M6 - Group 3 - Final Presentation

By Sergio Montoya, Laia Albors, Ibrar Malik, Adam Szummer

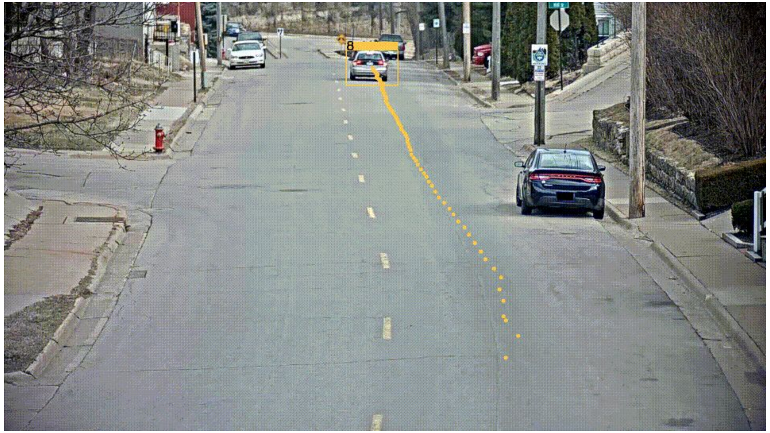
sergio.montoyadepaco, laia.albors, ibrar.malik, adam.szummer @e-campus.uab.cat

MTSC
Multi Target Single-Camera Tracking

Intro

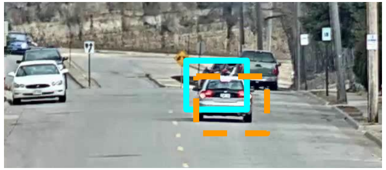
Goal: Assign unique IDs objects along the video.

Example of MTSC on camera c010 of sequence S03:



1. Detection at frame t

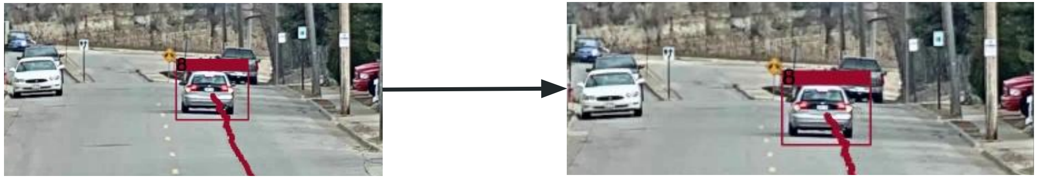
2. Update alive tracks, predicting the state for frame t.



3. Associate detections and tracks with Hungarian algorithm.

Track at frame t-1.

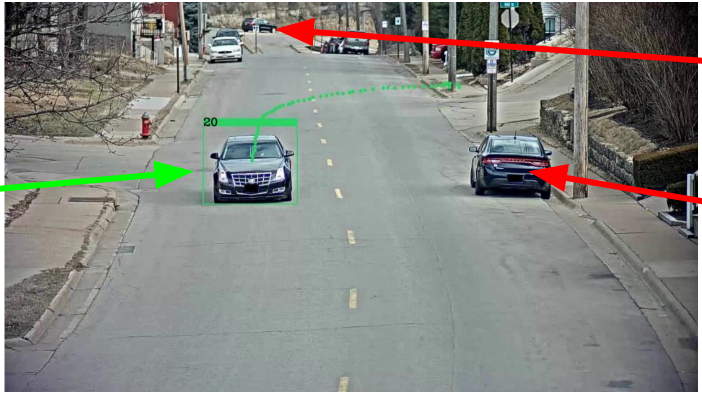
Track at frame t.



Start: Finetuning and what it does

- **Finetuning**: Only non-parked cars, that are near to the camera are annotated.

Annotated:
Non-parked cars,
near to the camera.



Small moving cars are not annotated.

Parked cars are not annotated.

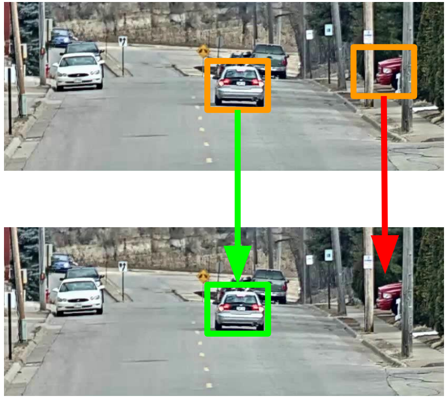
On sequence S03 c010:

- **Without** the finetuned network ➔ IDF1 of 0.357.
- **With** the finetuned network ➔ IDF1 of 0.664.

* Finetuned networks: FasterRCNN, RetinaNet, and SSD.

Tracking strategy + parameters

1. Keep only bounding boxes above threshold.



Parameters:

- Detection threshold

2. Update the state of tracks that are alive.

Trackers:

- KCF: OpenCV implementation.
- IoU (Static): Predicted bounding box is the bounding box from the previous frame.
- Kalman: With a constant velocity model.

- State:

$$\mathbf{x}_t = (\mathbf{p}_t, \mathbf{v}_t)^T$$

$\mathbf{p}_t$ is composed of the bbox center and <u>area</u> (which also grows or shrinks with constant velocity)

- Motion model:

we move in the direction of

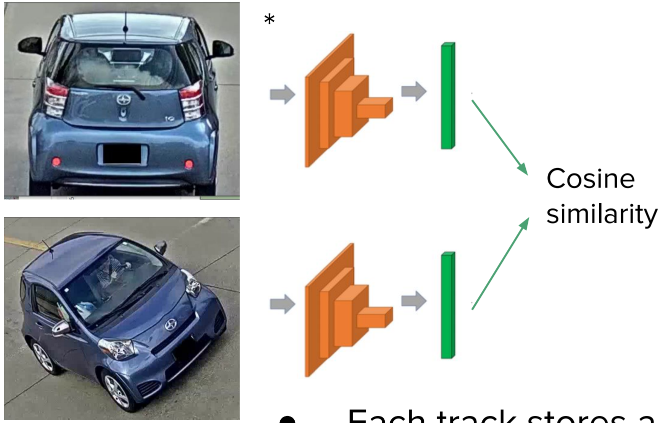
$$\mathbf{x}_t = \begin{bmatrix} \mathbf{1} & \Delta t \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \mathbf{x}_{t-1}$$

velocity remains constant

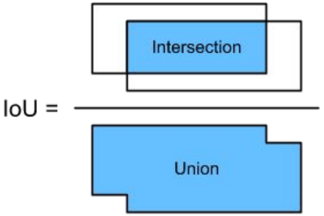
Tracking strategy + parameters

3. Association with Hungarian algorithm: bounding box similarity by IoU and identification template.

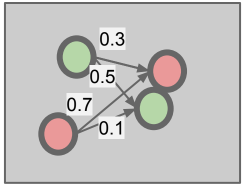
Problem: Which detection on frame *t+1* corresponds to every track in frame *t*.
Solution: For every candidate pair, we can compute some **similarity** value.



+



Contribution of each term weighted by a parameter called *beta*.



- Each track stores a running average of the feature vector that represents that track:

$$f_t = \alpha * f_{t-1} + (1 - \alpha) * f_i$$

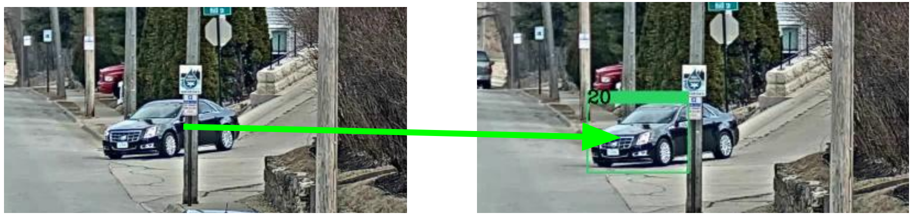
Parameters:
- Alpha
- Beta
- Minimum IoU for an association.

* Zheng et. al., Going Beyond Real Data: A Robust Visual Representation for Vehicle Re-identification
https://github.com/layumi/AICIty-reID-2020

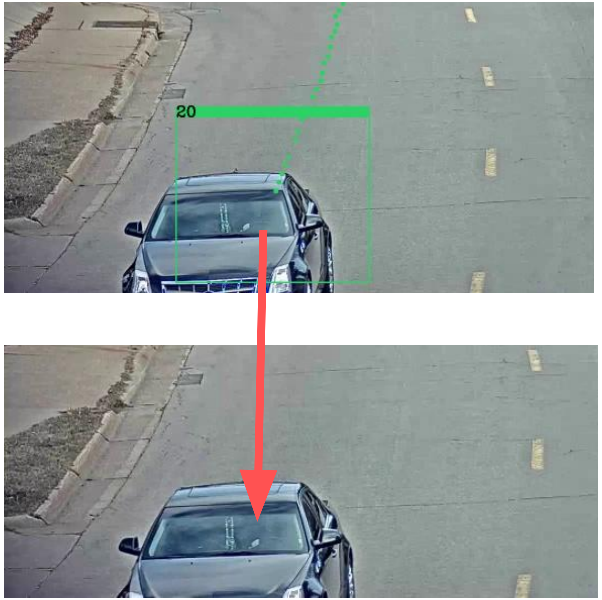
Tracking strategy + parameters

4. Remove and create new tracks.

- Create a new track for every unmatched detection.



- Remove those that have not been associated for the last n frames.

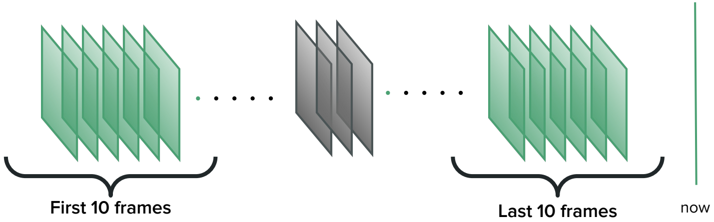


Parameters:

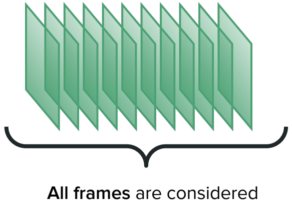
- How many frames to keep tracks without matches alive

Offline vs online post-processing

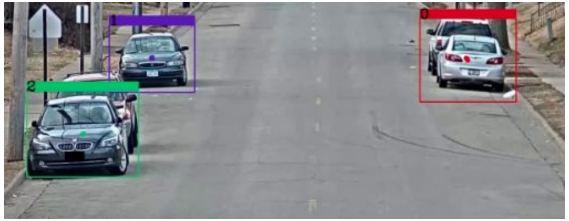
Online removal of parked cars, as tracking is in progress, allows to remove tracks for the parked cars as soon as the first and last 10 frames of a tracker do not update center of bounding box within threshold hence keeping only moving cars.

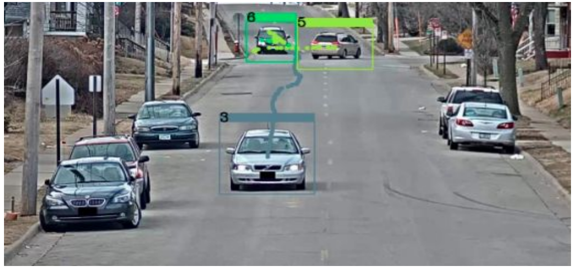


Offline removal of parked cars, can be done once the video is fully captured. In that approach we compare the std of all centers of track_history and remove tracks inside of a threshold.



Result





Group 3 - Adam Szummer

Offline vs online post-processing

For Sequence 03 we can see that offline post-processing yields better results giving 0.78 idf1 in average (without c015) in comparison to online with 0.74.

Possible disadvantage of online:

Certain cars on the camera might stop on the traffic or turn that results in deleting not updated tracker.

Possible disadvantage of offline:

You have to first capture the whole sequence for processing.

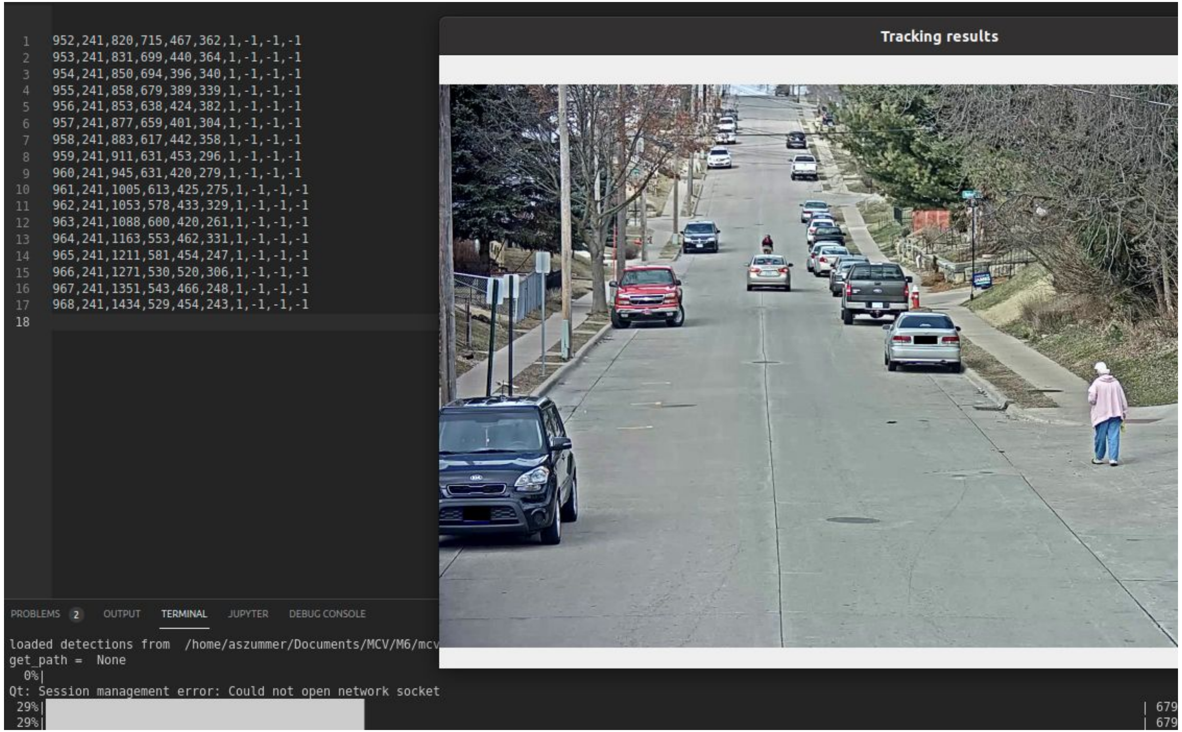
camera	Offline idf1	Online idf1
c010	0.88	0.86
c011	0.67	0.54
c012	0.88	0.82
c013	0.76	0.76
c014	0.73	0.76
c015	0.42	0.68

Note: ground truth for c015 is corrupted - example hidden slide

https://imatge.upc.edu/web/sites/default/files/projects/teaching/mcv-m6-video/aic19-track1-mtmc-train.zip

C015 corruption

Cars comes from the beginning of the camera. Data is downloaded directly from the server



Results grid search

Params:

Alpha: [0.2, 0.3, 0.4, 0.5, 0.6, 0.7]
Beta: [0.2, 0.3, 0.4, 0.5, 0.6, 0.7]
Skip_frames: [5, 10, 15, 20]
Conf_thresh: [0.4, 0.5, 0.6, 0.7]
Iou_thresh: [0.2, 0.3, 0.4, 0.5]
Tracker: [IoU, kalman, kcf]
Deep_sort: [True, False]

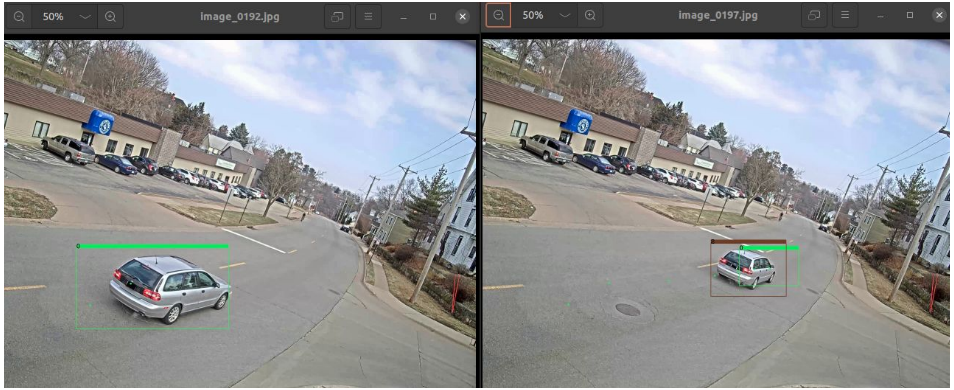
Alpha - describes the moving average between feature vectors
Beta - describes the importance of feature vectors for association
Skip_frames - describes max age of tracker
Conf_thresh - describes detection threshold
IoU_thresh - describes IoU for association
Tracker - describes different types of trackers
Deep_sort - whether the feature vectors are used

Best Results:

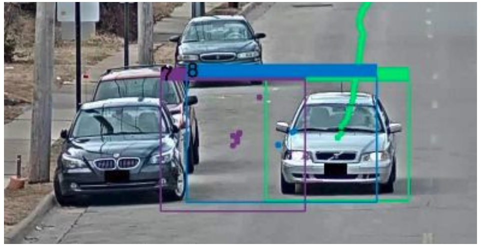
Group 3 - Adam Szummer

Sequence	IDF1	alpha	beta	skip_frames	conf_threshold	Iou_thresh	Tracker	Deep_sort
S03	0.74	0.7	0.7	20	0.4	0.2	kalman	False
S04	0.60	0.5	0.0	15	0.6	0.4	kcf	False
S01	0.71	0.2	0.5	10	0.6	0.4	kalman	True

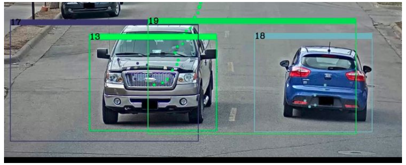
Qualitative analysis

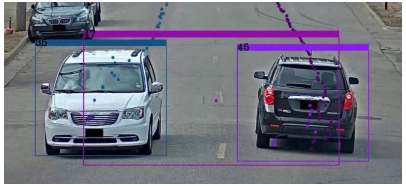


Car quickly entering the scene, changing the size of bounding box rapidly so that IoU is off



Getting some wrong detections along the parked cars, probably due to low conf threshold



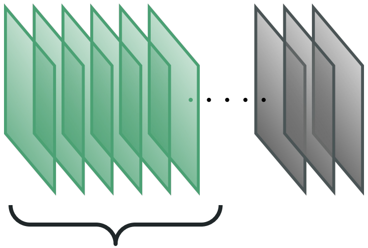


Merging two cars into one detection when one car enters the scene

MTMC
Multi Target Multi-Camera Tracking

MTMC: two approaches

ONLINE

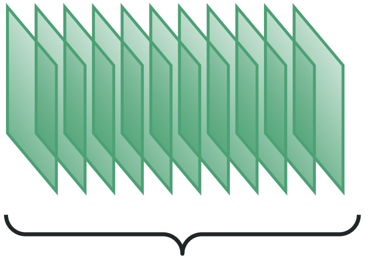


Only **past frames** are used

tracking + feature extraction* +
multi-camera matching

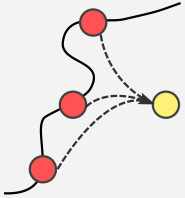
*Zheng et. al., Going Beyond Real Data: A Robust Visual Representation for Vehicle Re-identification [link]

OFFLINE



All frames are considered

tracking + feature extraction*

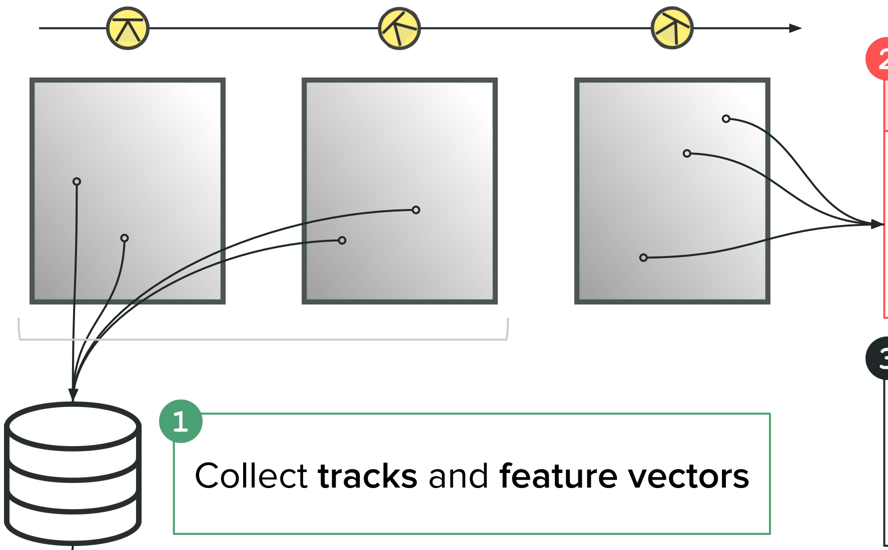


feature aggregation +
multi-camera matching

Online

for every camera video...

at every time step



2 New detections/tracks

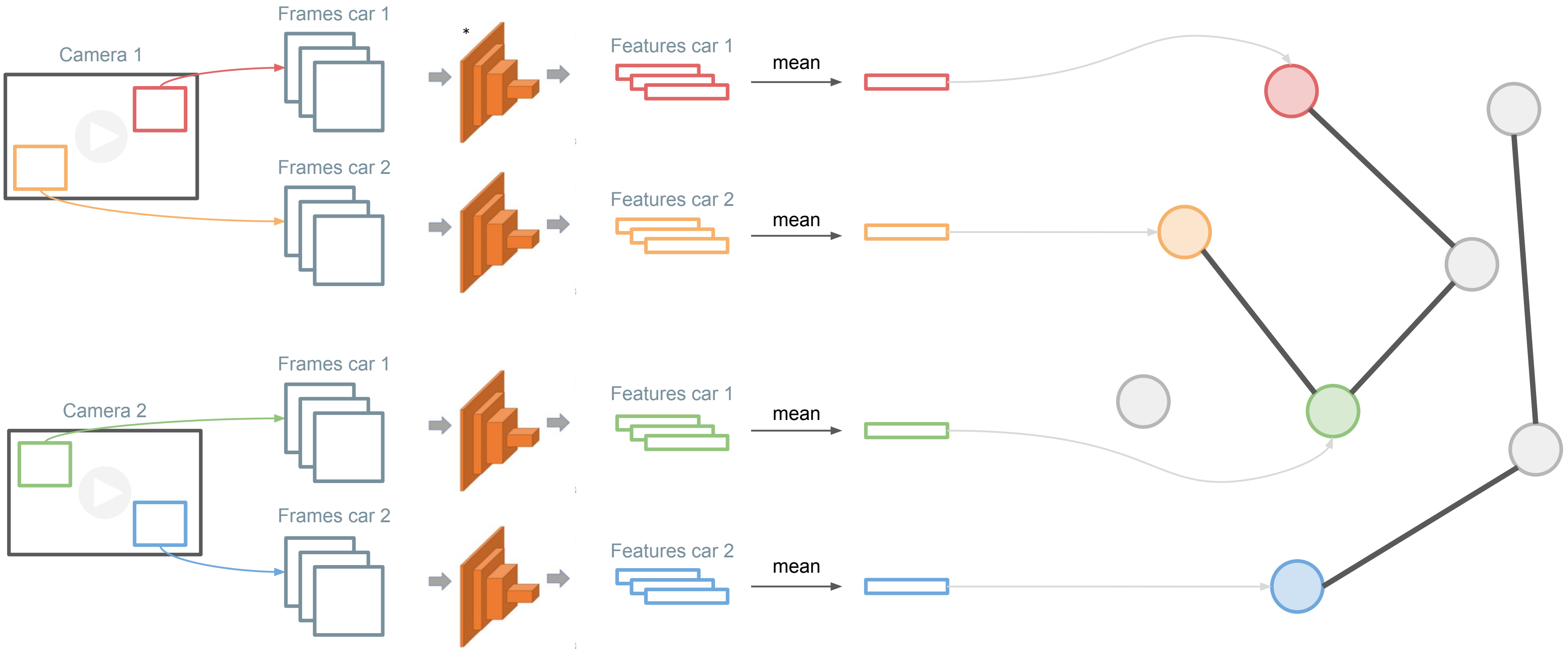
1. **Compare** with all **existing tracks** (similarity matrix).
2. **Linear assignment**

1 Collect **tracks** and **feature vectors**

3 Replace the local track id with the identifier of its most similar track

After *k* frames, if a track stays **static** we discard it

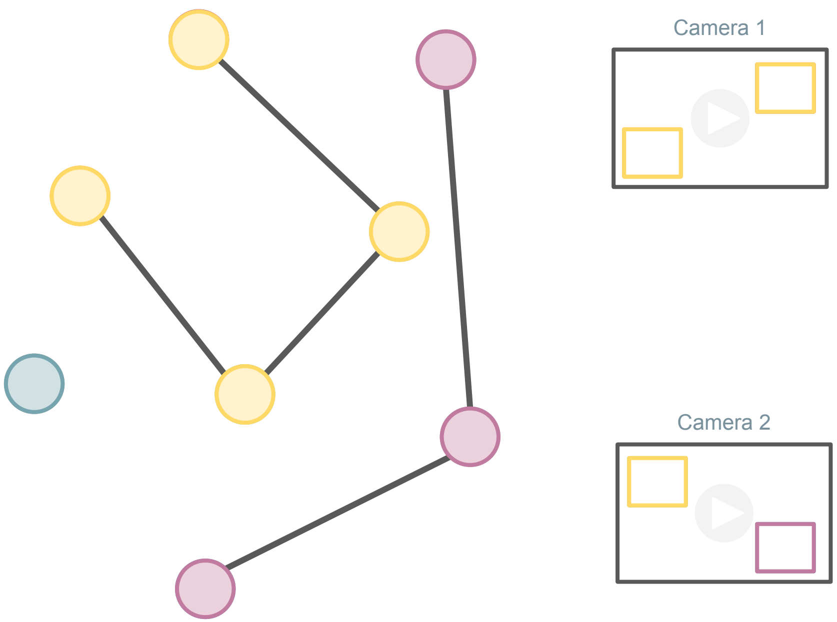
Offline



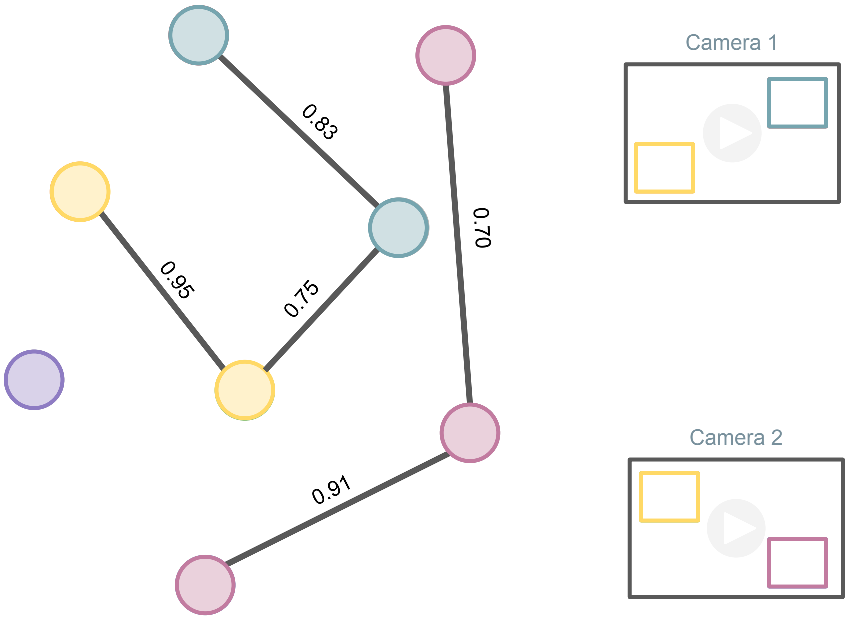
*Zheng et. al., Going Beyond Real Data: A Robust Visual Representation for Vehicle Re-identification [link]

Offline

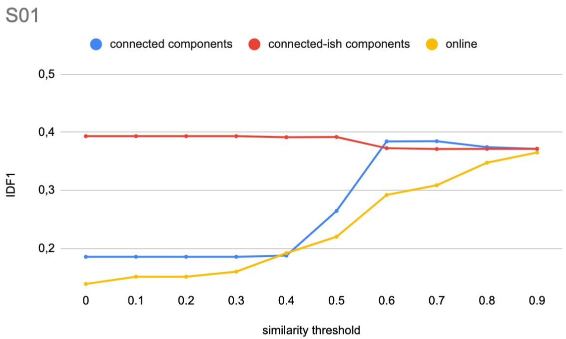
Connected components

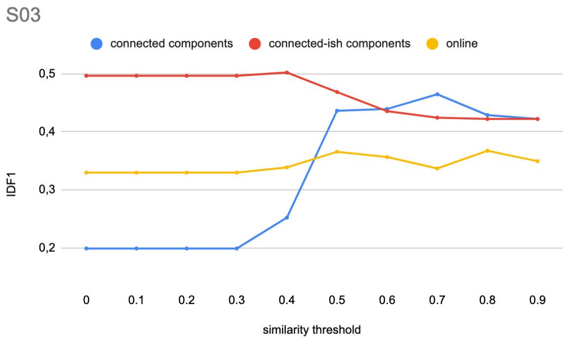


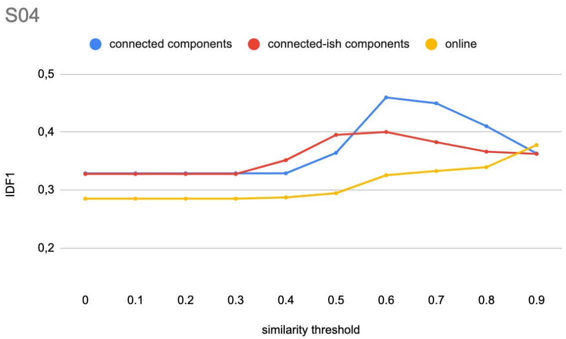
Connected-ish components



Final results







Metric	IDF1	IDP	IDR	Precision	Recall
SEQ 1	0,392	0,324	0,505	0,580	0,915
SEQ 3	0,502	0,441	0,590	0,703	0,948
SEQ 4	0,352	0,342	0,426	0,647	0,918
Average	0,415	0,369	0,507	0,643	0,927

Best: similarity threshold = 0.4 and the offline algorithm with our version of the connected components computation.

Qualitative results

=

Qualitative results

≠

Conclusions

- **Ground truth:** we must follow the dataset definition (**no parked cars**) and know its limitations (**no small detections**) ➔ **preprocessing** is required.
- **Tracking issues:** occlusions, cars crossing themselves, ...
- **Hard to generalize:** to get good scores, we require **hyper-parameter tuning** and **fine tuning** for every sequence.
- **Offline vs. online:** better results on both MTSC/MC with offline, but online seems more practical.

Future work

- Use insights from MTMC to solve issues on MTSC: merge interrupted tracks because of occlusions, better association of detections using learnt features.
- Further exploration of learnt features.

Conclusions

WIP:

- Important to understand ground truth so we can have correct evaluations (smaller cars not labelled, parked cars not labelled, ...).
- Fine Tuning is crucial for higher scores of tracking
- Tracking problems has lot of dependencies such as occlusions by cars passing by, feature detection from different angles, camera positioning
- It is hard to find best parameters for all cameras in the same time
- Post-processing for removing parked cars is key to improve results as the annotations don't consider them
- Offline tracker needs to process the whole sequence first but yields better results as it removes static tracks better
-

Future work:

- Merge the interrupted tracks that comes from the same car based on feature vectors from the re-identification network.

THANKS!

ANY QUESTIONS?