Master in Computer Vision Barcelona

Project Module 6 Coordination

Video Surveillance for Road Traffic Monitoring

J. Ruiz-Hidalgo / X. Giró

j.ruiz@upc.edu / xavier.giro@upc.edu

Final Report: Team 4

Maria Vila (mariava.1213@gmail.com)
Sara Lumbreras (jfslumbreras@gmail.com)
Yael Tudela (yaeltudelabarroso@gmail.com)
Diego Alejandro Velázquez (diegovd0296@gmail.com)

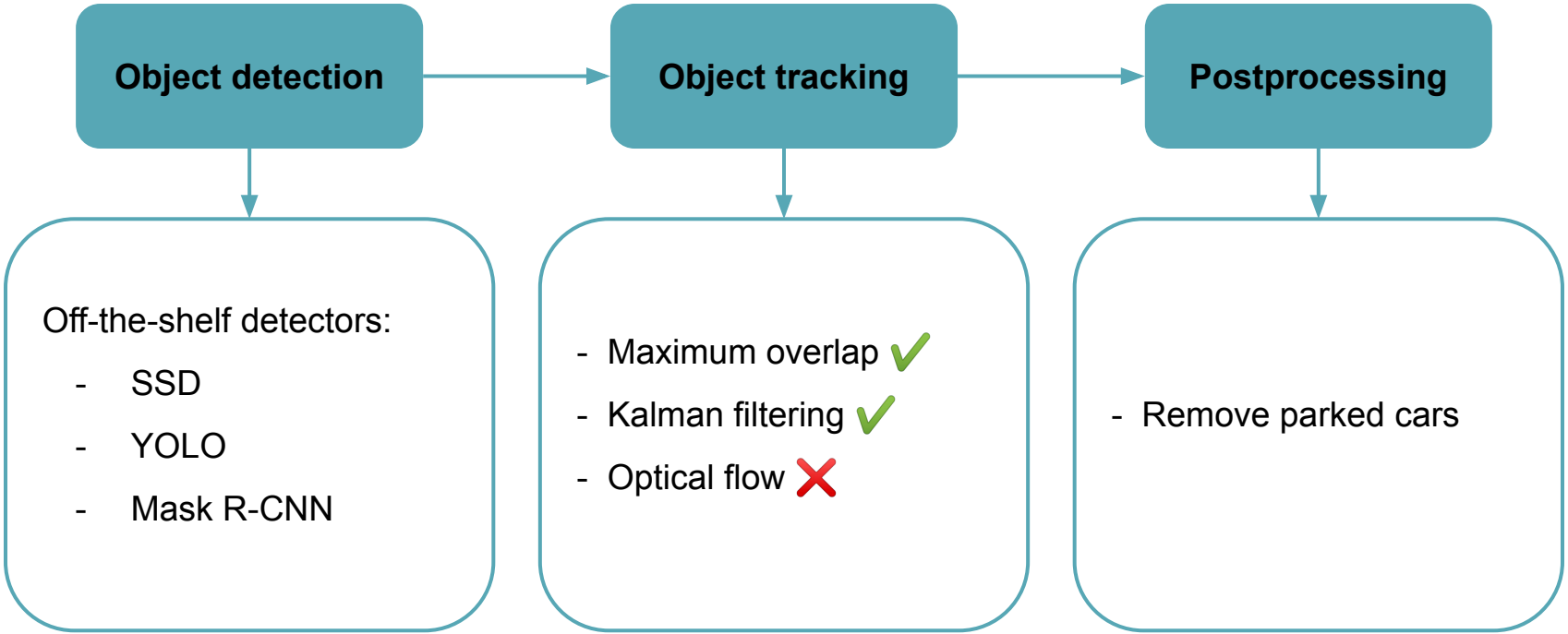
Master in Computer Vision Barcelona

UAB UOC UPC upf.

Outline

1. Multi Object Single Camera Tracking

1.1 Method


Object detection
Object tracking
Postprocessing
Off-the-shelf detectors:
- SSD
- YOLO
- Mask R-CNN
- Maximum overlap ✔
- Kalman filtering ✔
- Optical flow ✘
- Remove parked cars

1.1 Method

SORT algorithm [1]:

Uses Kalman filtering to track objects and refine bounding box predictions. Concretely, for each frame:

- Performs state predictions by propagating the state pdf
- Updates the prediction pfd with the measurements

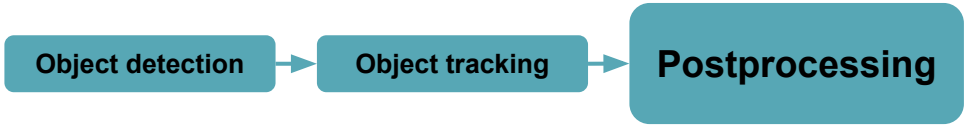
The statistical model used is a linear dynamic model, concretely a linear velocity model.

Maximum Overlap:

1. Remove overlapping detections
2. 1st frame: assign a new id for each detection
3. Frame N+1: Compare BB from frame N+1 to the ones in range {N - min(5, N), N}:
   a. If overlap of two consecutive BB > 0.4 → assign the same ID value
   b. For the other BB → assign a new ID value
4. If a detection appears in the first 5 frames of the video sequence → assign the same track value

[1] Bewley, Alex et al. “Simple Online and Realtime Tracking.” 2016 IEEE International Conference on Image Processing (ICIP) (2016).

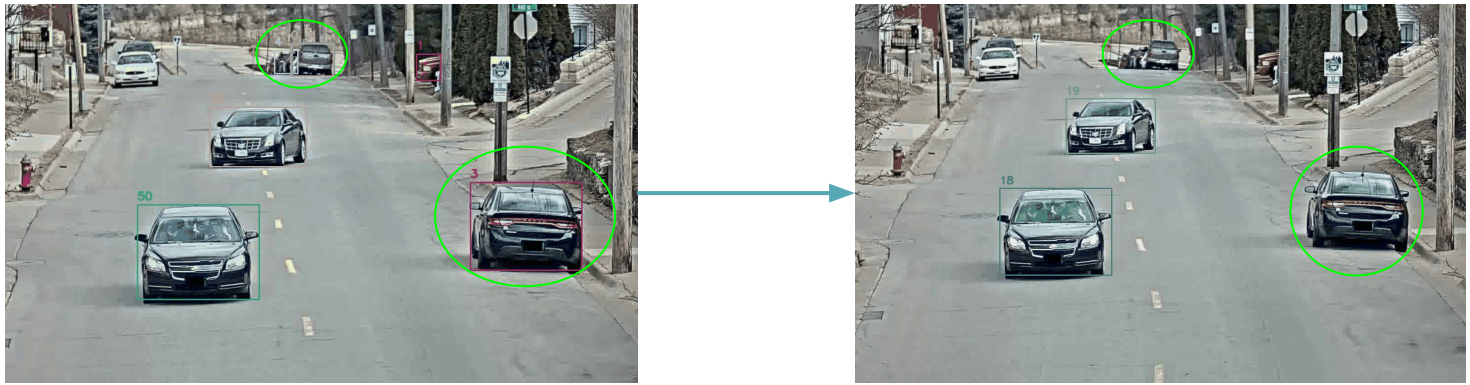
1.1 Method



Remove parked cars:

Compute the centroid of each track along the video sequence. Compute the variance of the centroid:

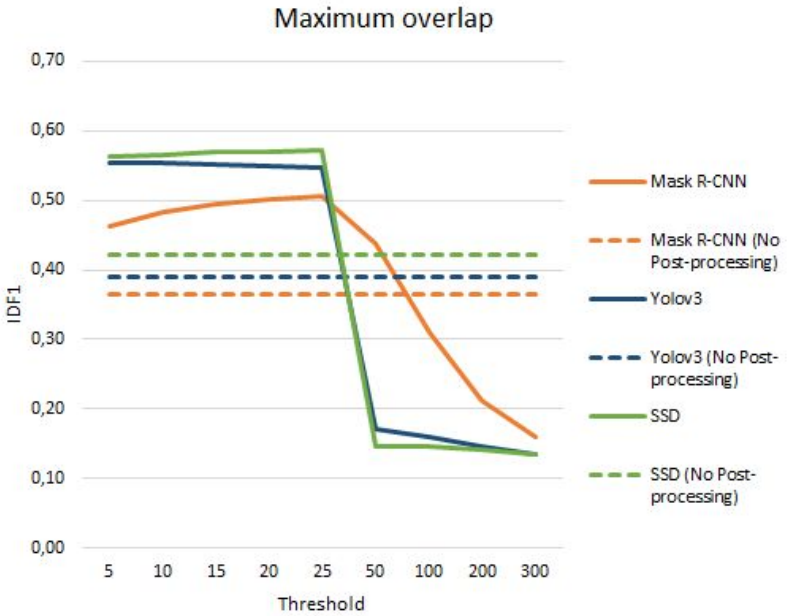
- If it is smaller than threshold mark as parked → remove the track
- If it is greater than threshold mark as moving → keep the track



1.2 Results

Used sequences S01 and S04 to obtain the best value of the threshold to remove parked cars.

- Low threshold: does not remove some parked cars
- High threshold: removes moving cars



1.2 Results

Sequence 3: IDF1 results

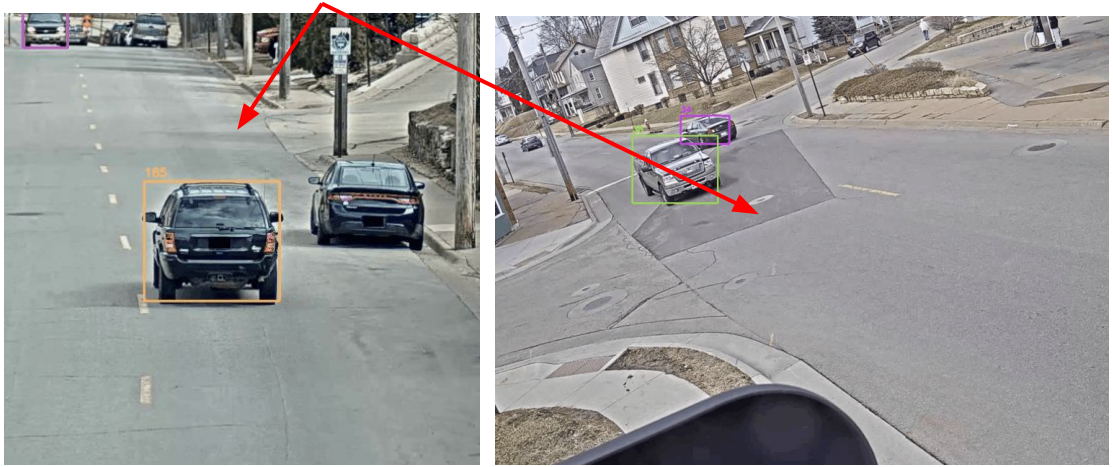
Detector	Tracking	Postprocessing	Camera							
			c010	c011	c012	c013	c014	c015	Average	Execution time*
Mask R-CNN	Kalman cte vel.	No	0.2505	0.0408	0.0254	0.3099	0.5880	0.0028	0.2029	~24.6 min
		Yes	0.7666	0.5241	**0.6263**	0.8184	**0.7467**	0.0699	0.5920	~25.5 min
SSD	Kalman cte vel.	No	0.4361	0.2418	0.0394	0.6247	0.5829	0.0152	0.3234	~36.3 min
		Yes	0.8353	0.6601	0.5156	0.6867	0.7361	**0.1611**	**0.5992**	~38.0 min
YOLO	Kalman cte vel.	No	0.3580	0.0911	0.0286	0.6600	0.5481	0.0050	0.2818	~33.5 min
		Yes	0.8415	**0.6839**	0.2291	0.7349	0.6908	0.0879	0.5446	~34.9 min
Mask R-CNN	Maximum overlap	No	0.2500	0.0419	0.0254	0.2964	0.4960	0.0029	0.1854	~21.0 min
		Yes	0.7972	0.4481	0.5689	0.3130	0.5734	0.0515	0.4587	~21.1 min
SSD	Maximum overlap	No	0.4286	0.2125	0.0423	0.7775	0.5491	0.0143	0.3374	~35.3 min
		Yes	**0.8756**	0.6082	0.4988	0.8161	0.6375	0.0982	0.5890	~36.7 min
YOLO	Maximum overlap	No	0.3358	0.0873	0.0243	0.7266	0.4387	0.0049	0.2696	~31.9 min
		Yes	0.8431	0.6737	0.4901	**0.8274**	0.4652	0.1014	0.5668	~32.8 min

*Intel(R) Core(™) i3-5010U CPU @ 2.10GHz 2.10 12GB RAM

1.2 Results

Here we show some detection problems we couldn't correct

Kalman filter: id swaps for the the same object.



Maximum Overlap: Overlapping objects are assigned the same Id.

1.2 Results

We compare the state of the art results with our best performing model: SSD with Kalman filtering and remove car postprocessing with a threshold of 25.

Method	IDF1	Recall
DS+YOLO	78.9%	67.6%
DS+SSD	79.5%	69.2%
DS+FRCNN	78.9%	66.9%
TC+YOLO	79.1%	68.1%
TC+SSD	**79.7%**	**70.4%**
TC+FRCNN	78.7%	68.5%
MO+YOLO	77.8%	69.0%
MO+SSD	72.8%	68.0%
MO+FRCNN	75.6%	69.5%

State of the art results

Metric	IDF1	Precision (detection)	Recall (detection)
SEQ 1	59.1%	94.5%	51.7%
SEQ 3	59.9%	72.3%	70.4%
SEQ 4	62.9%	88.9%	59.1%
AVERAGE	60.6%	85.2%	60.4%

Our results

1.3 Conclusions

Detectors:

- ❌ SSD is the slowest detector
- ✅ SSD is the best performing detector

Tracking:

- ❌ Maximum Overlap can assign same track value to multiple objects in the same frame
- ✅ Kalman with post processing outperforms Maximum Overlap
- ❌ Kalman qualitative results look worse

Postprocessing:

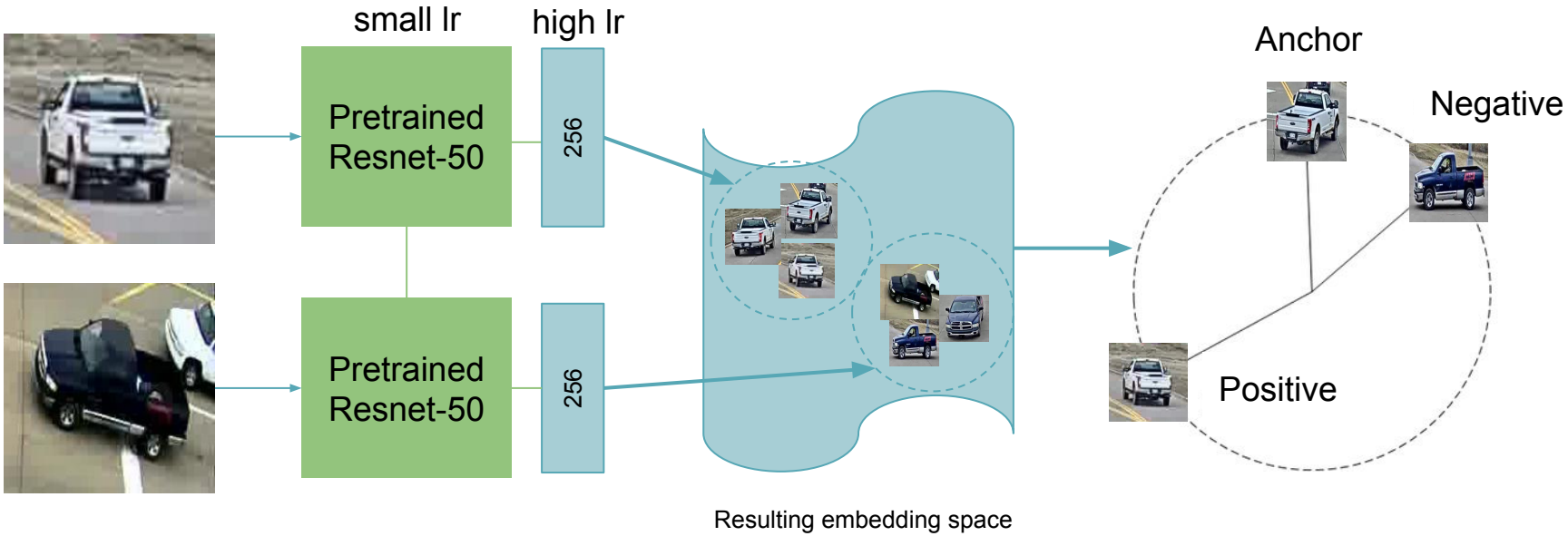
- ❌ Careful with the threshold of removing parked cars
- ✅ Improves the results

2. Multi Object Multi Camera Tracking

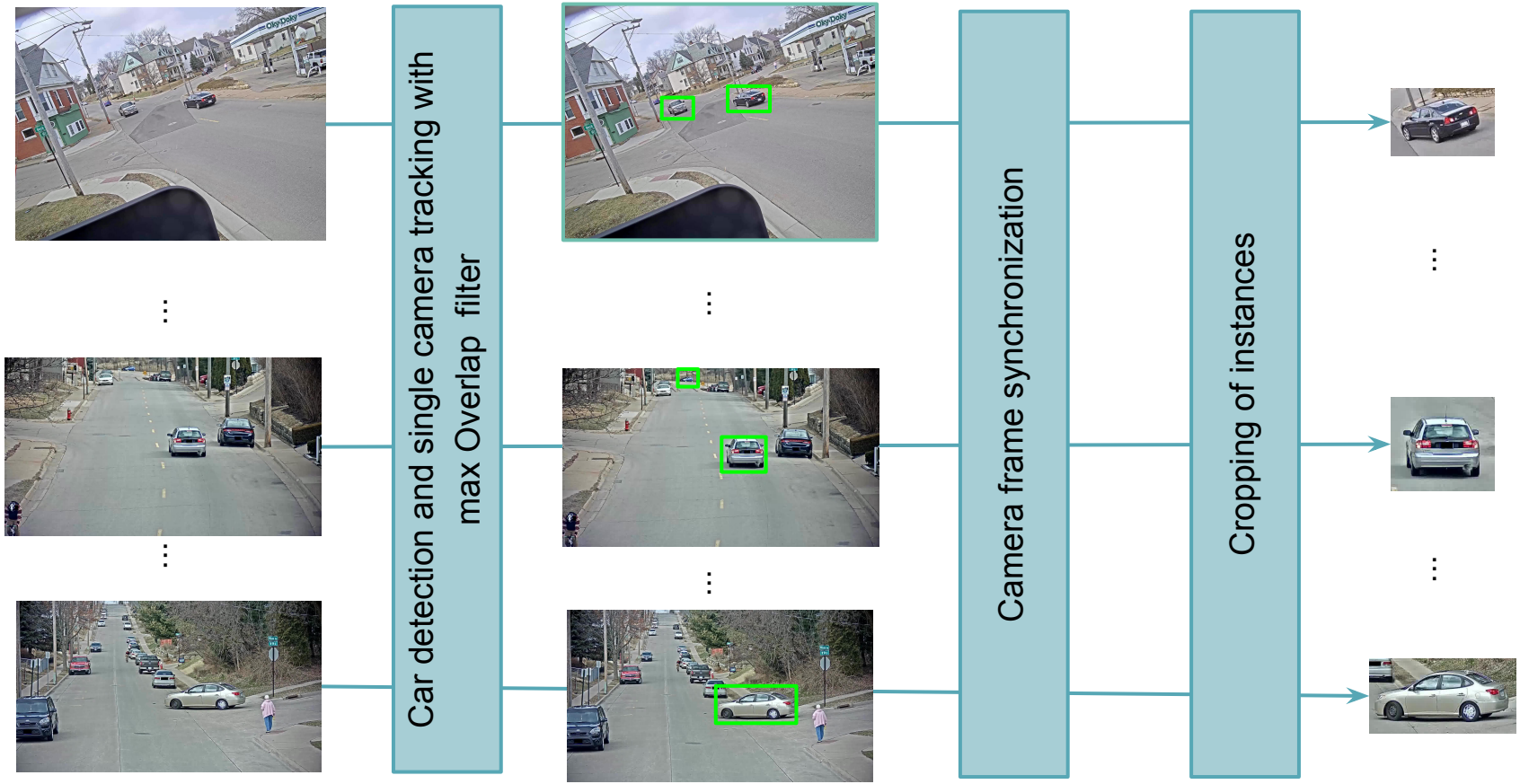
2.1 Method

Siamese network:

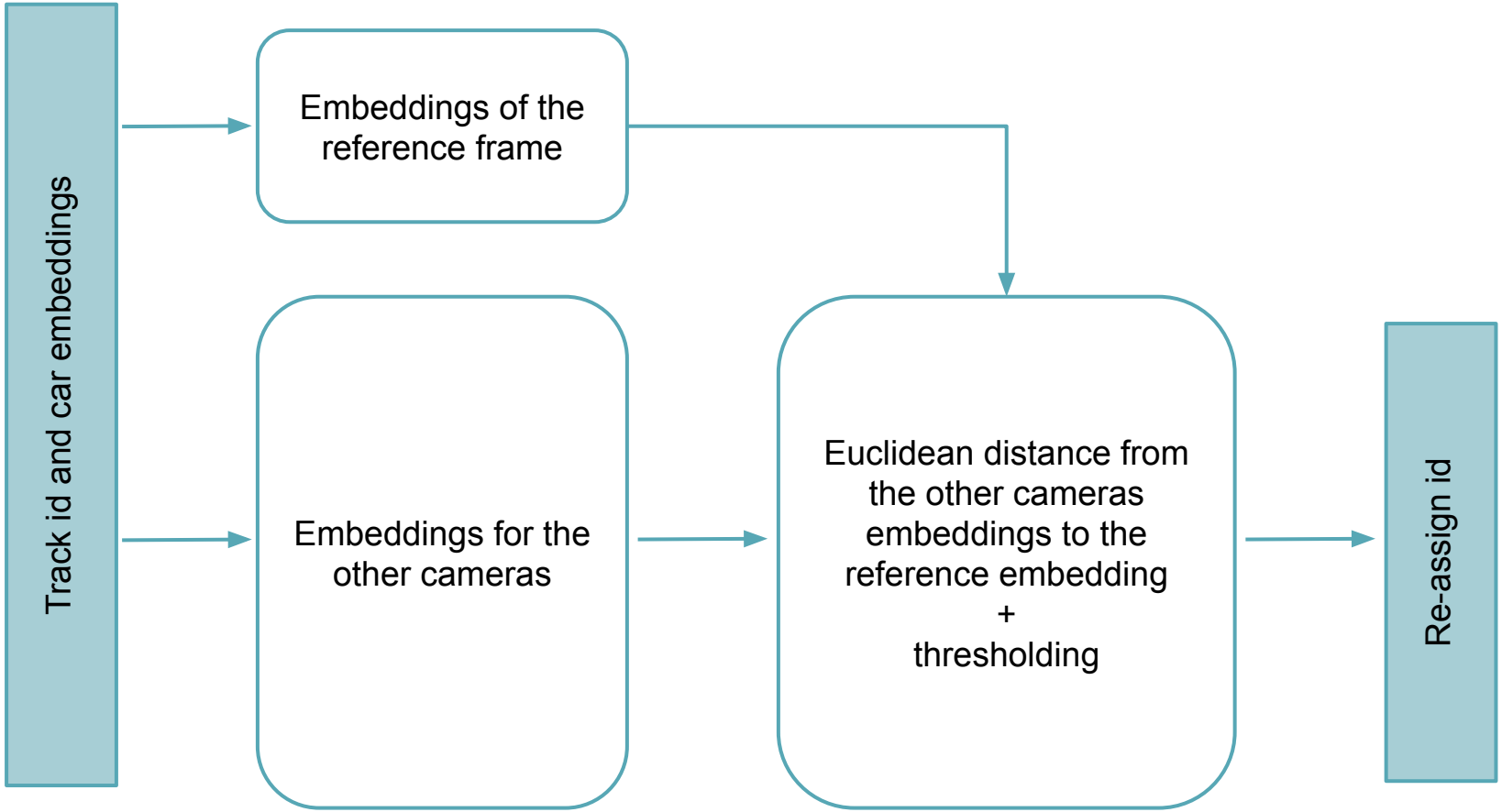
- We have used sequences S01 and S04 car crops as training data



2.3 Method

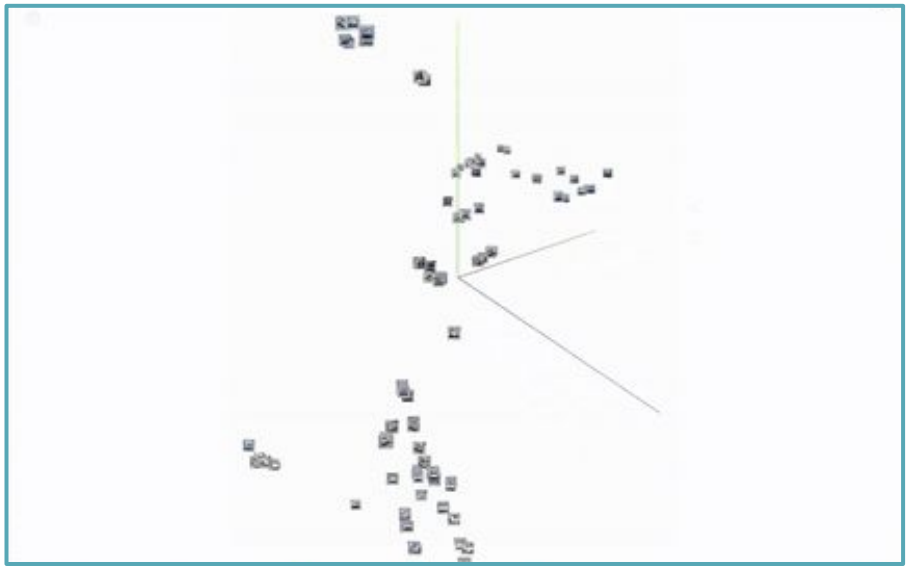

Car detection and single camera tracking with max Overlap filter
Camera frame synchronization
Cropping of instances

2.1 Method

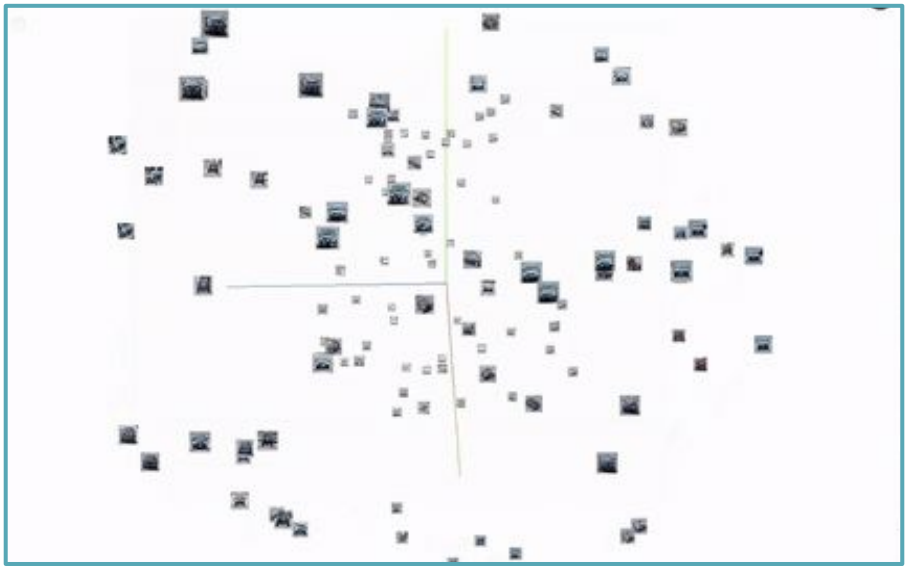

Track id and car embeddings
Embeddings of the reference frame
Embeddings for the other cameras
Euclidean distance from the other cameras embeddings to the reference embedding
+
thresholding
Re-assign id

2.2 Results

Visualization of the obtained embedding of sequence 3 for all the provided cameras.

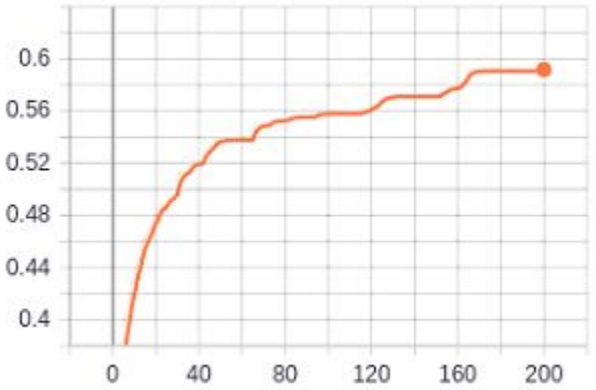


T-SNE

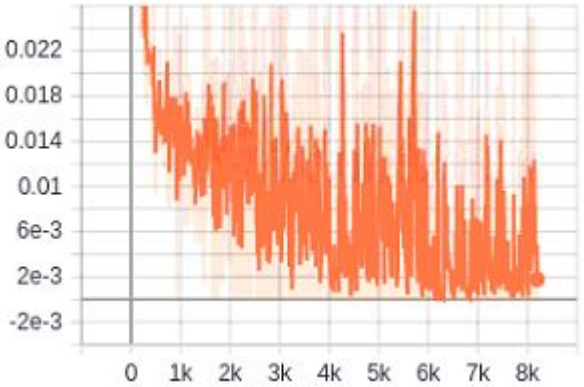


PCA

2.2 Results

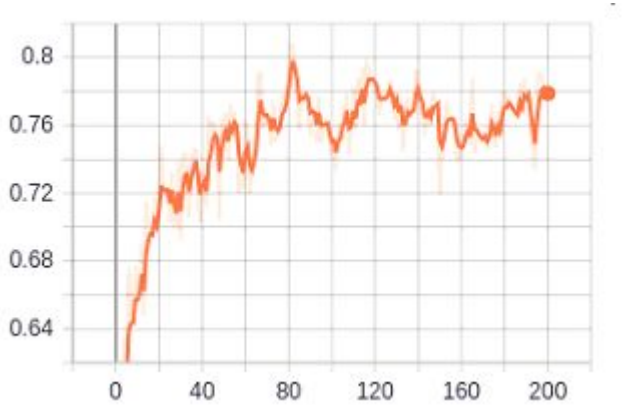


Accuracy



Metric Loss (triplet margin loss)

Here are some of the metrics extracted from the test set (sequence 3)



NMI (Normalized Mutual Information)

2.2 Results

			Camera						
Detector	**Tracking**	**Metric**	**c010**	**c011**	**c012**	**c013**	**c014**	**c015**	**Average**
Mask R-CNN	Max Overlap	IDF1	0.77	0.28	0.06	0.31	0.54	0.01	0.32
		Precision	0.63	0.17	0.03	0.23	0.58	0.01	0.27
		Recall	0.99	1	1	0.99	0.98	1	0.99
		IDP	0.63	0.16	0.03	0.19	0.43	0.01	0.24
		MOTA	0.42	-3.83	-28.24	-2.31	0.16	-158.5	-32.05
		MOTP	42.05	120.79	124.43	151.11	81.14	49.56	94.84

$$MOTP = \frac{\sum_{i,t} d_{i,t}}{\sum_t c_t}$$

MOTP: It's the average distance over number of assigned objects. It shows the ability of the tracker to estimate precise object positions, independent of its skill at recognizing object configurations, keeping consistent trajectories, etc. [1]

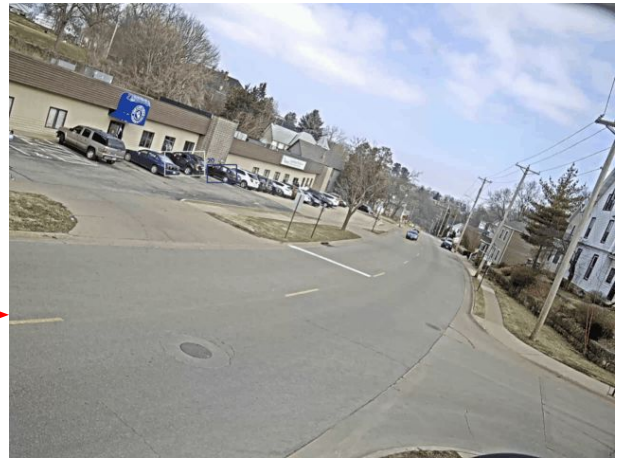
$$MOTA = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t}$$

MOTA: Overall performance metric. It takes into account the ratio of misses in the sequence, the ratio of false positives and the ratio of mismatches. [1]

[1] Bernardin, K., Elbs, A., & Stiefelhagen, R. (2006, May). Multiple object tracking performance metrics and evaluation in a smart room environment. In *Sixth IEEE International Workshop on Visual Surveillance, in conjunction with ECCV* (Vol. 90, p. 91).

2.3 Conclusions

Do not have the same ID



Embedding works but the multi tracking system does not assign the same id to the same cars.
Why we think it doesn't work?
Detections:

- Kalman distorts the detected bounding boxes and in some cases we get a negative coordinate
- Maximum overlap in some cases assigns same id to different objects in the same frame

Cameras:

- Hard to synchronize cameras, different frame rates and offsets
- Different cameras that point towards opposite directions

3. Conclusions

3. Conclusions

- We had to adjust the detections to the GT by removing parked cars.
- Multi object Single camera tracking provides decent results regarding the simplicity of the algorithms

- Multi object multi camera tracking is not accurate
- Try a different approach
- Improve the mentioned possible problems

Ground truth

Detections

Thank you for your attention,

do you have any questions?