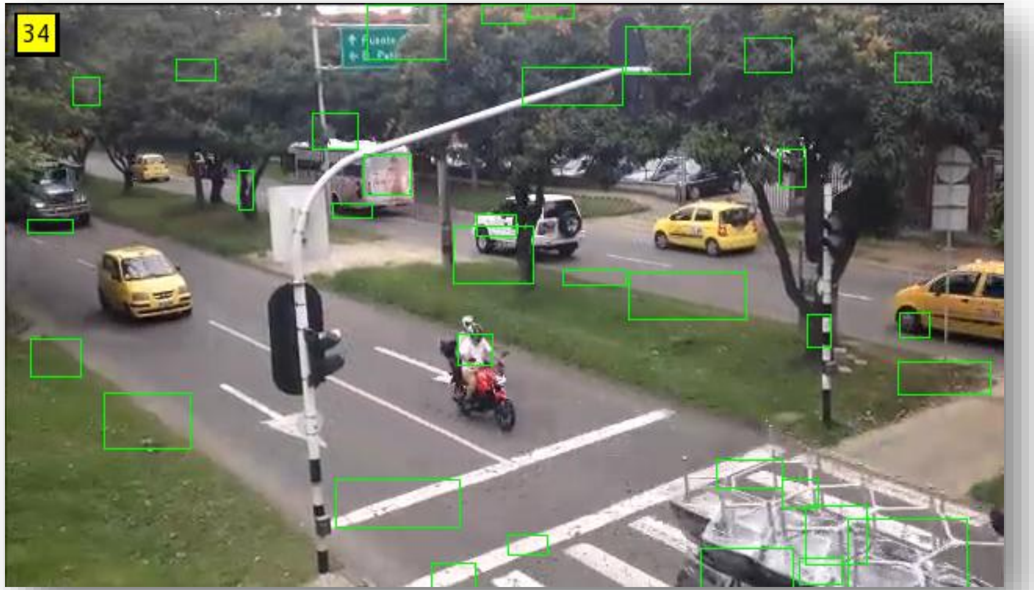
The Potential of Video Analysis to Improve Urban Traffic
MOVICi 20.April.2018



Jorge E. Espinosa, Sergio A. Velastin , and
John W. Branch

POLITÉCNICO COLOMBIANO
JAIME ISAZA CADAVID

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN

Motivation

- Congestion of roads (Travel times increased by 40%)
- Regulation and control
- Traffic Control Centers
- Difficulties in urban environments
- Vehicle Interaction (Multimodal Flow)

Imagen Diario ADN - Medellín

Imagen de Arizona Department of Transportation

Motivation

- Medellín
  - Second largest city in Colombia
  - 2.5 million inhabitants (3.4 in metropolitan area)
  - GDP per capita USD 8.489 (2014), Colombia 7,913 (feeds aspiration of private transport?)
  - 1 vehicle/3persons (including motorbikes)
  - Red Environmental Alerts

- WHO: 1.25 million traffic-related deaths (Colombia 8107)
  - Average 17.4 per 100.000 people
  - Colombia 16.8, UK 2.9, Spain 3.7
  - Fatalities per 100.000 vehicles: Colombia 83.3, UK 5.1 Spain 5.3
  - In 2015 only 28 countries (7% world population) had laws addressing all 5 risk factors (speed, drunk driving, helmets, seat-belts and child restraints)
  - 26% deaths in poorer countries are of pedestrians and cyclists

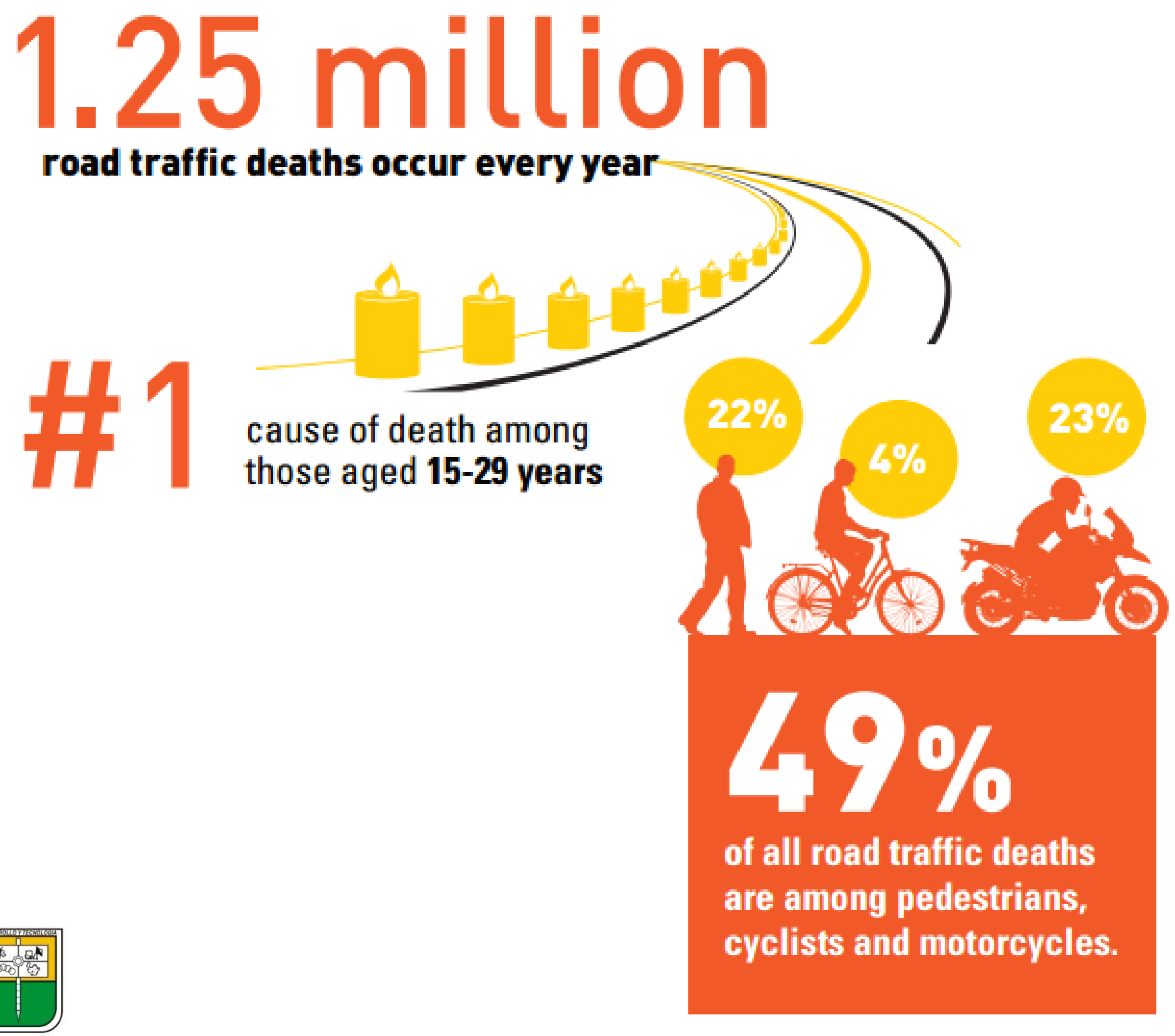
POLITÉCNICO COLOMBIANO
JAIME ISAZA CADAVID

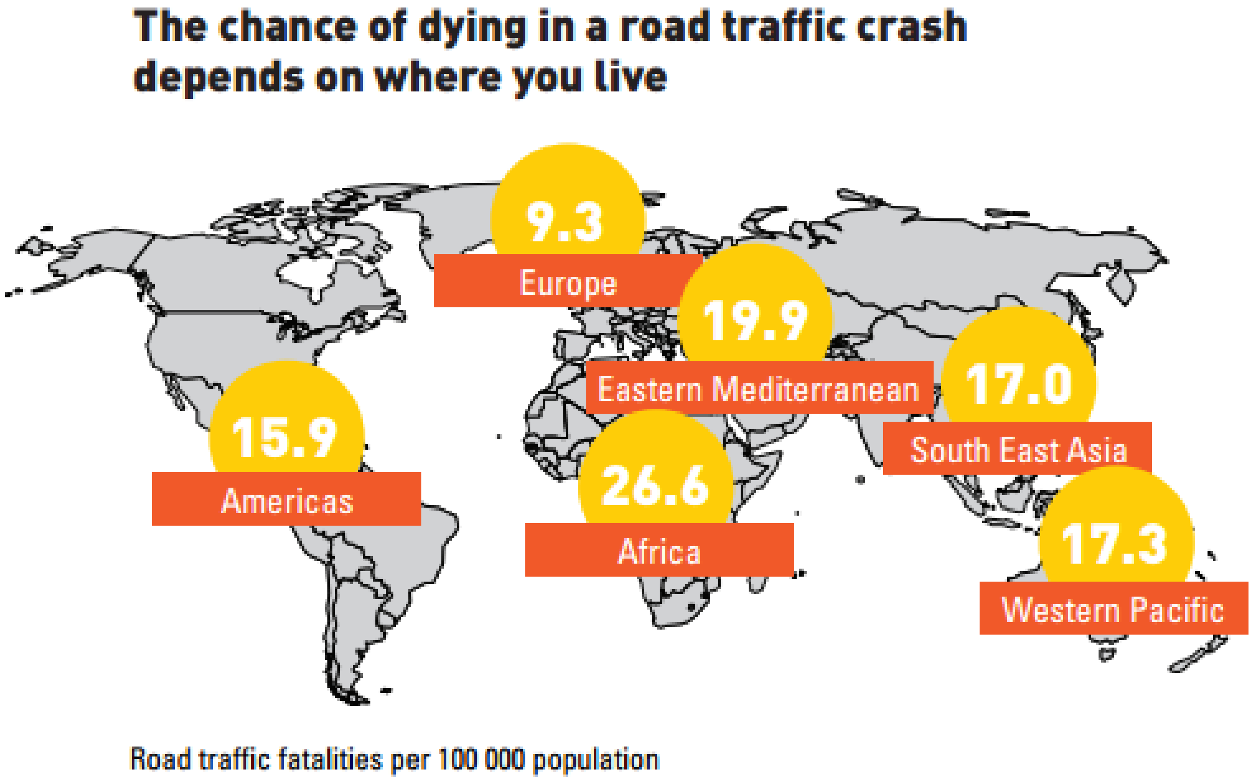
UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN

uc3m

Motivation

Exploring further ....





POLITÉCNICO COLOMBIANO
JAIME ISAZA CADAVID

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN

Enablers

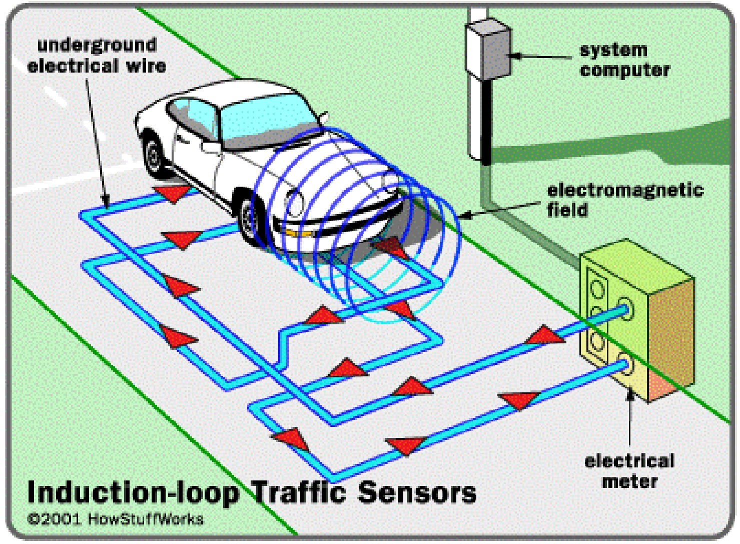
- **“Societal”:**
  - Education
  - Legislation
  - Safer vehicle standards
  - Effective enforcement
  - …
- **Technical:**
  - Safer Roads (surface, lighting, walkways, bike lanes …)
  - Traffic control centres
  - Smart video and other sensors
    - Computer Vision
    - Big data and data fusion
    - Artificial Intelligence
    - Cheaper hardware
    - Driver assistance (including autonomous vehicles)
    - BUT: can they reach “poorer” road users?
  - …

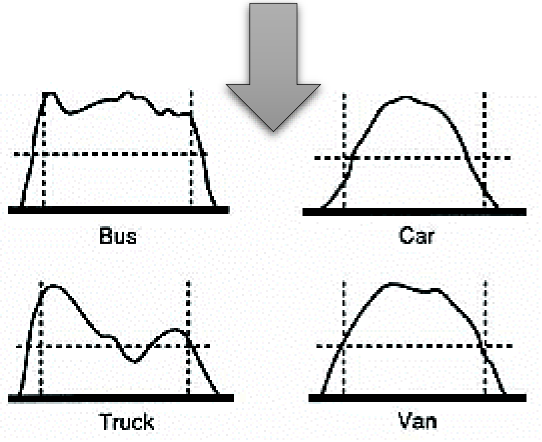
POLITÉCNICO COLOMBIANO JAIME ISAZA CADAVID

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN

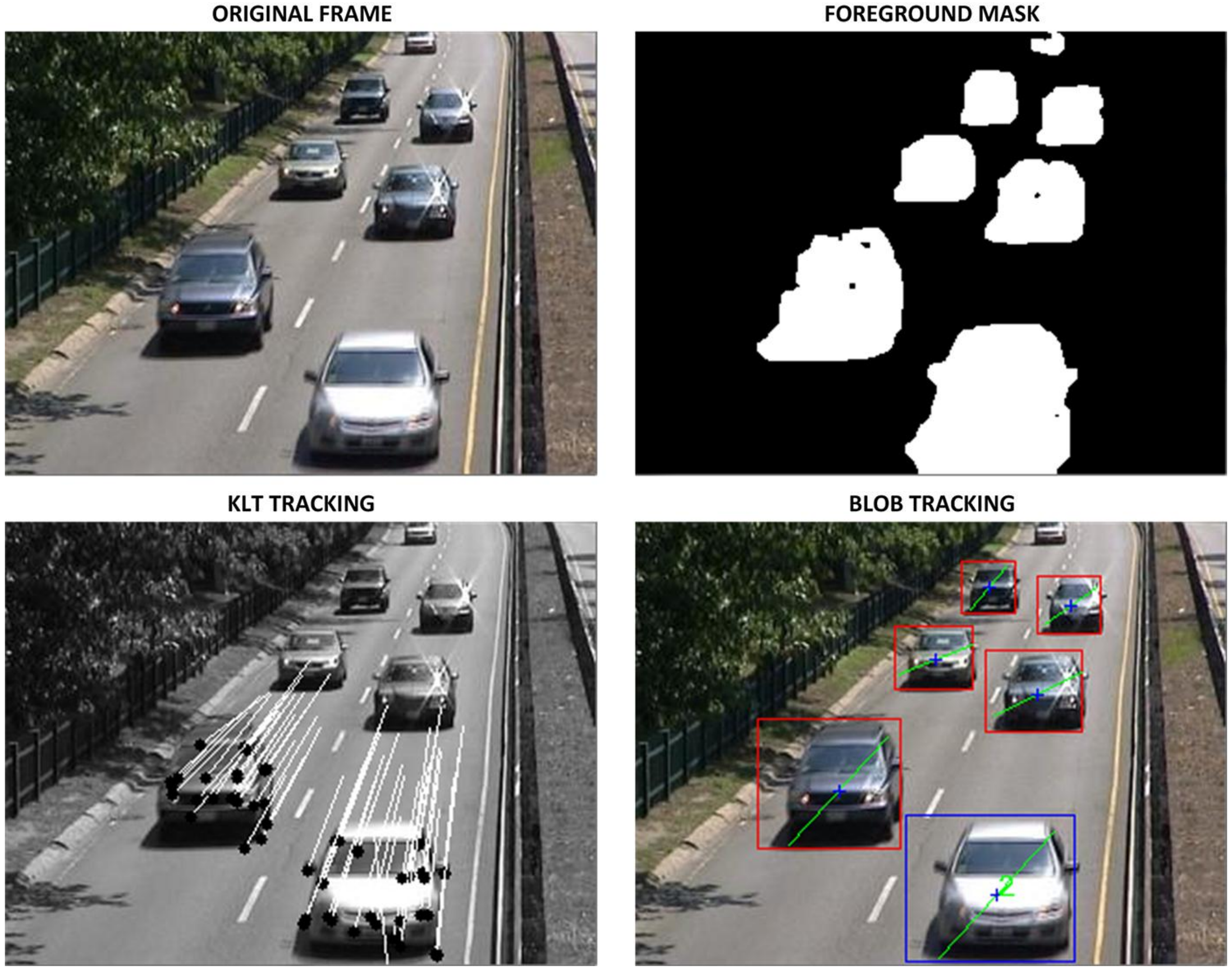
Introduction

Vehicle detection





Taken from Federal Highway Administration Research and Technology

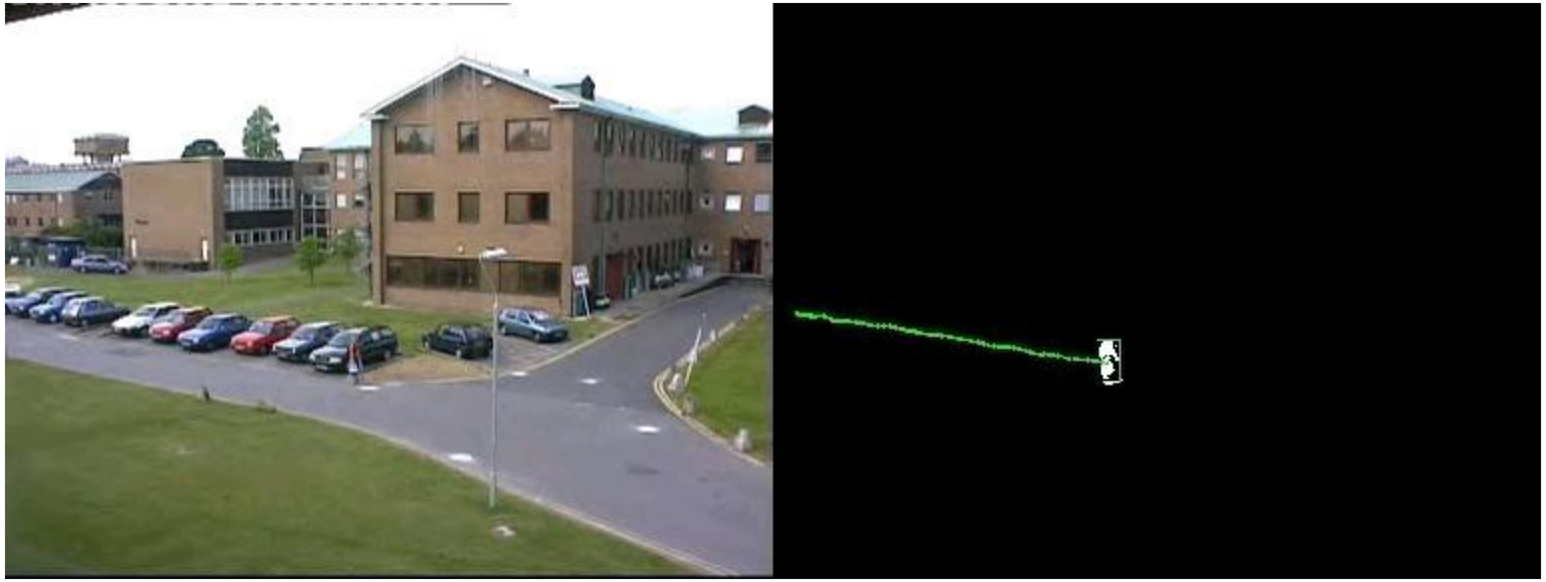


Taken from Vehicle detection, tracking and counting

Motion Tracking

Detection of moving objects -> Blobs

Blob matching -> Trajectories



Stationary background, mostly background,

Stationary objects tend to disappear!

A urban traffic environment (UK)

Well ordered, nice pictures

Data is collected

Bus and Motorcycle samples

Bus (290 samples)

Motorcycle (143 samples

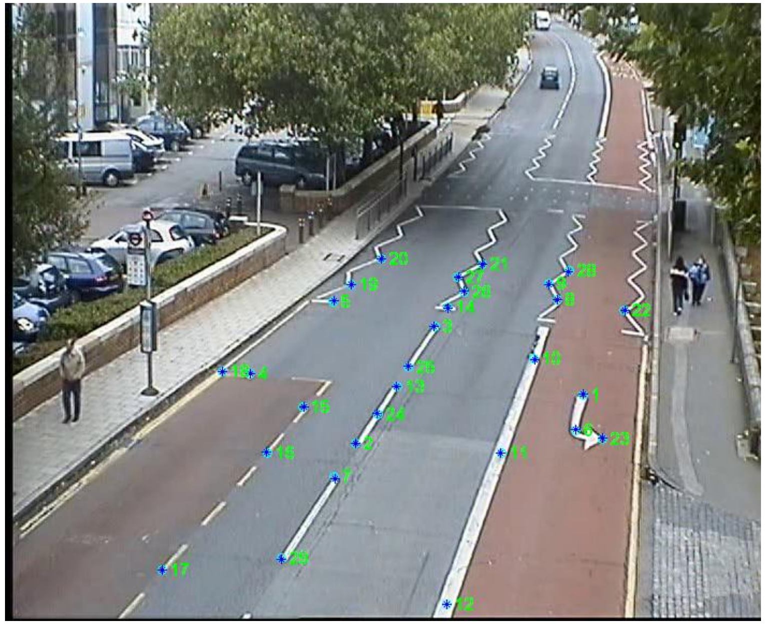
Feature Database

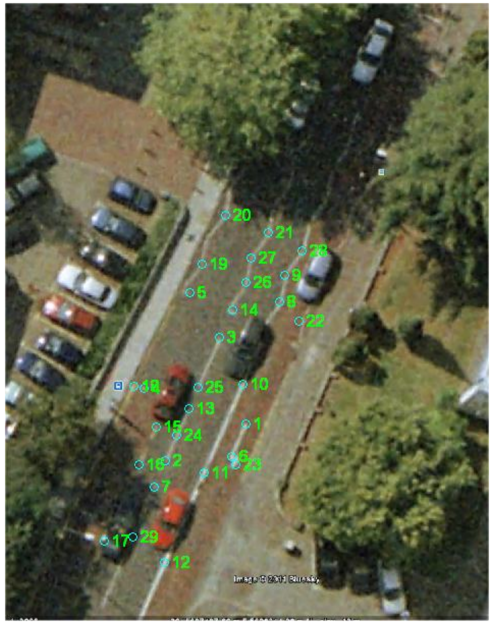
Car and Van samples

Car (1033 samples)

Van (589 samples)

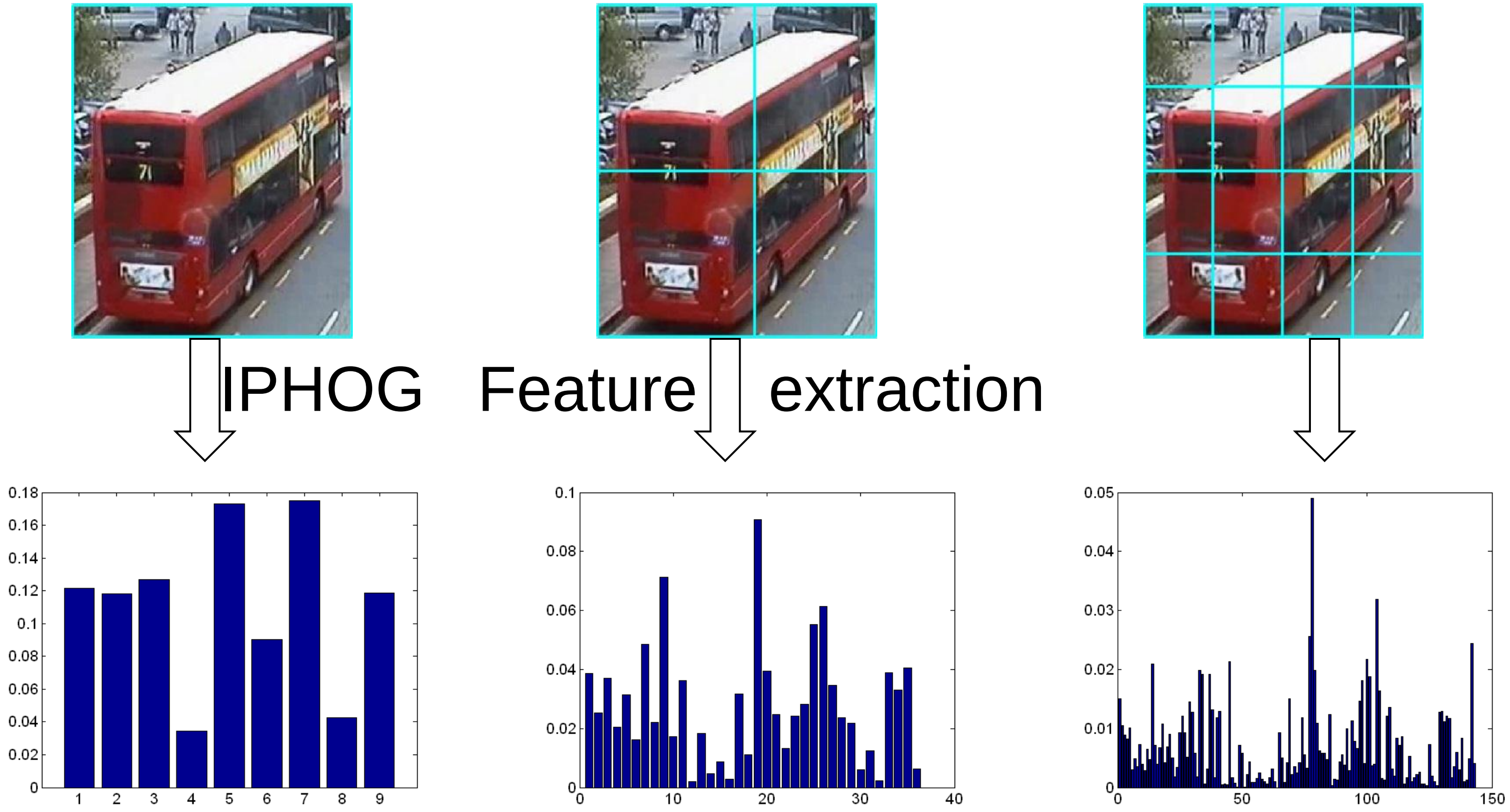
Camera needs calibrating!





Calibration reference image Plan view image from Google Earth

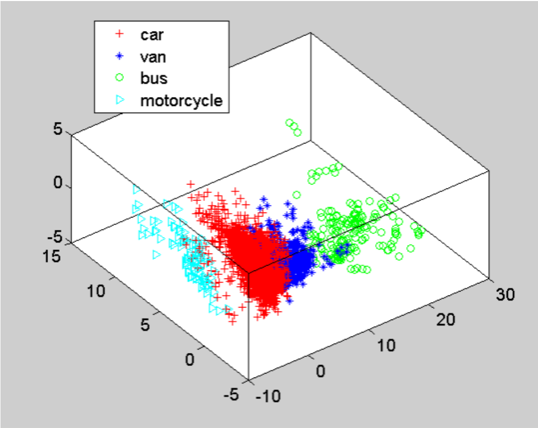
We extract “features” from images
(there is an infinite number of possible features!)



Level 0 + concatenate Level 1 + concatenate Level 2

We “train” a “classifier”

- Use the samples we collected and **manually** annotated
- Extract features we decided could discriminate different types
- Add some rules (e.g. proportion of length to width …)
- Hopefully we can distinguish different types of vehicles

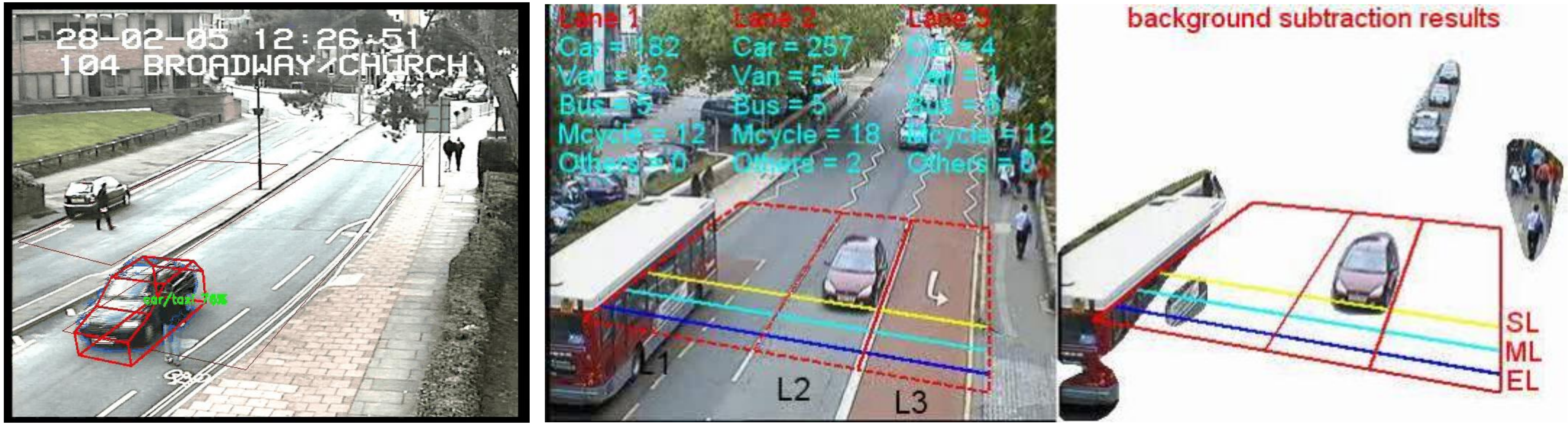


Automatic detected silhouette data

uc3m

Putting things together!

- Camera data comes in
- We use motion tracking to extract "blobs"
- For each "blob" we use the "classifier" to find out vehicle type
- We can then count, measure speeds, detect infringements, etc.



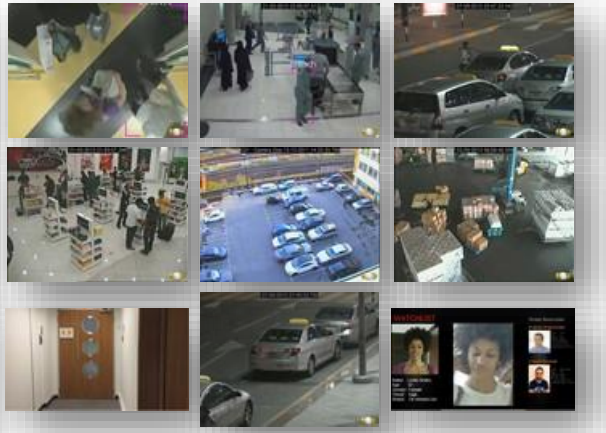
POLITÉCNICO COLOMBIANO JAIME ISAZA CADAVID

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN

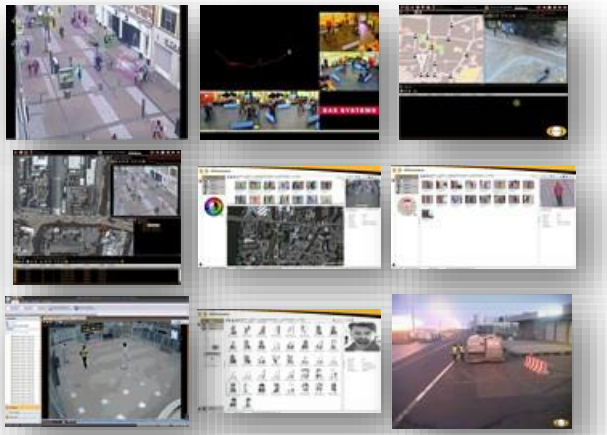
Example commercial system (Ipsotek UK)

Perimeter Protection

Intrusion Detection

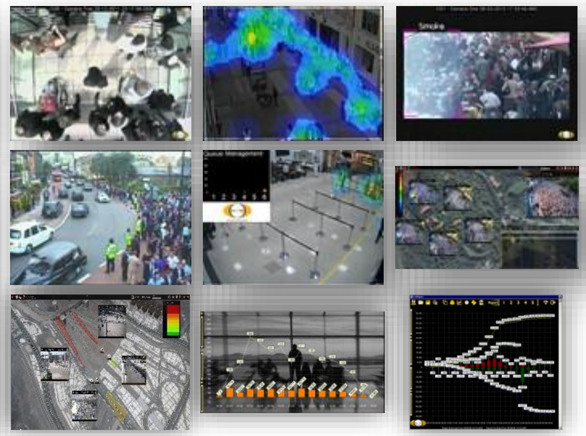


Investigation and Forensics

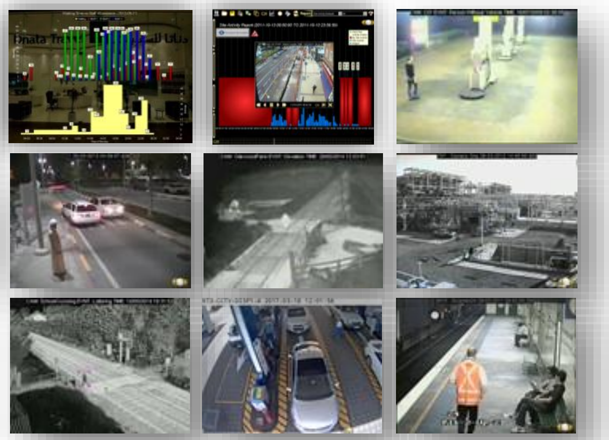


Traffic Management

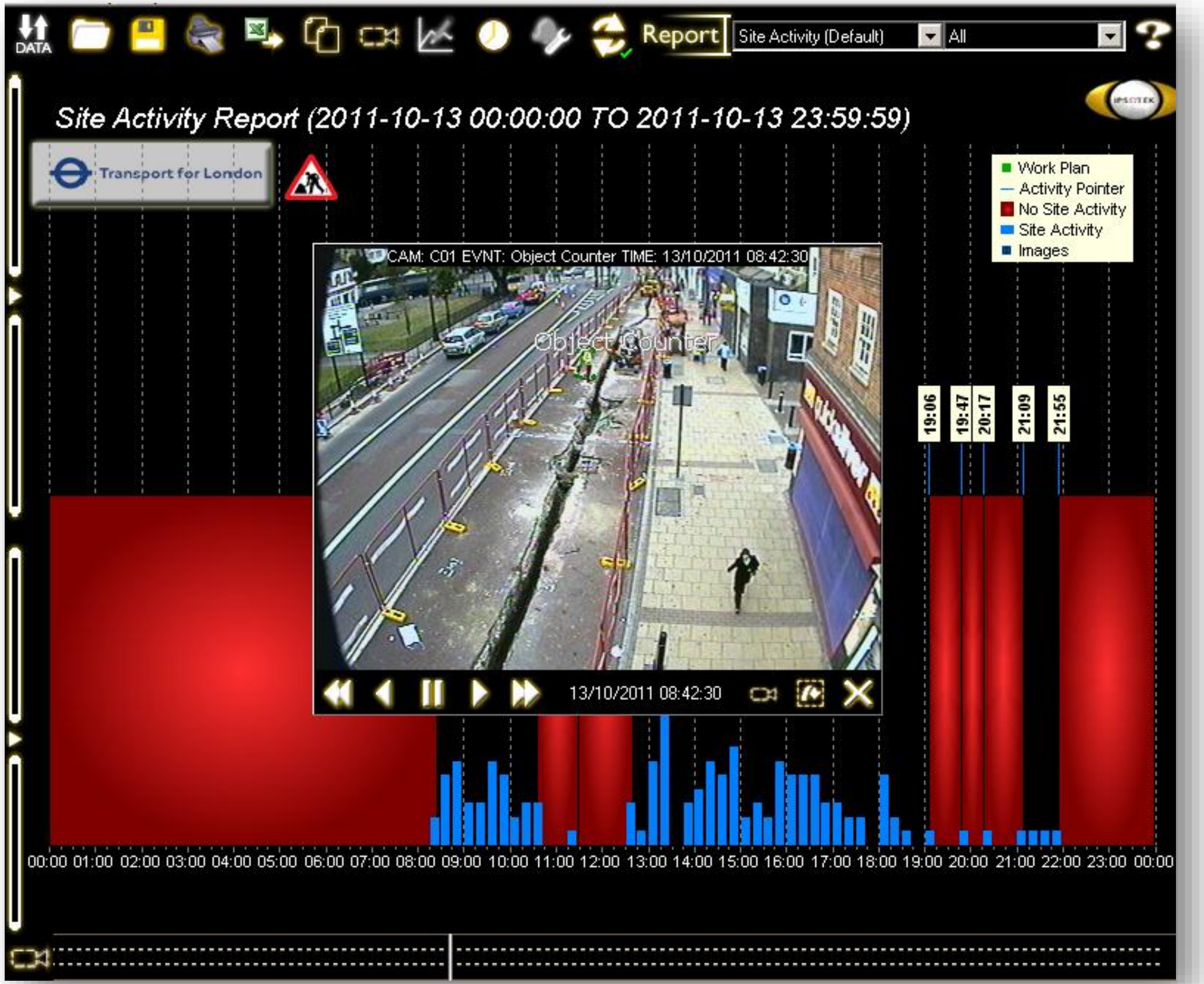
Crowd Management



Operations Management



Roadworks Monitoring – (Ipsotek)



Challenges

- Computational cost
- Use of available infrastructure - quality
- Camera angles, calibration
- Partial Occlusions
- Illumination variations, day/night
- Continuous changes in the background
- Road user behavior (people, bikes)

- **How do we know what are best features to extract?**
- **How do we best train a system?**

What do humans do? Can we emulate them?

We seem to be able to locate and label objects using a SINGLE image

A (nearly) new paradigm: **Deep Learning**

- Based on well-known “neural networks”
- Advances in hardware: multiple core graphics cards allow many fast simultaneous operations
- Same hardware allows building large networks with many “layers” i.e. *depth* (needed for much better classification)
- Particularly well-suited to images
- Internet enables very large repositories of images (e.g. do a google search for “car” images) providing the variability needed for generality
- These networks are able to compute best set of features that solve a problem by building them up hierarchically similar to brains

POLITÉCNICO COLOMBIANO JAIME ISAZA CADAVID

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN

Convolutional Neural Networks

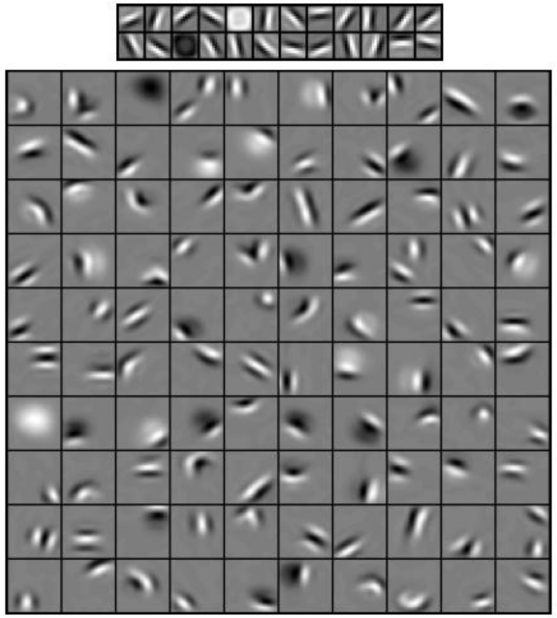
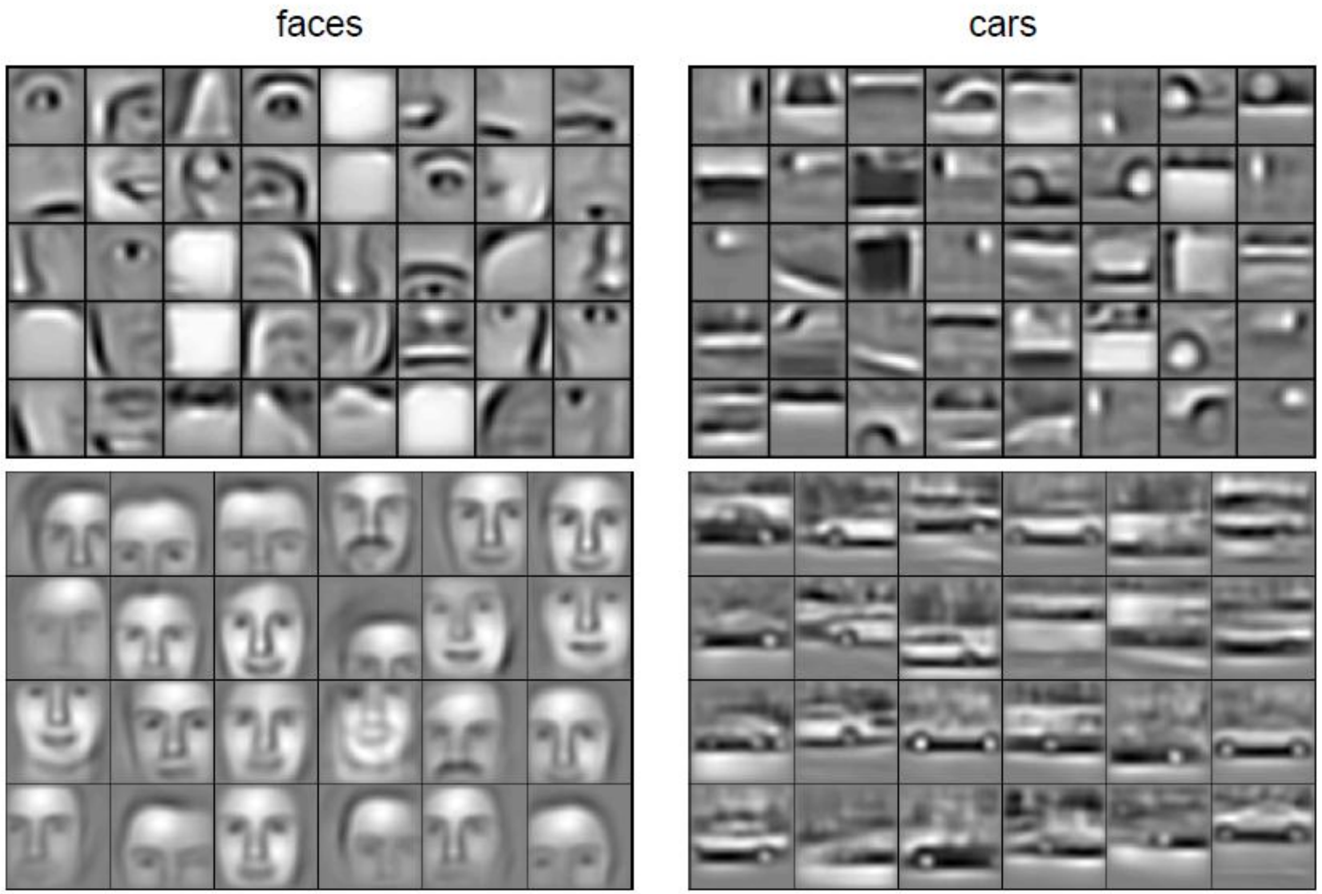


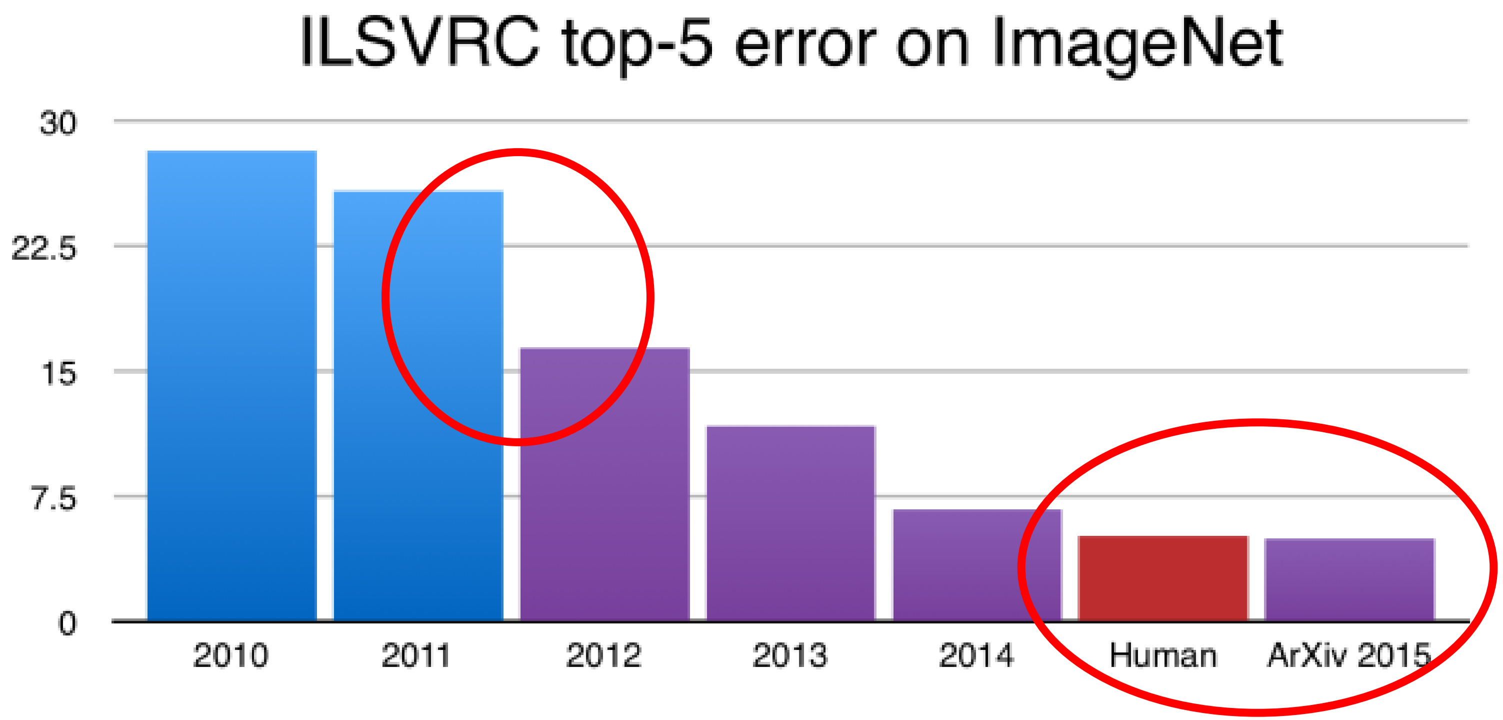
Figure 2. The first layer bases (top) and the second layer bases (bottom) learned from natural images. Each second layer basis (filter) was visualized as a weighted linear combination of the first layer bases.



H. Lee, R. Grosse, R. Ranganath, y A. Y. Ng,

Convolutional Neural Networks

IMAGENET Image Large Scale Visual Recognition Challenge



Source: http://image-net.org/

Our approach

- Successful deep networks (e.g. AlexNet, Faster-RCNN) have already been trained on millions of examples, so they have *learnt* how to extract good features
- So, we collect traffic data that is representative of our conditions
- And use the pre-trained networks with our data

- We still need to manually annotate our data to *evaluate* these nets

- And we also propose our own networks to compare with existing ones

- (cannot give technical details because of time limitations)

uc3m

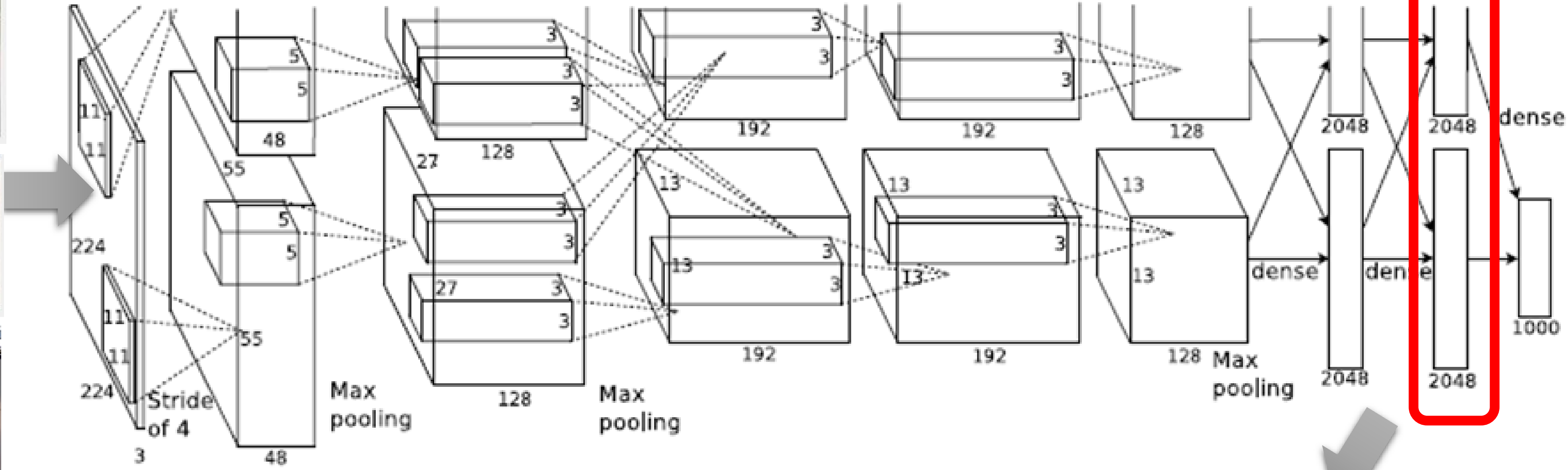
AlexNet for Vehicle Detection

The four categories Dataset created for classification used in **AlexNet model**
80 Examples per Category = 320 Total

POLITÉCNICO COLOMBIANO
JAIME ISAZA CADAVID

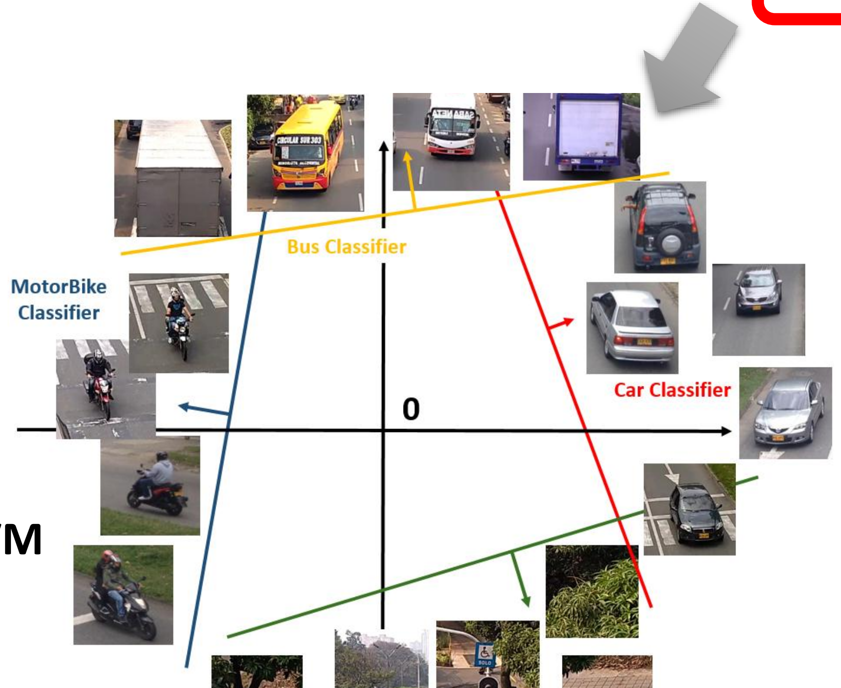
UNIVERSIDAD
NACIONAL
DE COLOMBIA
SEDE MEDELLÍN

AlexNet for Vehicle Detection



AlexNet

Extracted using motion tracking



Linear SVM

Car

Bus

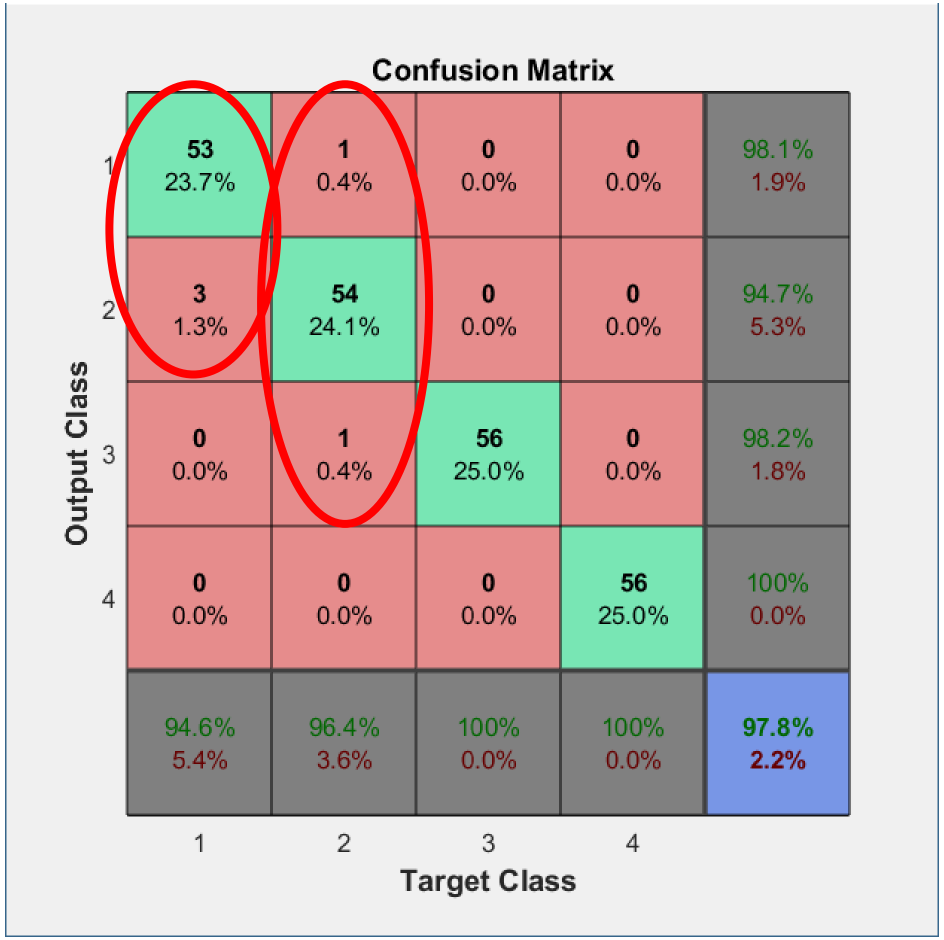
MotorBike

UrbTree

POLITÉCNICO COLOMBIANO
JAIME ISAZA CADAVID

AlexNet for Vehicle Classification

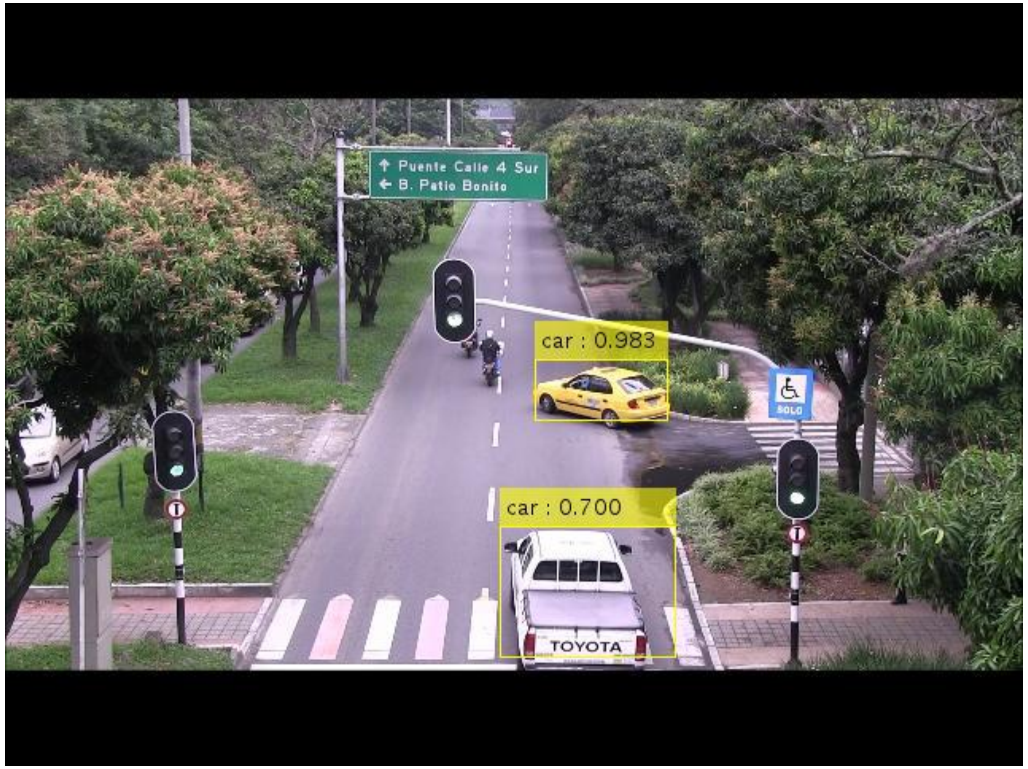
Classification Results



Confusion Matrix of the experiments.
(Class 1: Buses 2: Cars 3: Motorcycles 4: urbTree)

- Mean Accuracy: **97,80%**
  (Training 30% – Test 70 %)
- Cross Validated Mean Accuracy : 100%
  (k=10, Training 90% – Test 10 %)
- Cross Validated Mean Accuracy : 99,31%
  (k=10, Training 10% – Test 90 %)

Faster R-CNN Results



Detection and Classification **F1= 0.76**

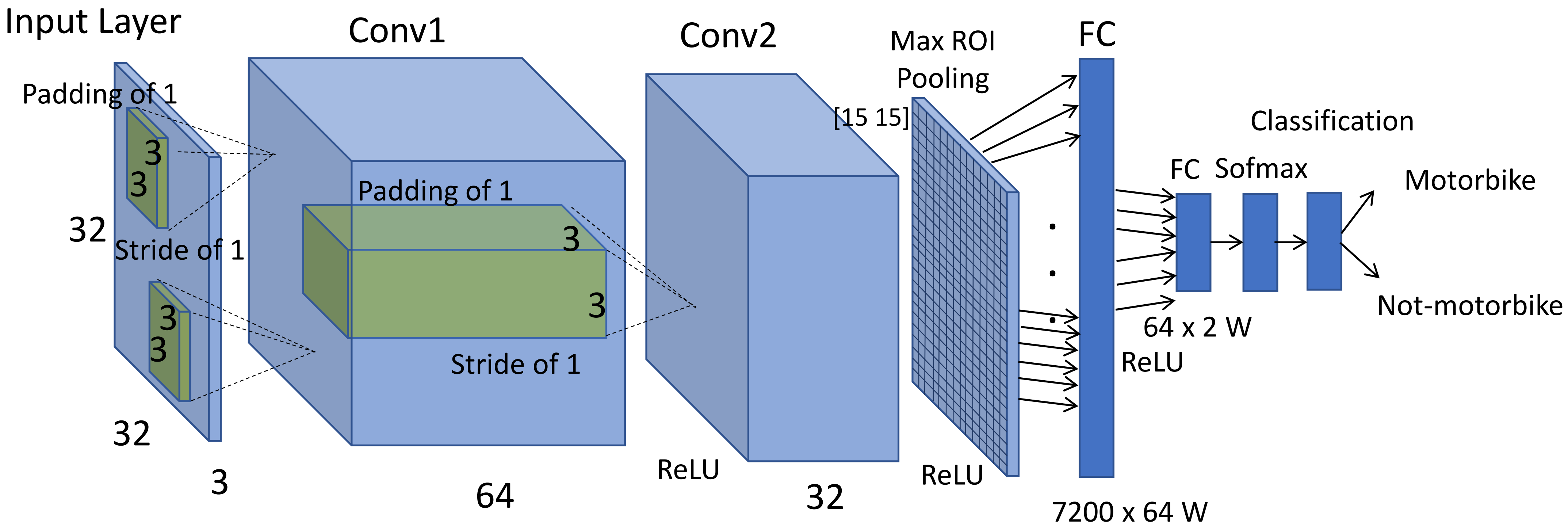
Experiments and Results

AlexNet Results

Detection and Classification **F1= 0.68**

Focusing on Motorbikes

CNN model inspired in Faster R-CNN



- Optimization Algorithm for training:
  Stochastic Gradient Descent with momentum (SGDM)

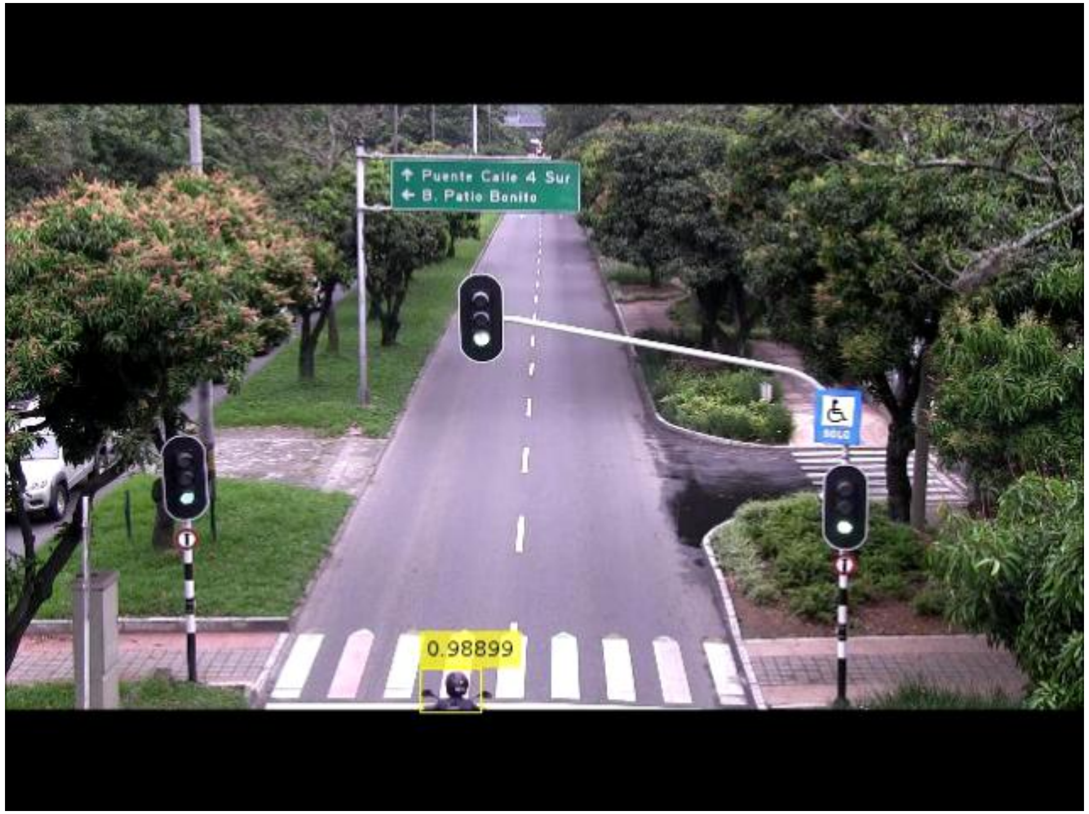
$$\theta_{\ell+1} = \theta_\ell - \alpha \nabla E(\theta_\ell) + \gamma(\theta_\ell - \theta_{\ell-1})$$

- Took 32 hours for training the dataset
  (50% Training – 30%Validating – 20%Testing)

POLITÉCNICO COLOMBIANO JAIME ISAZA CADAVID

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN

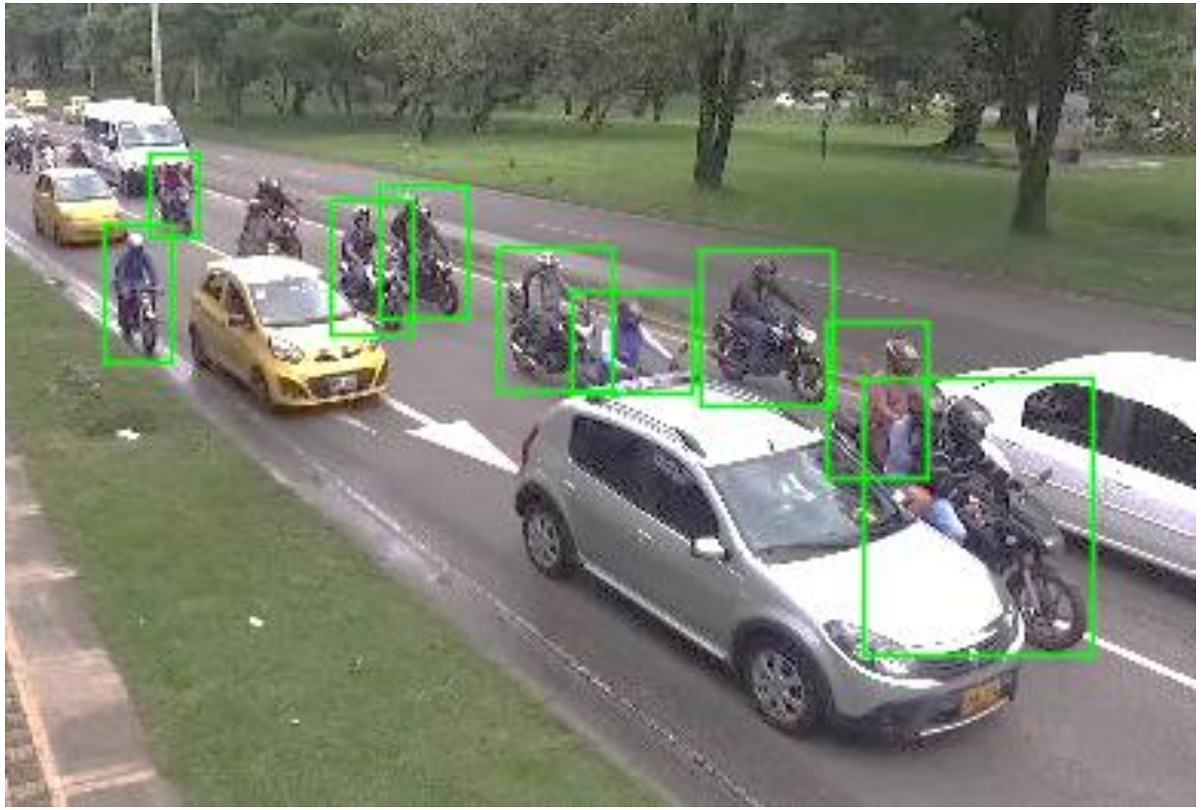
“Las Vegas” dataset



- 1812 annotated images
- 640 x 480
- Low occlusions
- **AP=92%**

uc3m

The Motorbike urban dataset

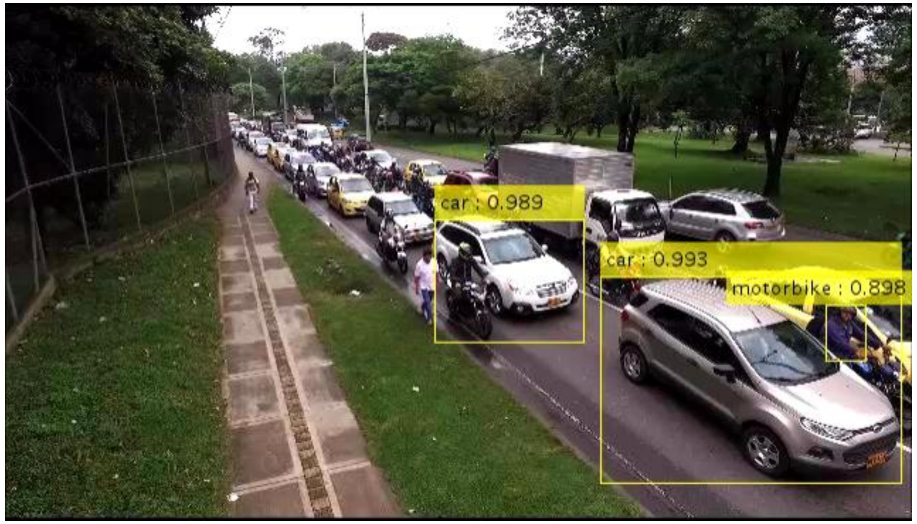


- Captured by drone
- 7,500 annotated images
- 220 motorcycles on urban traffic.
- 640 x 364 pixels
- 41,040 ROI annotated objects
- Minimum H size 25 pixels
- **60% Annotated object are occluded**

POLITÉCNICO COLOMBIANO JAIME ISAZA CADAVID

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN

Results on Motorbike Urban dataset



Faster R-CNN VGG16 based AP=23%

AlexNet + GMM AP=17%

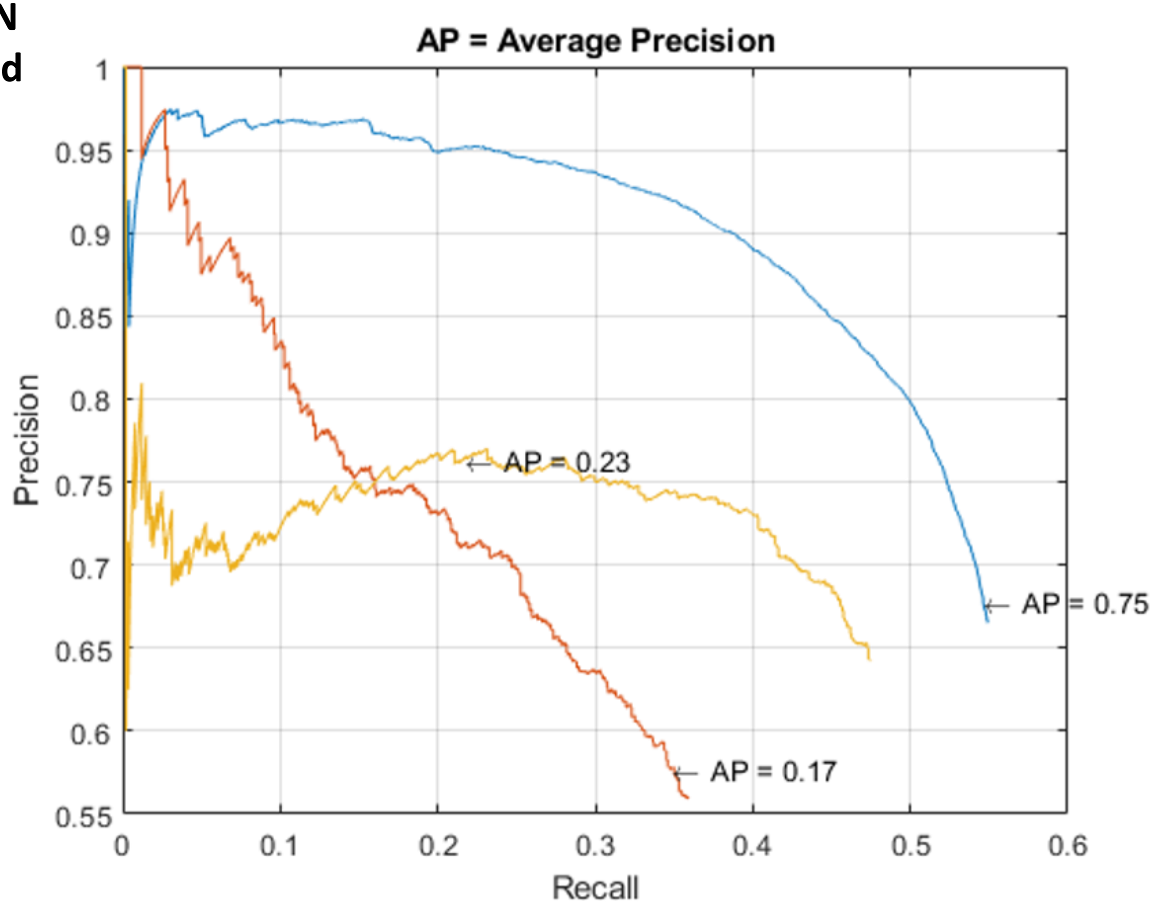
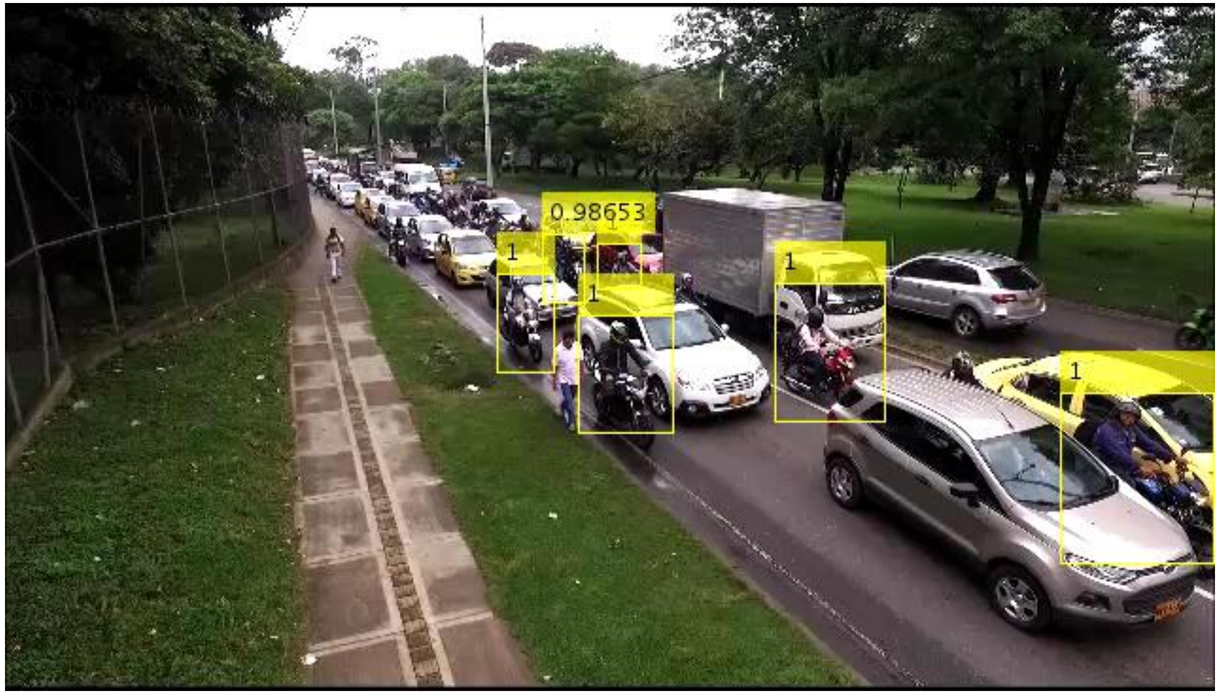
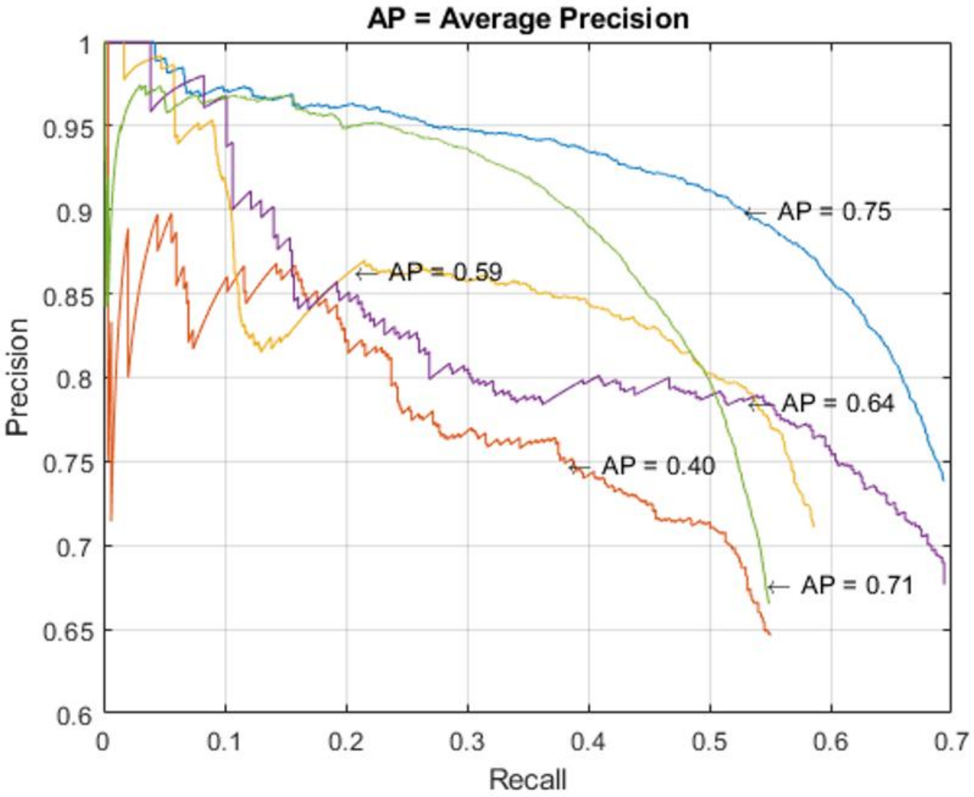


Figure 7 Average Precision (AP) of the model compared with AlexNet+GMM and Faster R-CNN VGG16 based.

“Motorbike Urban dataset”





- Our approach
- **AP=75% (vs. 23% and 17%)**

POLITÉCNICO COLOMBIANO
JAIME ISAZA CADAVID

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN

Conclusions and Future Work

- Computer vision is a promising technology with the potential to address the problem of monitoring traffic
- Urban traffic monitoring is a challenging problem especially when focusing on vulnerable road users (e.g. motorbikes in emerging countries)
- Commercial systems are becoming more robust, but still face challenges in cluttered urban environments
- Deep Learning has shown to be a “disruptive” approach and these initial results indicate that they have the potential to achieve acceptable results.
- Graphic GPU cards and conventional PCs already can achieve near real-time performance (and costs are likely to continue dropping)
- There is still much work to be done! e.g. to exploit the temporal properties of video sequences.

POLITÉCNICO COLOMBIANO
JAIME ISAZA CADAVID

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN

uc3m

Acknowledgements

NVIDIA

MARIE CURIE ACTIONS

CONEX
CONNECTING EXCELLENCE TO UC3M

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD

POLITÉCNICO COLOMBIANO JAIME ISAZA CADAVID

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN

uc3m

sergio.velastin@ieee.org

© Man Bouncing Question Mark Towards Doctor - Artist: [Art Glazer]

POLITÉCNICO COLOMBIANO
JAIME ISAZA CADAVID

UNIVERSIDAD
NACIONAL
DE COLOMBIA
SEDE MEDELLÍN