Supplement: A Continuous Occlusion Model for Road Scene Understanding

Vikas Dhiman[†] Quoc-Huy Tran[*] Jason J. Corso[†] Manmohan Chandraker[*]

[†] University of Michigan, Ann Arbor, MI, USA

[*] NEC Laboratories America, Inc. Cupertino, CA, USA

A. Computation of Ellipsoids

In this section, we present computation of the parameters of ellipsoids representing occupancy of traffic participants. For an object O_i, let $\mathbf{B}^i = [l\ w\ h]$ be its 3D dimensions, where l, w and h are its length, width and height on the ground plane, respectively. We wish to calculate the center $\boldsymbol{\mu}_c^i$ and spread $\boldsymbol{\Sigma}_c^{i^{-1}}$ of the ellipsoid representing the occupancy of O_i with respect to the camera coordinate system $\mathcal{C}$ of the current frame.

Consider an object coordinate system $\mathcal{O}$, which has the same orthonormal axes as the camera coordinate system and the origin at the projected point of the object center on the ground plane. For an object O_i, the center $\boldsymbol{\mu}_o^i$ and spread $\boldsymbol{\Sigma}_o^{i^{-1}}$ of the ellipsoid representing the occupancy of O_i in the object coordinate system are expressed as

$$\boldsymbol{\mu}_o^i = \begin{bmatrix} 0 & 0 & \frac{h}{2} \end{bmatrix}^\top, \tag{1}$$

$$\boldsymbol{\Sigma}_o^{i^{-1}} = \begin{bmatrix} \frac{4}{l^2} & 0 & 0 \\ 0 & \frac{4}{w^2} & 0 \\ 0 & 0 & \frac{4}{h^2} \end{bmatrix}. \tag{2}$$

In object coordinates, the ellipsoid equation is written as

$$(\mathbf{x}_t - \boldsymbol{\mu}_o^i)^\top \boldsymbol{\Sigma}_o^{i^{-1}} (\mathbf{x}_t - \boldsymbol{\mu}_o^i) = 1, \tag{3}$$

where $\mathbf{x}_t$ is a point in object coordinates. Given the relative pose of O_i in camera coordinates, we can extract the rotation $\mathbf{R}$ and translation $\mathbf{t}$ for converting points in camera coordinates to object coordinates and rewrite (3) as

$$(\mathbf{R}\mathbf{x}_c + \mathbf{t} - \boldsymbol{\mu}_o^i)^\top \boldsymbol{\Sigma}_o^{i^{-1}} (\mathbf{R}\mathbf{x}_c + \mathbf{t} - \boldsymbol{\mu}_o^i) = 1, \tag{4}$$

where $\mathbf{x}_c$ is a point in camera coordinates.

Let $\mathbf{t}' = \mathbf{t} - \boldsymbol{\mu}_o^i$, then (4) becomes

$$\mathbf{x}_c^\top \mathbf{R}^\top \boldsymbol{\Sigma}_o^{i^{-1}} \mathbf{R}\mathbf{x}_c + 2(\mathbf{R}^\top \mathbf{t}')^\top \mathbf{R}^\top \boldsymbol{\Sigma}_o^{i^{-1}} \mathbf{R}\mathbf{x}_c + \mathbf{t}'^\top \boldsymbol{\Sigma}_o^{i^{-1}} \mathbf{t}' = 1. \tag{5}$$

Next, denote $\boldsymbol{\Sigma}_c'^{i^{-1}} = \mathbf{R}^\top \boldsymbol{\Sigma}_o^{i^{-1}} \mathbf{R}$ and $\boldsymbol{\mu}_c^i = -\mathbf{R}^\top \mathbf{t}'$, then (5) becomes

$$(\mathbf{x}_c - \boldsymbol{\mu}_c^i)^\top \boldsymbol{\Sigma}_c'^{i^{-1}} (\mathbf{x}_c - \boldsymbol{\mu}_c^i) - \boldsymbol{\mu}_c^{i\top} \boldsymbol{\Sigma}_c'^{i^{-1}} \boldsymbol{\mu}_c^i + \mathbf{t}'^\top \boldsymbol{\Sigma}_o^{i^{-1}} \mathbf{t}' = 1. \tag{6}$$

Finally, denoting $\boldsymbol{\Sigma}_c^{i^{-1}} = \frac{\boldsymbol{\Sigma}_c'^{i^{-1}}}{1 - \mathbf{t}'^\top \boldsymbol{\Sigma}_o^{i^{-1}} \mathbf{t}' + \boldsymbol{\mu}_c^{i\top} \boldsymbol{\Sigma}_c'^{i^{-1}} \boldsymbol{\mu}_c^i}$, we have the ellipsoid equation in camera coordinates

$$(\mathbf{x}_c - \boldsymbol{\mu}_c^i)^\top \boldsymbol{\Sigma}_c^{i^{-1}} (\mathbf{x}_c - \boldsymbol{\mu}_c^i) = 1. \tag{7}$$

Therefore, the center $\boldsymbol{\mu}_c^i$ and spread $\boldsymbol{\Sigma}_c^{i^{-1}}$ of the ellipsoid representing the occupancy of O_i in camera coordinates are expressed as

$$\boldsymbol{\mu}_c^i = -\mathbf{R}^\top \mathbf{t}', \tag{8}$$

$$\boldsymbol{\Sigma}_c^{i^{-1}} = \frac{\boldsymbol{\Sigma}_c'^{i^{-1}}}{1 - \mathbf{t}'^\top \boldsymbol{\Sigma}_o^{i^{-1}} \mathbf{t}' + \boldsymbol{\mu}_c^{i\top} \boldsymbol{\Sigma}_c'^{i^{-1}} \boldsymbol{\mu}_c^i}, \tag{9}$$

where $\mathbf{R}$, $\mathbf{t}'$, $\boldsymbol{\Sigma}_c'^{i^{-1}}$, $\boldsymbol{\Sigma}_o^{i^{-1}}$ are derived as above.

B. Other Energies for 3D Object Localization

This section provides the details of other energies, namely dynamic and size energies (in addition to point track and detection bounding box energies already presented in the main paper), that are used in our localization experiment.

B.1. Dynamic energy

Dynamic energy imposes both temporal smoothness and holonomic constraints. In particular, holonomic constraints penalize changes in object position that are not in the direction of object orientation at the previous time step, as

$$\mathcal{E}_{\text{dyn-hol}}^{it} = 1 - \omega^i(t-1) \cdot (\mathbf{p}^i(t) - \mathbf{p}^i(t-1)). \tag{10}$$

Temporal smoothness constraints add a penalty for unsmooth changes in object orientation and velocity over consecutive time steps, as

$$\mathcal{E}_{\text{dyn-ori}}^{it} = \|\omega^i(t) - \omega^i(t-1)\|^2, \tag{11}$$

$$\mathcal{E}_{\text{dyn-vel}}^{it} = \|(\mathbf{p}^i(t) - 2\mathbf{p}^i(t-1)) + \mathbf{p}^i(t-2)\|^2. \tag{12}$$

The total dynamic energy is expressed as a weighted combination of holonomic and smoothness constraints

$$\mathcal{E}_{\text{dyn}}^{it} = \lambda_{\text{dyn-hol}} \mathcal{E}_{\text{dyn-hol}}^{it} + \lambda_{\text{dyn-ori}} \mathcal{E}_{\text{dyn-ori}}^{it} + \lambda_{\text{dyn-vel}} \mathcal{E}_{\text{dyn-vel}}^{it}, \tag{13}$$

where $\lambda_{\text{dyn-hol}}$, $\lambda_{\text{dyn-ori}}$ and $\lambda_{\text{dyn-vel}}$ are the weights of the component energies.

B.2. Size energy

Size energy imposes prior knowledge on object dimensions as

$$\mathcal{E}_{\text{size}}^{it} = (\mathbf{B}^i - \hat{\mathbf{B}})^\top \Sigma_{\hat{\mathbf{B}}}^{-1} (\mathbf{B}^i - \hat{\mathbf{B}}), \tag{14}$$

where $\hat{\mathbf{B}}$ and $\Sigma_{\hat{\mathbf{B}}}$ are the mean and covariance, respectively, of object dimensions obtained from the KITTI dataset.

C. Parameter settings

Parameter	k, k_u, k_d	λ_{track}	λ_{detect}	λ_{dyn}	λ_{size}
Value	$10 \ln(49)$	1	1	10	7

Table 1: Parameter settings for our experiments.