around each vehicle and identify its centroid



VEHICLE TRACKING: ASSOCIATION

- Region based (or centroid) tracking is implemented which computes an association score between regions in successive frames in order to identify matching vehicles
- Compute the Euclidean distance between centroids in frame i and all new centroids in $i+1$. The matching point, and therefore the point most likely to be associated with the same vehicle is the next frame is the one with the smallest Euclidean distance
- A vehicle's path can be tracked for the entirety of the video by connecting the matching centroids, enhancing the potential traffic parameters that can be measured; direction of traffic flow; lane detection; average trajectories



Mocked example of tracking through association

VEHICLE COUNTING: REGIONS

- The number of vehicles can also be computed as a vehicle contour (or centroid) enters a specified region
- Regions must be pre-specified and care must be taken not to double count the same vehicle if it appears in the 'count' region more than once
- This does require manually defining regions which is not easily generalizable
- Instead, a standard horizontal boundary line is defined halfway through the entire frame which can be generalised across most CCTV videos



VEHICLE COUNT

Vehicle Detection Rates

	Correctly Counted	False Positives (FP)	False Negatives (FN)
Training	83%	15%	8%
Validation	67%	25%	33%
Test	71%	60%	14%

FP – Multiple objects consistently detected on Vans

FN - Due to Occlusion of Vehicles and therefore merging of objects

FP - Affected by different frame rate, meaning vehicle appears on line multiple times.



FN: issue detecting Van



FP: issue with multiple detection on van



FN: merging of objects due to occlusion

IMPLICATIONS

- Better inform road infrastructure design by monitoring traffic behaviour over time
- Exploiting existing CCTV systems can reduce the need for expensive, dedicated hardware and manual operations
- Provide enhanced traffic parameters that are not possible by the functionality of current systems
- Increased frequency of traffic data to produce statistics
- The framework designed should be generalisable to other applications using CCTV footage, such as pedestrian analysis in urban locations like highstreets, stations and airports.

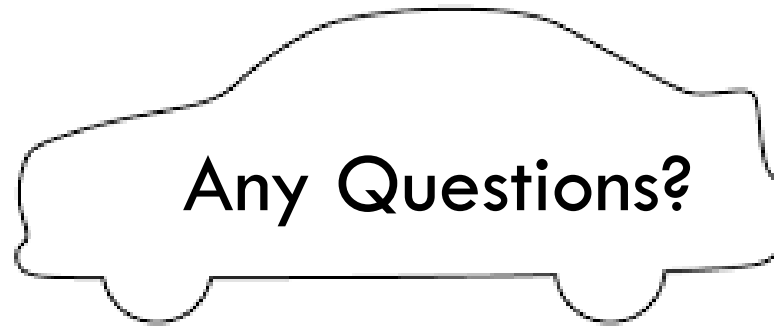
RECOMMENDED NEXT STEPS

- Add image stabilization at the start of the pipeline to deal with poor quality CCTV feeds (shakiness, swaying). This will improve generalizability
- Explore other traffic parameters, such as vehicle speed
- There are various characteristics of camera in terms of angle, perspective and distances to road which makes generalizing this system difficult
 - Potentially cluster camera by characteristics and optimize parameters separately for each group
- Predict vehicle trajectories using historical paths to deal with the occlusion issue
- Develop statistical methodology to make useful inferences from CCTV data, i.e. gross traffic parameters to reflect hourly, weekly, monthly statistics

CONCLUSION

- Proof of concept of computer vision system applied to raw CCTV footage from open access, non-calibrated cameras
- Calculating basic vehicle parameters, though this can be expanded once the vehicle tracking algorithm is fully implemented
- Potential exploitation of the existing CCTV camera network to automate traffic monitoring. This can allow more frequent statistics and reduce the need for expensive manual operations
- This work provides a basis on which to develop further work around the stability of the algorithm when it is generalized to a wider range of cameras

THANK YOU



Acknowledgements

Sponsorship: ONS Data Science Campus

Supervisor: Dr Gregory Slabaugh