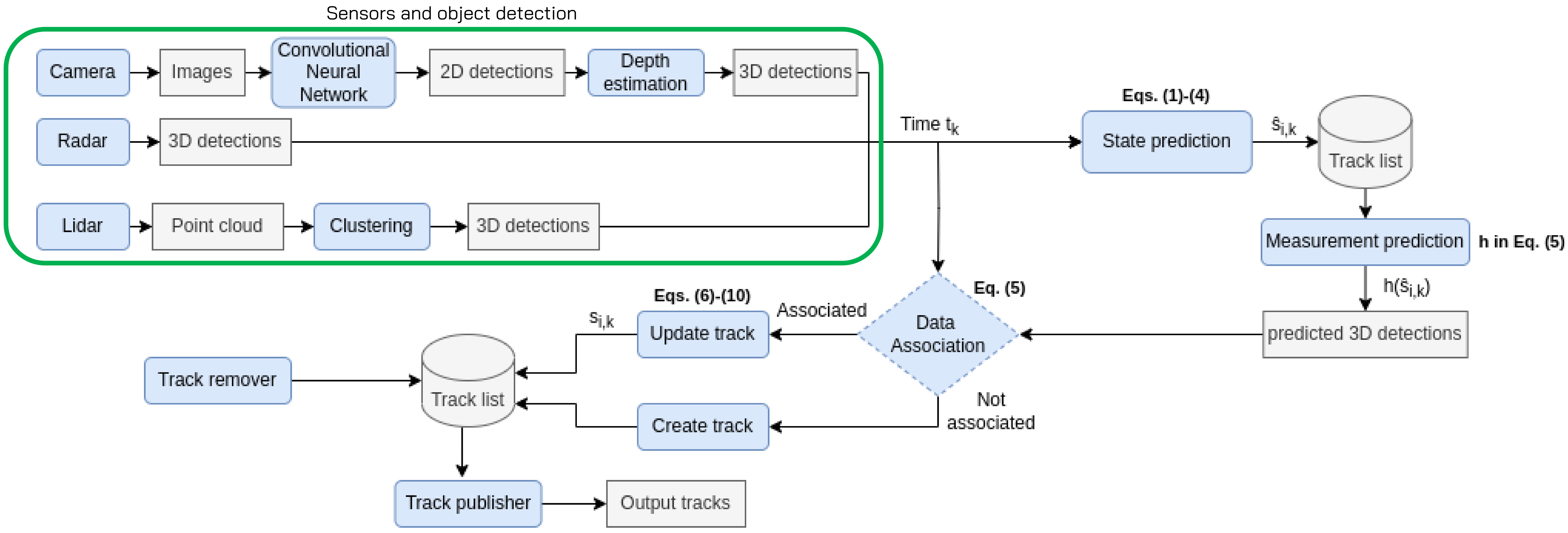
Robust Multimodal and Multi-Object Tracking for Autonomous Driving Applications

Applus+ IDIADA

Marc Pérez Quintana, Nil Munté Guerrero

- Sensor-agnostic multimodal fusion framework for multiple object tracking that can seamlessly integrate information coming from different object detectors, sensors, and vehicle-to-everything messages, either from other road users or from the infrastructure.
- All the information received is converted to a standardized set of detections that are then combined using a Kalman Filter with a constant velocity model.
- To ensure robustness, we propose methods to handle errors in classification and incorrect bounding box reconstruction.



The i-th tracklet has the state $\mathbf{s}_{k,i} = [r_x, r_y, v_x, v_y]^\top \in \mathbb{R}^4$ and its covariance $\mathbf{P}_{k,i} \in \mathbb{R}^{4\times4}$.

State prediction

$$\hat{\mathbf{s}}_{k,i} = \mathbf{F}(\Delta t)\mathbf{s}_{k-1,i} \quad (1)$$

$$\mathbf{F}(\Delta t) = \begin{bmatrix} \mathbf{I}_2 & \Delta t \mathbf{I}_2 \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \quad (2)$$

$$\hat{\mathbf{P}}_{k,i} = \mathbf{F}(\Delta t)\mathbf{P}_{k-1,i}\mathbf{F}(\Delta t)^\top + \mathbf{Q}(\Delta t) \in \mathbb{R}^{4\times4} \quad (3)$$

$$\mathbf{Q}(\Delta t) = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & b\Delta t \mathbf{I}_2 \end{bmatrix} \quad (4)$$

Measurement prediction

From state to measurement space: $h\colon \mathbb{R}^4 \to \mathbb{R}^m$

Data association

Mahalanobis distance $d(\hat{\mathbf{s}}_{k,i}, \mathbf{d}_{k,j})\colon \mathbb{R}^4 \times \mathbb{R}^m \to \mathbb{R}$ defined by:

$$d = \sqrt{(\mathbf{d}_{k,j} - h(\hat{\mathbf{s}}_{k,i}))^\top \mathbf{\Sigma}_{k,i,j}^{-1} (\mathbf{d}_{k,j} - h(\hat{\mathbf{s}}_{k,i}))} \quad (5)$$

Update track

Mean and covariance of the innovation:

$$\mathbf{y}_{k,i} = \mathbf{d}_{k,j} - h(\hat{\mathbf{s}}_{k,i}) \in \mathbb{R}^m \quad (6)$$

$$\mathbf{\Sigma}_{k,i,j} = \mathbf{H}\hat{\mathbf{P}}_{k,i}\mathbf{H}^\top + \mathbf{R}_{k,j} \in \mathbb{R}^{m\times m} \quad (7)$$

Where:

$\mathbf{H} \in \mathbb{R}^{m\times4}$ the Jacobian of h

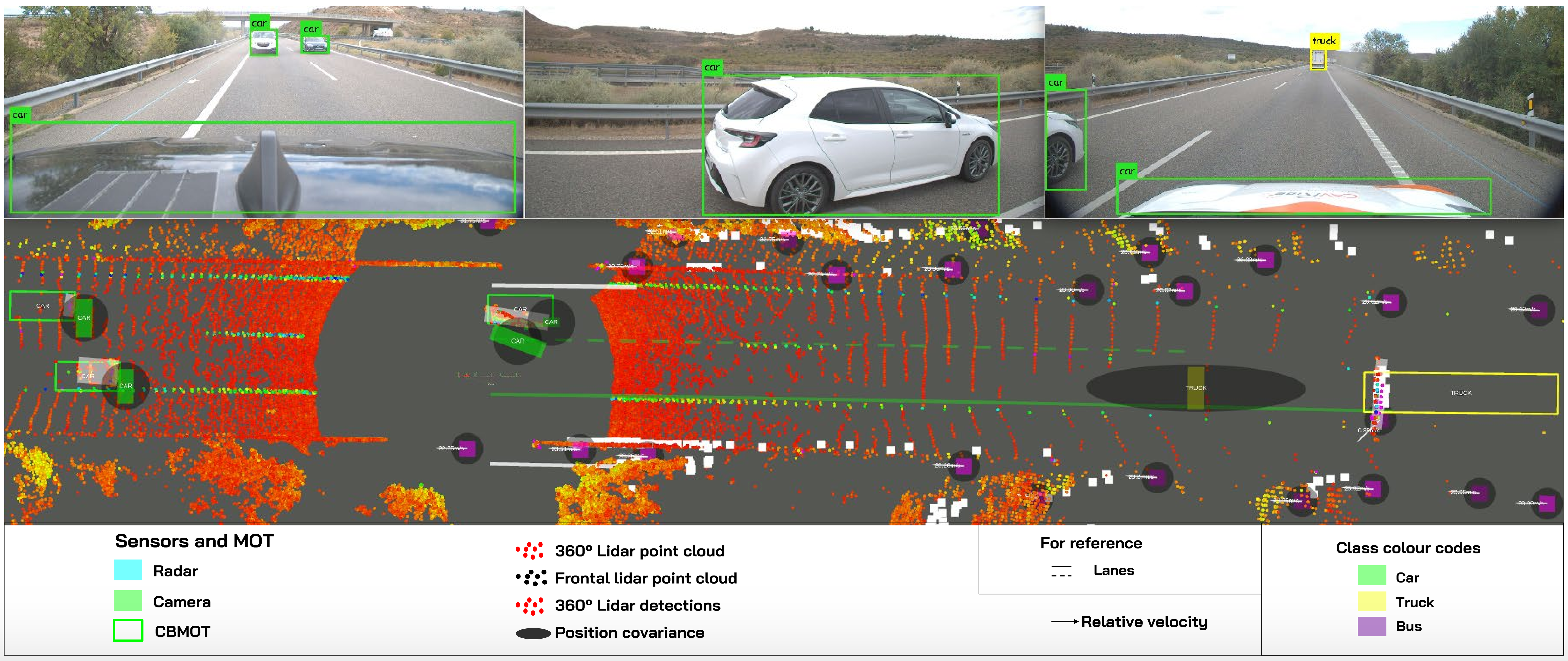
$\mathbf{R}_{k,j} \in \mathbb{R}^{m\times m}$ the covariance of the detection

Detection used to update the associated tracklet following:

$$\mathbf{K}_{k,i} = \hat{\mathbf{P}}_{k,i}\mathbf{H}^\top \mathbf{\Sigma}_{k,i,j} \in \mathbb{R}^{4\times m} \quad (8)$$

$$\mathbf{s}_{k,i} = \hat{\mathbf{s}}_{k,i} + \mathbf{K}_{k,i}\mathbf{y}_{k,i} \in \mathbb{R}^4 \quad (9)$$

$$\mathbf{P}_{k,i} = (\mathbf{I} - \mathbf{K}_{k,i}\mathbf{H})\hat{\mathbf{P}}_{k,i} \in \mathbb{R}^{4\times4} \quad (10)$$



[1] M. Pérez, A. Agudo. **Sensor-Agnostic Multimodal Fusion for Multiple Object Tracking from Camera, Radar, Lidar and V2X.** FISITA World Congress 2023

[2] M. Pérez, A. Agudo. **Robust Multimodal and Multi-Object Tracking for Autonomous Driving Applications.** ICAR 2023