3D Road Scene understanding with soft continuous occlusion modeling

Vikas Dhiman

NEC Laboratories America

September 19, 2014

Aim of the project

- Use output of various sub-systems
  - Ego motion
  - Detection bounding boxes
  - Detection orientation
  - Maps, GPS
  - Point tracks on objects

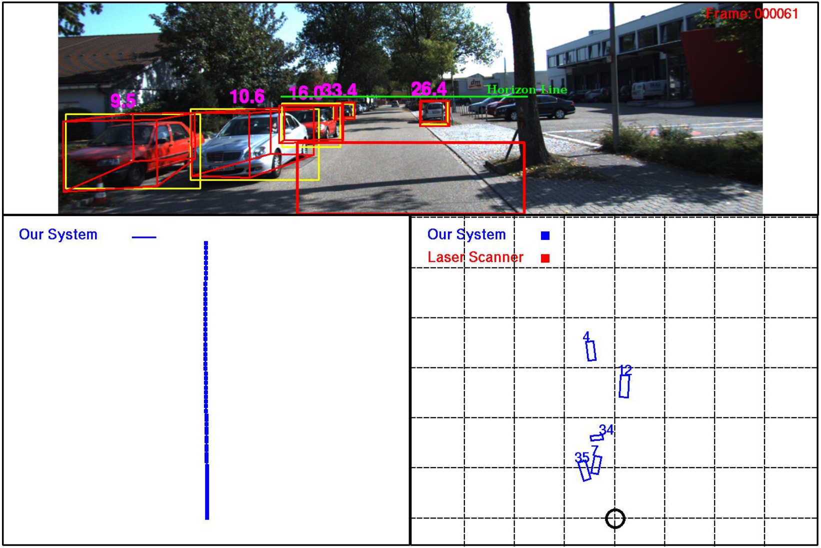
Aim of the project

- Use output of various sub-systems
  - Ego motion
  - Detection bounding boxes
  - Detection orientation
  - Maps, GPS
  - Point tracks on objects
- along with heuristics like size prior, cars usually align with lanes etc

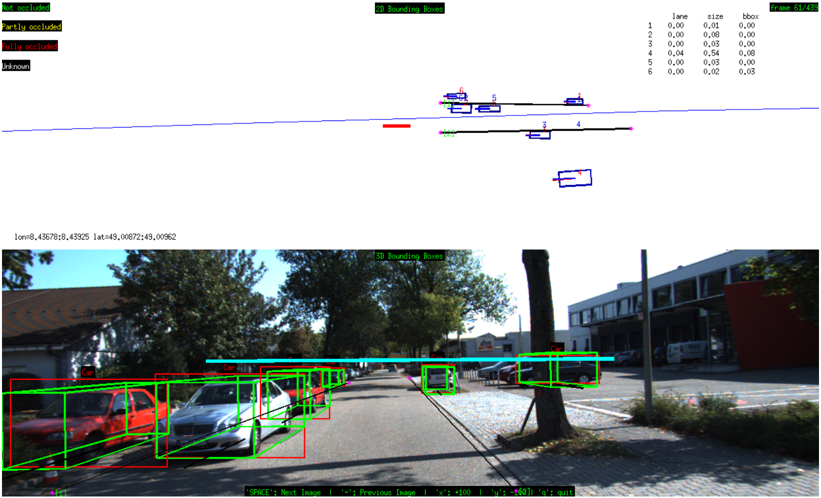
Aim of the project

- Use output of various sub-systems
  - Ego motion
  - Detection bounding boxes
  - Detection orientation
  - Maps, GPS
  - Point tracks on objects
- along with heuristics like size prior, cars usually align with lanes etc
- Output the most likely configuration of scene that best fits the output of various sub-systems and our heuristic knowledge

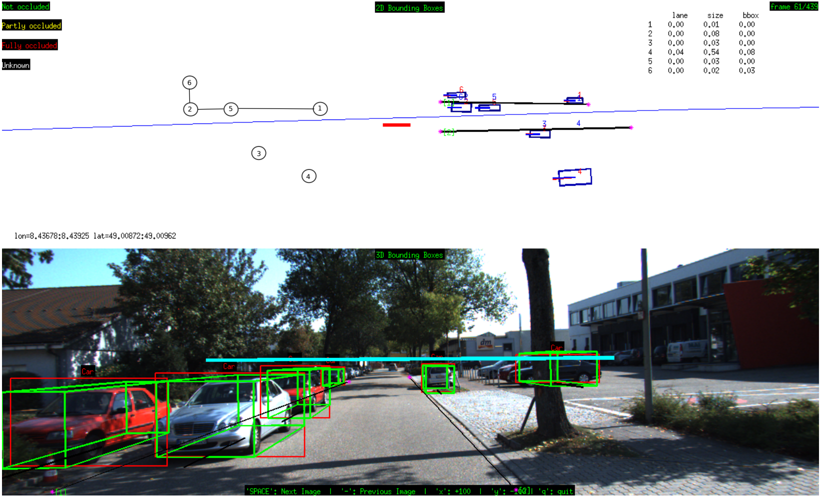
Before

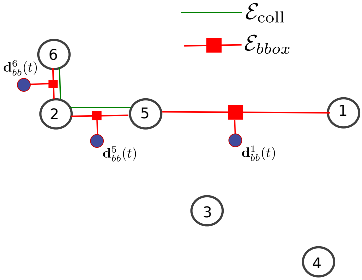

Frame: 000061
9.5
10.6
16.0
33.4
26.4
Horizon Line
Our System
Our System
Laser Scanner
4
12
34
7
35

After

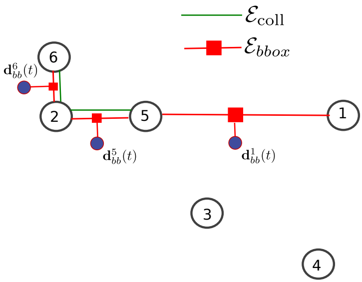


Not occluded
Partly occluded
Fully occluded
Unknown
2D Bounding Boxes
Frame 61/439
lane size bbox
1 0.00 0.01 0.00
2 0.00 0.08 0.00
3 0.00 0.03 0.00
4 0.04 0.54 0.08
5 0.00 0.03 0.00
6 0.00 0.02 0.03
lon=8.43678:8.43925 lat=49.00872:49.00962
3D Bounding Boxes
Car
'SPACE': Next Image | '-': Previous Image | 'x': +100 | 'y': -100 | 'q': quit


Not occluded
Partly occluded
Fully occluded
Unknown
2D Bounding Boxes
frame 61/439
lane size bbox
1 0.00 0.01 0.00
2 0.00 0.08 0.00
3 0.00 0.03 0.00
4 0.04 0.54 0.08
5 0.00 0.03 0.00
6 0.00 0.02 0.03
lon=8.43678:8.43925 lat=49.00872:49.00962
3D Bounding Boxes
Car
'SPACE': Next Image | '-': Previous Image | 'x': +100 | 'y': -100 | 'q': quit



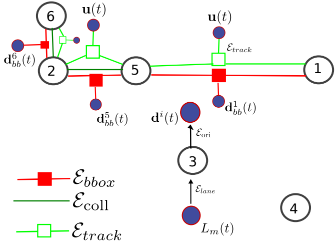
$\mathcal{E}_{\mathrm{coll}}$
$\mathcal{E}_{bbox}$
6
$\mathbf{d}_{bb}^{6}(t)$
2
5
1
$\mathbf{d}_{bb}^{5}(t)$
$\mathbf{d}_{bb}^{1}(t)$
3
4



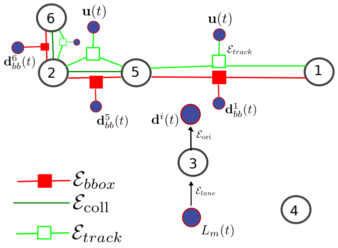
$$\mathcal{E}_{\mathrm{col}}^{ijt} = \frac{\left|\Sigma_i\right|^{\frac{1}{4}}\left|\Sigma_j\right|^{\frac{1}{4}}}{\left|\frac{1}{2}\Sigma_i + \frac{1}{2}\Sigma_j\right|^{\frac{1}{2}}} e^{-\frac{1}{8}\left(\mathbf{p}^{(i)}(t)-\mathbf{p}^{(i)}(t)\right)^{\top}\left(\frac{1}{2}\Sigma_i+\frac{1}{2}\Sigma_j\right)^{-1}\left(\mathbf{p}^{(i)}(t)-\mathbf{p}^{(i)}(t)\right)} \tag{1}$$

$$\mathcal{E}^{ijt}_{\text{occ}}(\Omega^i(t), \mathbf{B}^i, \Omega^j(t), \mathbf{B}^j) = \sum_k p^{\text{track}}_{ik}(\pi_{\Omega^i(t)}(\mathbf{B}^i) - \mathbf{d}^k(t))^\top \boldsymbol{\rho}(i, j, t) \tag{2}$$

where p^{track}_{ik} is the probability of kth detection being the right track for i hypothesis. (to be computed from tracklet score)



$$\mathcal{E}_{\text{ori}}^{it} = (1 - \cos(\omega_{\text{det}}^{i} - \omega^{(i)}(t)) \qquad (3)$$



$$\mathcal{E}^{it}_{\mathrm{ori}} = (1 - \cos(\omega^i_{\mathrm{det}} - \omega^{(i)}(t)) \tag{3}$$

$$\mathcal{E}^{it}_{\mathrm{track}}(\{\Omega^k(t)\}_k, \{\Omega^k(t-1)\}_k, \mathbf{B}^i, \mathbf{B}^j,) = \sum_{i=1}^{N}\sum_{j=1}^{M} a^i_j(t)\|u_j(t) - \pi_{\Omega^i(t)}(\pi^{-1}_{\Omega^i(t-1)}(u_j(t-1)))\|^2 \tag{4}$$

where $a^i_j(t)$ is the probability of association of $u_j(t)$ with tracklet i.

$a_j^i(t)$ is soft occlusion dependent probability. Assume the projection of tracklet is given by 2D bounding box $[u_l^i, v_t^i, u_r^i, v_b^i] = \pi_{\Omega^i(t)}(\mathbf{B}^i)$. Approximate bounding box by ellipse

$$\mu_i = \frac{1}{2}\begin{bmatrix} u_l^i + u_r^i \\ v_t^i + v_b^i \end{bmatrix} \tag{5}$$

$$\Sigma_i = \begin{bmatrix} \frac{2}{(u_l^i - u_r^i)^2} & 0 \\ 0 & \frac{2}{(v_t^i - v_b^i)^2} \end{bmatrix} \tag{6}$$

The occlusion function

Following [?] we model occlusion as soft continuous probablistic function that varies with distance from camera and distance of the center of projection of the object.

$$f^i_{\text{occ}}(u, v, \lambda) = \frac{N(u, v; \mu_i, \Sigma_i)}{1 + e^{-\frac{\lambda - \mu_i^{(d)}}{\beta}}} \text{where} \tag{7}$$

$$\beta = \frac{\mathbf{B}^i_z}{2 \log(49)} \tag{8}$$

$$\mu_d = \Omega^i(t)_z \tag{9}$$

Soft occlusion modeling using can be viewed as if the traffic participant are being viewed as translucent objects with the probability of reflection as $P_{\text{reflection}} = f^{i}_{\text{occ}}$. Given a point (u, v) on the image, the probability of its association to the i tracklet is given by

$$a^{i}_{j}(t) = f^{i}_{\text{occ}}(u, v, \mu^{(d)}_{i}) \prod_{k:\mu^{(d)}_{k} < \mu^{(d)}_{i}} (1 - f^{k}_{\text{occ}}(u, v, \mu^{(d)}_{i})) \qquad (10)$$

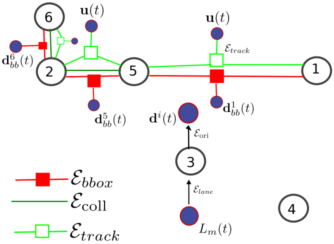
where the second term determines the probability of ray being *not occluded* by the k objects in front of the ith object and the first term is the probability of the ray being finally *occluded* by the i object.

Since our occlusion function f_{occ}^i is dependent upon depth as well, we can remove the explicit condition over $k : \mu_k^{(d)} < \mu_i^{(d)}$.

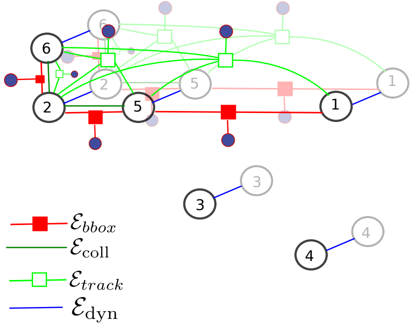
$$a_j^i(t) = f_{\text{occ}}^i(u, v, \mu_i^{(d)}) \prod_k (1 - f_{\text{occ}}^k(u, v, \mu_i^{(d)})) \tag{11}$$

Also, in practice we compute this association only for the tracklets that have overlapping 2D bounding boxes.

$$a_j^i(t) = \begin{cases} f_{\text{occ}}^i \prod_{k:(u,v) \in \pi_{\Omega^k(t)}(\mathbf{B}^k)} (1 - f_{\text{occ}}^k) & \text{if } i : (u, v) \in \pi_{\Omega^i(t)}(\mathbf{B}^i) \\ 0 & \text{otherwise} \end{cases} \tag{12}$$



$$\mathcal{E}^{it}_{\text{lane}} = \sum_{m:\text{DIST}(L_m(k),\mathbf{p}^{(i)}(t))<50} \frac{1-\omega^{(i)}(t)\cdot \text{TAN}(L_m(k),\mathbf{p}^{(i)}(t))}{1+exp(-q(w_{\text{road}}-\text{DIST}(L_m(k),\mathbf{p}^{(i)}(t))))} \tag{13}$$



$$\mathcal{E}^{it}_{\text{dyn-hol}} = 1 - \omega^{(i)}(t-1) \cdot (\mathbf{p}^{(i)}(t) - \mathbf{p}^{(i)}(t-1)) \tag{14}$$

$$\mathcal{E}^{it}_{\text{dyn-ori}} = \|\omega^{(i)}(t) - \omega^{(i)}(t-1)\|^2 \tag{15}$$

$$\mathcal{E}^{it}_{\text{dyn-vel}} = \|(\mathbf{p}^{(i)}(t) - 2\mathbf{p}^{(i)}(t-1)) + \mathbf{p}^{(i)}(t-2)\|^2 \tag{16}$$

Inference

- Most previous works use MCMC which is slow

Inference

- Most previous works use MCMC which is slow
- The number of edges $\approx$ 10-20 per frame $\approx$ 500-1000 for 50 frame sequence

Inference

- Most previous works use MCMC which is slow
- The number of edges $\approx$ 10-20 per frame $\approx$ 500-1000 for 50 frame sequence
- The described graph is mostly trees with a few loops

Inference

- Most previous works use MCMC which is slow
- The number of edges $\approx$ 10-20 per frame $\approx$ 500-1000 for 50 frame sequence
- The described graph is mostly trees with a few loops
- Variants of Belief propagation with damping seems a good bet.

Inference

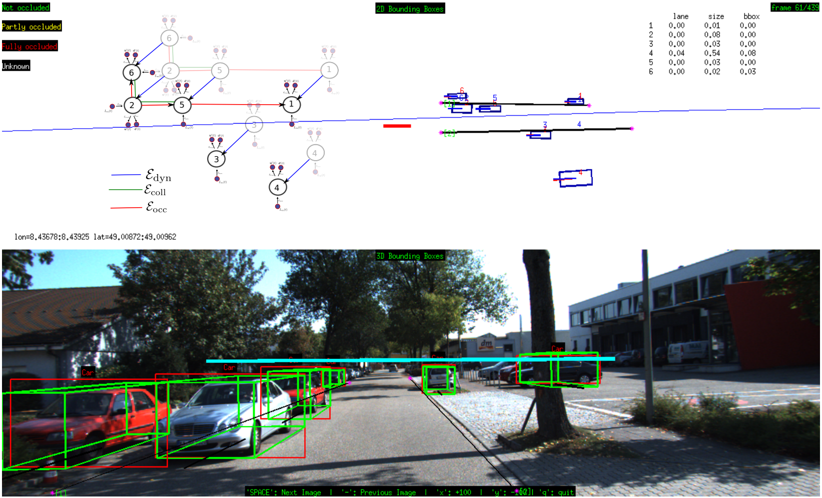
- Most previous works use MCMC which is slow
- The number of edges $\approx$ 10-20 per frame $\approx$ 500-1000 for 50 frame sequence
- The described graph is mostly trees with a few loops
- Variants of Belief propagation with damping seems a good bet.
- However, traditional version of inference algorithms including BP, TRW-S, Dual decomposition discrete state space.

Inference

- Most previous works use MCMC which is slow
- The number of edges $\approx$ 10-20 per frame $\approx$ 500-1000 for 50 frame sequence
- The described graph is mostly trees with a few loops
- Variants of Belief propagation with damping seems a good bet.
- However, traditional version of inference algorithms including BP, TRW-S, Dual decomposition discrete state space.
- We will need simpler smooth analytical approximation of messages for continuous state spaces.

Inference

- Most previous works use MCMC which is slow
- The number of edges $\approx$ 10-20 per frame $\approx$ 500-1000 for 50 frame sequence
- The described graph is mostly trees with a few loops
- Variants of Belief propagation with damping seems a good bet.
- However, traditional version of inference algorithms including BP, TRW-S, Dual decomposition discrete state space.
- We will need simpler smooth analytical approximation of messages for continuous state spaces.
- The approximation will need to be done for each message passing step


Not occluded
Partly occluded
Fully occluded
Unknown
2D Bounding Boxes
frame 61/439
lane size bbox
1 0.00 0.01 0.00
2 0.00 0.08 0.00
3 0.00 0.03 0.00
4 0.04 0.54 0.08
5 0.00 0.03 0.00
6 0.00 0.02 0.03
$\mathcal{E}_{\text{dyn}}$
$\mathcal{E}_{\text{coll}}$
$\mathcal{E}_{\text{occ}}$
lon=8.43678:8.43925 lat=49.00872:49.00962
3D Bounding Boxes
Car
'SPACE': Next Image | '-': Previous Image | 'x': +100 | 'y': -100 | 'q': quit