3D Multi-Object Tracking for Autonomous Driving

Kemiao Huang

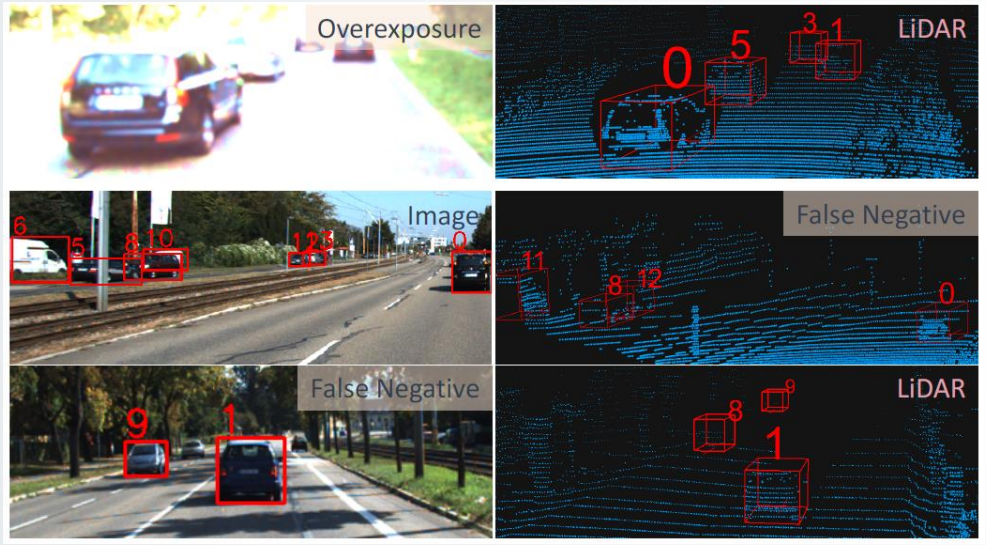
Department of Computer Science and Engineering

SUSTech Southern University of Science and Technology

1 Introduction

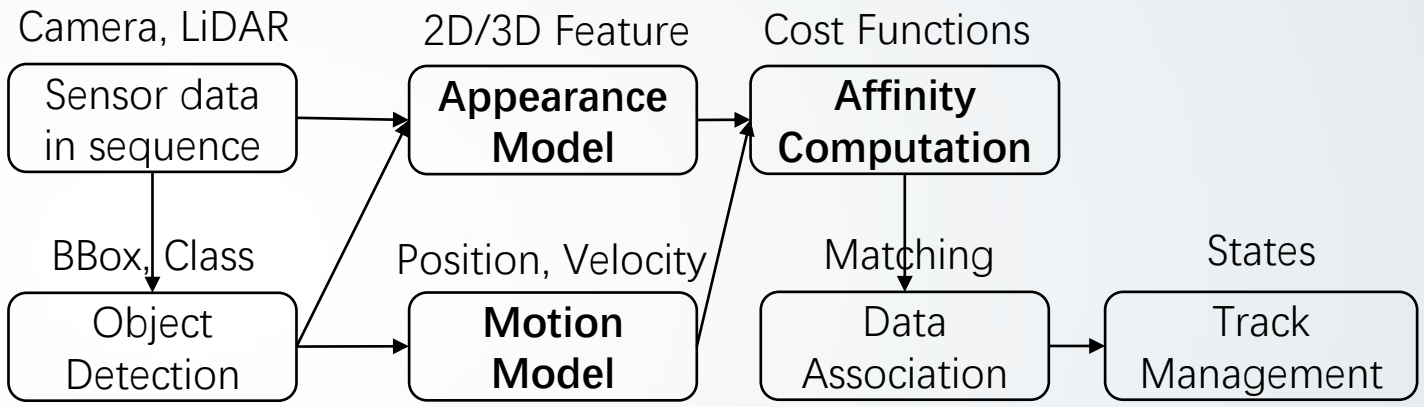
Background for autonomous system and multiple object tracking (MOT)

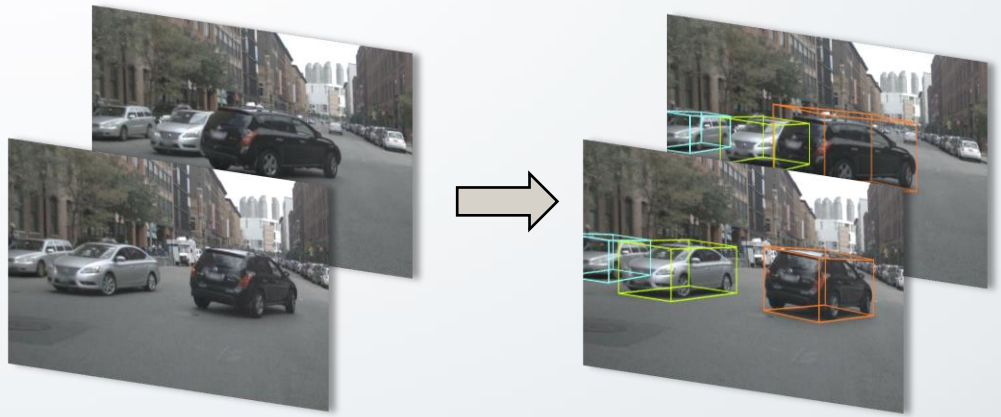
01 Background



Sensor Reliability

Multi-Object Tracking Pipeline





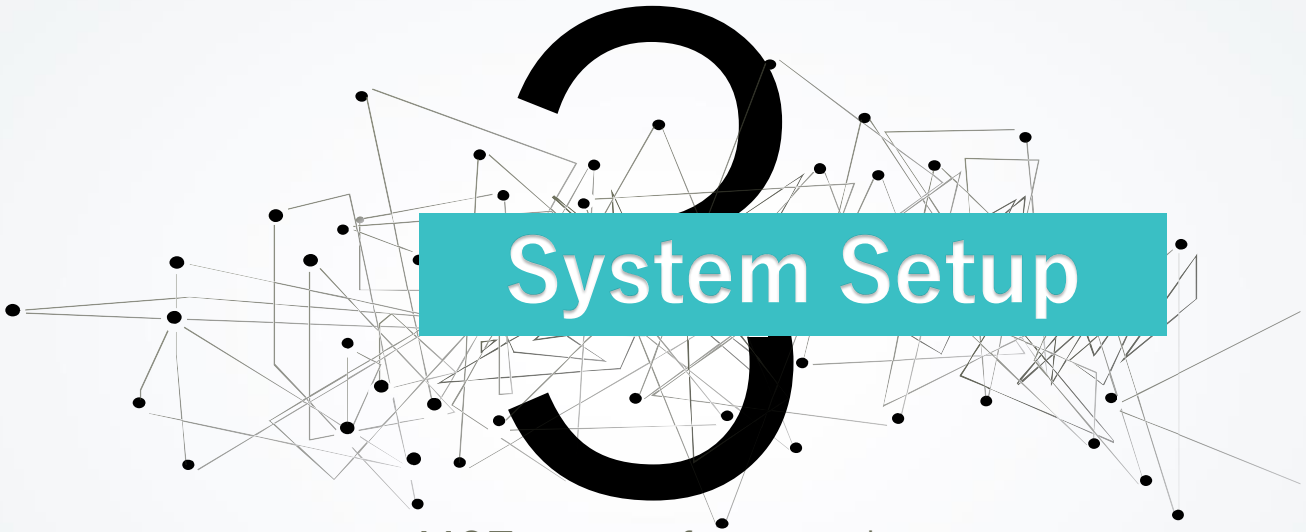
2 Related Work

State-of-the-art MOT approaches

02 Related Work

SUSTech Southern University of Science and Technology

Method		Appearance		Motion		Dissimilarity Computation (Total Cost)
		Feature Extractor	Branch Cost	Motion Model	Branch Cost	
Filter Tracker	PMBM [1]	X	-	Poisson Multi-Bernoulli Mixture	Multi-Bernoulli	-
	Deep-SORT [2]	ResNet	Cosine Distance	Kalman Filter	IoU + Mahalanobis	Motion Cost
DNN Tracker	MASS [3]	SIFT	Masked Correlation	Custom	IoU + Euclidean	Weighted Sum
	End-to-End [4]	Siamese Network (VGG16)	-	Fully Connected (Occupancy Grid)	-	Fully Connected
	Hu et al. [5]	Faster RCNN	Exponential L1 Norm	Two LSTMs	IoU	Weighted Sum
Multi-Modality Tracker	mmMOT [6]	VGG16 + PointNet + Attention Map	Absolute Subtraction	X	-	Appearance Cost
Ours		ResNet + PointNet + Fusion Net	Absolute Subtraction	Kalman Filter	IoU + Mahalanobis	Weighted Sum

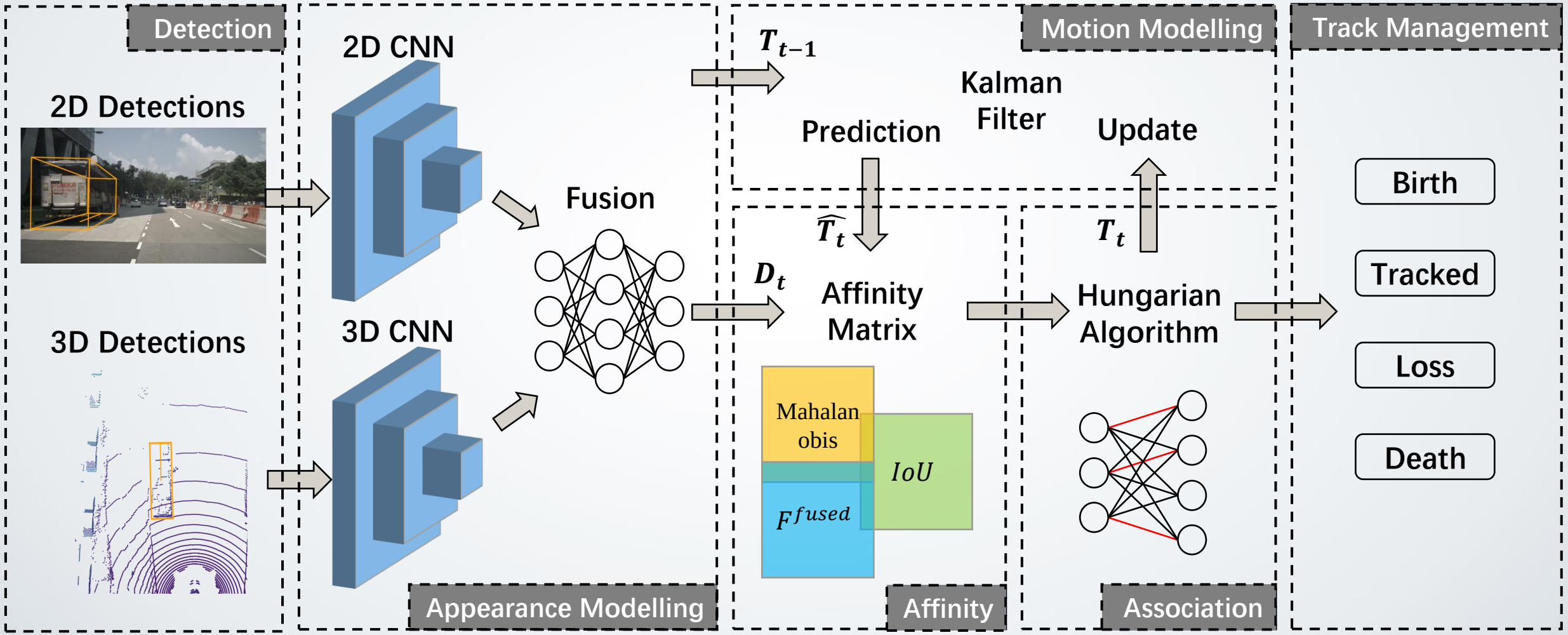


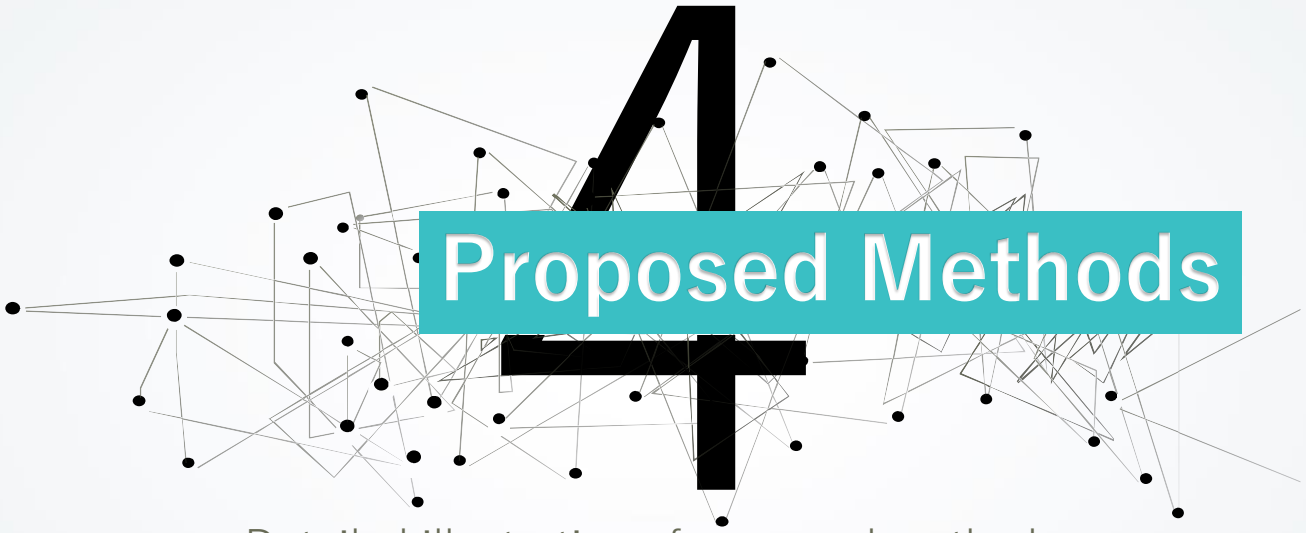
3 System Setup

MOT system framework

03 System Setup

SUSTech Southern University of Science and Technology





4 Proposed Methods

Detailed illustration of proposed methods

04 Proposed Methods

SUSTech Southern University of Science and Technology

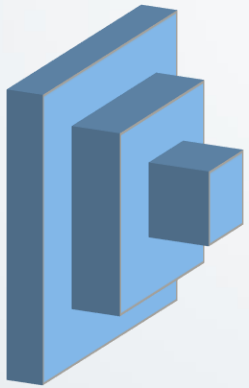
Image Feature Extractor

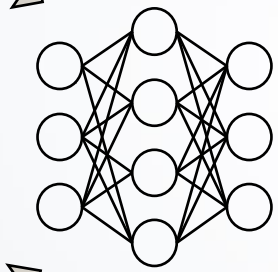
Finetune:
ResNet-50 /
VGG-16

Siamese Networks

Point Cloud Feature Extractor

Finetune:
PointNet (skip pooling)





Feature Fusion

Fully Convolution Layer or Attention map

$$G = sigmoid(conv(F, k))$$

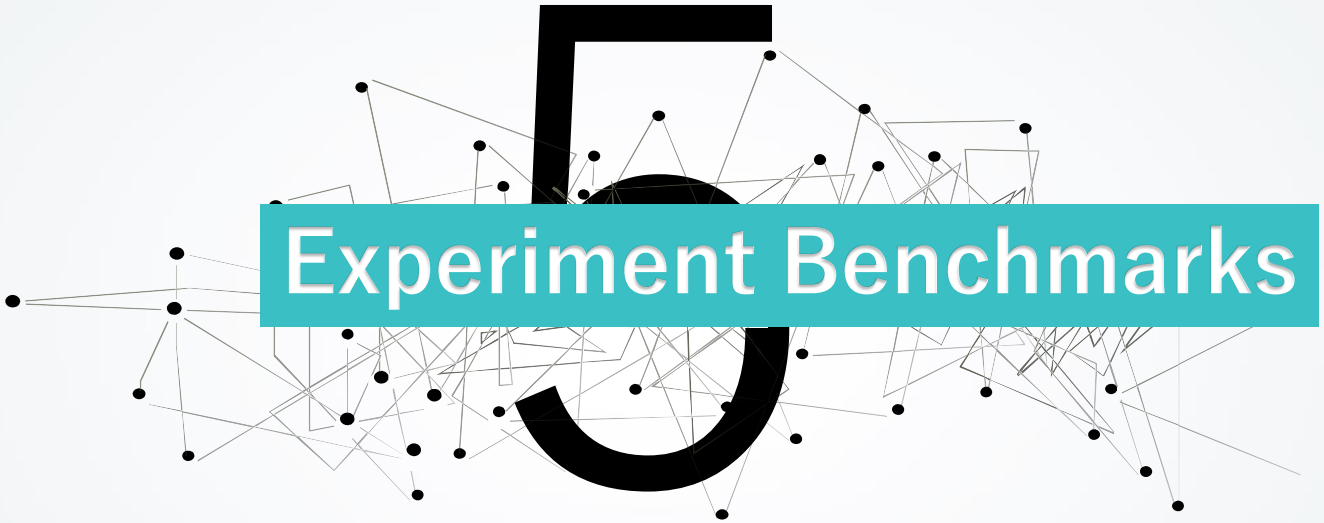
$$F_{fused} = [G1 * conv(F1, k1) + G2 * conv(F2, k2)] / (G1 + G2)$$

Kalman Filter

P: state covariance
Q: process covariance
R: measurement covariance

Affinity Matrix

$$A(i,j) = w_{loc} \cdot A_{loc}(i,j) + w_{app} \cdot A_{app}(i,j) + w_{iou} \cdot A_{iou}(i,j)$$



5 Experiment Benchmarks

05 Experiment Benchmarks

SUSTech Southern University of Science and Technology

MOT Benchmarks

- **Kitti**
  - 21 training sequences
  - 29 test sequences
- **NuScenes**
  - 1000 scenes
  - 40k keyframes

Metrics:

- Mostly Tracked (MT)
- Fragments (FM)
- Mostly Lost (ML)
- ID switches (IDSW)
- $MOTA = 1 - \frac{FN+FP+IDSW}{GT} \in (-\infty, 1]$
- $MOTP = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t} \in [0,1]$

Method	MOTA↑	MOTP↑	Prec.↑	Recall↑	FP↓	FN↓	ID-s↓	Frag↓	MT↑	ML↓	FPS↑
AB3DMOT	69.22	**86.76**	85.78	89.18	1975	1446	**0**	**71**	75.46	3.7	203.4
mmMOT	**79.46**	85.44	93.45	**88.84**	791	**1417**	75	272	**76.85**	**2.32**	13.34
ours(motion)	74.34	85.72	**94.73**	81.93	**572**	2270	10	75	59.26	18.52	**240.2**
ours(motion+appearance)	77.71	85.68	94.12	85.80	676	1791	10	96	63.43	5.56	42.71