THE ANDREW & ERNA VITERBI
FACULTY OF
ELECTRICAL
ENGINEERING

SIPL
Signal and Image Processing Lab

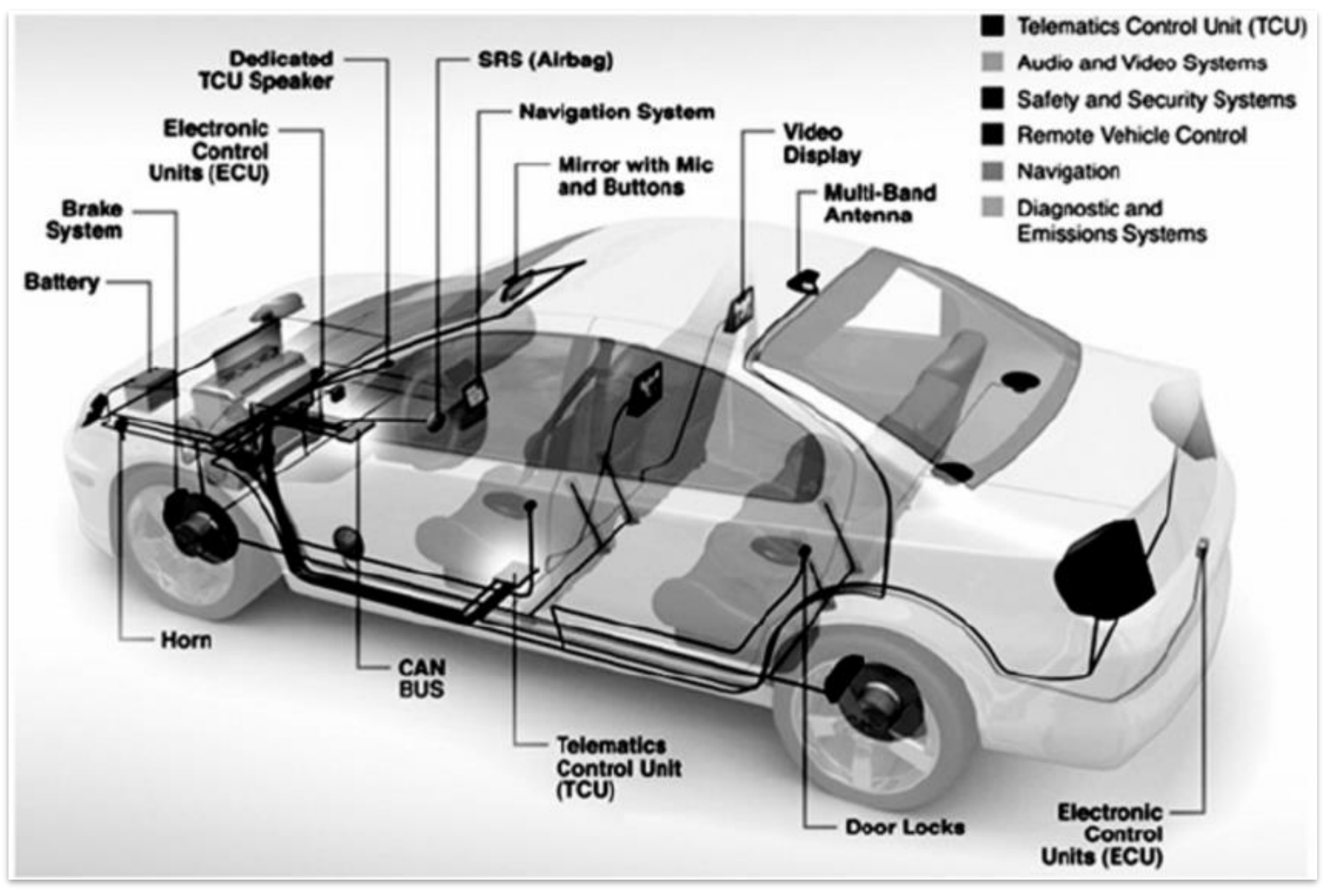
TECHNION
Israel Institute
of Technology

Automobile Driver Fingerprinting

Lana Brik and Nir Voloshin, Supervised by Pavel Lifshits

Introduction

- In advanced car systems, information passes through many units (ECUs) using CAN-Bus protocol. It's possible to extract wide range of data from many sensors.
- Using the data to identify the drivers, applying feature extraction with non-linear methods of dimension reduction.


Telematics Control Unit (TCU)
Audio and Video Systems
Safety and Security Systems
Remote Vehicle Control
Navigation
Diagnostic and Emissions Systems
Dedicated TCU Speaker
SRS (Airbag)
Navigation System
Electronic Control Units (ECU)
Mirror with Mic and Buttons
Video Display
Multi-Band Antenna
Brake System
Battery
Horn
CAN BUS
Telematics Control Unit (TCU)
Door Locks
Electronic Control Units (ECU)

Car units connected to CAN-Bus

Goals

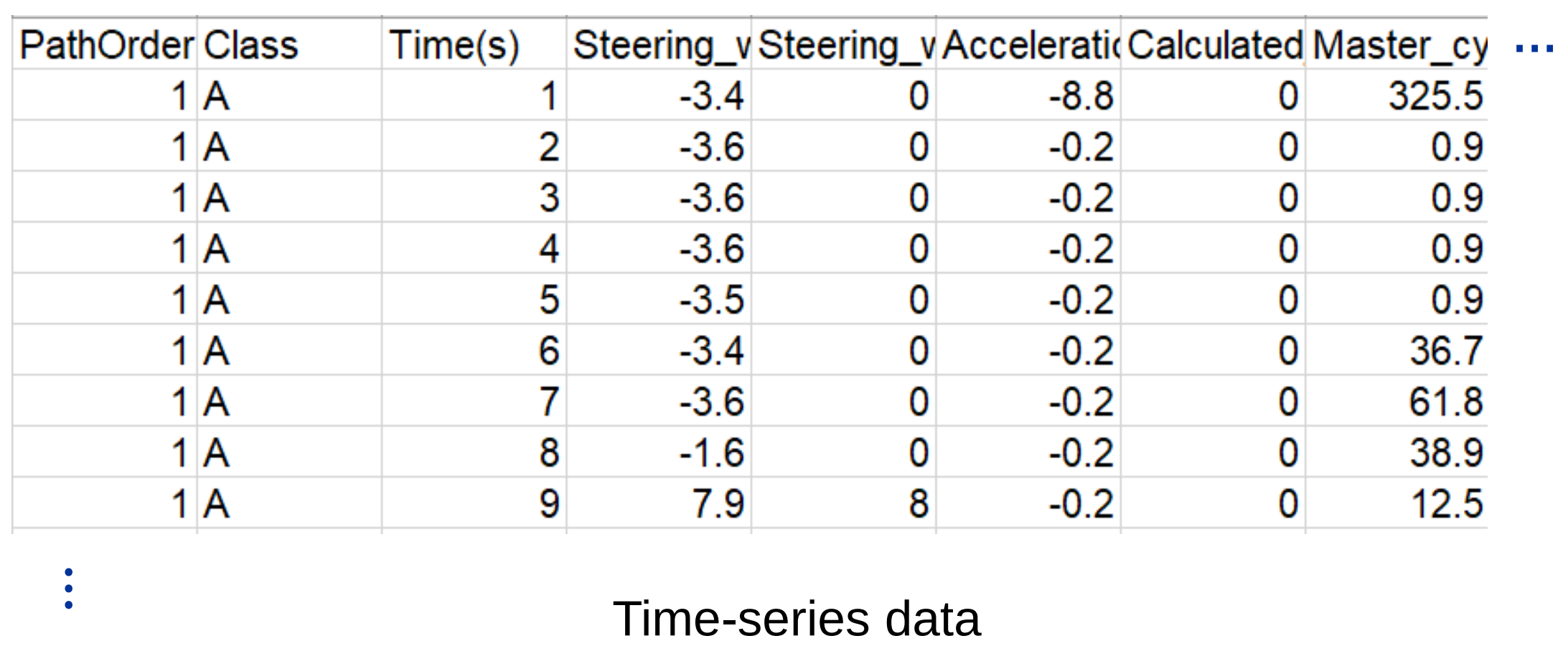
- Classify the drivers
  - dimension reduction utilizes the data
  - Within minimal driving time
  - Computationally efficient algorithm

Challenges

- Drivers with similar style
- Environment influences and dependencies
  - Traffic, car accidents, random pedestrians
- Finding dimension reduction method which utilizes the data
- Processing data according to hyper-parameters

Data Preprocessing

- The data from the CAN-Bus is recorded in packets form
- Transforming the data into time-series form
- Receiving 51 sensors for each recorded driver



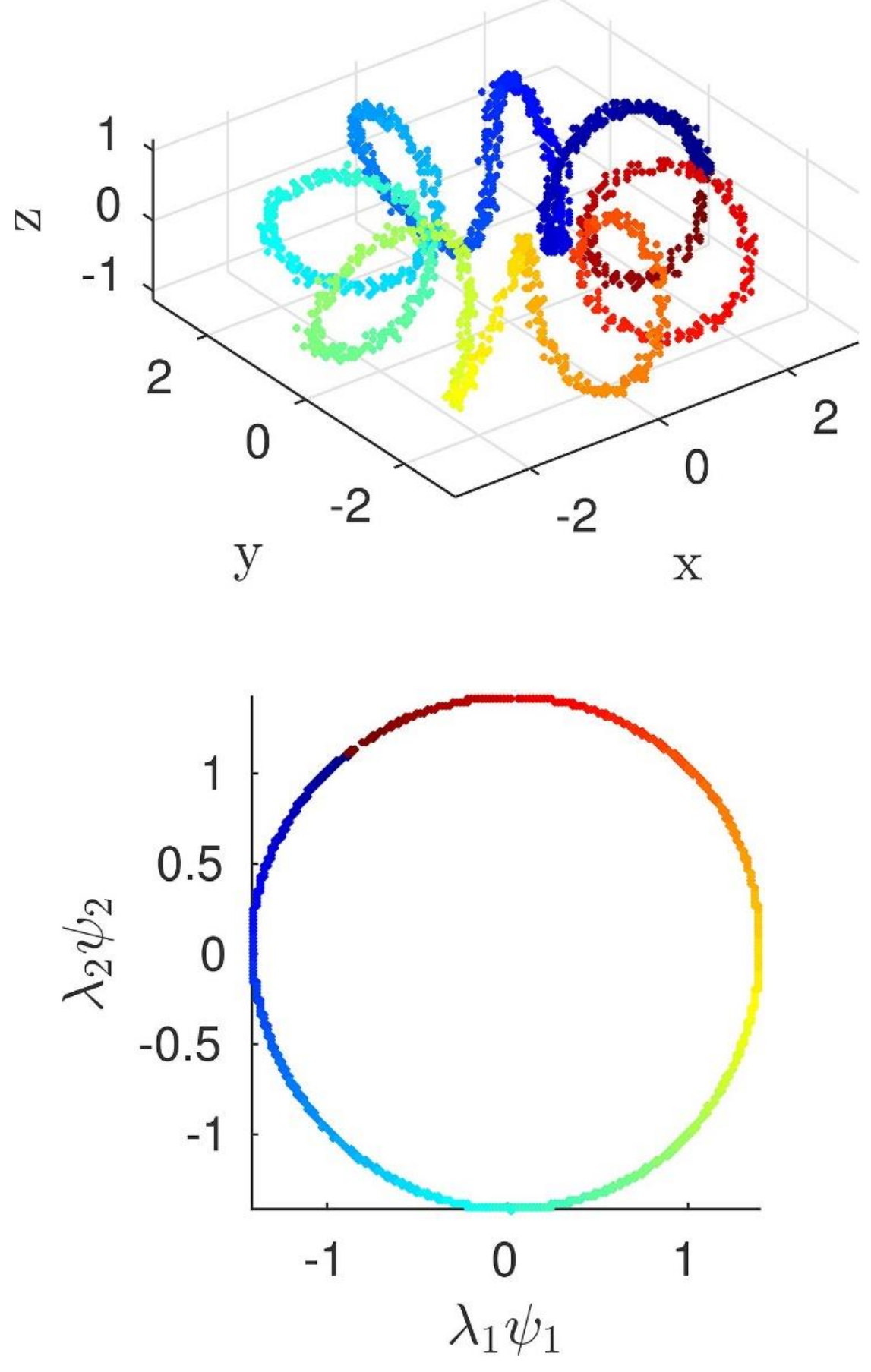
PathOrder	Class	Time(s)	Steering_v	Steering_v	Acceleratio	Calculated	Master_cy	…
1	A	1	-3.4	0	-8.8	0	325.5	
1	A	2	-3.6	0	-0.2	0	0.9	
1	A	3	-3.6	0	-0.2	0	0.9	
1	A	4	-3.6	0	-0.2	0	0.9	
1	A	5	-3.5	0	-0.2	0	0.9	
1	A	6	-3.4	0	-0.2	0	36.7	
1	A	7	-3.6	0	-0.2	0	61.8	
1	A	8	-1.6	0	-0.2	0	38.9	
1	A	9	7.9	8	-0.2	0	12.5	
⋮								

Time-series data

- Filtering sensors according to their variance
- Dividing the data to windows according to hyper-parameters
  - Window length and overlap
- Getting windows for each driver

Dimension Reduction

- Chosen non-linear method is Diffusion Map
- The goal of the algorithm is to find the structure on which the data lays on and representing it in fewer dimensions space
- The diffusion distance between samples is the Euclidian distance in that space
- The new space composed of multiplication between eigen values and vectors obtained by decomposing the Markov Matrix

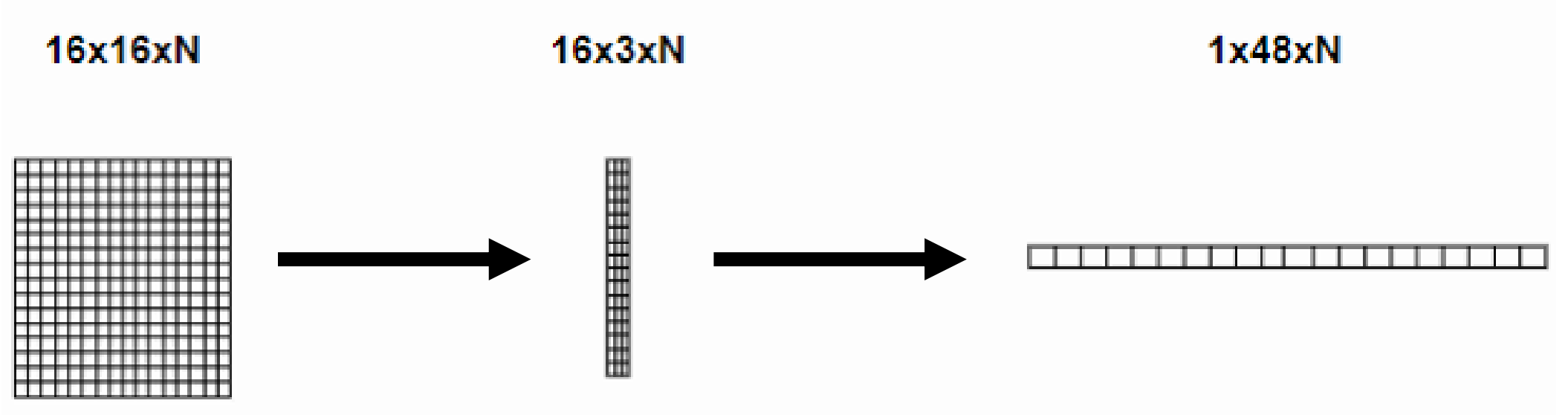


Example for 3-dimensional helix represented in 2-dimensional diffusion map

Features Extraction

- Applying the algorithm on the drivers' windows
  - Getting the new space matrix and choosing 3 vectors with biggest eigen values-
    $L \times 17 \Rightarrow 16 \times 16 \Rightarrow 1 \times 48$

  L represents the window length hyper-parameter



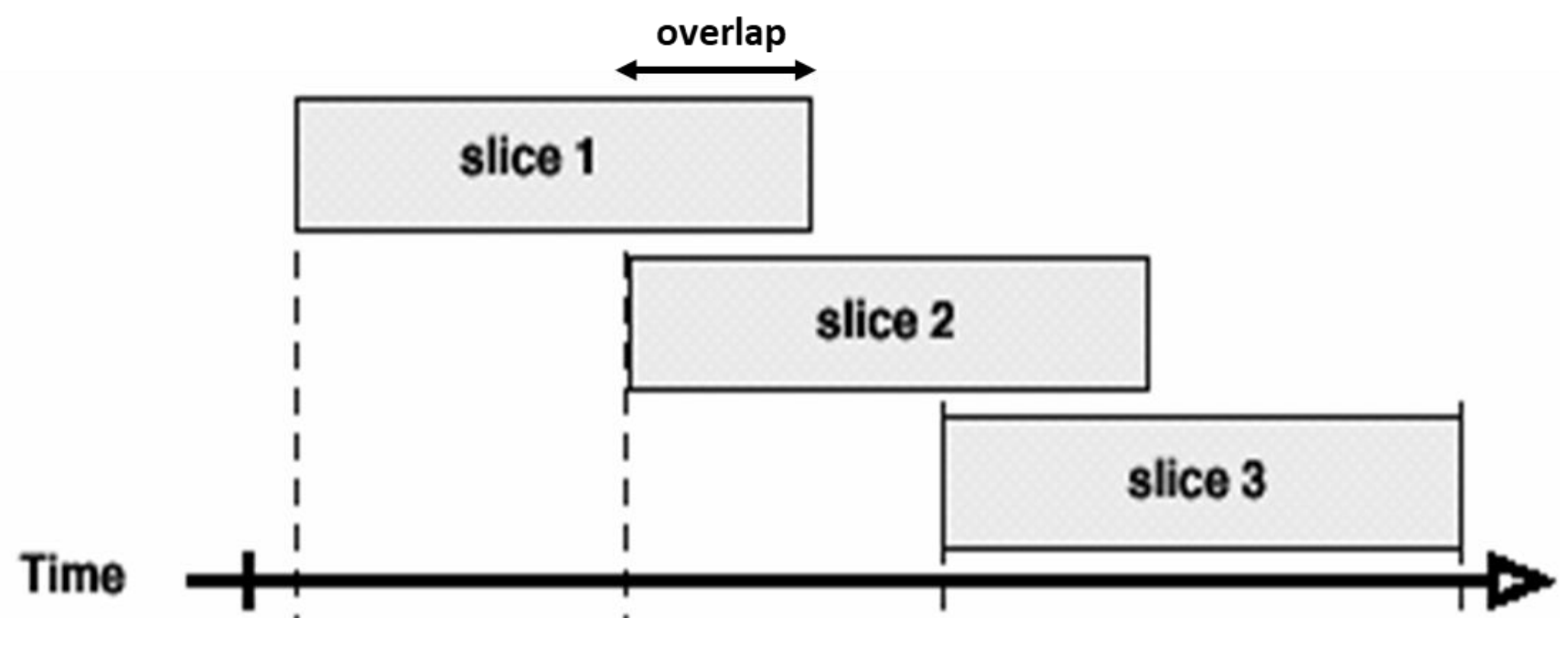
For each driver getting N windows, applying the algorithm resulting in vector for each window

Classification

- Chosen multi-Class classifier is Random Forest
- Training the classifier with feature vectors
- Optimize by cross-validation to determine hyper-parameters

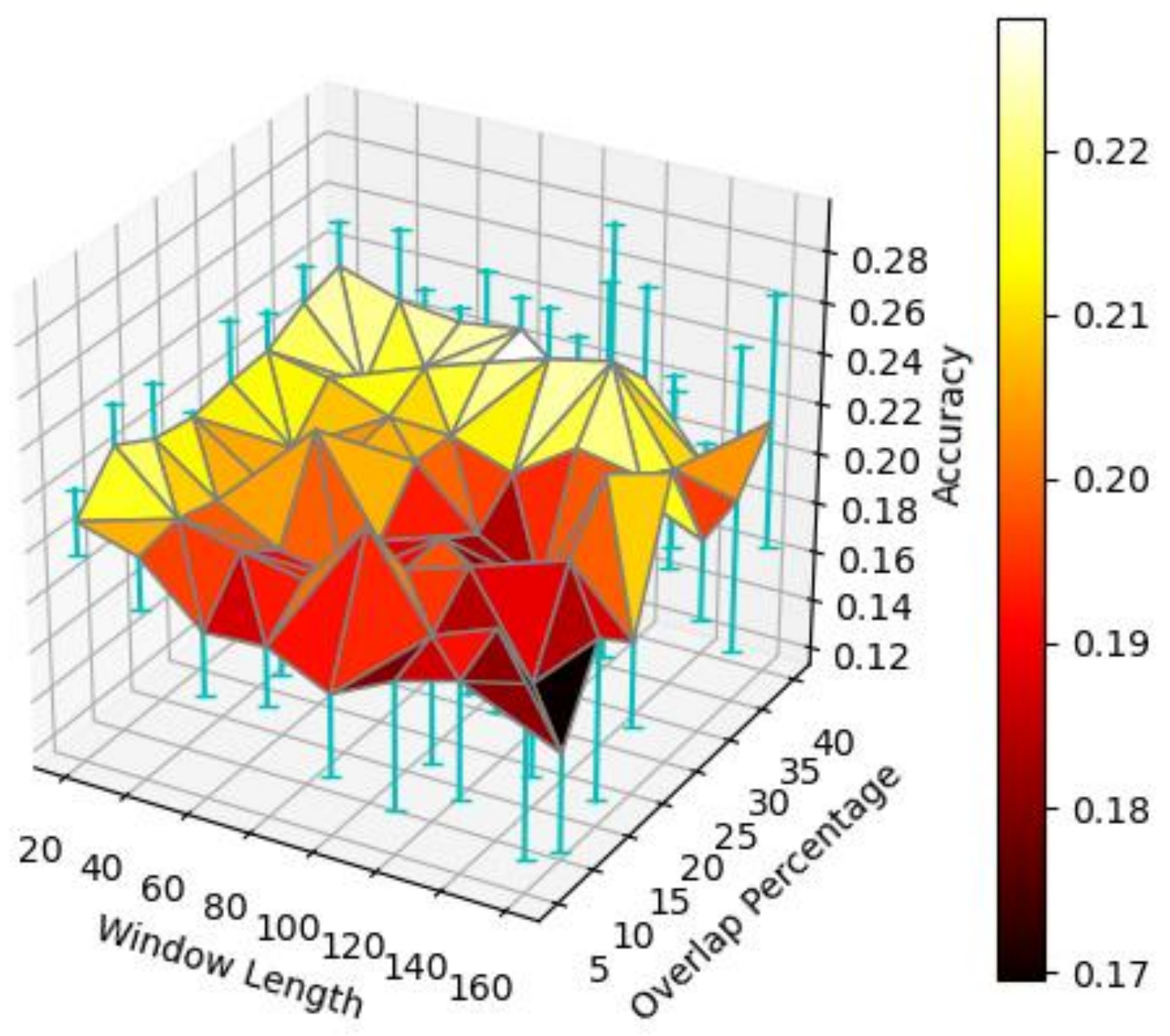
$$Window\ Length = [20,40,60,80,100,120,140,160]$$

$$Window\ Overlap = [5,10,15,20,25,30,35,40]$$



Visualization of the hyper-parameters

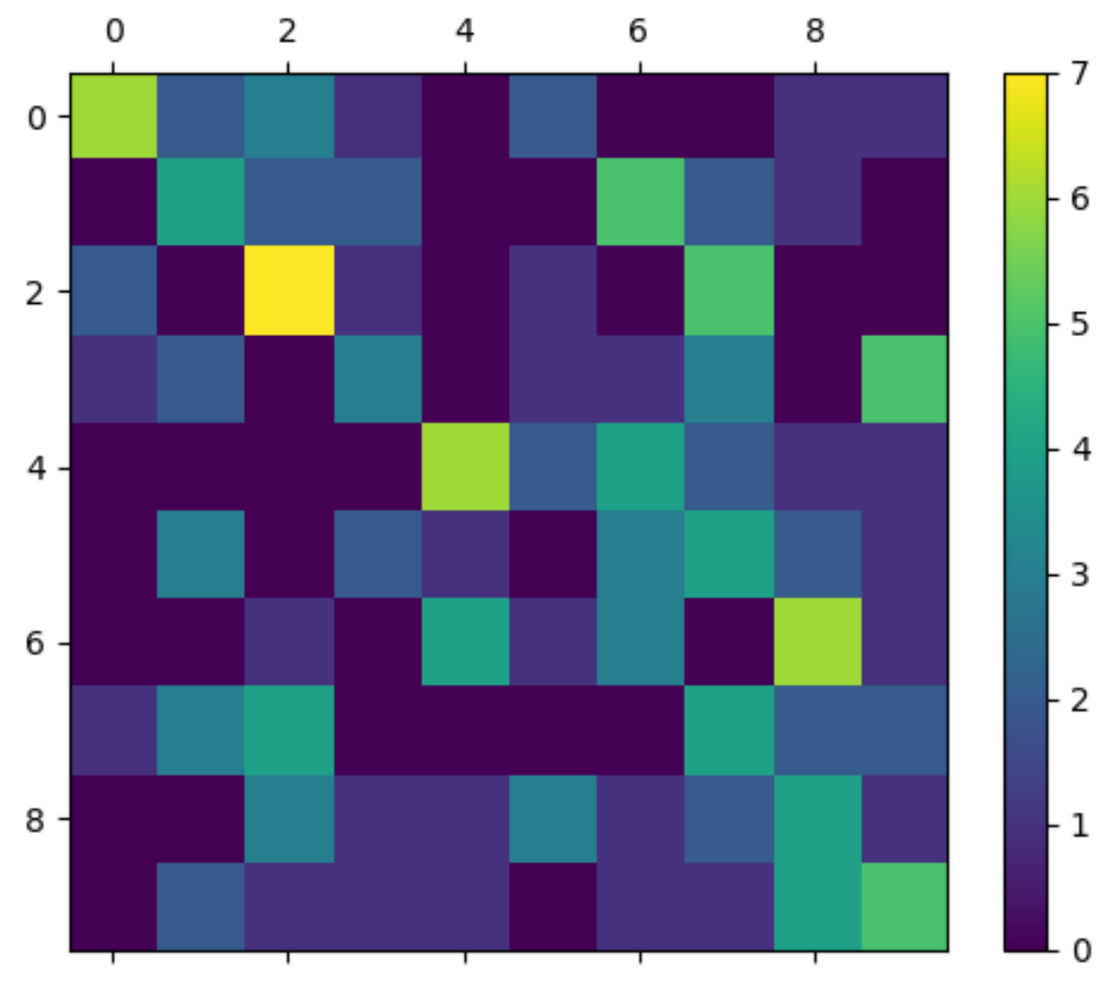
- Optimization results-



Window Length=120 Window Overlap=35%
Resulting 27%±5% accuracy

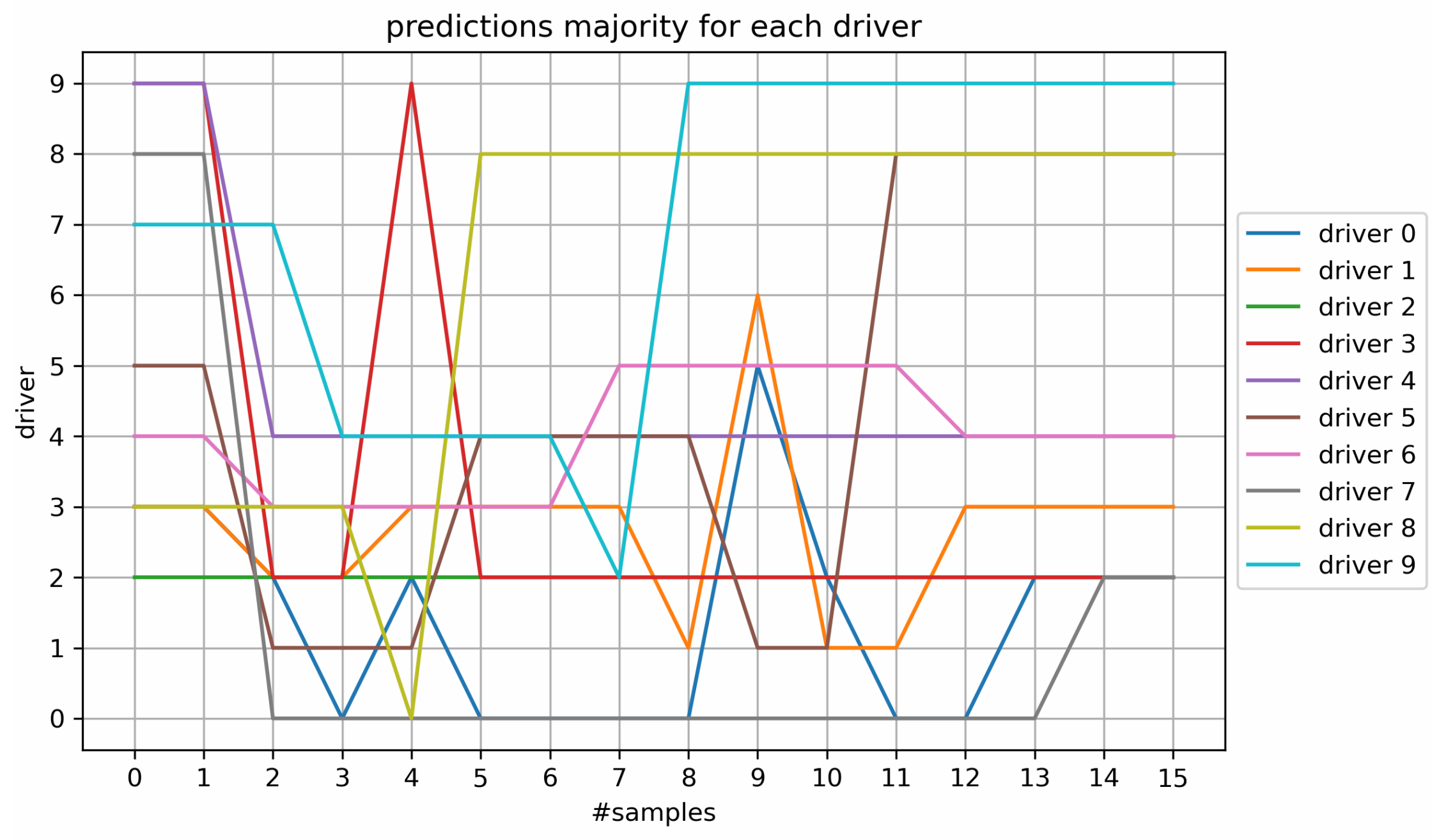
Results

- Using the results of the optimization
- Results on the test set-



Confusion Matrix on test set
Accuracy = 26.25%

- The classifier distinguishes better between drivers 0,2,4 with accuracy of 40%, and worse for others
- Simulation of real event- for each driver, taking the majority of the predictions through time-



- Drivers 4,8,9 converge to the right values

Conclusions

- Classifier results with optimized hyper-parameters with accuracy of 26.25% aren't satisfying
- Simulation of real event results aren't satisfying, reaching only 3 out of 10 drivers prediction right
- The diffusion space isn't allowing distinction between drivers
- Diffusion Map isn't a good fit for this data dimension reduction