

# 3D Reconstruction Meets Semantics

## Reconstruction Challenge

Radim Tyleček, Torsten Sattler,  
Thomas Brox, Marc Pollefeys, Robert B. Fisher

EU project TrimBot2020

ICCV 2017 Workshop, Venice



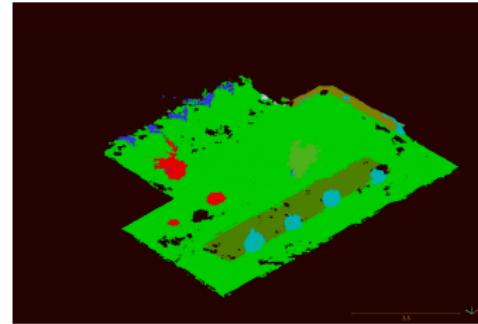
# Outline

- Challenge Goals
- Garden Dataset
- Evaluation
- Results



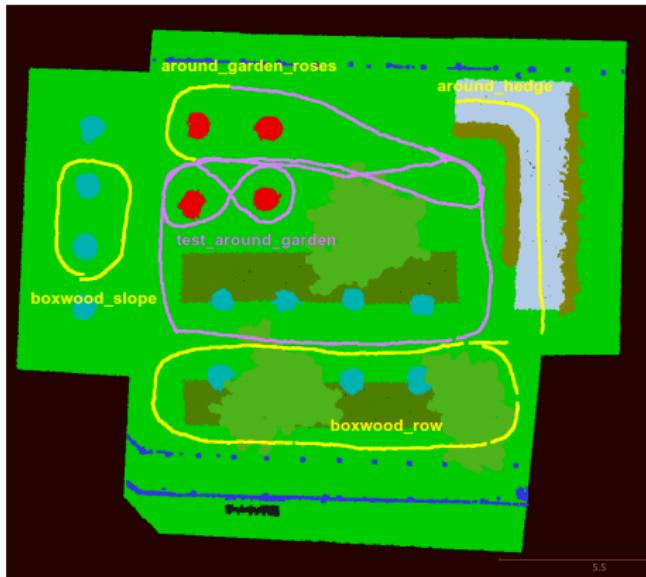
# 3DRMS Challenge Goals

- **Input:** Set of images and their known camera poses
- **Goal:** Create a semantically annotated 3D model of the scene
  - Compute depth maps for the images
  - Fuse them together into a single 3D model
  - Incorporate information from the semantics



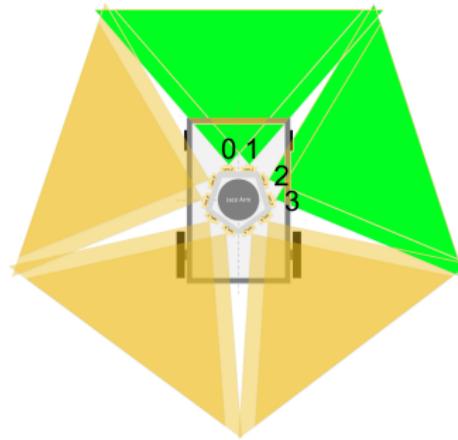
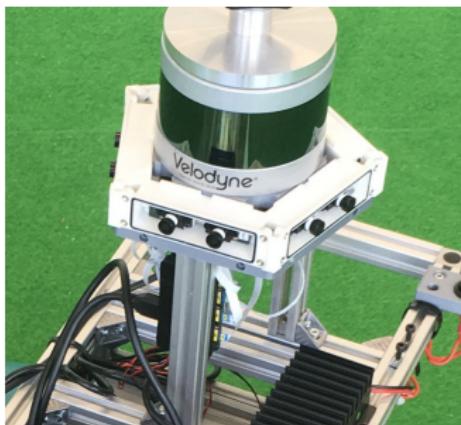
# Garden Dataset

- A set of **training** sequences
  - Calibrated images with their camera poses
  - Ground truth semantic annotations for a subset of these images
  - A semantically annotated 3D point cloud depicting the area of the training sequence
- A **testing** sequence
  - Calibrated images with their camera poses
  - Ground truth for evaluation only



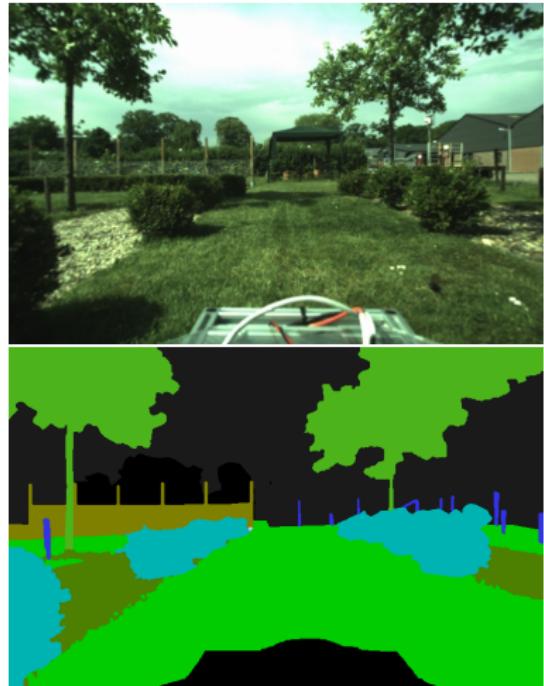
# Garden Dataset: Image Data

- Four cameras (two stereo pairs) 0-1, 2-3
  - WVGA resolution (752x480), wide FOV
  - Color (cams 0&2) and greyscale (1&3)
- SfM pose estimation and registration to GT point cloud



# Garden Dataset: Semantics

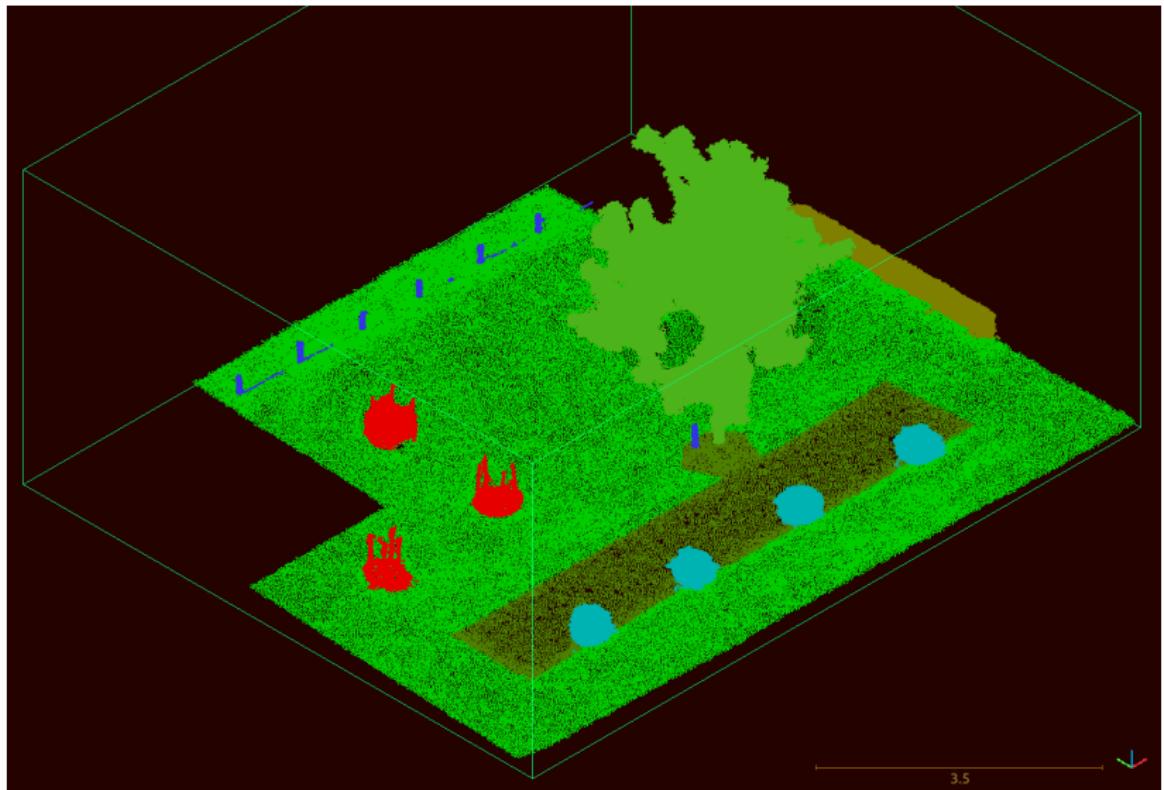
- Set of 9 classes we distinguish
  - Grass (light green)
  - Ground (brown)
  - Pavement (grey)
  - Hedge (ochre)
  - Topiary (cyan)
  - Rose (red)
  - Obstacle (blue)
  - Tree (dark green)
  - Background (black)



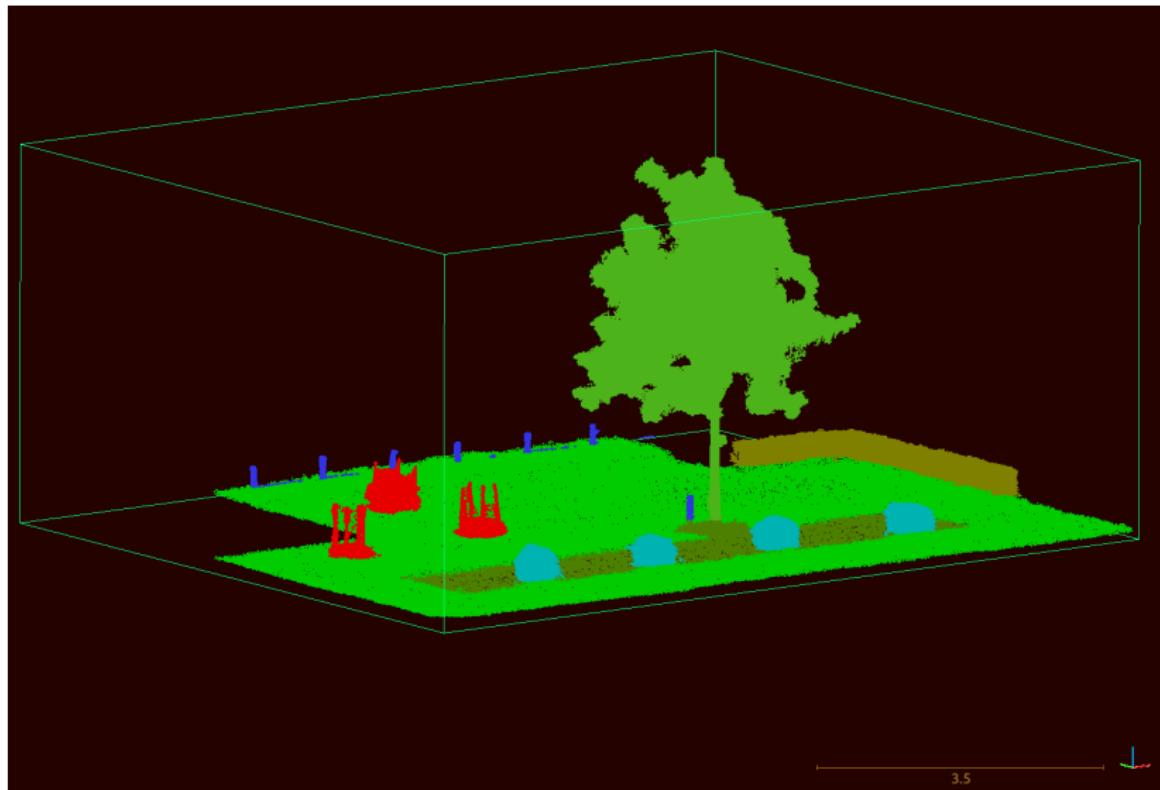
# Garden Dataset: Training Point Cloud



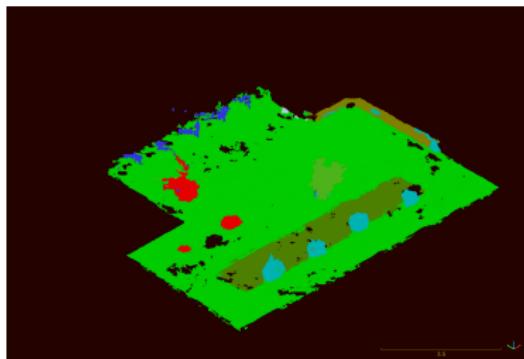
# Garden Dataset: Test Point Cloud



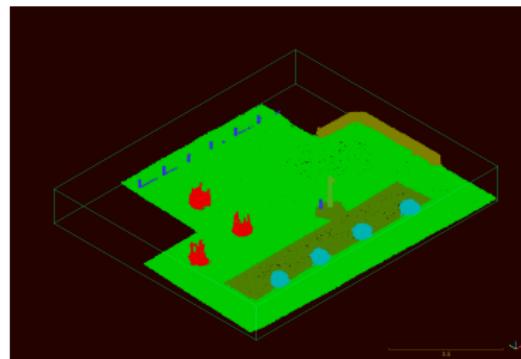
# Garden Dataset: Test Point Cloud



# Submissions vs. Ground Truth

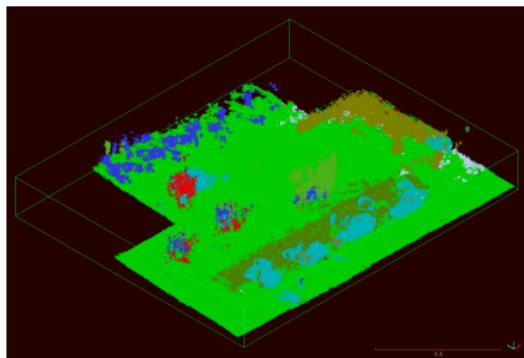


Taguchi

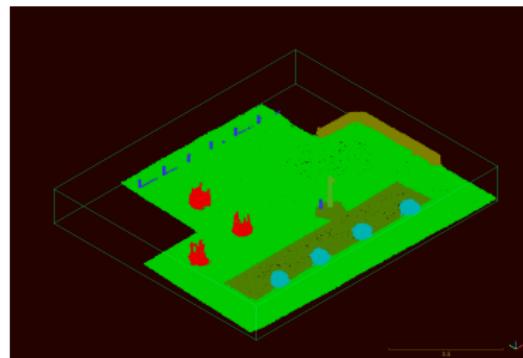


GT

# Submissions vs. Ground Truth



Moras



GT

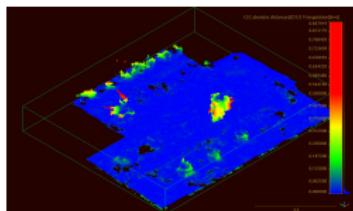


# Evaluation Methodology

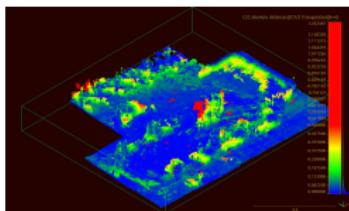
- Test set GT
  - XY: delimited by the perimeter of the test area
  - Z: limited to 1 m high section above the ground
- 3D geometry
  - Cumulative histograms of distances (mesh  $\rightleftarrows$  GT points)
  - **Accuracy** is distance  $d$  (in m) such that 90% of the reconstruction is within  $d$  of the ground truth mesh
  - **Completeness** is the percent of points in the GT point cloud that are within 5 cm of the reconstruction
- Semantics
  - Labels assigned to vertices or faces of the 3D model
  - Projected to all test images
  - **Prediction accuracy** of the pixels corresponding to the 3D test part



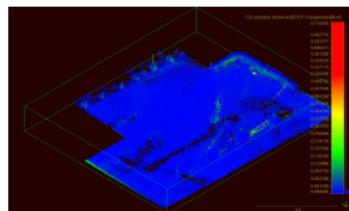
# Evaluation: 3D Geometry



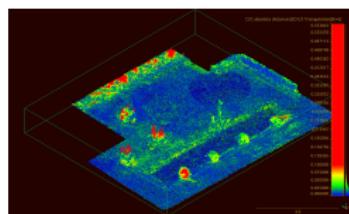
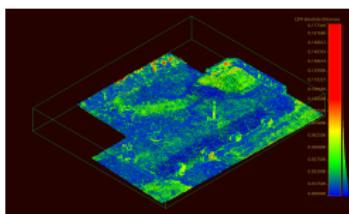
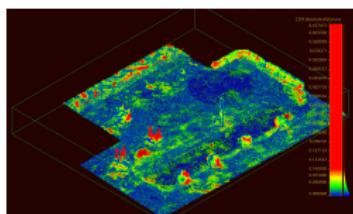
Taguchi



Moras



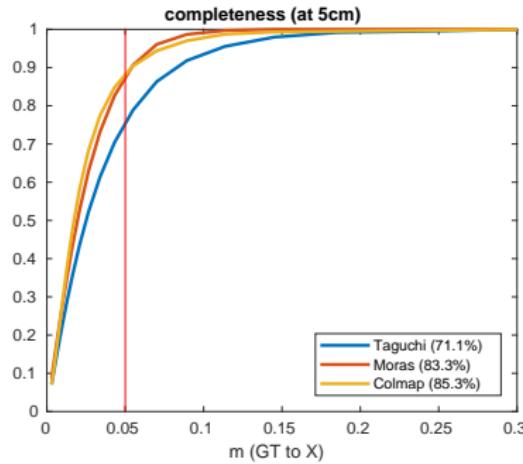
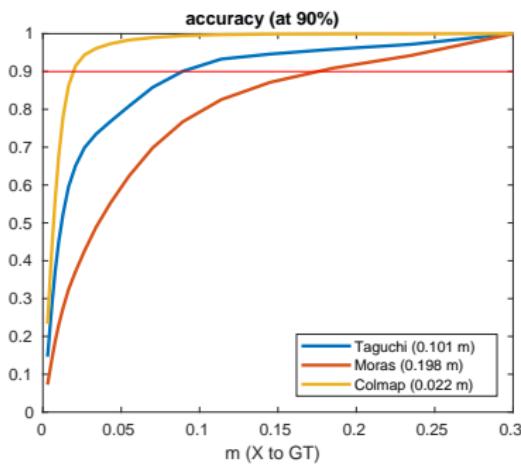
ColMap<sup>1</sup> (baseline)



- Distances: Cold colors indicate well reconstructed segments
- Hot colors indicate hallucinated surface (accuracy) or missing parts (completeness).

<sup>1</sup>Schönberger et al.: Structure-from-motion revisited. CVPR, 2016.

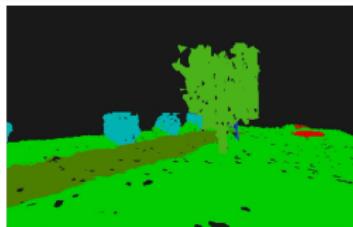
# Evaluation: 3D Geometry



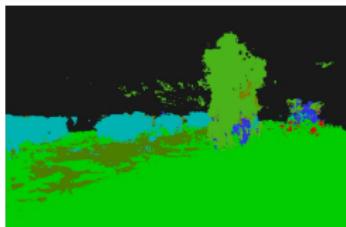
- Cumulative plots of distances (mesh  $\leftrightarrow$  GT points)



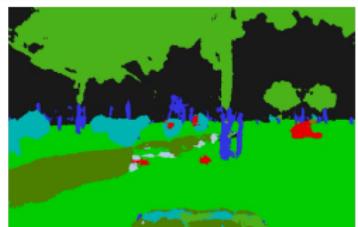
## Evaluation: Semantics



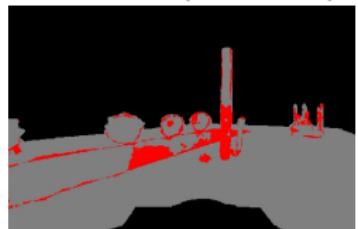
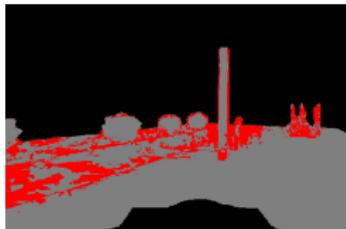
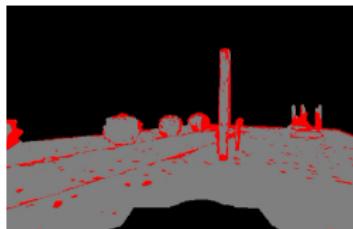
Taguchi



Moras



SegNet<sup>2</sup> (baseline)



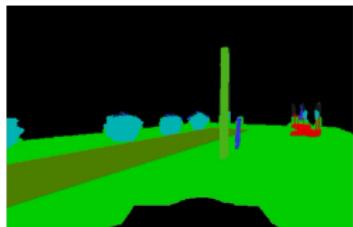
Prediction

Error mask

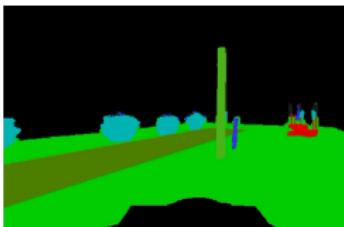
- Error: *red* incorrect pixels, *grey* correct, *black* not evaluated
- SegNet: pre-trained with ImageNet, 20k synthetic garden images + challenge train set

<sup>2</sup>Badrinarayanan et al.: SegNet: A deep convolutional encoder-decoder architecture for image segmentation. PAMI, 2017.

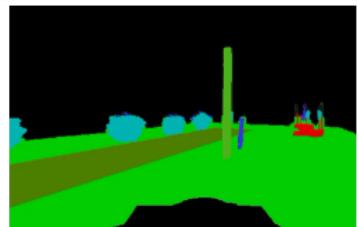
## Evaluation: Semantics



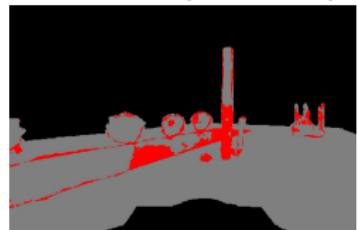
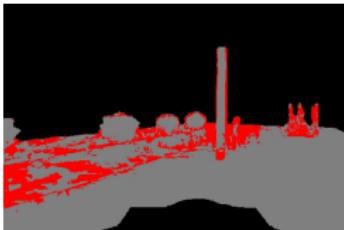
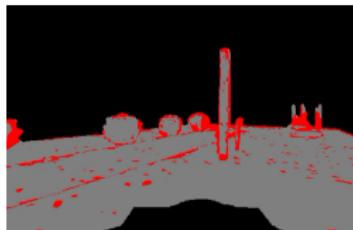
Taguchi



Moras



SegNet<sup>3</sup> (baseline)



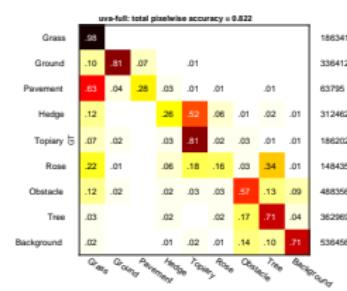
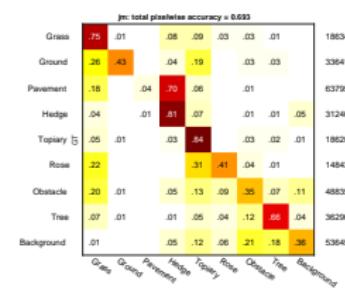
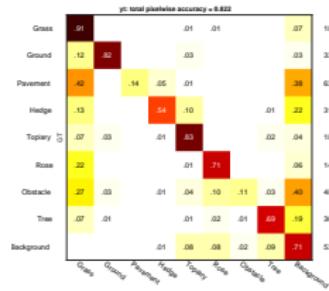
GT

Error mask

- Error: *red* incorrect pixels, *grey* correct, *black* not evaluated
- SegNet: pre-trained with ImageNet, 20k synthetic garden images + challenge train set

<sup>3</sup>Badrinarayanan et al.: SegNet: A deep convolutional encoder-decoder architecture for image segmentation. PAMI, 2017.

# Evaluation: Semantics



82.2%

Taguchi

69.3%

Moras

82.2%

SegNet (baseline)

- Confusion matrix: *dark* on diagonal indicates good match of the prediction with GT labels
- Semantic accuracy: pixelwise ratio of correct predictions over all test images



# Evaluation: Summary of Performance

<i>Method</i>	3D Reconstruction		<i>Semantic</i>
	<i>Accuracy</i>	<i>Completeness</i>	<i>Accuracy</i>
Taguchi	<b>0.101 m</b>	71.1 %	<b>82.2 %</b>
Moras	0.198 m	<b>83.3 %</b>	69.3 %
Baseline	<i>0.022 m</i>	85.3 %	82.2 %



# Challenge Results

Method	3D Reconstruction		Semantic
	Accuracy	Completeness	Accuracy
Taguchi	<b>0.101 m</b>	71.1 %	<b>82.2 %</b>
Moras	0.198 m	<b>83.3 %</b>	69.3 %
Baseline	0.022 m	85.3 %	82.2 %

---

**Congratulations and thanks to all participants!**

We are open to additional submissions.

