Reviewer 1:

- Add a table of notations defining the input and output of the problems (object-point association and 3D object localization)
- Include in Table 2 the errors (t, dim) corresponding to the initial estimates
- Convert Equation 4 to probability domain
- Equation 13 is computed numerically using importance sampling.
- Equation 8 can be interpreted as a combination of soft-max functions.

Reviewer 2:

- Instead of using binary functions, we model objects by translucent functions, which not only accounts for object shape uncertainty but also allows a continuous occlusion model that remains amenable to continuous optimization tools.
- 3D information, i.e., 3D bounding boxes, is used for computing 3D ellipsoids which represent object occupancy in our occlusion model. We use the 3D bounding boxes estimated by Song et al. for computing the object-point association probabilities in Section 5.1 and for initializing all localization methods in Section 5.2.
- Update Equation 4.2 to include all objects
- The number of objects is known a priori in our occlusion model and is given by Song et al. This prior knowledge is also provided to other methods such as BBox and RAS.
- Outliers in point trajectories have high reprojection errors and hence low object-point association probabilities.
- Update Equation 19 to exclude the object poses at the previous time-step
- Add a new baseline to the association experiment **(?)**
  + Reconstruct a 3D point cloud within each 2D bbox track
  + Run majority voting on the 3D point cloud w.r.t. all 3D bbox centers
- Add a new baseline to the localization experiment **(?)**
  + Reconstruct a 3D point cloud within each 2D bbox track
  + Fit a 3D bbox on the reconstructed 3D point cloud
- Ellipsoids have been used in many previous works for modeling different object categories such as humans, cars, etc.
- Explain or provide references to Sentence 287-289 or Equation 4
- Provide numbers of sequences, frames, objects of the KITTI dataset
- Add additional citations

Reviewer 3:

- Transmission probability only indicates whether a ray is transmissible to a particular depth but doesn't provide information on which object obstructs the ray at that depth. Alternatively, this information is given by the reflection probability. Therefore, a combination of transmission and reflection probabilities is more suitable for explaining the image observations.
- Comment on the results of the localization experiment **(?)**