iQmulus/TerraMobilita benchmark on Urban Analysis

Bruno Vallet, Mathieu Brédif, Béatriz Marcotegui, Andres Serna, Nicolas Paparoditis
Introduction
Introduction

- Mobile laser scanning (MLS) generates massive amount of data
- Urban cores are objects of utmost interest:
  - Urban planning
  - Inventory and maintenance
  - Accessibility diagnostic
- Need for tools to analyse MLS data acquired in urban cores
- Need for a benchmark of existing tools
Benchmark objectives

- Trigger interest on MLS in scientific communities:
  - Computer vision
  - Photogrammetry/remote sensing
  - Geometry processing

- Provide reliable and large scale ground truth for works on MLS

- Define an ambitious goal for MLS based urban analysis

- Provide an objective tool to compare the qualities of urban analysis algorithms
Guidelines

- Fully controlled annotation of the data. For each point:
  - object/segment id
  - class label

- Very generic semantic tree to provide an ontology for urban scenes

- Evaluation:
  - Multicriteria: not a ranking but an evaluation of the pros and cons of each benchmarked algorithm
  - Objective: no parameters/thresholds
Outline

- Dataset
- Analysis problem statement
- Ground truth production
- Evaluation metrics
- Participants & results
- Conclusion
Dataset
Data

- Acquisition with Stereopolis MLS:
  - 360° Riegl sensor, multiecho
  - Applanix georeferencing

- Anisotropic resolution:
  - Across trajectory: Constant angular resolution (0.03°) => distance dependent geometric resolution
  - Along trajectory: Constant time resolution (10ms) => speed dependent geometric resolution
Data attributes

- Attributes:
  - $X,Y,Z$: coordinates of the echo in a geographical frame
  - $X_0,Y_0,Z_0$: coordinates of the laser center at the time this echo was acquired
  - Reflectance: backscattered intensity corrected for distance
  - num_echo: number of the echo in case of multiple returns
  - Time: time at which the point was acquired

- Data provided in ply file format for easy and generic attribute handling.
Data
Area

- 10+1 zones in the center of Paris (6ème arrondissement)
- Each zone has 30 (12) million points corresponding to 2 minutes of acquisition each and around 500m (depending on vehicle speed)
Analysis problem statement
Scene analysis

- We call scene analysis the combination of:
  - A **segmentation** of the scene in individual objects surfaces
  - A **semantic labellisation** (classification) of these objects

- Participants are asked to provide a ply file, adding for each point:
  - A segment identifier **id** (defining the segmentation)
  - A class label **class** (defining the classification)
Introduction: segmentation
Introduction : classification
Targeted Communities

- Classification specialists:
  - Interested in classification ground truth
  - Not interested in object individualization
  - Growing interest in contextual classification

- Segmentation specialists:
  - Growing interest in semantics to assist the segmentation

- Detection specialists:
  - Detectors for specific object types

- The semantic and geometric problems are connected
Scene analysis: semantics

- The semantic tree is very detailed:
  - Surface classes:
    - Road
    - Curb
    - Sidewalk
    - Facade/building
  - Objects classes:
    - Dynamic/static
    - Natural/man made
    - Punctual/linear/extended

- Participants can go as deep as they wish in the semantics tree
- Evaluation will be performed accordingly
Scene analysis: semantics
Scene analysis : semantics

- Other object
  - Static
    - Punctual
    - Linear
    - Extended
  - Other static
    - Other punctual
      - Post
      - Bollard
      - Floor lamp
      - Traffic light
      - Traffic sign
      - Signboard
      - Mailbox
      - Trash can
      - Parking meter
      - Bicycle terminal (vélib)
  - Other linear
    - Barrier
    - Roasting
    - Grid
    - Chain
    - Wire
    - Low wall
  - Other extended
    - Shelter
    - Kiosk
    - Scaffolding
    - Bench

- Dynamic
  - Other dynamic
  - Pedestrian
    - Other pedestrian
      - Static
      - Walking
      - Running
      - With stroller
      - Holding
      - Leaning
      - Skater
      - Rollerskater
      - Wheelchair
  - Other 2 wheels
    - Bicycle
    - Scooter
    - Moped
    - Motorbike
  - Other 4 wheels
    - Car
    - Van
    - Truck
    - Bus
  - Other furniture
    - Table
    - Chair
    - Stool
    - Trash can
    - Waste

- Natural
  - Other natural
    - Tree
    - Bush
    - Potted plant
    - Hedge
Ground truth production
Ground truth production tool

- Requirements
  - Fast and easy navigation and annotation
  - Segmentation at point level
  - Interactivity/editability

- We designed an interface in sensor geometry:
  - Columns are points acquired consecutively
  - Consecutive columns correspond to points acquired at a time interval equal to the time for the laser beam to finish a 360° sweep
Data: sensor space
Segmentation

- In 2D, segmentation is created and maintained by a partition graph.

- User is provided with graph editing tools:
  - Create a node (at a pixel corner) possibly on an existing edge.
  - Create an edge (along pixel boundaries):
    - A straight line (Brezenham).
    - A minimal path for the cost:

      \[
      cost = \max \left( 1, \lambda \left( 1 - \max \left( \frac{\Delta \alpha}{\alpha_0}, \frac{\Delta d}{d_0}, \frac{\Delta r}{r_0} \right) \right) \right)
      \]

- Parameters = weights for Normal/Depth/Intensity difference term.
- User can interactively tune these parameters.
- Move an existing node (recomputes all adjacent edges).
Other features

- A segment can be split by any plane defined by:
  - Three points
  - Two points (vertical)
  - One point (vertical and orthogonal to beam direction)
  - Plus an offset

- Segments can be merged (necessary in case of occlusions)

- Segments can be tagged by a label from the semantic tree

- Zooming, Panning

- Snapping

- Import/Export point clouds with label/ids per points

- Web based (javascript+webGL)
Example
Production details

- Production of the learning dataset (12Mpts) with an alpha version of the tool

- For the 10 zones of the benchmark:
  - 10 participants
  - 2 days production each
  - Around 60% of the 300 Mpts annotated
  - Easy production management thanks to the web based tool:
    - Each participants gets a unique link allowing them to process a 30 Mpts block
    - Their work is simply stored as a graph
  - Graphs are controlled and final ground truth ply files exported
Evaluation metrics
Multicriteria evaluation

- Evaluate the algorithm result:
  - As a classification algorithm: confusion matrix
  - As a detection algorithm:
    - precision/recall for object classes
    - No notion of object for surface classes
Precision/Recall

- Need to answer the questions
  - Is a **Ground truth** (GT) object detected?
  - Is an **Algorithm result** (AR) a good detection?

- Answer (and evaluation) requires to match objects from the GT to objects from the AR

- This matching allows to define:
  - **Precision** = \#(GT match AR)/\#GT
  - **Recall** = \#(AR match GT)/\#AR

- Thus precision/recall is defined on a subjective matching criterion
Delocalisation

Ground truth

Algorithm result
Dilatation/Erosion

Ground truth

Algorithm result

Dilatation

Erosion
Scission/fusion

Ground truth

Split (N to 1)

Algorithm result

Merge (1 to M)
N to M associations

Ground truth

Algorithm result
Intersection/Union Ratio

\[ R = \frac{S(VT \cap RD)}{S(VT \cup RD)} \]

- Gives a « distance » between objects:
  - 0 = no intersection
  - 1 = perfect match
- Matching often defined by a threshold on R
- Above 0.5, no N to M matchings
- But 0.5 is very strict
- Precision/recall depends highly on this threshold
Proposition

- Give precision and recall as a function of this threshold:
  - No arbitrary (subjective) choice of a threshold
  - Compare algorithms by comparing curves
- For thresholds below 0.5, also give the number of N to 1 and 1 to M pairings
Participants & results
Participants

- CMM - MINES ParisTech (Andres Serna, Beatriz Marcotegui):
  - Based on elevation images
  - Mathematical Morphology based image processing
  - Machine learning techniques
  - Does the full analysis (segmentation and classification)

- Institute of Photogrammetry and Remote Sensing (IPF) – KIT (Martin Weinmann):
  - Extract a variety of low-level geometric features
  - Supervised classification based on careful feature selection
  - Only classification evaluated
Ground truth
CMM result
IPF result
Results for CMM

- Classification (one 30 Mpts zone):
  - Surface/object: 92.6%
  - Building/ground surface: 98.3%
  - Curb/sidewalk/road: 98.4%

<table>
<thead>
<tr>
<th>GT/AR</th>
<th>Road&amp;side</th>
<th>Curb</th>
</tr>
</thead>
<tbody>
<tr>
<td>road</td>
<td>71.8348</td>
<td>0.684865</td>
</tr>
<tr>
<td>sidewalk</td>
<td>25.7088</td>
<td>0.687476</td>
</tr>
<tr>
<td>curb</td>
<td>0.187855</td>
<td>0.896216</td>
</tr>
</tbody>
</table>
Results for CMM

- Classification (one 30 Mpts zone):
  - Surface/object: 92.6%
  - Building/ground surface: 98.3%
  - Curb/sidewalk/road: 98.4%

<table>
<thead>
<tr>
<th>GT/AR</th>
<th>Road&amp;side</th>
<th>Curb</th>
</tr>
</thead>
<tbody>
<tr>
<td>road</td>
<td>71.8348</td>
<td>0.684865</td>
</tr>
<tr>
<td>sidewalk</td>
<td>25.7088</td>
<td>0.687476</td>
</tr>
<tr>
<td>curb</td>
<td>0.187855</td>
<td>0.896216</td>
</tr>
</tbody>
</table>
Results for CMM

- Classification (one 30 Mpts zone):
  - Surface/object: 92.6%
  - Building/ground surface: 98.3%
  - Curb/sidewalk/road: 98.4% but curb (2.3%) confused for sidewalk (0.7%) and road (0.7%) because of rasterization.
  - Static/mobile object: 91.8%
  - Pedestrian/2/4 wheelers: 99.3%

<table>
<thead>
<tr>
<th>GT/AR</th>
<th>pedestrian</th>
<th>2 wheelers</th>
<th>4+ wheelers</th>
</tr>
</thead>
<tbody>
<tr>
<td>pedestrian</td>
<td>1.63193</td>
<td>0.00262888</td>
<td>0.123962</td>
</tr>
<tr>
<td>2 wheelers</td>
<td>0.388468</td>
<td>0.653378</td>
<td>0</td>
</tr>
<tr>
<td>4+ wheelers</td>
<td>0.112435</td>
<td>0.0281088</td>
<td>97.0591</td>
</tr>
</tbody>
</table>
Results for CMM

- Detection (one 30 Mpts zone) : All objects

![Graph showing Precision and Recall for Detection]
Results for CMM

- Detection (one 30 Mpts zone):
  - Static objects:
  - Dynamic objects:

![Graphs showing precision and recall for static and dynamic objects.](image)
Results for IPF

- Classification (learning dataset only)
  - Surface/object : 87.8%
  - Ground/Building surface : 93.7%
  - Static/mobile object : 91.5%
  - Pedestrian/2/4 wheelers : 68.5%

<table>
<thead>
<tr>
<th>GT/AR</th>
<th>pedestrian</th>
<th>2 wheelers</th>
<th>4+ wheelers</th>
</tr>
</thead>
<tbody>
<tr>
<td>pedestrian</td>
<td>4.06508</td>
<td>0.401652</td>
<td>0.0832806</td>
</tr>
<tr>
<td>2 wheelers</td>
<td>0.397657</td>
<td>8.72356</td>
<td>1.02395</td>
</tr>
<tr>
<td>4+ wheelers</td>
<td>10.4307</td>
<td>19.173</td>
<td>55.7012</td>
</tr>
</tbody>
</table>
Conclusion
Conclusion

- Very challenging benchmark:
  - Large dataset, requiring a large amount of work for ground truth production
  - Very detailed semantic tree
  - Difficult data:
    - Vehicle stops (point accumulations)
    - Transversal roads (very different scanning geometry)
  - Objectivity:
    - Manual production of the ground truth
    - Parameter free evaluation
Perspectives

- Releasing a larger part of the ground truth for learning
- More targeted benchmarks (car type determination, static/mobile object determination, ...)
- Benchmark will stay open for future participants
- Having the participants provide an executable instead of a result:
  - Comparison of timings
  - More validity to the benchmark results (no fine parameter tuning)
- Vector evaluation for surface limits
- Correcting the anisotropy in pointwise evaluation
Thank you for your attention
Visit us at
data.ign.fr/benchmarks/UrbanAnalysis