Building large urban environments from unstructured point data

Florent Lafarge, INRIA and Clément Mallet, IGN

Context
The 3D modeling of urban environments from airborne data is a topic of major interest in many applications such as urban planning, navigation aids or wireless communications. Reconstructing such scenes is a difficult task because the urban objects significantly differ in term of complexity, diversity, scale and density.

Motivations
Existing methods have been designed in specific contexts. Our goal is to propose a more global model which can handle:
- multiple types of objects: not just buildings, but also trees and complex ground
- hybrid representation: combination of surface primitives and mesh patches
- any kind of urban scenes: dense, peri-urban, old fashion objects, countryside
- various types of data: point clouds from Lidar or Multi-View Stereo

Our solution

A/ Point cloud classification
- Four classes of interest are distinguished: Building (roof components), Tree (excluding shrubs), Ground clutter (outliers, facades and insignificant objects as cars, fences, wires, antennas or cranes)

The input point cloud is classified by MRF. The formulation takes into account different local point descriptors and spatial regularization contraints. The optimal solution is found by Graph-Cuts.

B/ Geometric shape extraction
- Various types of 3D-primitives are detected by a region growing procedure from points labeled as building. We impose 2 constraints on the primitives: a minimal size and a maximal quadratic error.
- 3D-segments for building contours (accuracy not completeness)
- Planes for planar roof sections
- Cylinders, spheres & cones for other roof parts

C/ Planimetric labeling
- 3D-primitives and the other urban components are arranged in a planimetric map using a multi-label energy minimization. Point labels are projected in a 2D-grid G, such that the configuration space L=G ground, vegetation, plane, cylinder, sphere, cone, roof planes
- energy: standard form $\mathcal{E}(l) = \sum_{i \in G} \beta_i + \sum_{j \in G} \gamma_j l_{ij}$
- data term: Z-error between the surface from $l_j$ and the highest point in the cell $i$ if $l_j$ roof, and constant penalization $c$ otherwise
- propagation constraints: combination of breakline-dependent neighborhood, label smoothness and structure arrangement law
- optimization: $\alpha$-expansion embedded in a parallelization scheme

D/ Object representation in 3D
- The 3 types of elements contained in the scenes are modeled in 3D by different geometric structures
  - Buildings: hybrid combining 3D primitives and mesh patches
  - Trees: template matching with ellipsoidal shapes
  - Ground: meshing procedure

Experiments

Accuracy and performances
The mean error depends on the data type and the point density. For a 2pts/m² point cloud, the typical error is contained in [0.2 m, 0.35 m]. High local errors correspond to outlers, trees, roof superstructures and undesirable points referring to cars, fences or wires.

Our hybrid representation provides a 3D-model where the building is accurately modeled by planes and a sphere for the regular parts, and by mesh-patches for the atypical surfaces (i.e. the undulating roofs). The fine (resp. coarse) 3D-model has 46K facets (resp. 864 facets) and a 0.24 m (resp. 0.33 m) mean error to the input data

Lidar vs MVS

Large scenes

Marseille, France
Input: 39 M points, 20 km²
result: 109K primitives, 36K trees, 3hrs