OBJECT GEOLOCATION USING MRF BASED MULTI-SENSOR FUSION

Vladimir A. Krylov and Rozenn Dahyot
ADAPT Centre, School of Computer Science & Statistics, Trinity College Dublin, Ireland

Multi-Sensor Fusion Pipeline
- We perform optimization for object discovery and geolocation based on the following input estimates:
  ✓ Individual discovered objects (semantic segmentation)
  ✓ Monocular depth estimated for each discovered object $\Delta_i$
  ✓ LiDAR candidate matches
- We define a Markov Random Field (MRF) model over the space $X$ of all view-rays intersections:
  - label $z=0$ if not occupied by object
  - label $z=1$ if occupied
- MRF configuration is characterized by its corresponding energy $U$. Optimal = minimum of $U$. Optimization: ICM.

Energy terms:
- Unary consistency terms:
  $u_D(z_i|X, Z) = z_i \sum_{j=1,2} ||\Delta_{ij} - d_{ij}\|^2$
  $u_L(z_i|X, Z) = z_i l_i^2$
- Pairwise term. No occlusions. No spread:
  $u_0(R_i|X, Z) = \prod_{x_i \in R_i} (1 - z_i)$
- High-order term. Penalty not matched rays.:
  $u_C(R_i|X, Z) = \sum_{x_m, x_n \in R_i} z_m z_n \|x_m - x_n\|^2$

Total energy:

Post-processing strategies:
- Clustering via averaging (lower precision);
- Snapping to LiDAR positions (lower recall).

The same MRF-based architecture can incorporate:
- Further image and data modalities;
- Geometrical and position assumptions;
- SLAM-like formulations.

Experimental Results
- We introduce a dataset with 209 pole-mounted traffic lights in central Dublin.
- We employ:
  ✓ Google Street View imagery: 1307 panoramas
  ✓ Airborne LiDAR scan reporting 12300 pole-like matches

Conclusion
- The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

References

Crowdsourced Imagery
In [3] we use crowdsourced Mapillary images for object discovery and geolocation.
We employ a modified pipeline:

We rely on Laplacian filtering for thresholding and Structure from Motion to estimate bearing, adjust GPS.

Conclusion
- Fully automated
- Customisable to various objects
- Efficient multi-sensor fusion

Relies on Street Level Imagery as primary detection source: performance validated on expert (Google Street View) and crowdsourced (Mapillary) imagery.

References