Progress report

Multiple Camera Tracking

3rd stage

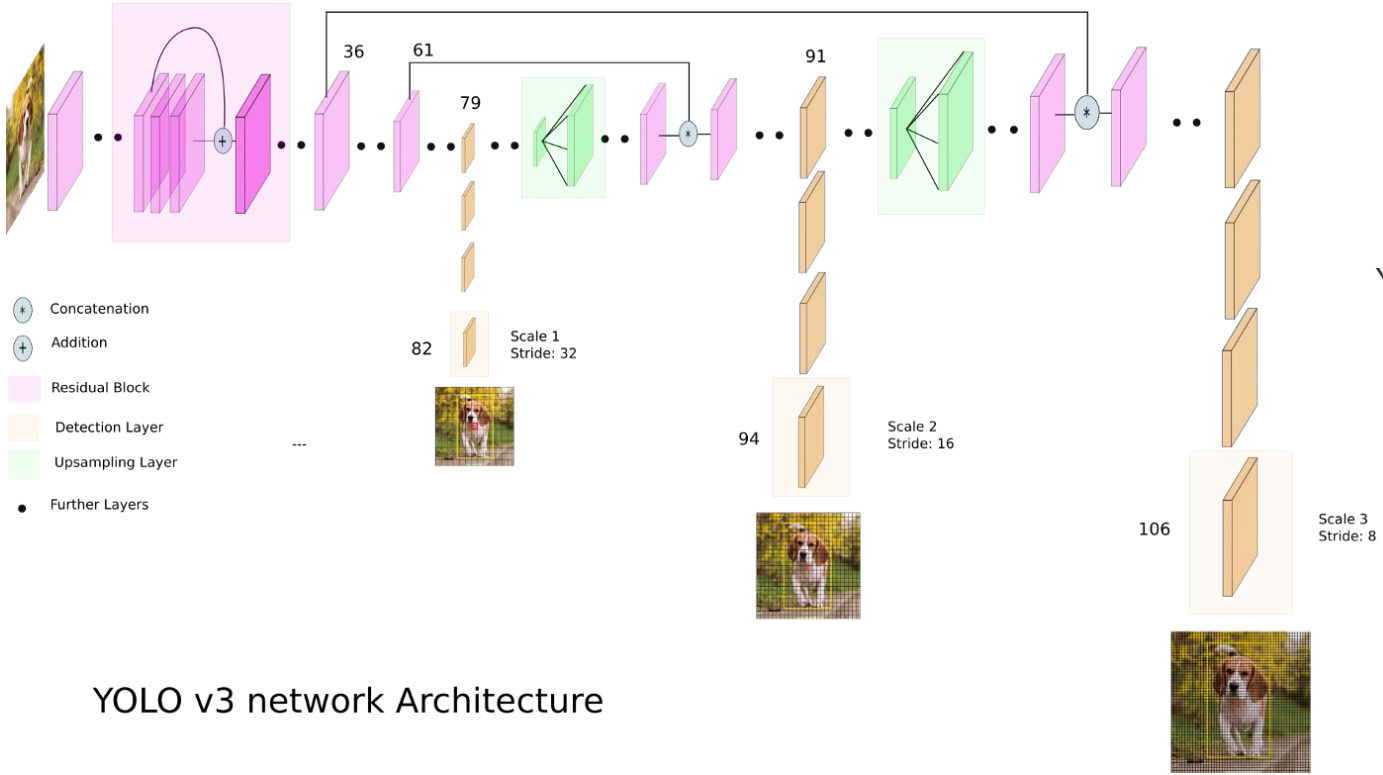
Content

1. SCT overview

SCT = Detection + Tracking

1.1. Detection

Object Detection = Object Localization + Object Classification



YOLO v3 network Architecture

YOLO:

- One-stage detector
- Detection is made on feature maps of different scales
- Anchor-based

YOLOv5:

- Backbone: CSPResBlock reduces FLOPs, while yield more informative gradient
- Inference in fp16
- Neck: PANet allow information propagation between detection layers more efficient
- Augmentation: Scaling, Color adjustment, Mosaic
- Auto learn anchor boxes

1.1. Person Detection

Dataset

- Class *Person* from COCO 2017
- 64,115 for training, 2,693 for validation

Model

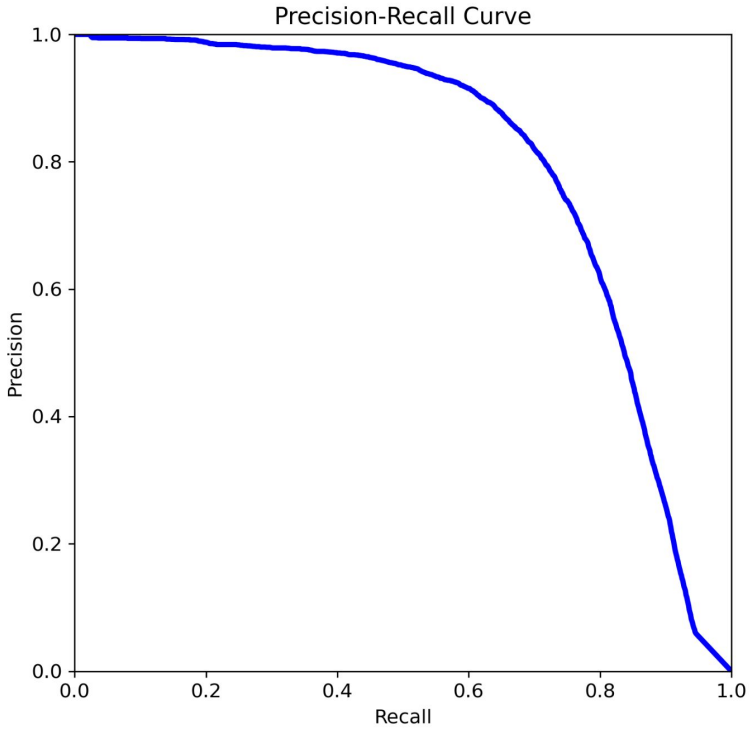
- YOLOv5s
- Use default parameters
- Input size: 640

Training time: 100 epochs

- AP@*0.5*: 0.79836
- AP@*0.5:0.95*: 0.54389

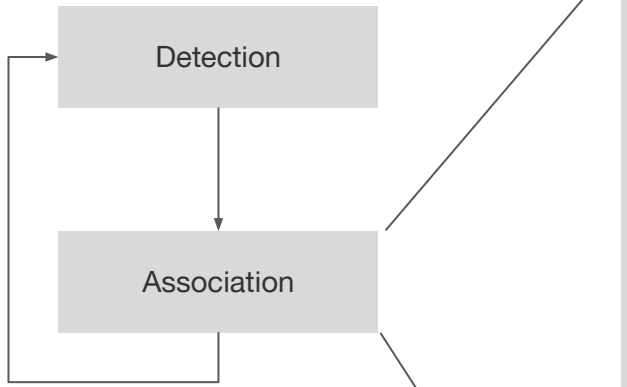
Assessment on MOT17 (1920x1080):

- False Negative is very frequent
- False Positive is rare



1.2. Tracking (SORT)

Tracking-by-Detection approach



SORT

Given:

- Detections D at frame t
- Prediction P using state from frame t-1 (Kalman)

Associate D with P using Hungarian algorithm

Properties:

- State:
  - Include motion information
  - Not include appearance information
- Not focus on solving occlusion issue
- Only use 1 previous frame for prediction
- Kalman assumes constant velocity

1.2. Tracking

Evaluation on MOT17

- Using previous trained person detection of YOLOv5
  - confidence threshold: 0.5
- iou_threshold: 0.3
- Using default Kalman Filter

1.3. Experiments

Detection		Tracking		Evaluation						
Input size	IoU thresh	max_age (s)	min_hits (s)	HOTA	DetA	AssA	GT_Dets	GT_IDs	Dets	IDs
640	0.45	1	1	31.103	20.637	47.117	336891	1638	90195	957
1280	0.45	1	1	35.668	26.745	47.899			117807	1215
1280	0.50	1	1	35.688	26.699	48.038			117693	1218
1280	0.55	1	1	35.833	26.76	48.312			118065	1224
1280	0.60	1	1	35.479	26.512	47.808			117474	1224
1280	0.55	0.5	1	36.021	27.468	47.591			120159	1347
1280	0.55	1.5	1	36.064	27.228	48.113			119547	1296
1280	0.55	2	1	36.043	27.195	48.114			119379	1290
1280	**0.55**	**1.5**	**0.15**	**41.625**	**36.583**	**47.88**			**169167**	**2343**
1280	0.55	1.5	0.25	40.728	34.763	48.18			157383	1974
1280	0.55	1.5	0.5	38.789	31.384	48.329			138270	1563

2. MCT research problem proposal

2.1. MCT overview

1 Re-ID:

Use appearance information (visual feature) to determine the same or different persons.

1. Feature learning: **OSNet**, FastReID
2. Metric Learning: **Triplet Loss**, **Cross Entropy Loss**, Center Loss
3. Matching to define same or different person based on gallery and query method.

2 **Spatio-Temporal Association**

Use spatial (camera layout, adjacent areas, ...) and temporal information (sequential of moving time) to matching person through many cameras.

2.1. MCT overview

STA with non-overlapping field of view

[1] build assumption from statistic (histogram): 2 images are likely the same when ST difference is small, they are likely different when ST difference is large. Do Re-ID, then re-ranking.

$$ST(i,j) = \frac{|T_i - T_j|}{T_{max}} \times \frac{\delta(C_i, C_j)}{D_{max}}$$

[2] combines ST and visual feature for the ranking.

$$C(I_i, I_j) = \frac{\|T_i - T_j\|}{T_{max}} \times \frac{\delta(D_i - D_j)}{D_{max}} \times d(f(I_i), f(I_j))$$

With known physical distance between cameras, [3] estimate 3D speed of object from SCT then derive the travelling time between camera for each object.

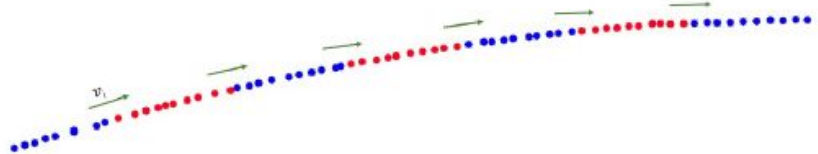
STA with overlapping field of view

Mapping strategies:

- Without calibration: Homography
- With calibration:
  - Epipolar line constraint
  - Pixel to Physical coordinate

[4] use GMM and EM algorithm to learn the entry/exit zone between cameras, then learn the probability of someone moving from exit of camera A to entry of camera B.

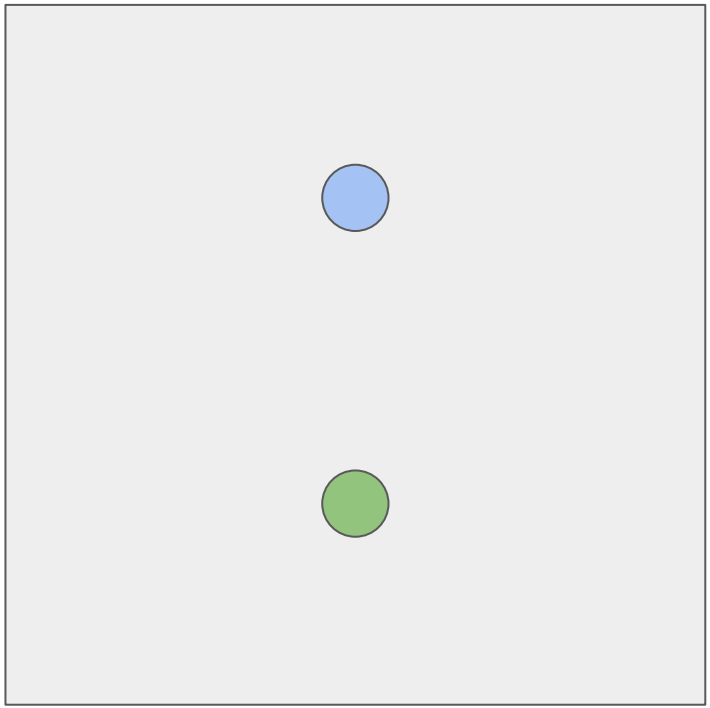
[5] map trajectories from different image planes to a same physical coordinate, then compute **direction similarity** and distance among them, combine both for the ranking.



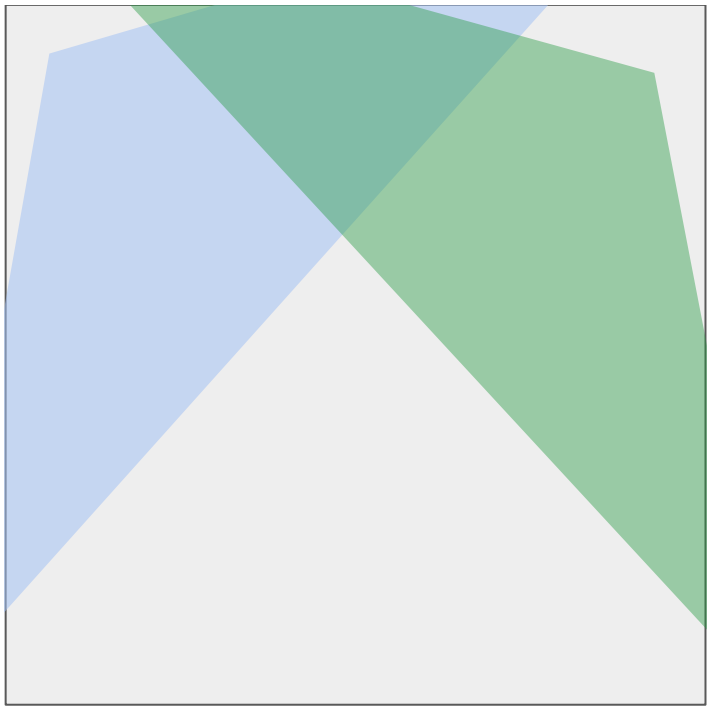
Goal: Research and develop solutions for STA with overlapping field of view.

2.2. Proposal: Camera Setup

360 degree camera



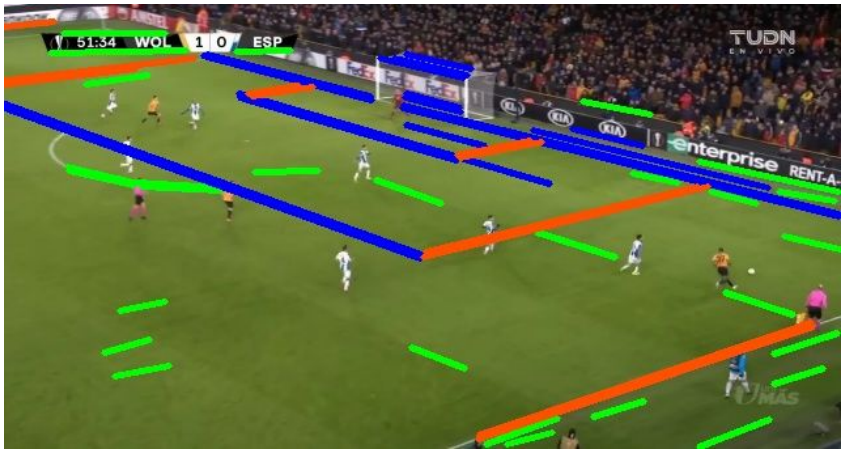
2D camera

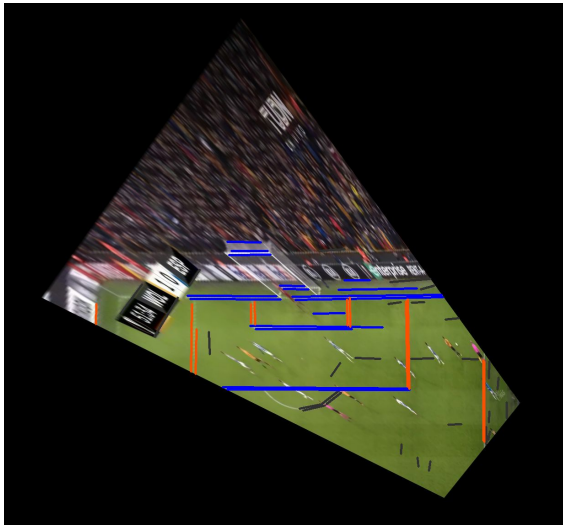


2.2. Proposal: Approaches

1 Homography (without calibration)

1. Find a homography H between 2 cameras in the overlapping area.
2. For track T, transform its trajectory in overlapping area of camera 1 to the view of camera 2 using H.
3. Filter all the tracks in camera 2 that is created before t, take the part of it in the overlapping area
4. Match the closest valid track as correspondence of T.

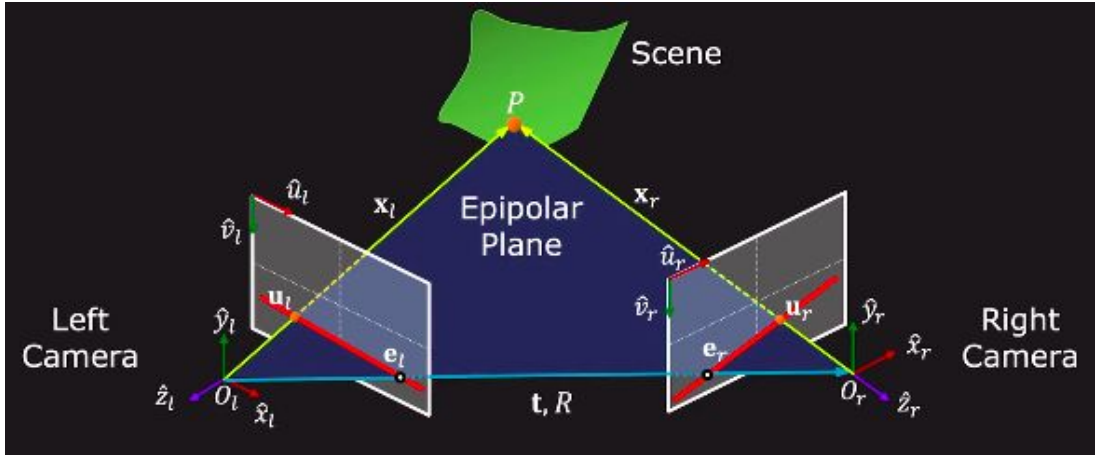




2.2. Proposal: Approaches

2 Epipolar line constraint (with calibration)

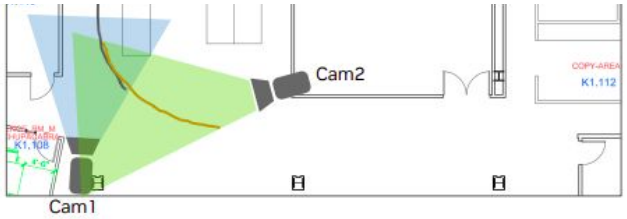
1. Find intrinsic parameter of camera 1 and camera 2.
2. Calculate the transition and orientation of the 2 cameras with a specific world coordinate frame.
3. For every tracklet in camera 1:
    a. Find the part of its trajectory in the overlapping area
    b. Its correspondence must lies on (or very close to) an epipolar line in camera 2, and vice versa. We apply this constraint for all points in a track.

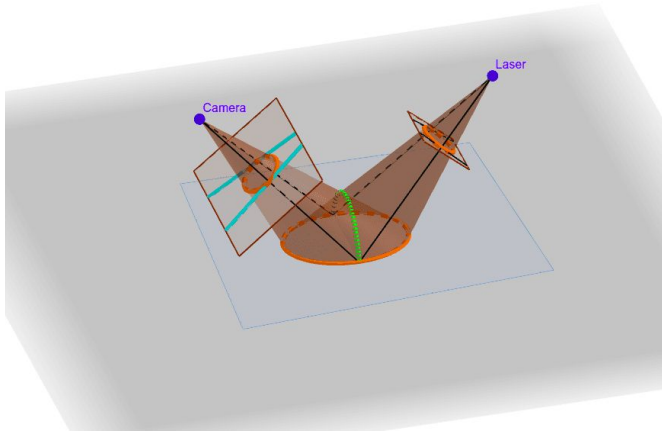


2.2. Proposal: Approaches

3 Pixel to Physical coordinates (with calibration)

1. Find intrinsic parameter of camera 1 and camera 2 and their extrinsic parameters w.r.t a specific world coordinate frame.
2. Calculate the equation of floor in space.
3. For every bounding box in image plane, derive the foot pixel. We know this pixel's correspondence in space must satisfy the equation of floor.
4. Transform the trajectory part in the overlapping area to physical coordinates.
5. Do assignment as in approach 1.





3. Application scenarios

2.1. Application scenarios

1 Crowd Insights

Recognize abnormal event in public places (company, schools, parks, events, etc.) in COVID-19:

- Trace trajectory and contacts of a person (positive case)
- Alert when someone does not wear mask
- Alert when a group exceeds a recommended number of people

2 Retail Insights

Build customer profile in retail stores:

- Record customer staying time at some specific areas
- Record customer attributes: age, gender, etc
- Recognize customer behaviours, e.g touching, buying preferences

3 Manage staff

Performance tracking for employee management:

- Record staff's staying time in areas of store
- Record staff's behaviours

4. Detailed plan for phase 1

Task	Week 1 (24/10 - 30/10)	Week 2 (31/10 - 06/11)	Week 3 (07/11 - 13/11)	Week 4 (14/11 - 20/11)	Week 5 (21/11 - 27/11)	Week 6 (28/11 - 04/12)	Week 7 (05/12 - 11/12)	Week 8 (12/12 - 18/12)	Week 9 (19/12 - 25/12)	Week 10 (26/12 - 30/12)
Prerequisites for Spatio-Temporal Association										
Setup camera, collect video										
Run SCT, save database										
Make evaluation metrics and ground-truth for MCT										
Do experiments with Spatio-Temporal Association										
Homography w/o calib, cam 2D							Report			
Homography w/o calib, cam 360										
Epipolar line constraint, cam 2D										
Pixel to physical coord, cam 2D										
Report										

References

[1] PROVID: Progressive and Multimodal Vehicle Reidentification for Large-Scale Urban Surveillance
[2] Multi-attribute driven vehicle re-identification with spatial-temporal re-ranking
[3] Single-camera and inter-camera vehicle tracking and 3D speed estimation based on fusion of visual and semantic features
[4] An Adaptive Learning Method for Target Tracking across Multiple Cameras
[5] Spatio-temporal consistency and hierarchical matching for multi-target multi-camera vehicle tracking

Progress report

Multiple Camera Tracking

3rd stage