IGN
l'information grandeur nature
INSTITUT NATIONAL
DE L INFORMATION
GÉOGRAPHIQUE
ET FORESTIÈRE

IQmulus

iQmulus/TerraMobilita benchmark on Urban Analysis

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

MATIS
METHODES D'ANALYSE ET DE TRAITEMENT D'IMAGES POUR LA STEREO-RESTITUTION

MINES ParisTech

TerraMobilita

IGN

Introduction

Introduction

- Mobile laser scanning (MLS) generates massive amount of data
- Urban cores are obects 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

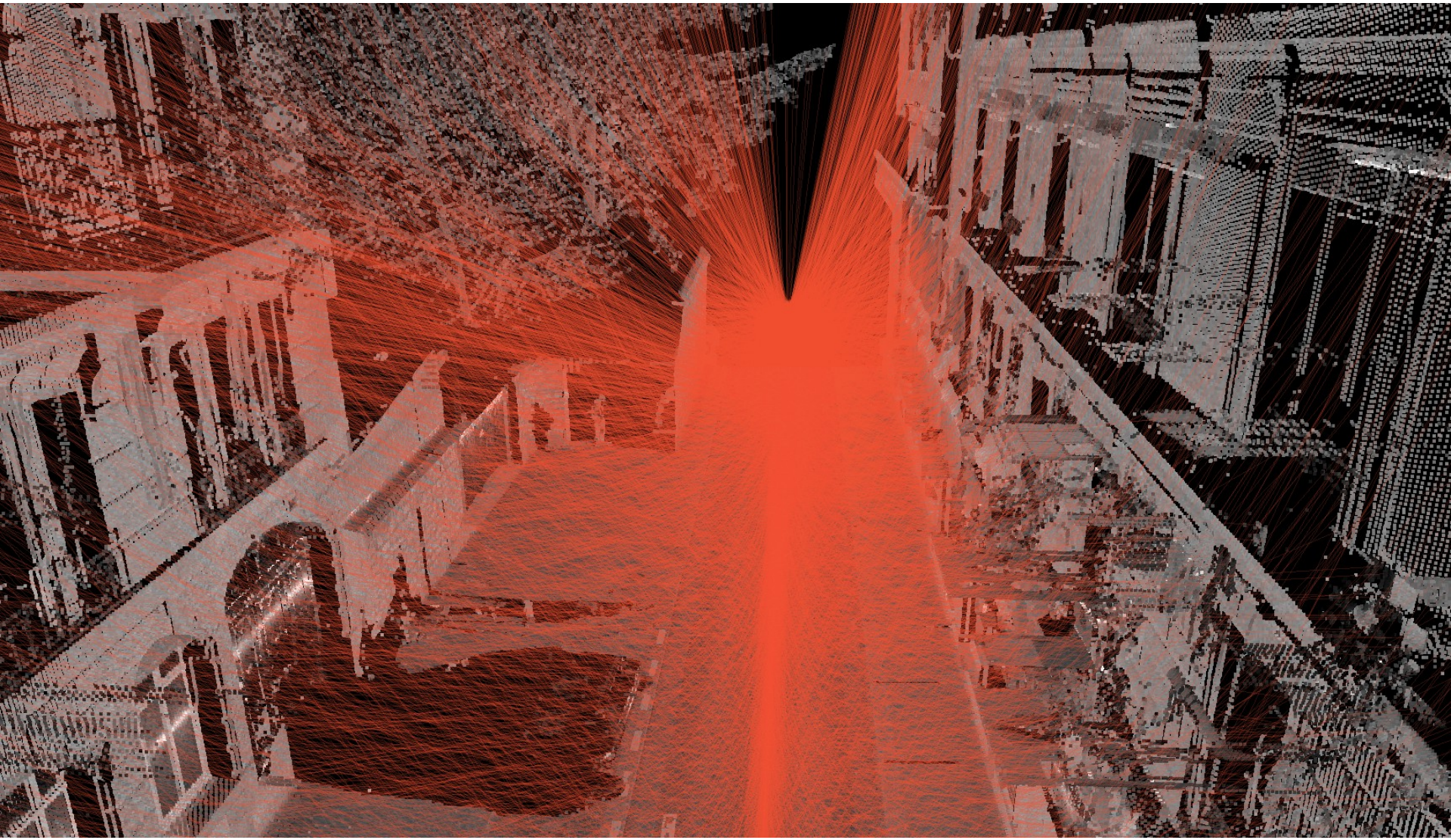
Data

- Acquisition with Stereopolis MLS :
  - **360° Riegl sensor, multiecho**
  - **Applanix georeferencing**
- Anisotropic resolution :
  - **Across trajectory : Constant angular resolution (0.03°) => distance dependant geometric resolution**
  - **Along trajectory : Constant time resolution (10ms) => speed dependant geometric resolution**

Data attributes

- Attributes :
  - **X,Y,Z : coordinates of the echo in a geographical frame**
  - **X0,Y0,Z0 : 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

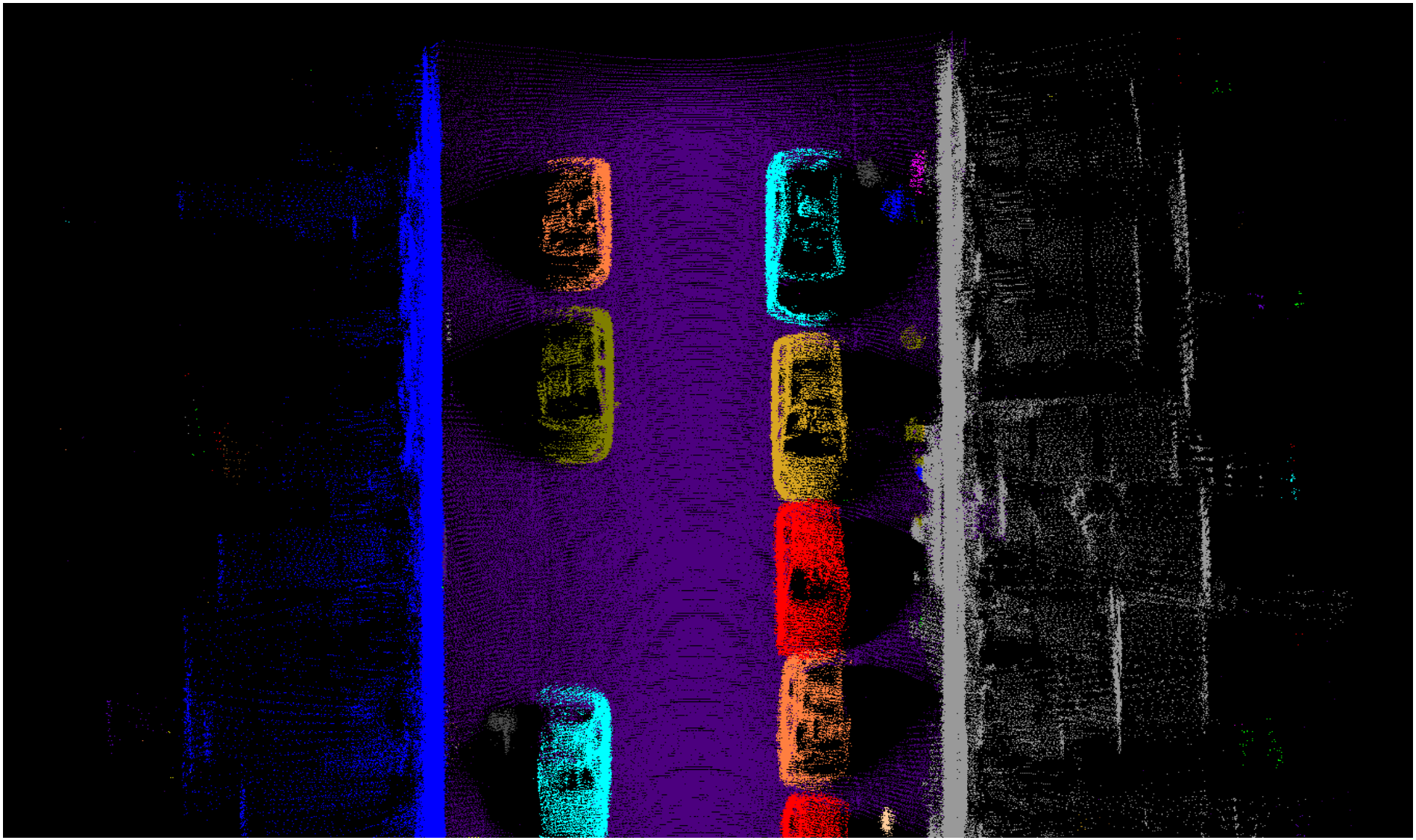
- 10+1 zones in the center of Paris (6ème arrondissement)
- Each zone has 30 (12) million points corresponding to 2 minutes of acquisition each and around 500m (depending on vehicle speed)

Analysis problem statement

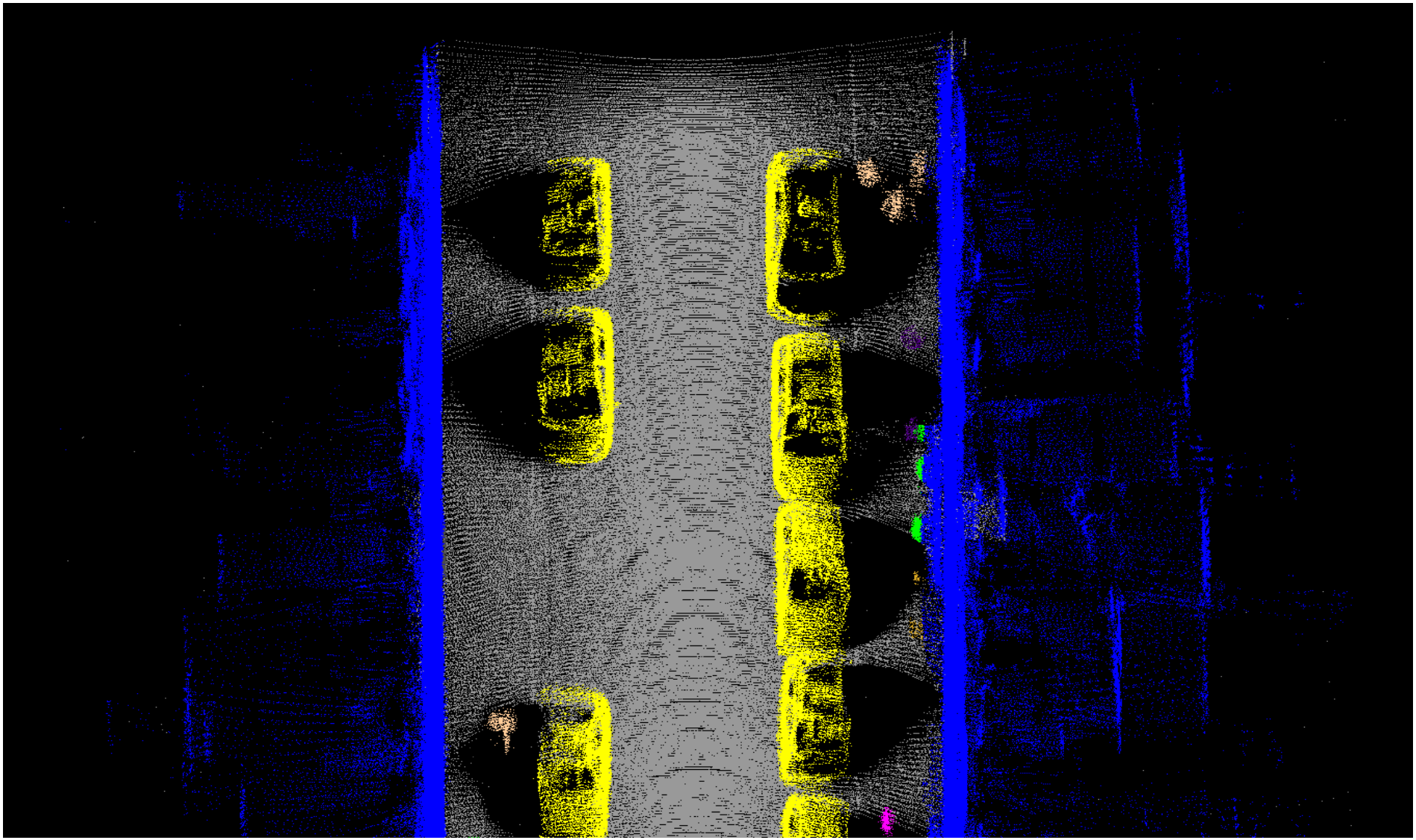
Scene analysis

- 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



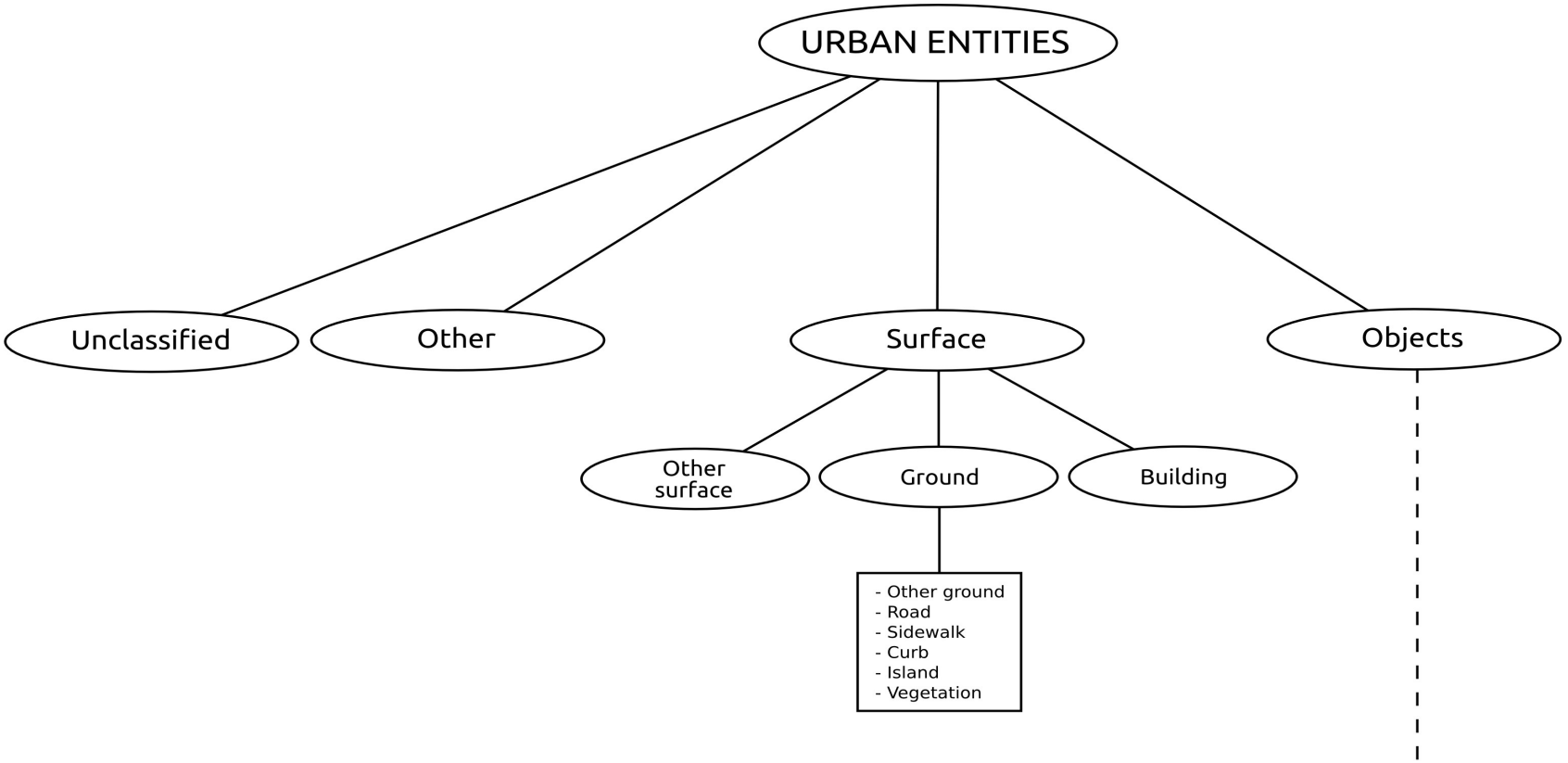
Targeted Communities

- 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

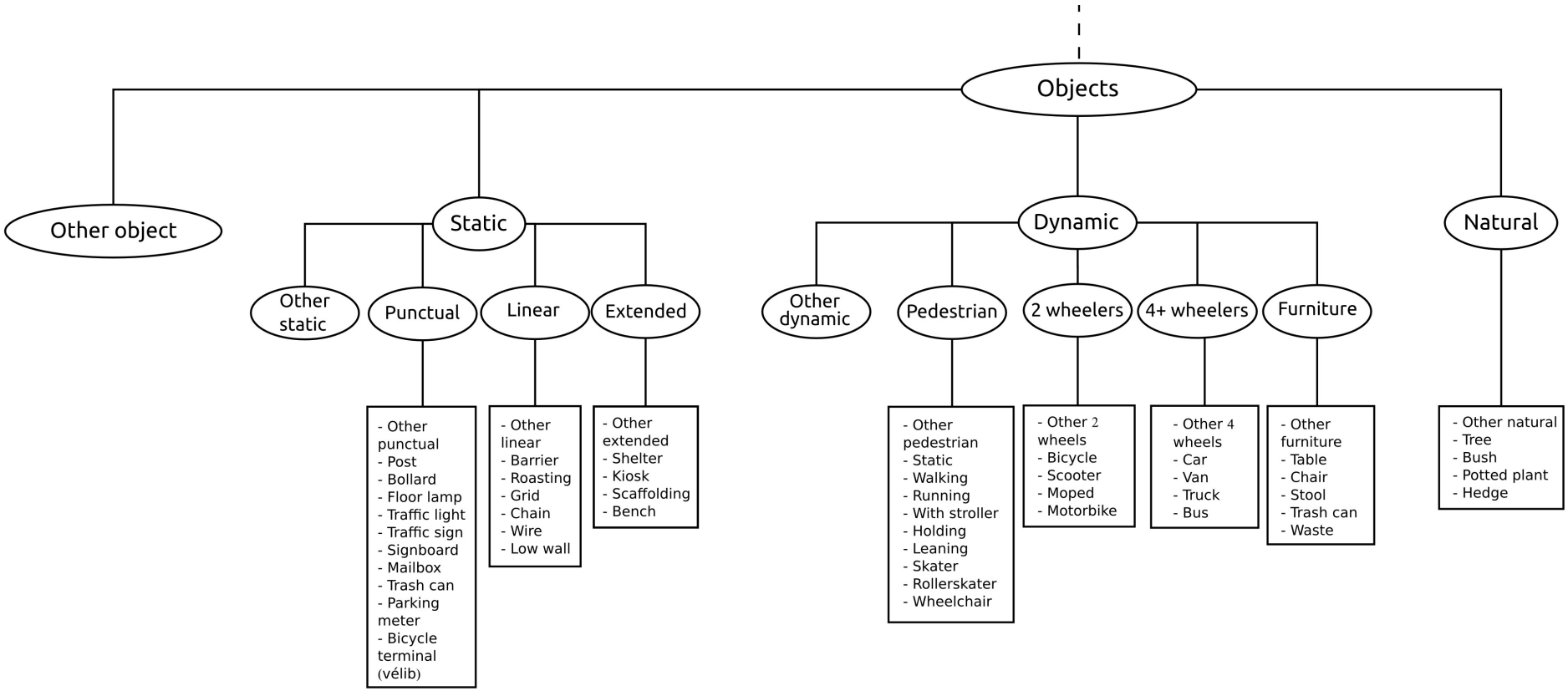
Scene analysis : semantics

- 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

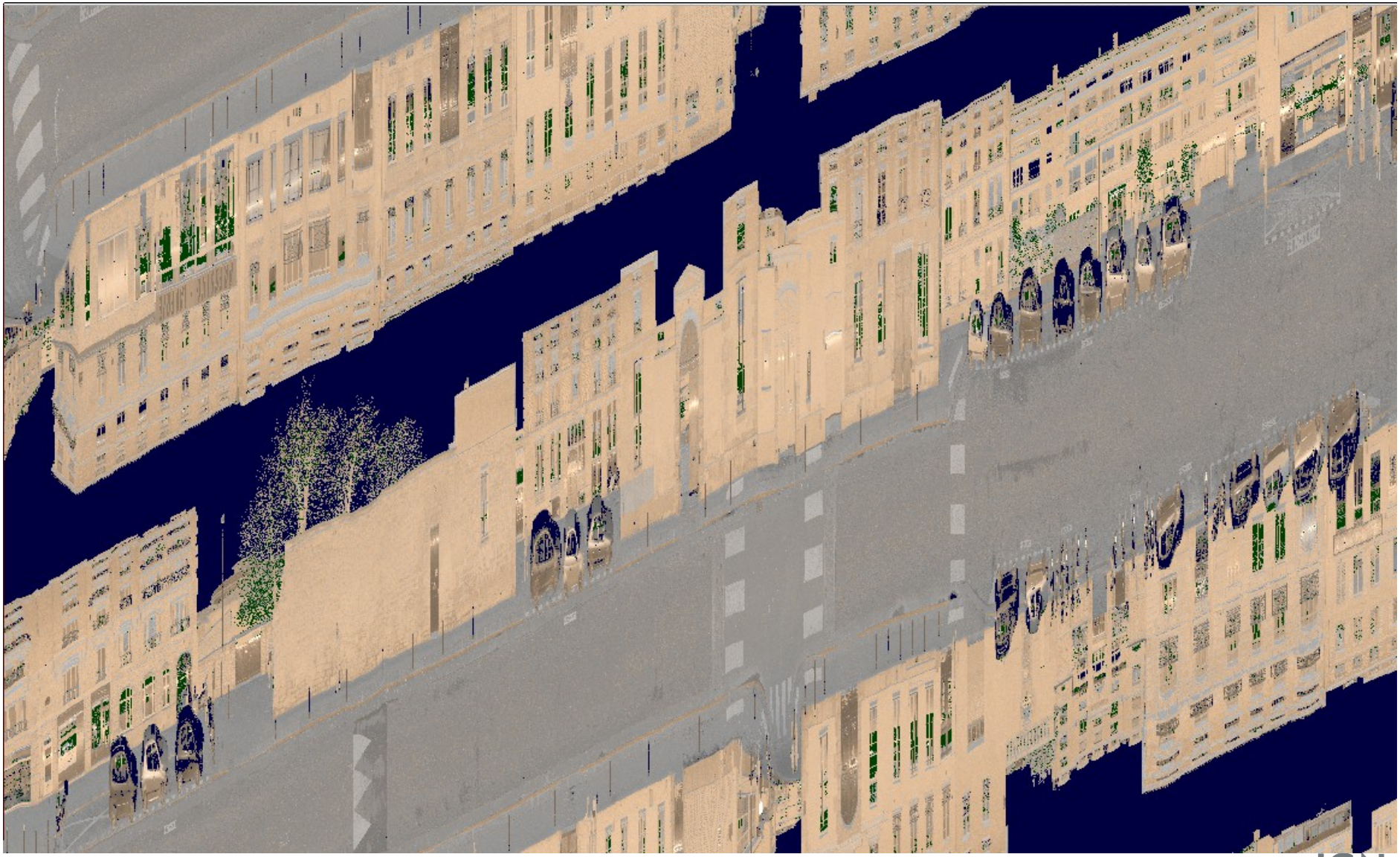


Ground truth production

Ground truth production tool

- Requirements
  - **Fast and easy navigation and annotation**
  - **Segmentation at point level**
  - **Interactivity/editability**
- We designed an inteface 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 :**

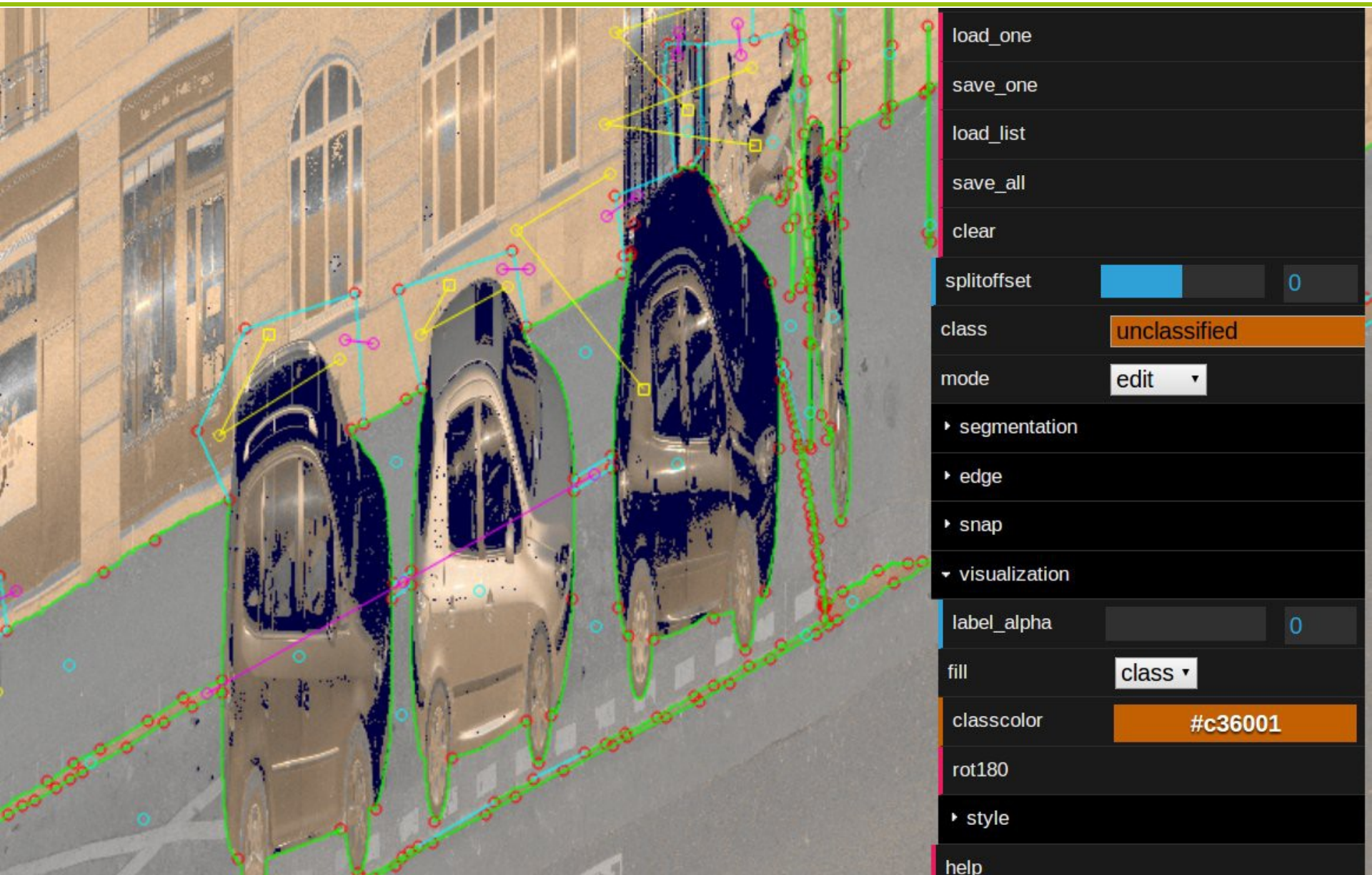
$$cost = \max\left(1, \lambda\left(1 - \max\left(\frac{\Delta_\alpha}{\alpha_0}, \frac{\Delta_d}{d_0}, \frac{\Delta_r}{r_0}\right)\right)\right)$$

    - **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


load_one
save_one
load_list
save_all
clear
splitoffset
0
class
unclassified
mode
edit
segmentation
edge
snap
visualization
label_alpha
0
fill
class
classcolor
#c36001
rot180
style
help

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 alowing 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

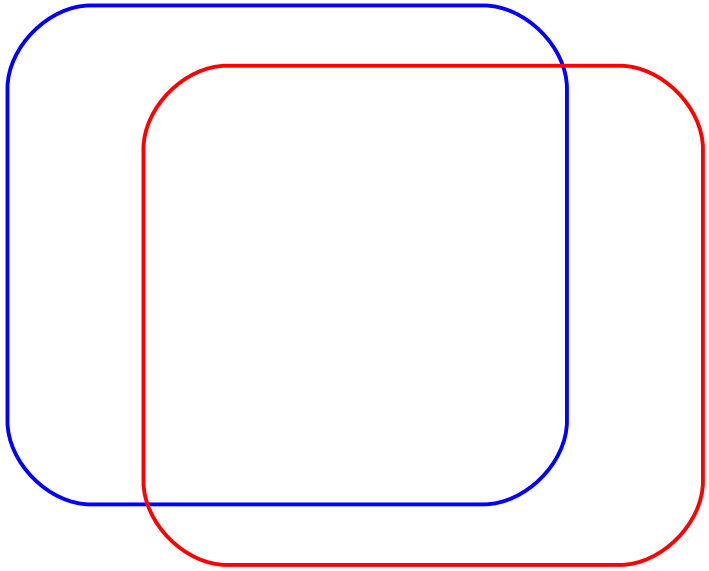
- Evaluate the algorithm result:
  - **As a classification algorithm: confusion matrix**
  - **As a detection algorithm :**
    - **precision/recall for object classes**
    - **No notion of object for surface classes**

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

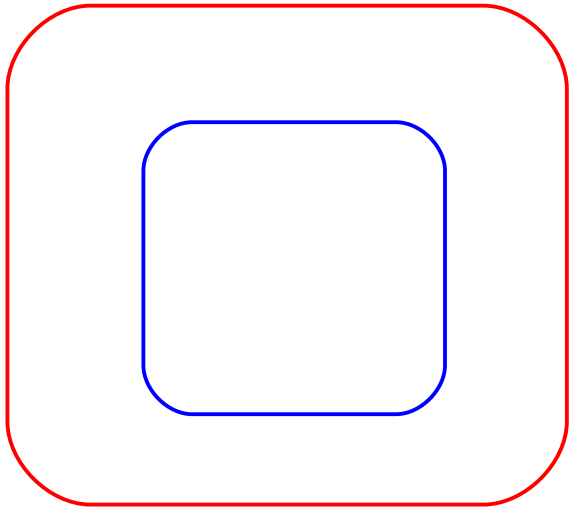
Delocalisation

Ground truth **Algorithm result**



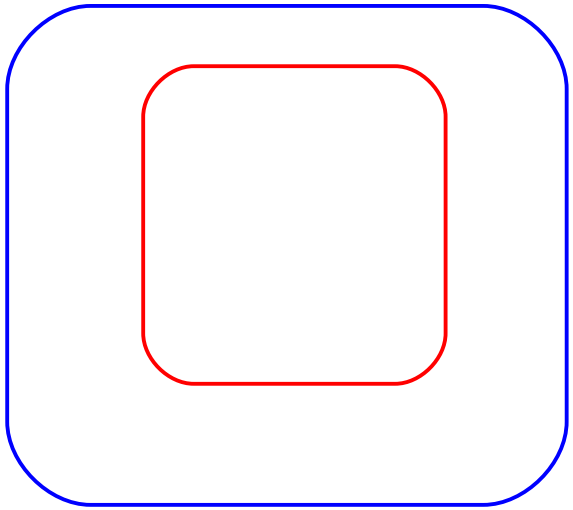
Dilatation/Erosion

Ground truth



Dilatation

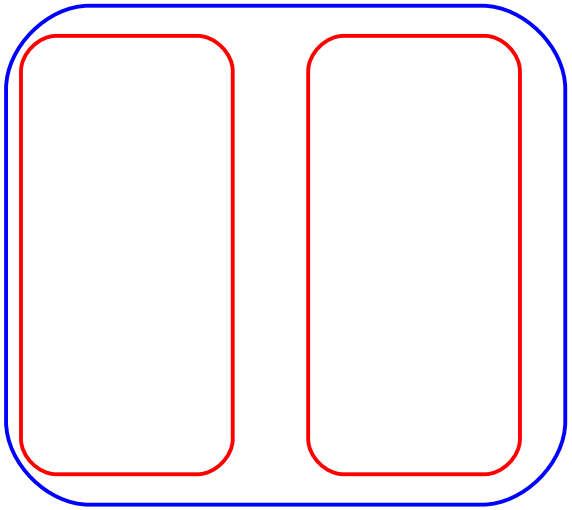
Algorithm result



Erosion

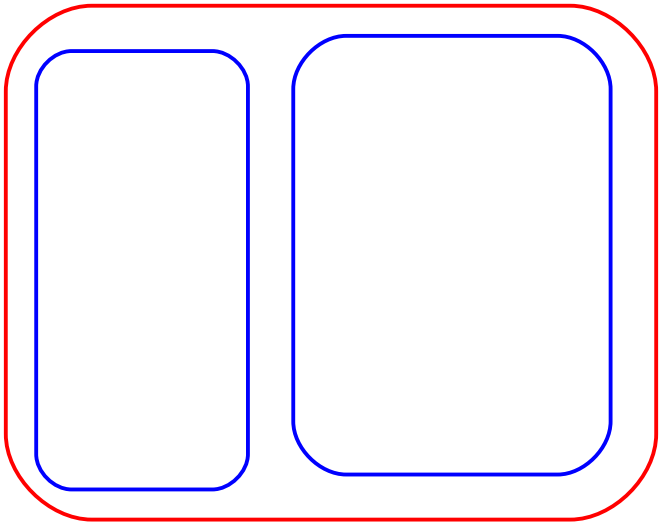
Scission/fusion

Ground truth



Split (N to 1)

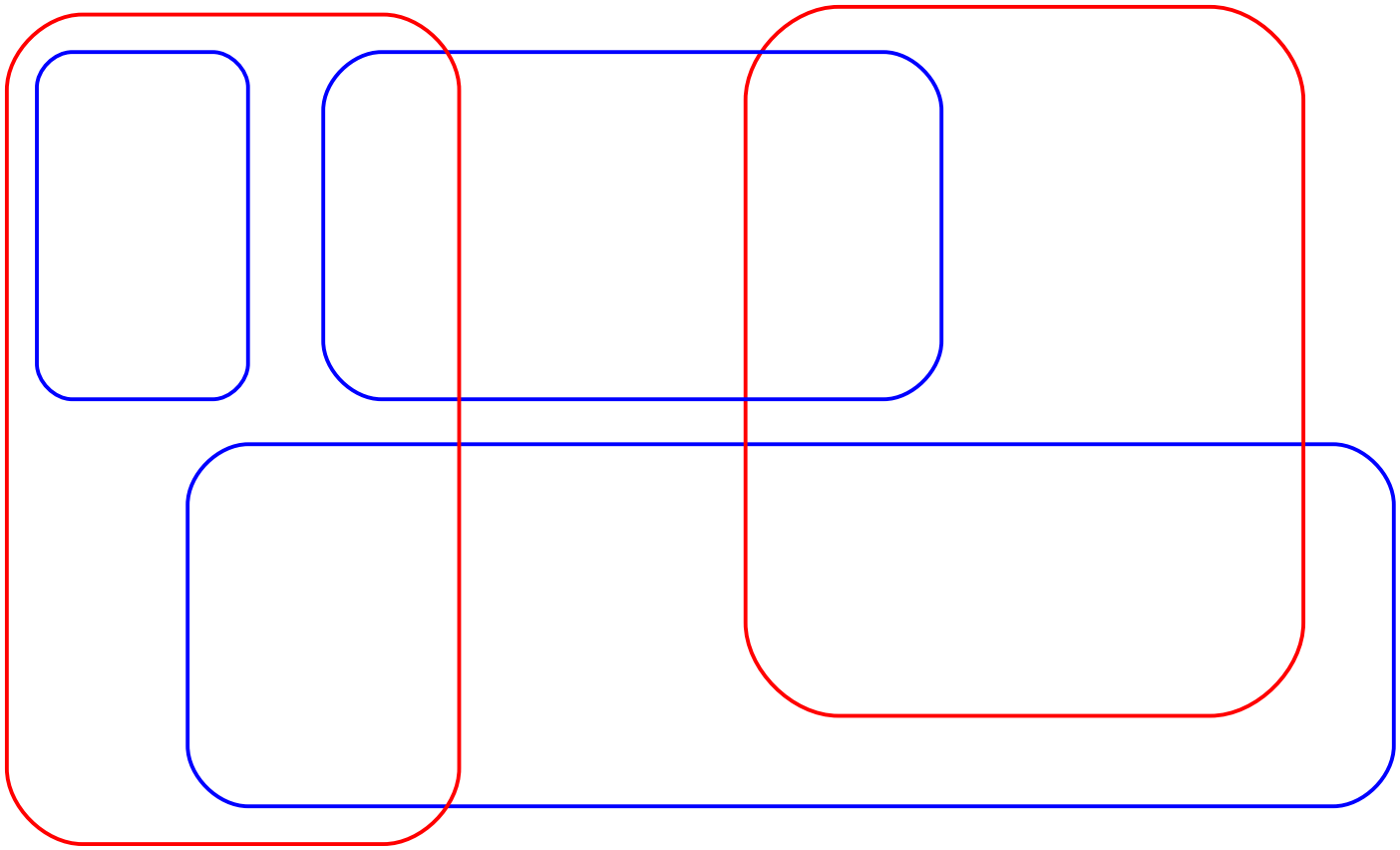
Algorithm result



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 objets :
  - **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

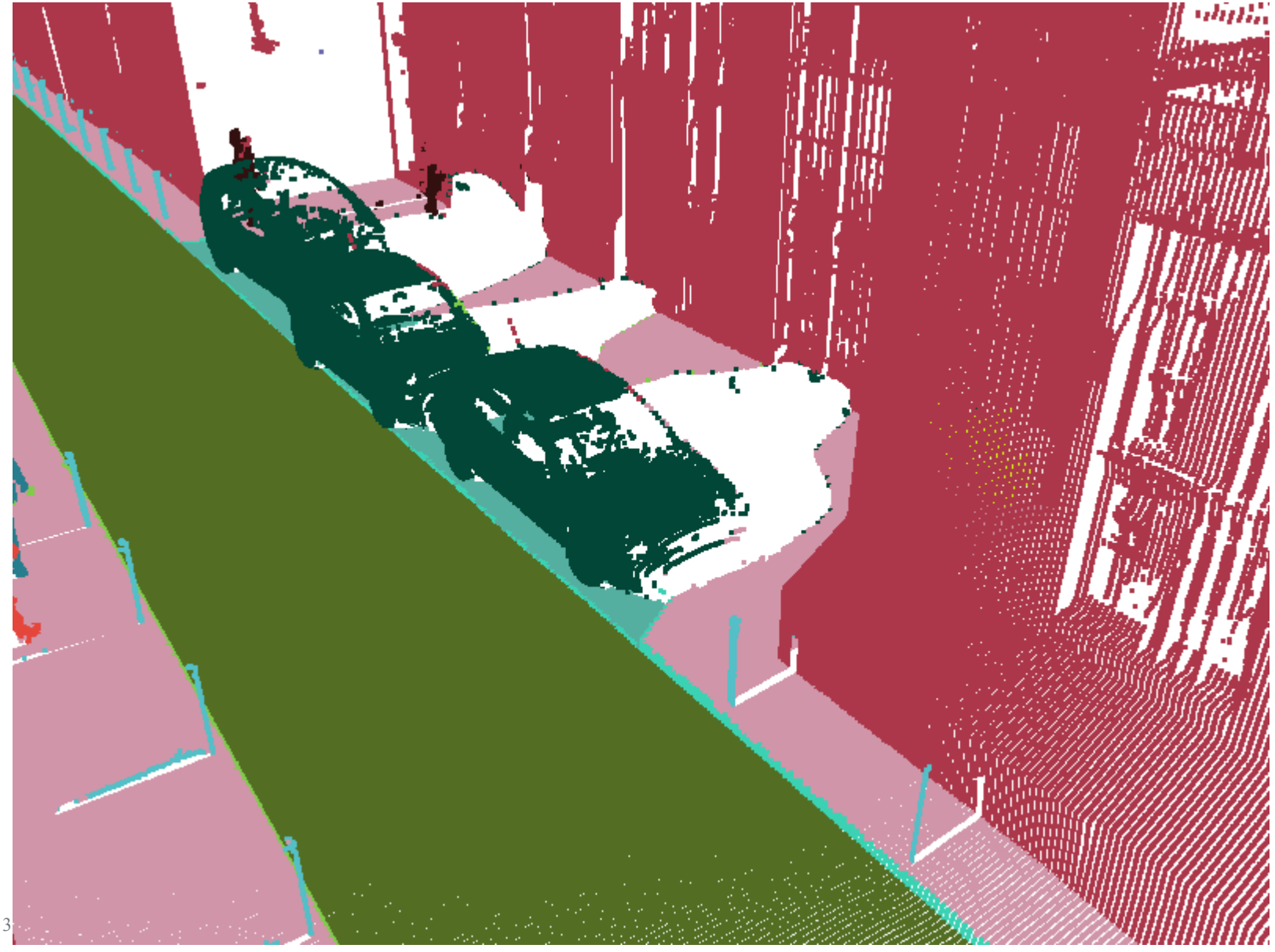
- Give precision and recall as a function of this threshold :
  - **No arbitrary (subjective) choice of a threshold**
  - **Compare algorithms by comparing curves**
- For thresholds below 0.5, also give the number of N to 1 and 1 to M pairings

Participants & results

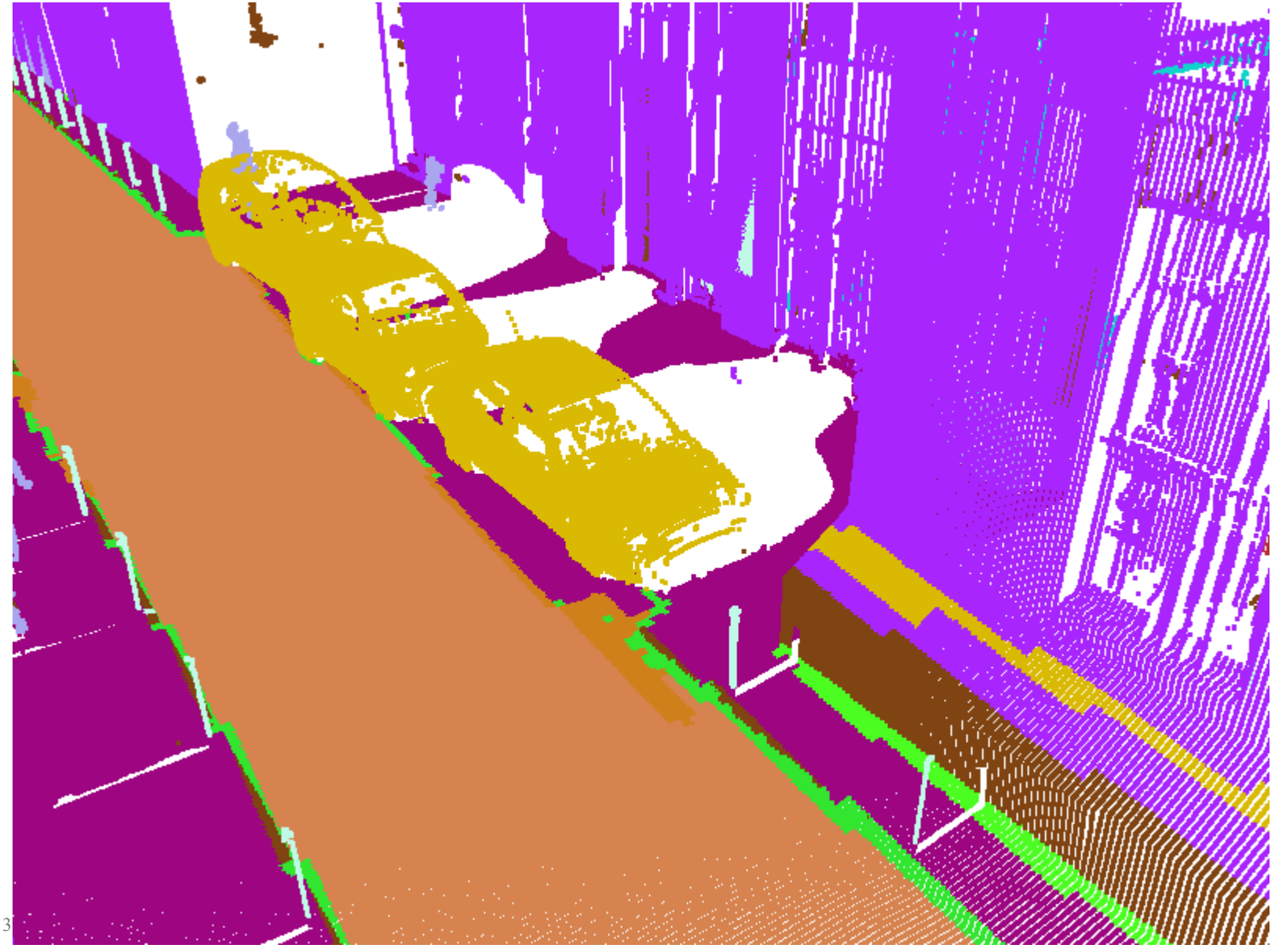
Participants

- 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

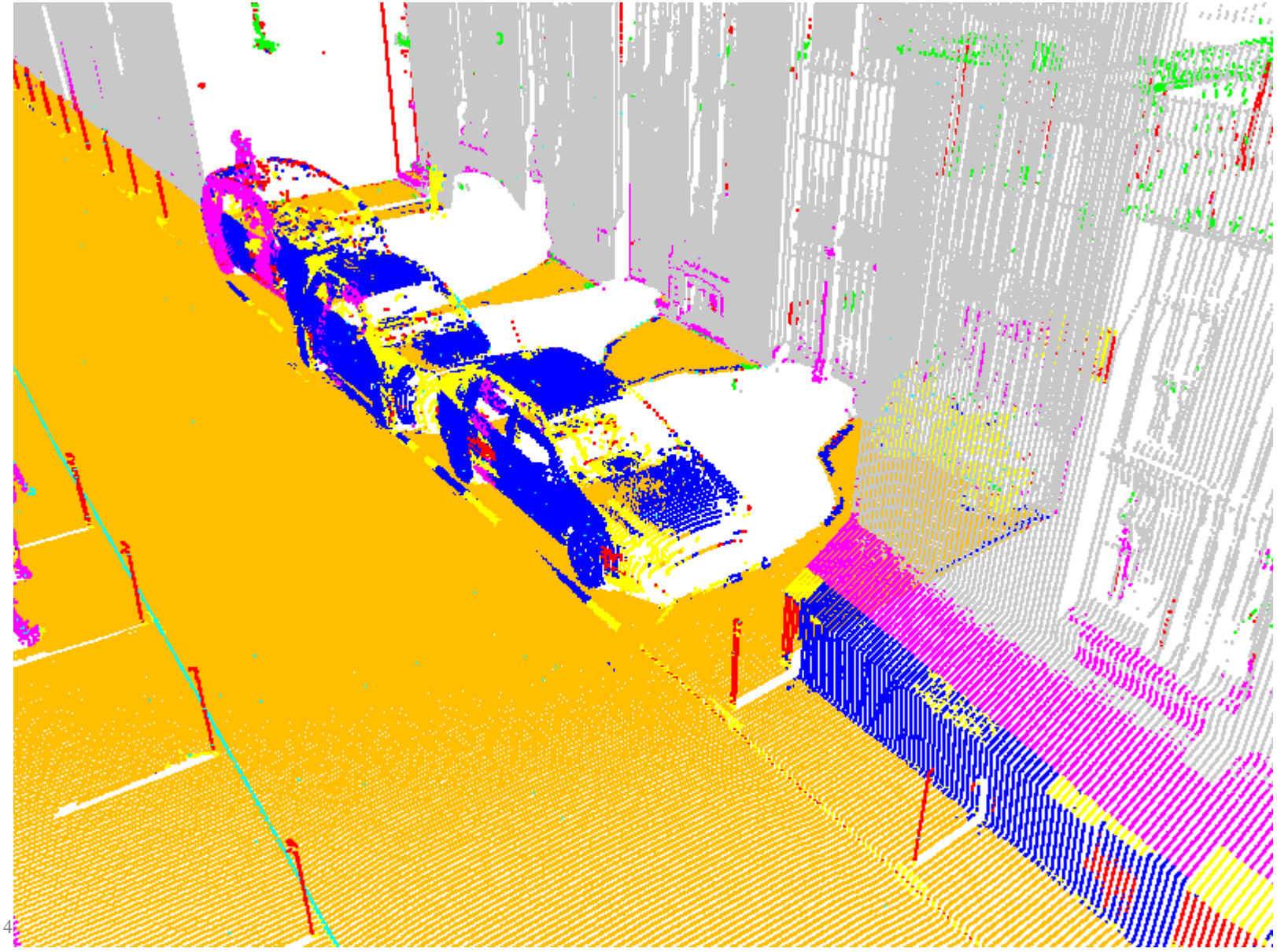
Ground truth



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%**

GT/AR	Road&side	Curb
road	71.8348	0.684865
sidewalk	25.7088	0.687476
curb	0.187855	0.896216

Results for CMM

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

GT/AR	Road&side	Curb
road	71.8348	0.684865
sidewalk	25.7088	0.687476
curb	0.187855	0.896216

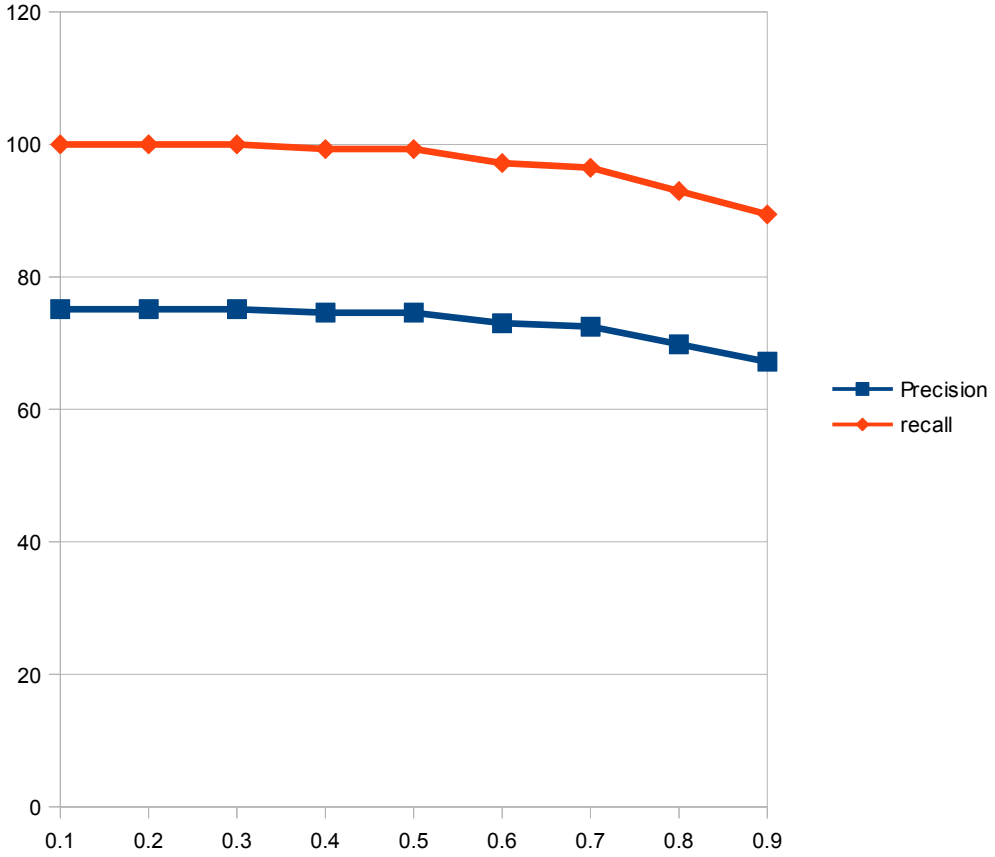
Results for CMM

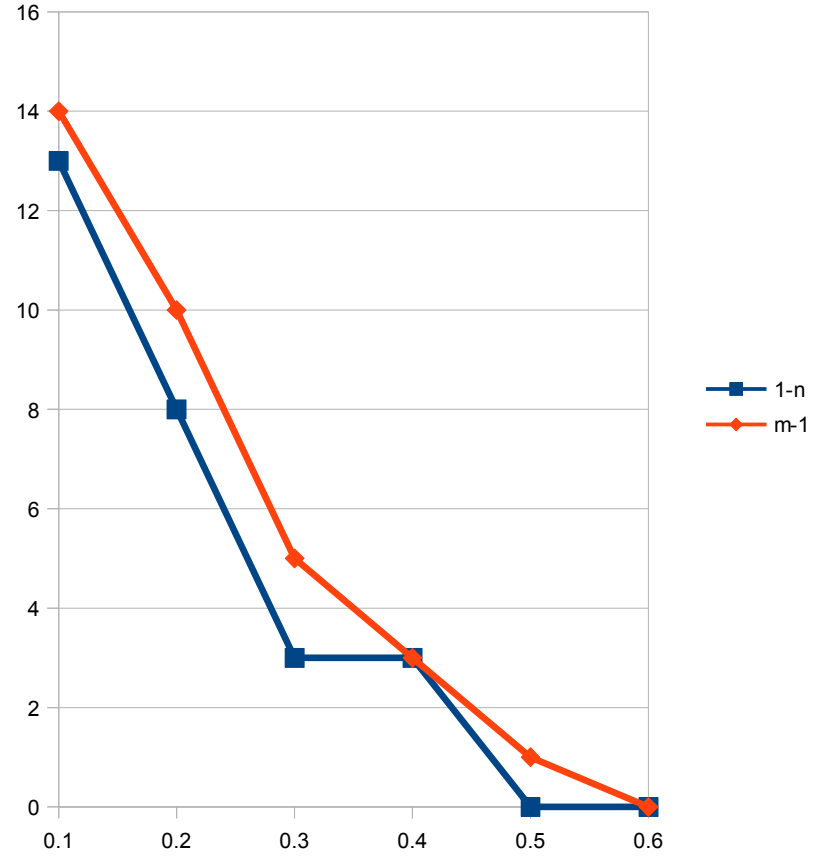
- Classification (one 30 Mpts zone) :
  - **Surface/object : 92.6%**
  - **Building/ground surface : 98.3%**
  - **Curb/sidewalk/road : 98.4% but curb (2.3%) confused for sidewalk (0.7%) and road (0.7%) because of rasterization.**
  - **Static/mobile object : 91.8%**
  - **Pedestrian/2/4 wheelers : 99.3%**

GT/AR	pedestrian	2 wheelers	4+ wheelers
pedestrian	1.63193	0.00262888	0.123962
2 wheelers	0.388468	0.653378	0
4+ wheelers	0.112435	0.0281088	97.0591

Results for CMM

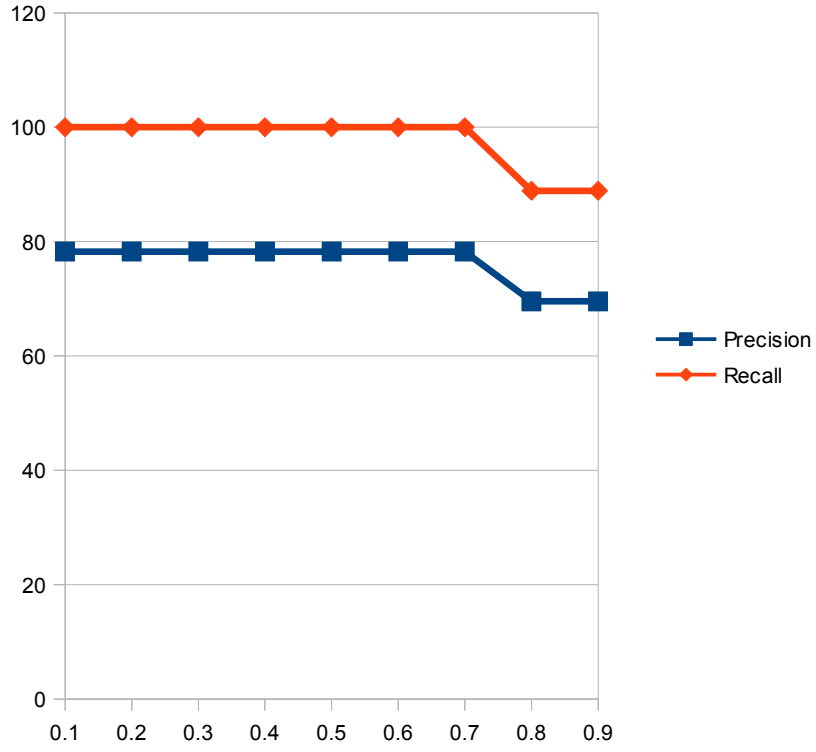
- Detection (one 30 Mpts zone) : All objects



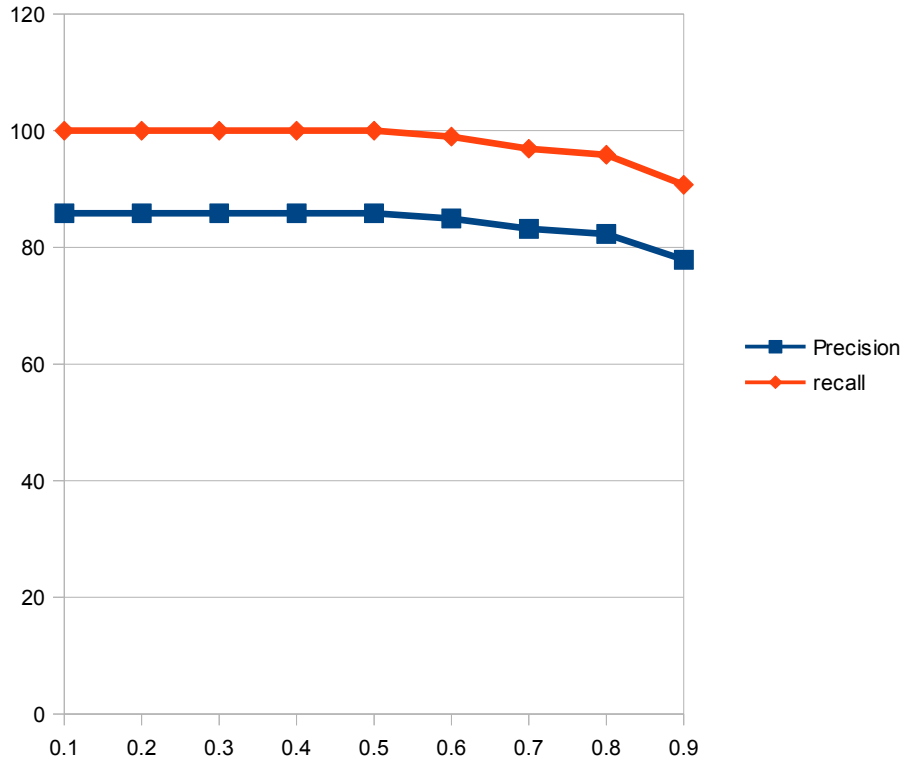


Results for CMM

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



Dynamic objects



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%**

GT/AR	pedestrian	2 wheelers	4+ wheelers
pedestrian	4.06508	0.401652	0.0832806
2 wheelers	0.397657	8.72356	1.02395
4+ wheelers	10.4307	19.173	55.7012

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

- Perspectives :
  - Realeasing 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

IGN
INSTITUT NATIONAL
DE L'INFORMATION
GÉOGRAPHIQUE
ET FORESTIÈRE

l'information grandeur nature

IQmulus

Thank you for your attention
Visit us at
data.ign.fr/benchmarks/UrbanAnalysis

IGN