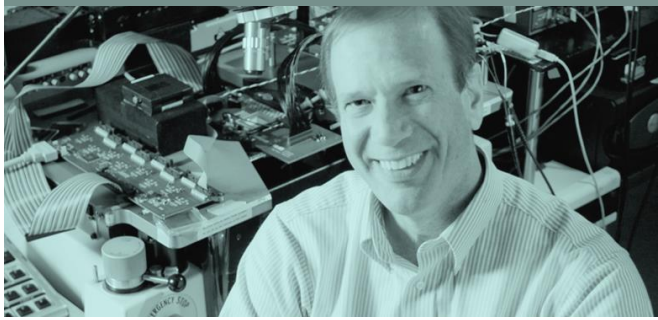


Semantically-Aware Visual Navigation

Han-Pang Chiu

Center for Vision Technologies
SRI International, Princeton, NJ, USA.
Email: han-pang.chiu@sri.com

SRI International

© 2017 SRI International. All Rights Reserved. Confidential

Agenda

- Plug-and-Play Metric Navigation
- Semantically-Aware Visual Navigation
  - Hybrid Navigation, Mapping, and Geo-Registration
  - SAANE: Semantically-Aware Attentive Neural Embeddings
- Computation Hardware – NVIDIA JETSON TX2

© 2017 SRI International. All Rights Reserved. Confidential

SRI International®

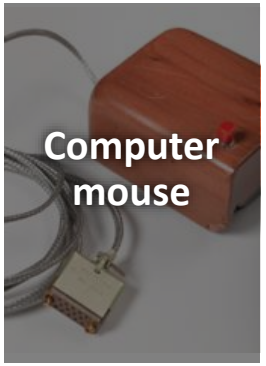
About SRI International

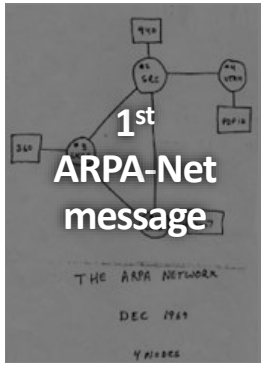
INVENT | APPLY | TRANSITION

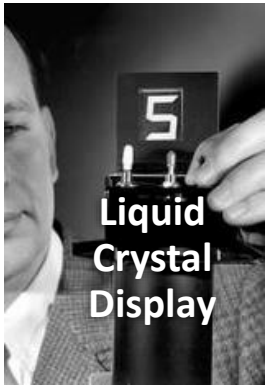
- An independent nonprofit research center merging two world-class research institutions
  - Stanford Research Institute founded in 1946
  - RCA/Sarnoff Labs founded in 1941
- $450+ million in annual revenue
- 1,800 employees, 40 labs, 20 locations
- 4000+ patents
- 500 research projects annually
- 100+ commercial projects/year
- 70+ startups spun-out in 10 years

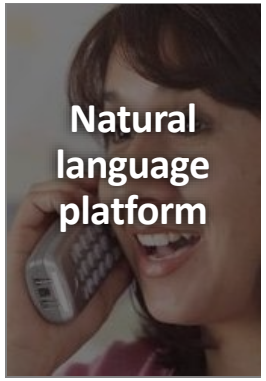
© 2017 SRI International. All Rights Reserved. Confidential

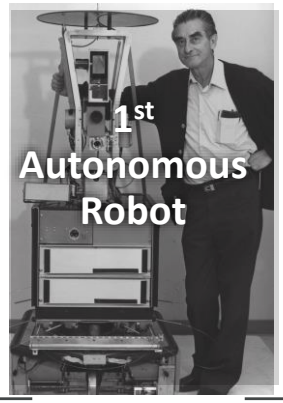
SRI's innovations…

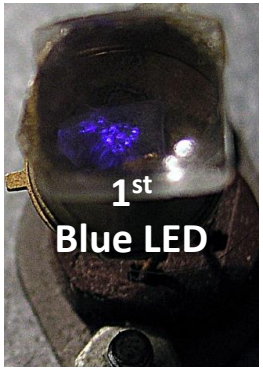


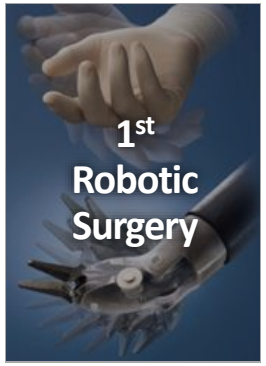


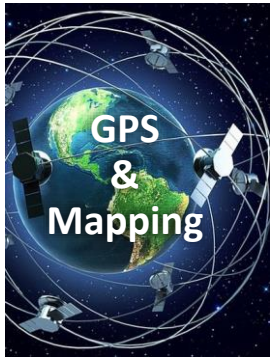


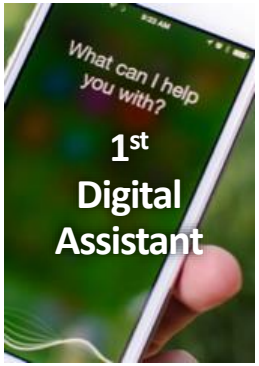


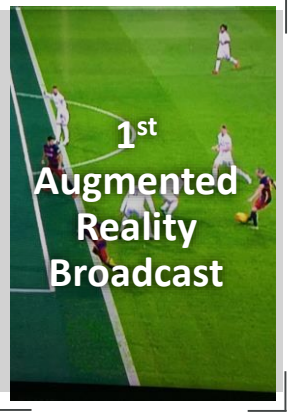












© 2017 SRI International. All Rights Reserved. Confidential

Navigation & Mapping

Autonomous Vehicles

Cars

Drones

Robots

Appliances

Augmented & Virtual Reality

VR Headsets

AR Headsets

DOD Operations in GPS-Denied

Drones

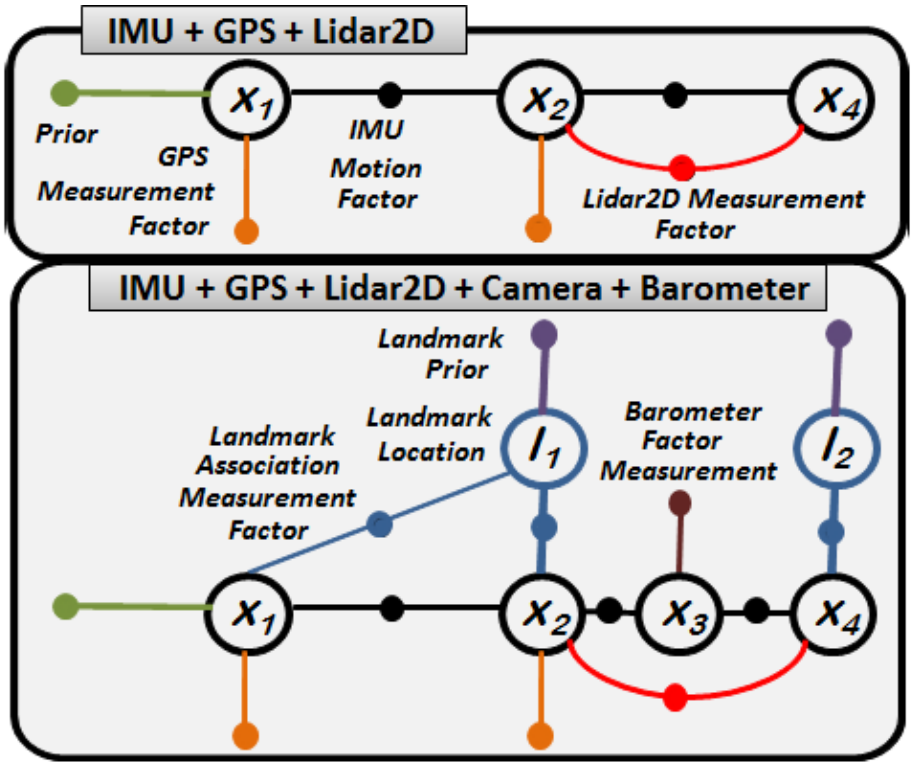
Urban Operations

3D Modeling Outdoors & Indoors

© 2017 SRI International. All Rights Reserved. Confidential

Factor Graphs for Multi-Sensor Navigation and Mapping

- We introduce a plug-and-play factor graph framework to encode sensor measurements with different frequencies, latencies, and noise distributions



* Note this formulation covers full-SLAM problems, which keep all past states for optimization.

- **Unary Factor:** $z = h(x_i) + v$
- **Binary Factor:** $z = h(x_{i-1}, x_i) + v$
- **Extrinsic Factor:** $z = h(x_i, l_j) + v$
- **IMU Factor:** Summarize multiple readings between two pose states.

>30 sensor types, > 60 sensors

Han-Pang Chiu et al., **Constrained Optimal Selection for Multi-Sensor Navigation Using Plug-and-Play Factor Graphs**, *IEEE ICRA*, 2014.
Han-Pang Chiu et al., **Robust Vision-Aided Navigation Using Sliding-Window Factor Graphs**, *IEEE ICRA*, 2013.

© 2017 SRI International. All Rights Reserved. Confidential

GPS Denied Navigation – Field Tests in 2015~2016

GOAL: Achieve GPS-level performance under all application scenarios (DARPA ASPN)

- Eliminate GPS as single point of failure
- Provide redundant capabilities and architectures with no single point of failure
- Provide Plug-and-Play, adaptable solution for multiple platforms and sensors

Shipboard (Jan. 2015)

Demo Lead: Navy: SSC-PAC

- 3 week test in Pacific Ocean aboard the USNS Waters

USNS Waters

Leader-Follower Ground Vehicles (December 2015)

Demo Lead: Army CERDEC

- 3 day test on two Humvees at Aberdeen
- Demonstrate GPS denied collaborative navigation.

Dismount (Feb. 2015)

Demo Lead: Army CERDEC

- 3 day outdoor/ indoor tests at Aberdeen
- Demonstrated scalability to low-SWAP, wearable processor platforms

Aerial (Oct. 2015,Jun. 2016)

Demo Lead: AFRL

- Three-week live DC3 and S3 flights at Ohio.
- Demonstrated GPS-denied navigation over different heights and terrains.

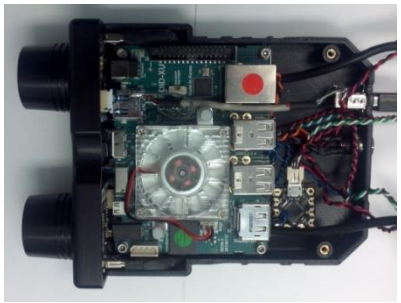
Han-Pang Chiu et al., **Precise Vision-Aided Aerial Navigation**, IEEE IROS, 2014.

© 2017 SRI International. All Rights Reserved. Confidential

GPS-Denied Collaborative Navigation and Mapping

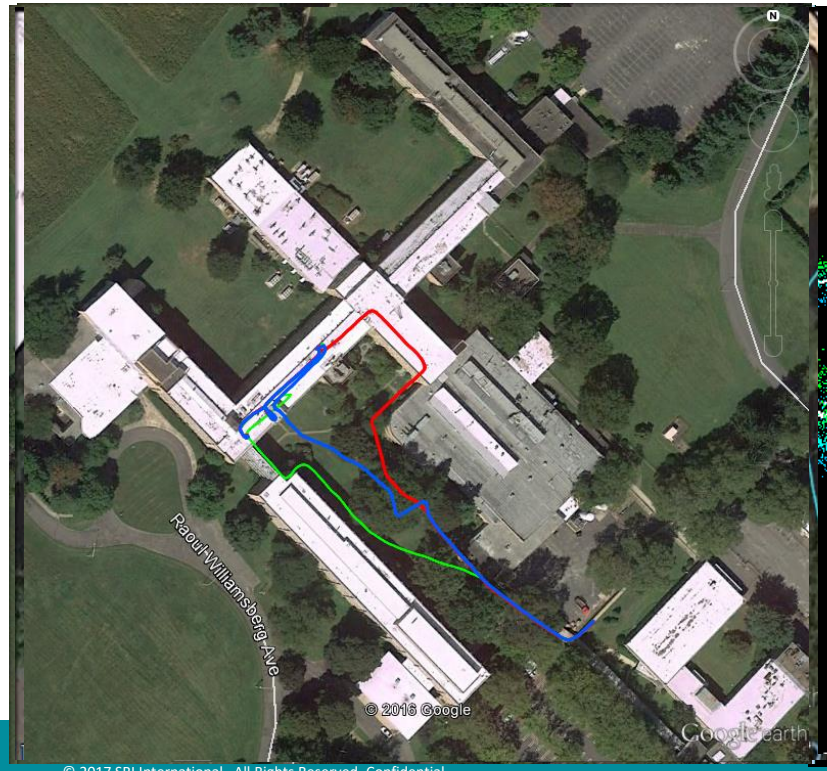
Dynamically created maps allows the system to ensure low drift errors in subsequent runs

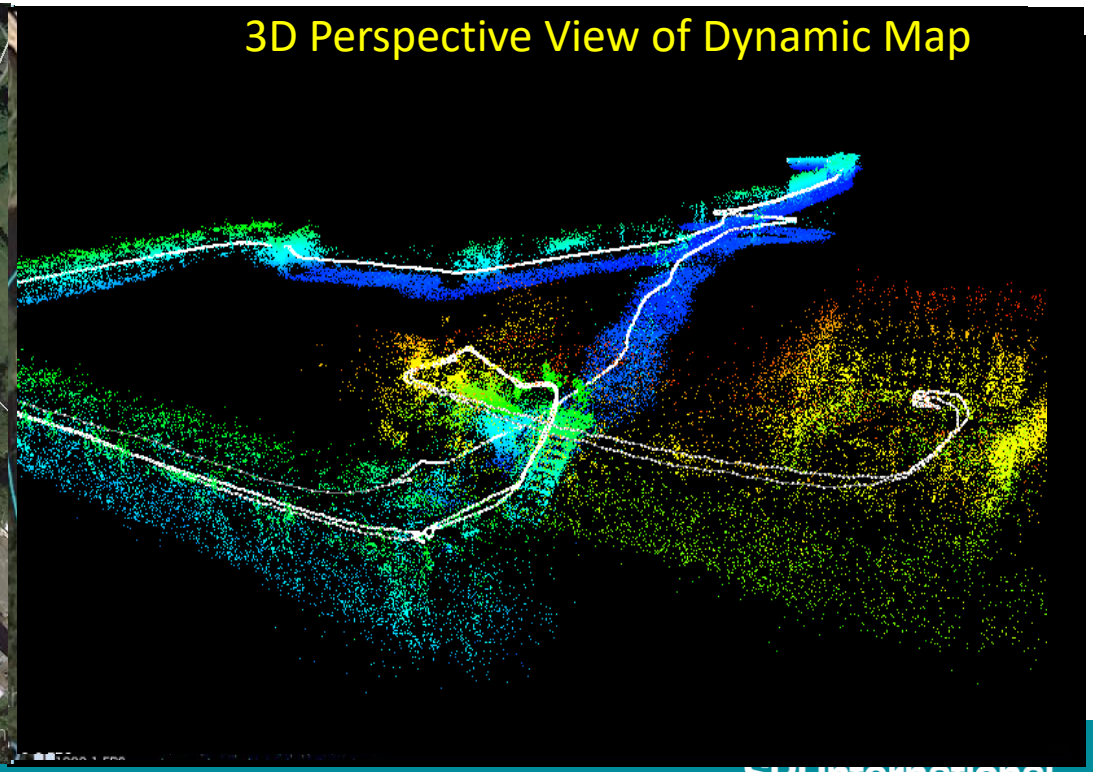
- **On the fly Creation of Visual Maps**
- **Match to Dynamic Maps to reset Drift**
- **Create Maps Across Multiple Platforms or Runs**



Navigation sensors: IMU, Cameras, GPS, Magnetometer, Barometer, Ranging Radios

SRI CAMSLAM

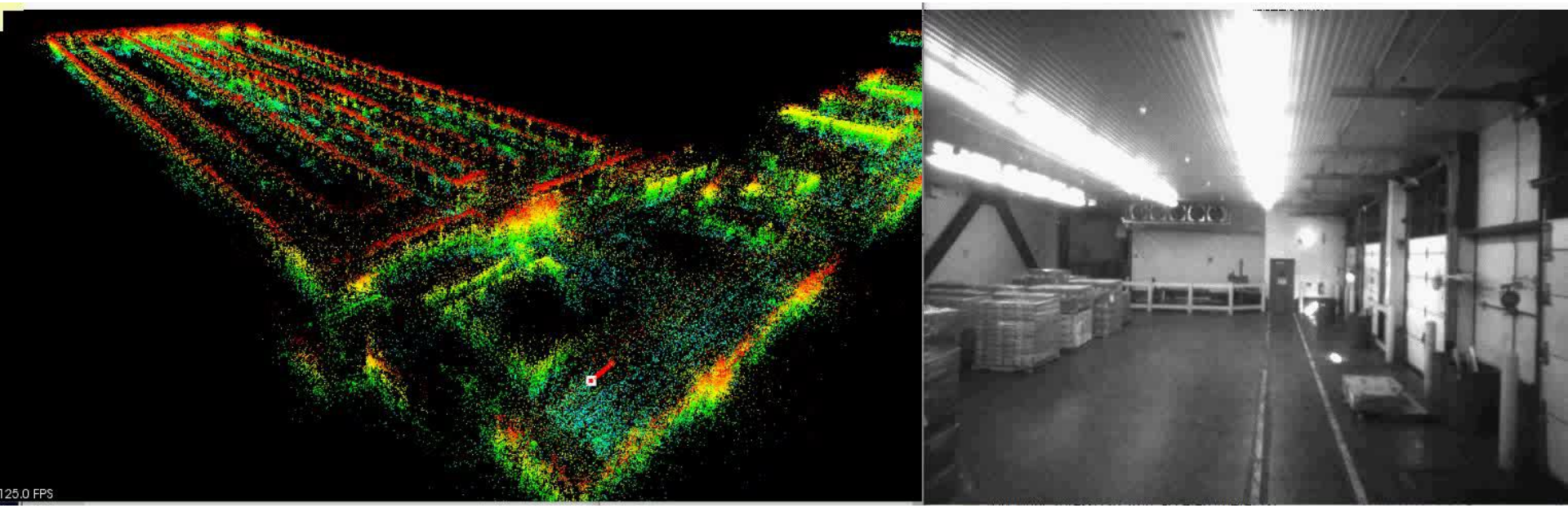




© 2017 SRI International. All Rights Reserved. Confidential

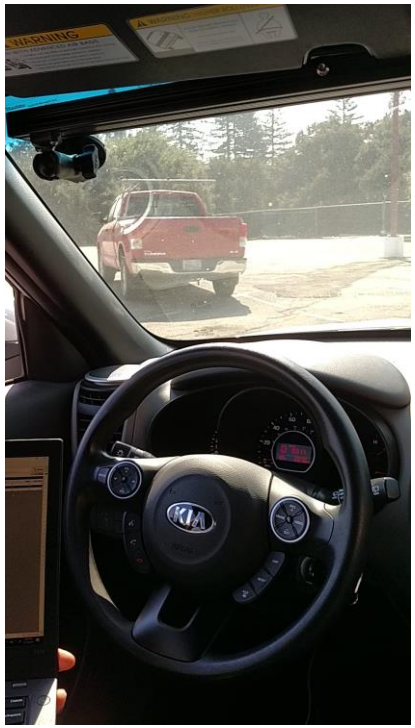
Robot Localization in Warehouse Environment

SRI CAMSLAM



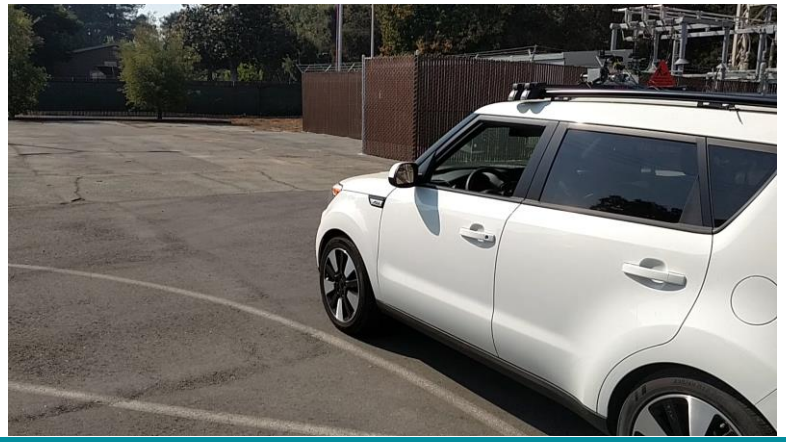
© 2017 SRI International. All Rights Reserved. Confidential

Robotic Applications

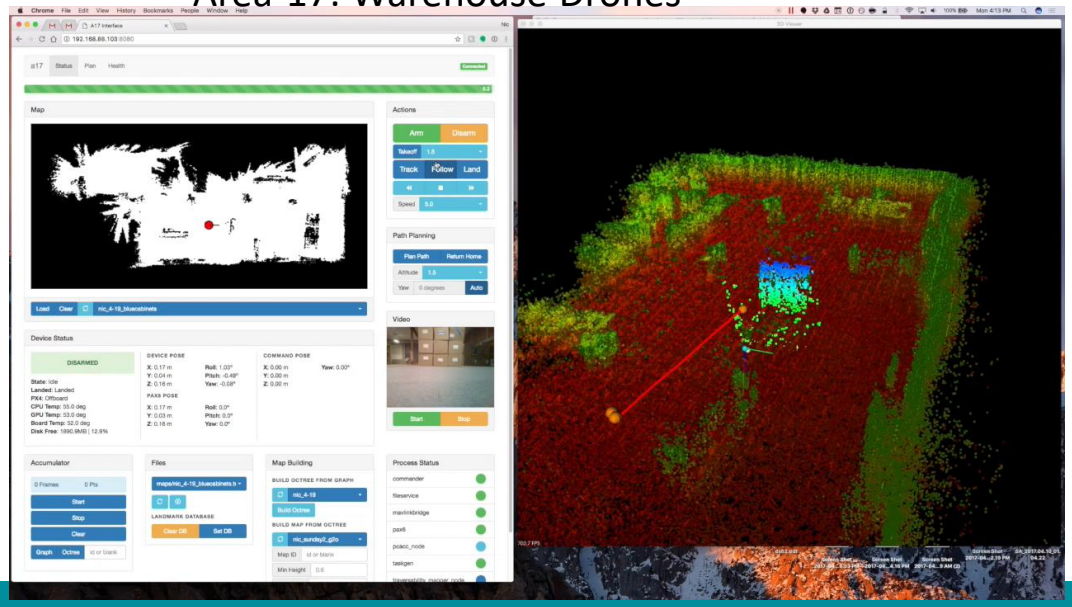


NIOSH Visual-Inertial Drone Navigation - Recorded Live Demo Video (8x speed)

Scotty Labs (bought by DoorDash): Autonomous Cars



Area-17: Warehouse Drones



© 2017 SRI International. All Rights Reserved. Confidential

Visual Navigation

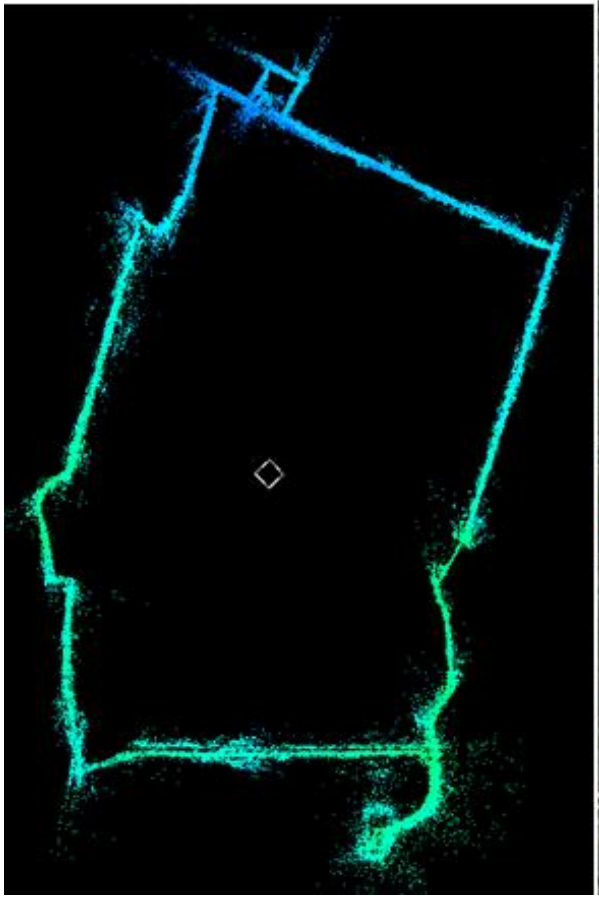
- Existing large-scale navigation solutions:
  - High-Precision Differential GPS with High-End IMU
  - Non-Differential GPS
  - Pre-Built Geo-Referenced Map Using LIDAR

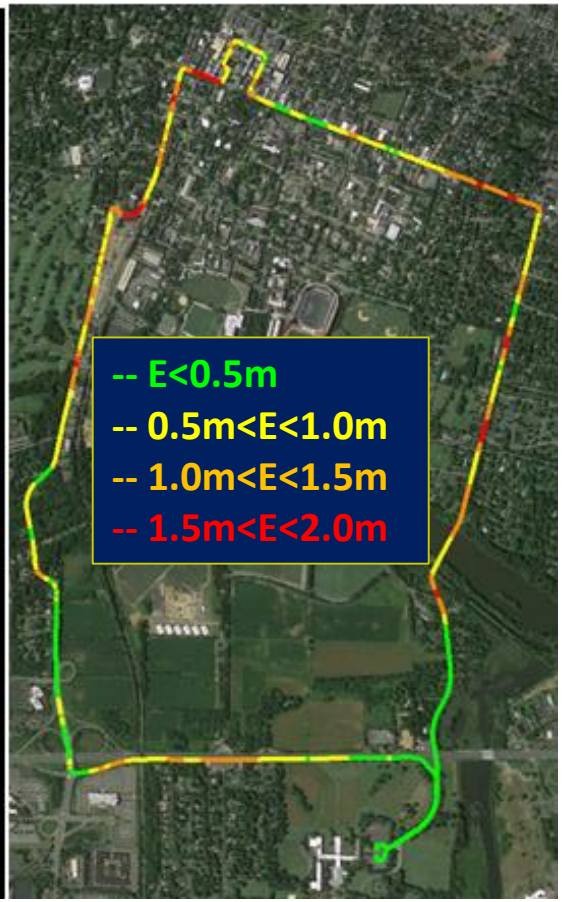
- Limitation to visual navigation:
  - The real world is very dynamic with lots of moving platforms and continuous scene changes.
  - Maps need to be continuously updated.
  - System needs to be adaptive to appearance changes across weather, time, and illumination variations.

© 2017 SRI International. All Rights Reserved. Confidential

Navigation Using Pre-Mapped Visual Landmarks

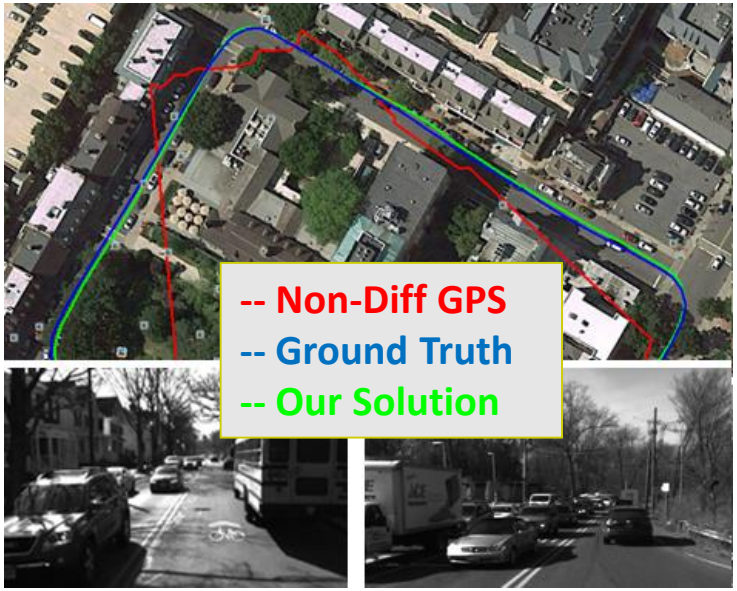
3D RMS Error (meter)	2015 Winter Partly Cloudy Noon	2016 Spring Cloudy Afternoon	2016 Spring Sunny Morning
Our Navigation Solution	**0.5378**	**0.9440**	**1.1200**





- Note we use INS solution (IMU, high-precision differential GPS) as ground truth.

Princeton City: 10.14km, 25 min



Han-Pang Chiu et al., **Sub-Meter Vehicle Navigation Using Efficient Pre-Mapped Visual Landmarks**, IEEE ITSC, 2016.

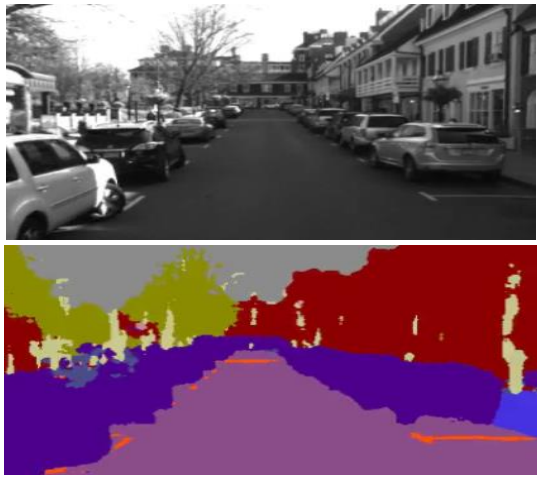
© 2017 SRI International. All Rights Reserved. Confidential

Advantage from Semantic Inference

A hybrid approach: *Develop and integrate both* ***semantic-inference*** *and metric-inference for navigation and mapping*

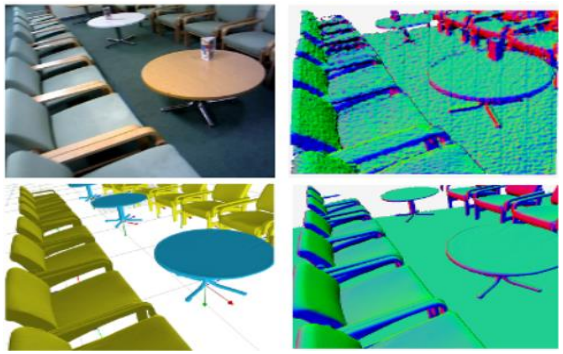
Semantic Inference

Accuracy



High-level objects are more robust to scene changes, and can be matched across time/space/platforms.

Efficiency



Sharing semantic information reduces bandwidth required for collaboration/ storage.

Applicability

It enables semantic scene representation and natural human-machine interaction for more applications.

© 2017 SRI International. All Rights Reserved. Confidential

Semantic Segmentation

- Our framework is designed to incorporate semantic information from any pre-trained networks that generate dense segmentation pixel labels on video frames.
  - We take the popular trained SegNet Encoder-Decoder Network as an example.

- We improve the efficiency of semantic segmentation while still maintaining its accuracy by converting the model into a low rank approximation of itself.
  - We extend the method* to apply both convolution and deconvolution layers.
  - The segmentation time is improved from 160 ms to 89 ms (almost 2x performance for >10Hz speed) to process one image on a single Nvidia K40 GPU.

CamVid DataSet	Global Accuracy	Class Average	Mean IOU	Sky	Building	Pole	Road	Tree	Vehicle	Sign	Fence	Pedestrian	Bike
SegNet	90.0207	72.097	60.15	90.43	93.72	56.72	95.88	93.04	89.66	56.24	61.54	98.52	86.23
Low Rank SegNet	86.965	74.292	56.03	93.68	79.90	43.74	95.00	92.96	81.01	61.31	56.19	87.22	74.47

* **Convolutional neural networks with low-rank regularization**, C. Tai et al., *Intl. Conf. on Learning Representations, 2016.*

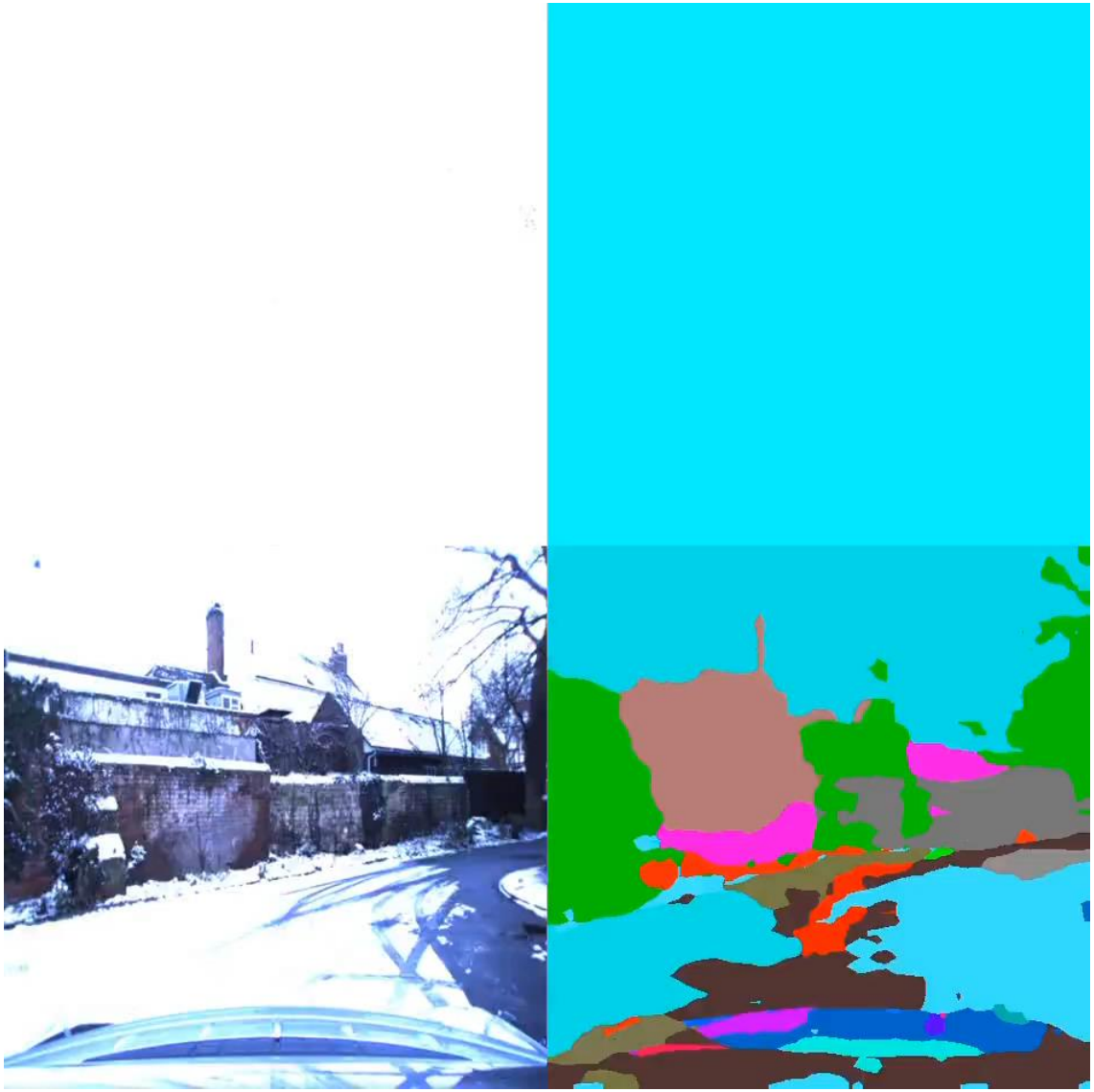
© 2017 SRI International. All Rights Reserved. Confidential

SRI International®

14

Semantic Visual Landmarks

- Using mapped landmark from permanent semantic classes is able to **<u>improve overall navigation accuracy between 8%~20% while reducing 30%~50% of storage.</u>**

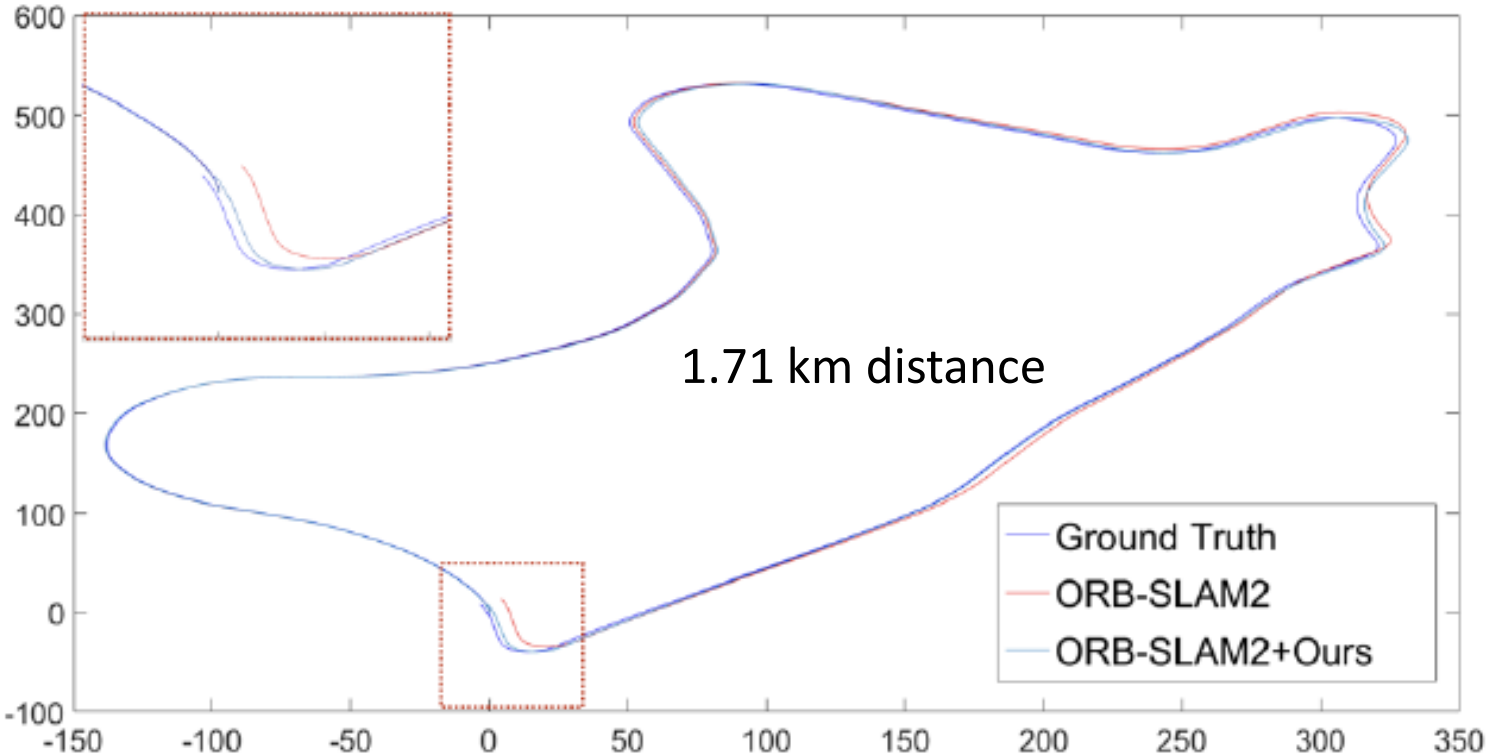


Varun Murali, Han-Pang Chiu, Supun Samarasekera, and Rakesh "Teddy" Kumar, **Utilizing Semantic Visual Landmarks for Precise Vehicle Navigation**, *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2017.

© 2017 SRI International. All Rights Reserved. Confidential

Improved Feature Tracking for Navigation

- Using consistent feature tracks from static semantic classes reduces 0.1~0.5% drift rate using ORB-SLAM2 over 11 KITTI sequences.
- The following example shows final error is reduced from 12.64 meters to 3.93 meters over 1.71 km.



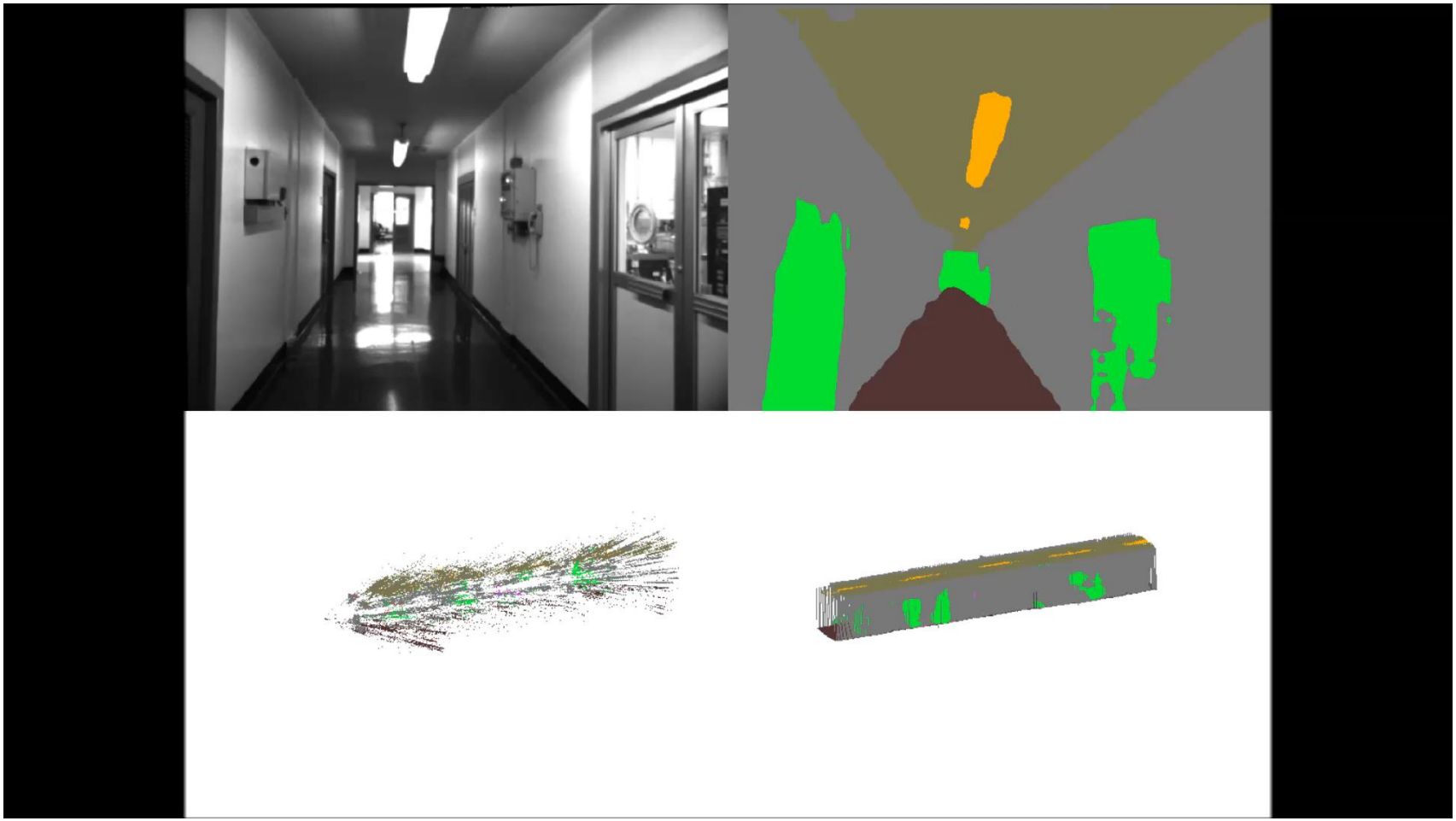
* **Utilize Semantic Visual Landmarks for Precise Vehicle Navigation**, V. Murali, H. Chiu, S. Samarasekera, R. Kumar, *ITSC, 2017.*

© 2017 SRI International. All Rights Reserved. Confidential

SRI International®

Real-Time Semantic Mapping and Modeling

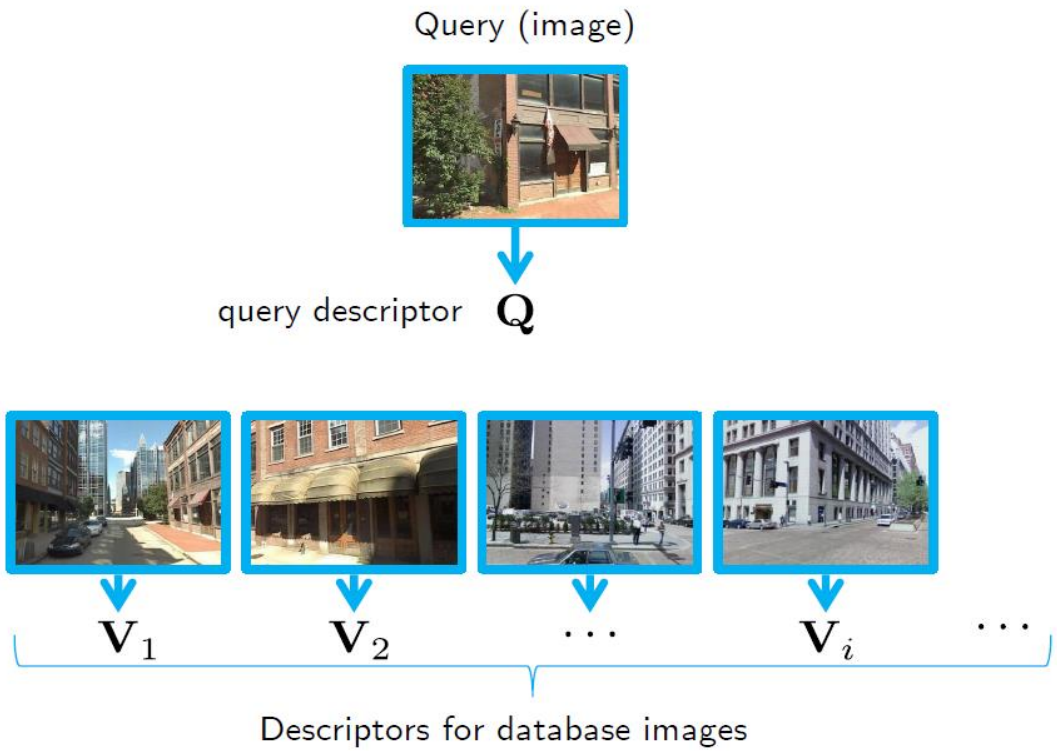
Fuse semantic labels for 3D dynamic mapping and modeling in new environments.



© 2017 SRI International. All Rights Reserved. Confidential

Revisit – Search Visual Map Database

Matching query images to geo-referenced databases across appearance changes, viewpoint changes, and scene changes.



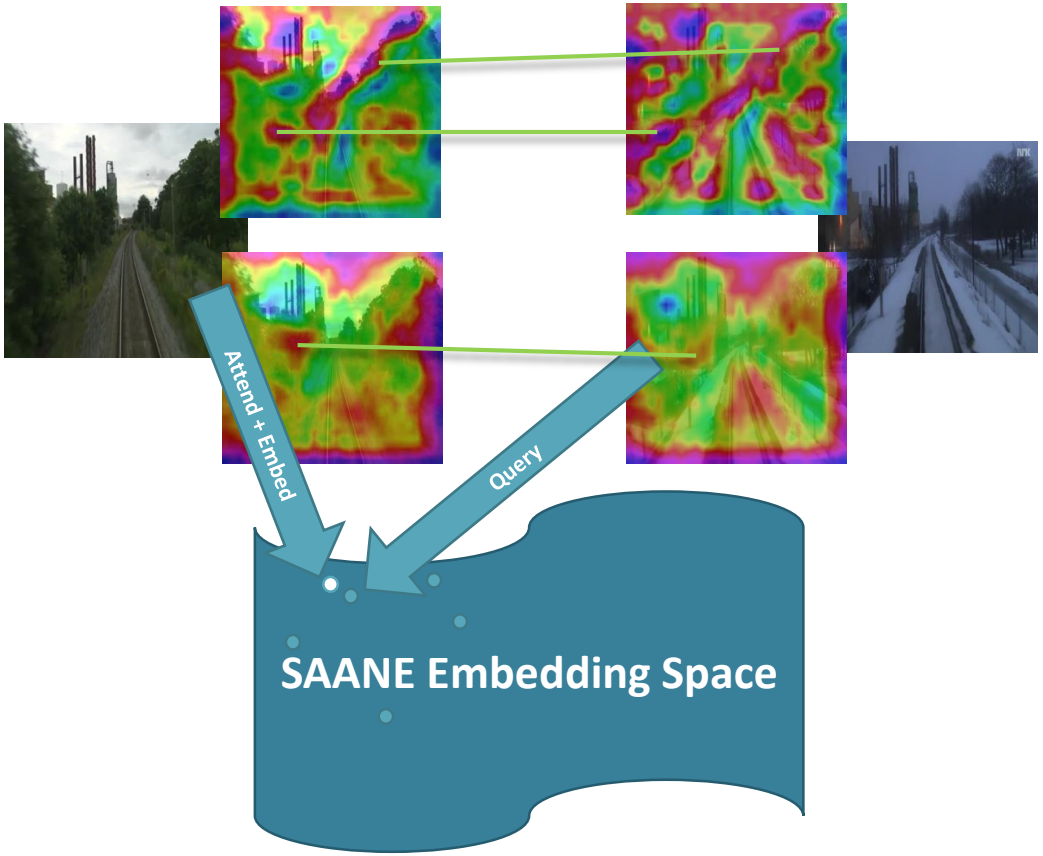
$$i^\star = \arg\max_{i \in 1 \ldots N} \; s(\mathbf{Q}, \mathbf{V}_i)$$

Nearest Neighbor search

© 2017 SRI International. All Rights Reserved. Confidential

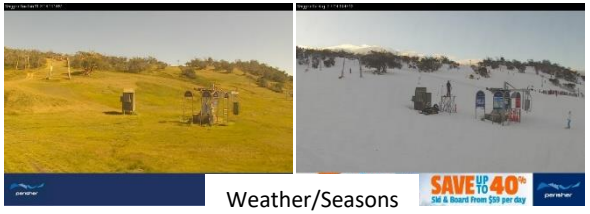
Image-Based Visual Localization

- **Goal:** Estimate position for a given monocular **query image** by matching to a database of **images of known locations,** in the presence of large appearance changes across weather and illumination variations.
- We propose to use embedding for this problem: A deep-learned compact Euclidean space where distances directly correspond to a measure of data similarity.




Day/Night


Day/Night


Weather/Seasons

Zachary Seymour, Karan Sikka, Han-Pang Chiu, Supun Samarasekera, Rakesh Kumar, **Semantically-Aware Attentive Neural Embeddings for Image-based Visual Localization**. British Machine Vision Conference (BMVC), 2019.

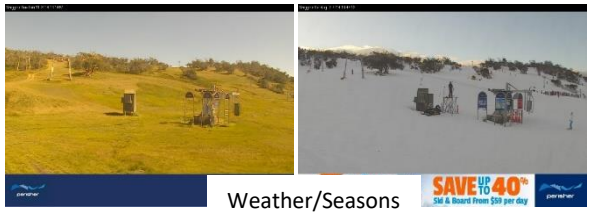
© 2017 SRI International. All Rights Reserved. Confidential

Semantically-Aware Visual Localization

- Estimate position for a given monocular **query image** by matching to a database of **images of known locations.**

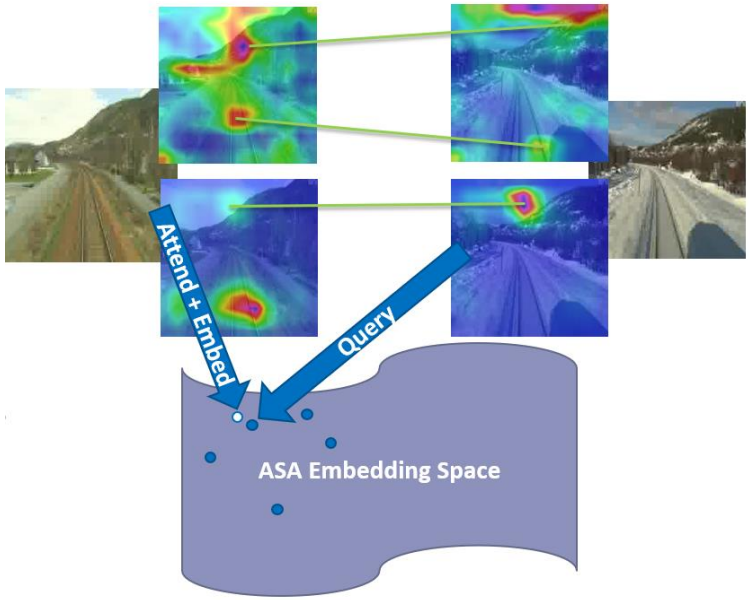
Day/Night

Day/Night



Weather/Seasons

- **<u>Semantic-Aware:</u>** Our deep network model incorporates pixel-wise semantic features in learning the image embeddings.
- **<u>Attention-Based:</u>** We train self-attention modules to encourage our model to focus on semantically consistent spatial regions.
- Training data: ~2 million images collected from 2,685 static webcams.
- **<u>Our accuracy outperforms STOA methods by 19%.</u>**

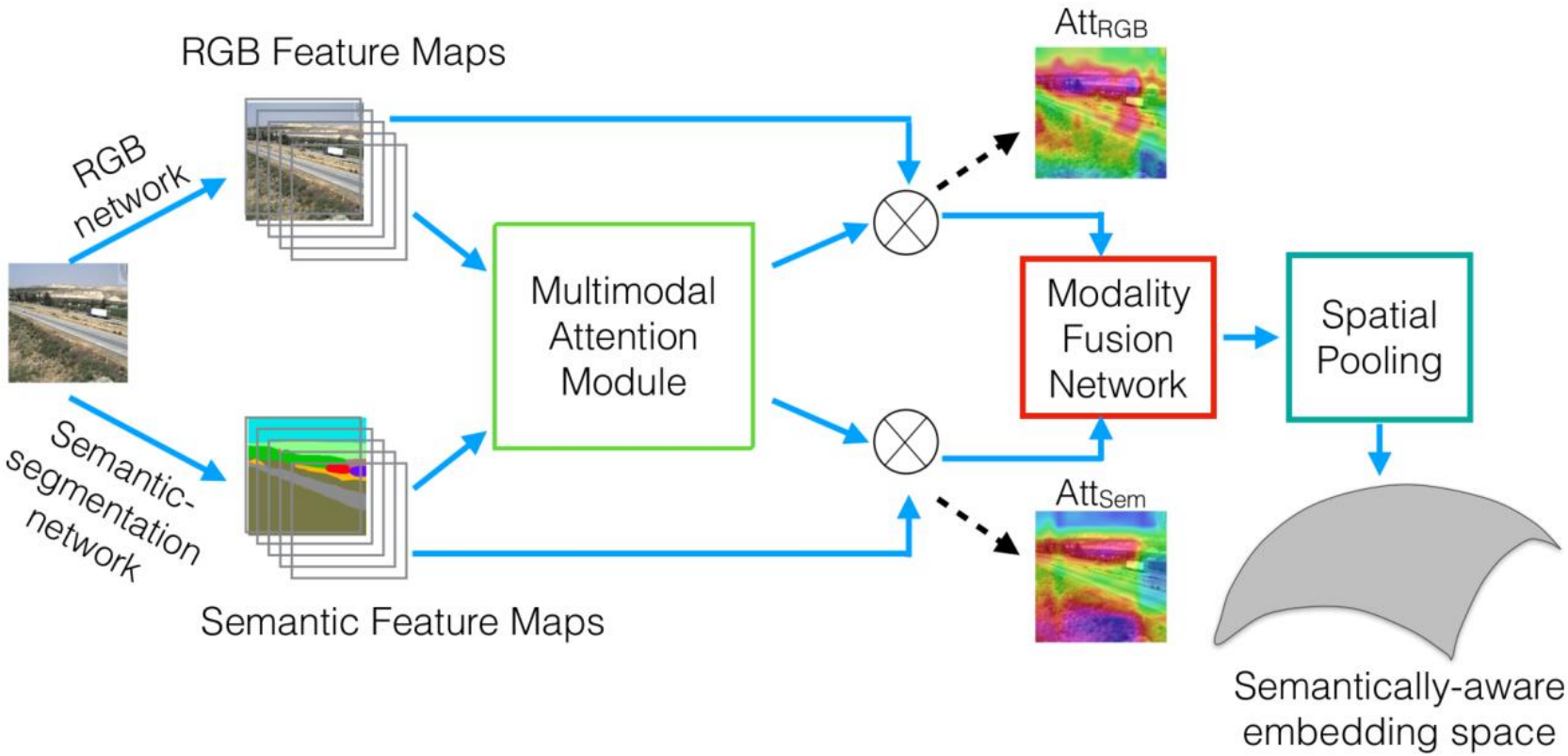


Zachary Seymour, Karan Sikka, Han-Pang Chiu, Supun Samarasekera, Rakesh Kumar, **Semantically-Aware Attentive Neural Embeddings for Image-based Visual Localization**. British Machine Vision Conference (BMVC), 2019.

© 2017 SRI International. All Rights Reserved. Confidential

SAANE: Semantically-Aware Attentive Neural Embeddings

- **SAANE:** Modeling a joint appearance-semantic image embedding space.
- **Training:** Collect ~2,000,000 images from 2,079 webcams taken at multiple times of year.
- **Test:** Several public navigation datasets, wholly distinct from the training data.
- **Evaluation:** Area Under PR Curve: A curve of precision % across different recall % for all data.



Zachary Seymour, Karan Sikka, Han-Pang Chiu, Supun Samarasekera, Rakesh Kumar, **Semantically-Aware Attentive Neural Embeddings for Image-based Visual Localization**. British Machine Vision Conference (BMVC), 2019.

© 2017 SRI International. All Rights Reserved. Confidential

Evaluation Results

- Our baseline achieves comparable results to state-of-the-art methods.
- Adding semantic cue helps challenging tasks, such as 25% at Day->Night.
- Multi-modal attention gives best performance.

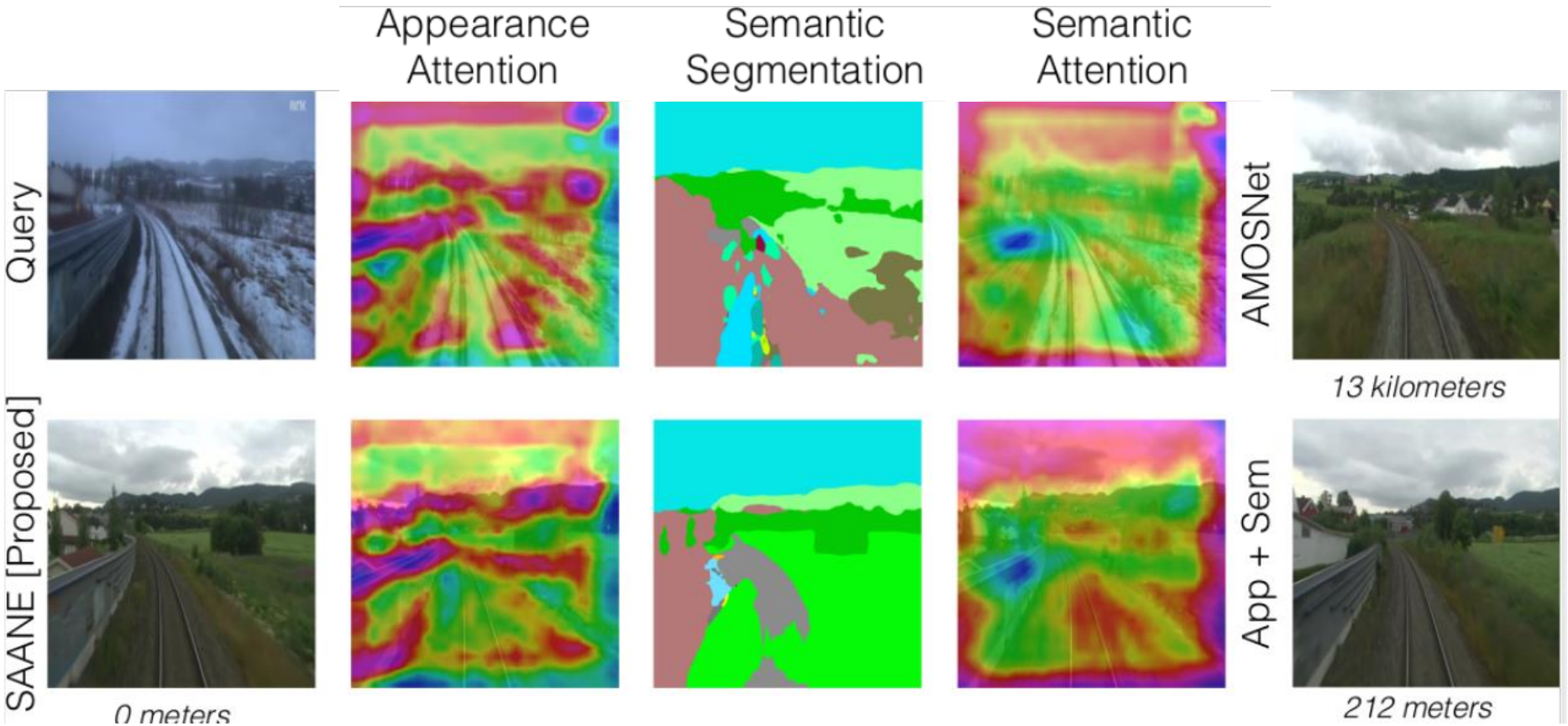
Area Under PR Curve: A curve of precision % across different recall % for all data.

Model	NordLand: Sum.->Win. 51 km	St. Lucia 17 km	RC: Sum.->Win. 2 km	RC: Day->Night
Appearance (Our Baseline)	58.0	71.4	79.6	36.6
DenseVLAD*	19.2	64.5	78.6	22.9
AMOSNet*	45.6	75.3	75.1	45.3
Appearance + Semantic	74.6	74.9	87.9	61.4
Appearance w/ individual Attention	74.2	74.0	86.5	60.9
Appearance + Semantic w/ Individual Attention	68.5	73.9	83.6	62.9
SAANE (Appearance + Semantic w/ Multi-modal Attention)	**77.3**	**78.3**	**88.5**	**67.3**

*DenseVLAD and AMOSNet are state-of-the-art 2D image-based methods for geo-localization and database retrieval.

© 2017 SRI International. All Rights Reserved. Confidential

Visualization of Localization Results - NordLand



© 2017 SRI International. All Rights Reserved. Confidential

Localization Across Seasons

- 2km database, Accuracy can be further improved by position prior, sequential verification, and 2D-3D refinements.

© 2017 SRI International. All Rights Reserved. Confidential

Localization Across Day & Night

- 2km database, Accuracy can be further improved by position prior, sequential verification, and 2D-3D refinements.

© 2017 SRI International. All Rights Reserved. Confidential

Refining with Location Priors

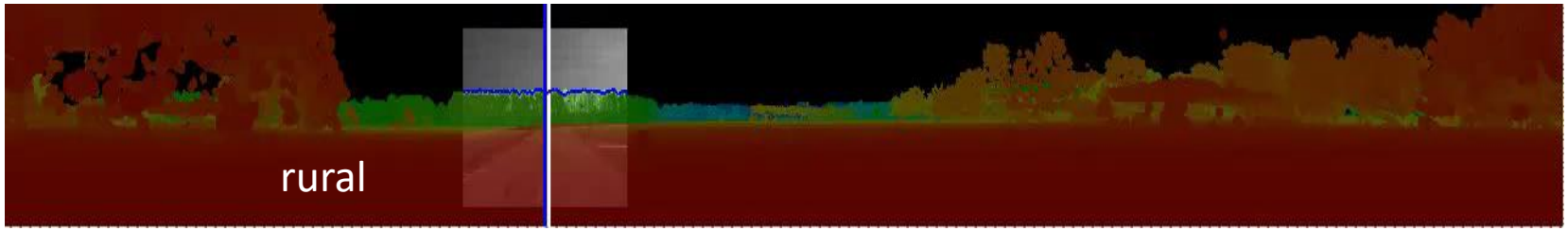
- **Recall at K (R@K)-** Percentage of test samples for which correct result is ranked within top-K retrieved results to the query. **Median Rank (MedR)-** Median of ground-truth matches in the ranking.

	Nordland (51 km)			RobotCar (2 km)		
	R@1	R@10	MedR	R@1	R@10	MedR
No prior	47.08	85.16	11	61.84	81.27	25
5 km	49.79	91.58	4	-	-	-
1 km	52.14	93.79	3	62.38	81.52	24
500 m	52.92	87.66	3	64.38	86.52	13
100 m	56.13	97.93	2	68.66	92.38	4
50 m	63.34	99.22	1	69.91	94.84	3
10 m	95.93	100.	1	78.19	100.	1
5 m	100.	100.	1	100.	100.	1

© 2017 SRI International. All Rights Reserved. Confidential

Semantic Geo-Registration

We perform 2D-3D geo-registration continuously between the input video frame and the matched geo-registration 2D-3D database (from cameras or LIDAR).



Han-Pang Chiu et al., **Augmented Reality Driving Using Semantic Geo-Registration**, *IEEE International Conference on Virtual Reality (VR)*, 2018.

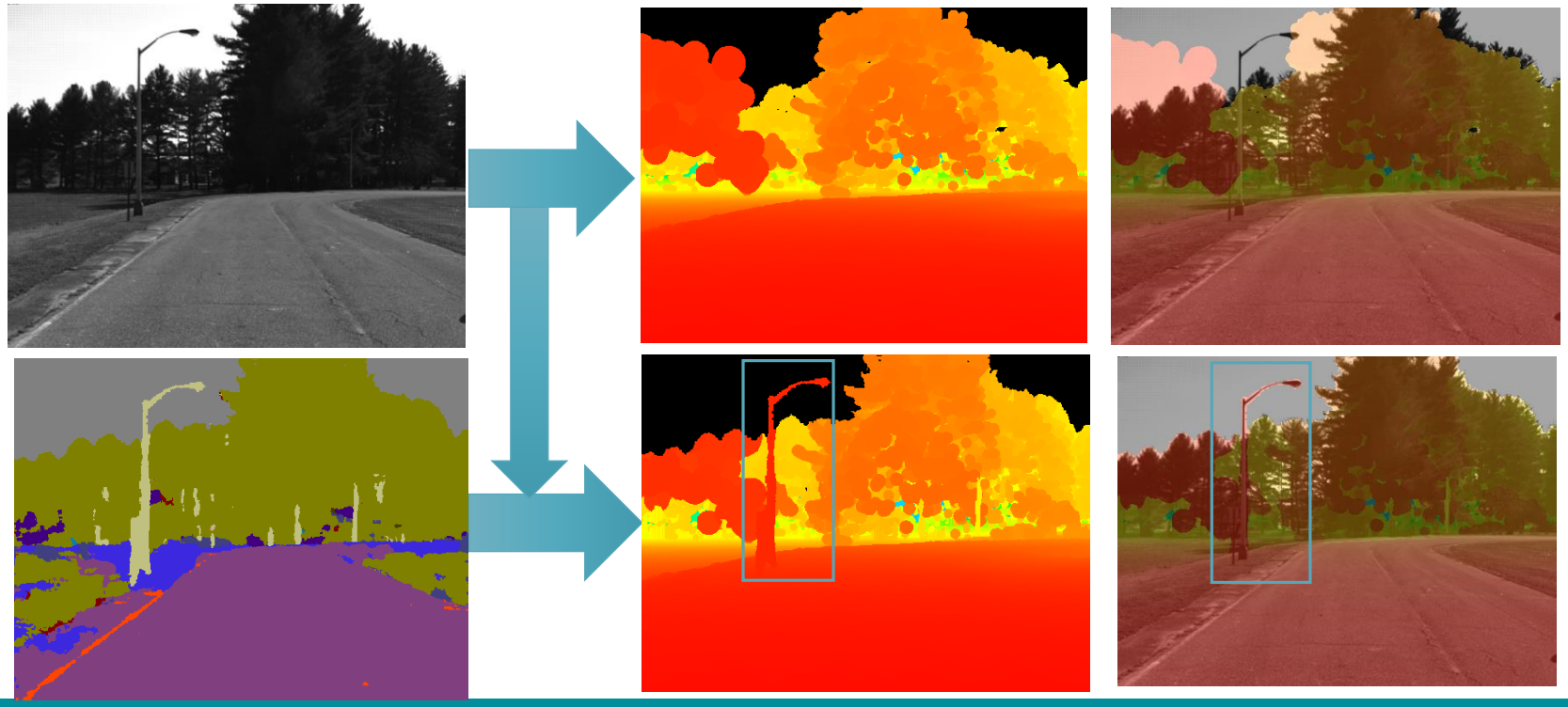
© 2017 SRI International. All Rights Reserved. Confidential

3D Object Hypothesis Verification

- 3D ground plane can be fitted based on depth map by propagating road labels.
- 3D information of objects are estimated based on verified 3D class hypothesizes from segmented regions on 3D ground plane.

$$p(t=true|m,G) \propto p(t=true|d) = N(d;0,\sigma_d), \quad r \simeq \frac{c*f}{I}, \quad o \simeq \frac{r}{\sqrt{(\frac{u}{f})^2+(\frac{v}{f})^2+1}} \begin{pmatrix} \frac{u}{f} \\ \frac{v}{f} \\ 1 \end{pmatrix} \quad d = O^T n + h,$$

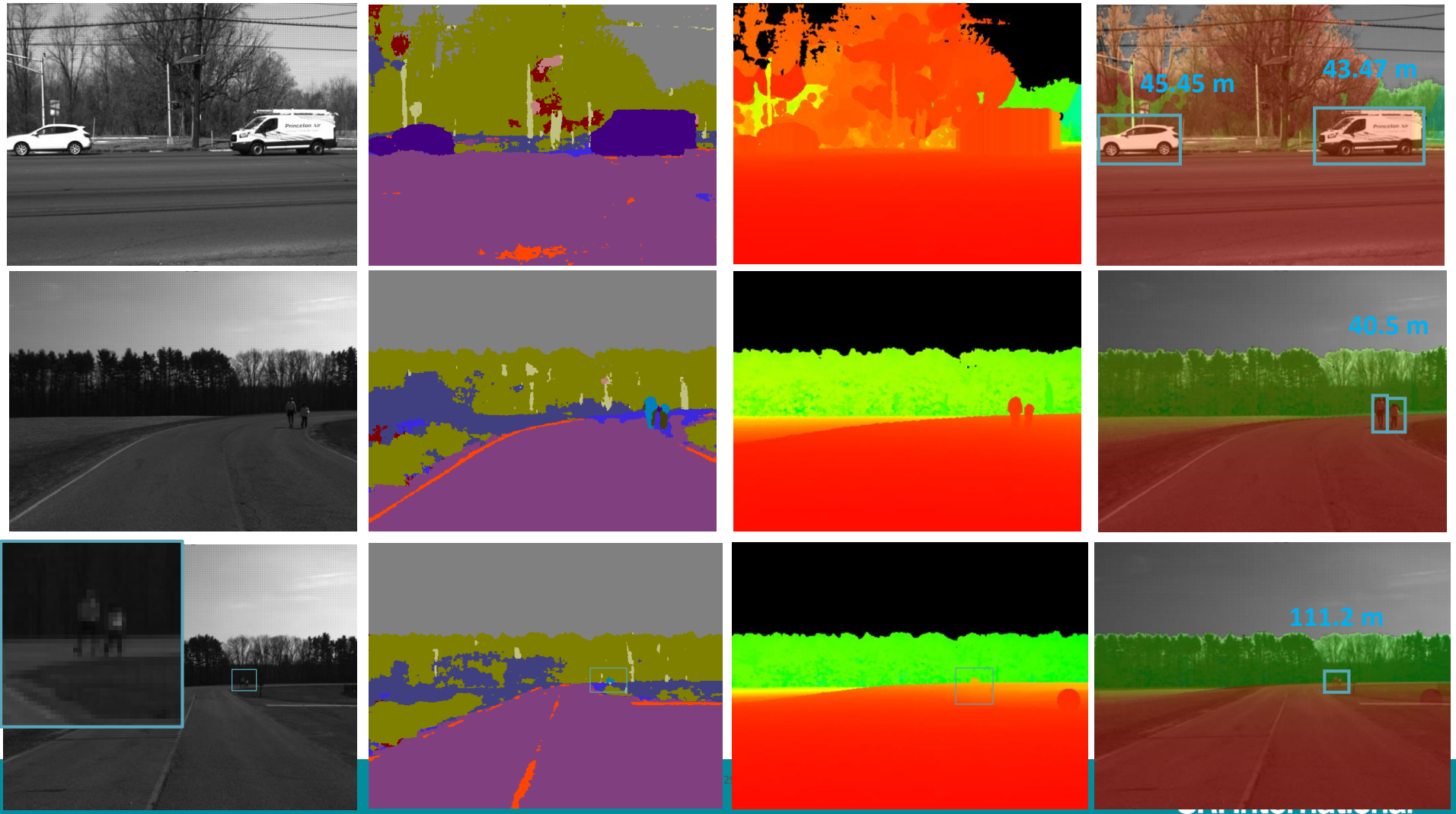
d: 3D dist. of obj. to ground, I: 2d obj. height, O: 3d obj. position, n: ground normal, h: cam. height



© 2017 SRI International. All Rights Reserved. Confidential

Depth Estimation and Occlusion Reasoning for Distant Objects

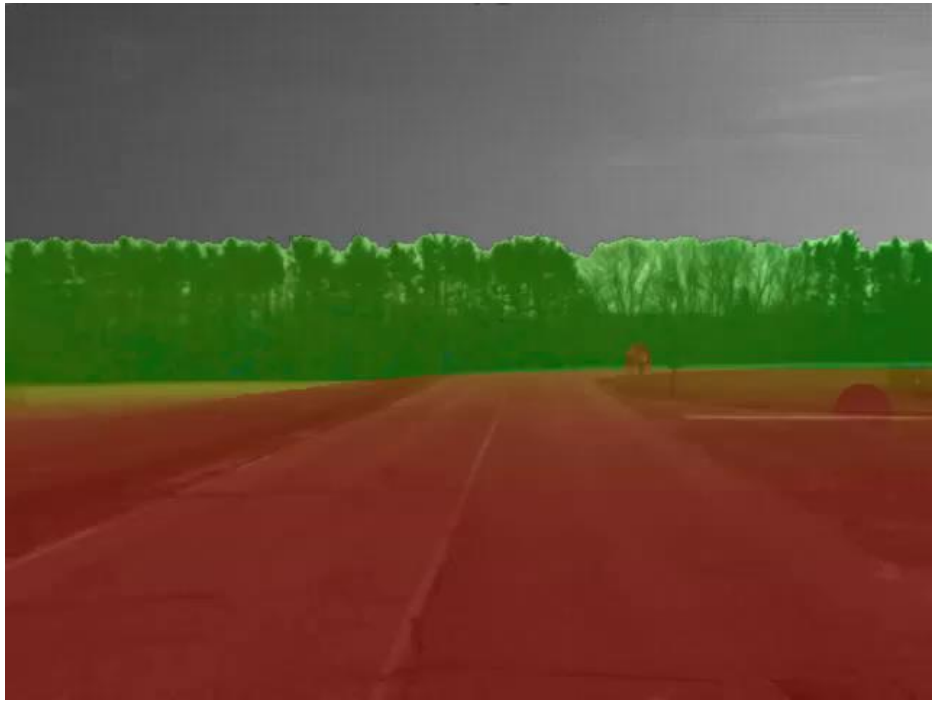
- Depth information for distant objects can be estimated (up to 120 meters based on minimum detection). This method can be used by propagating on-the-fly nearby stereo depth maps or prior depth maps.



Semantic Depth Reasoning for Static Background and Dynamic Objects

Input: Monocular Video

Output: Overlaid Depth Video



Han-Pang Chiu et al., **Augmented Reality Driving Using Semantic Geo-Registration**, *IEEE International Conference on Virtual Reality (VR)*, 2018.

© 2017 SRI International. All Rights Reserved. Confidential

Demonstration: Augmented Reality on Training Field

Han-Pang Chiu et al., **Augmented Reality Driving Using Semantic Geo-Registration**, *IEEE International Conference on Virtual Reality (VR)*, 2018.

© 2017 SRI International. All Rights Reserved. Confidential

Computation Hardware: NVIDIA JETSON TX2

- Demonstrated GPS-denied navigation
- Past Performance:
  - DARPA ASPN, Army AR-Stryker:
    - plug-and-play sensing
    - visual tracking and dynamic mapping
  - ONR AITT & WAR, DARPA Squad-X:
    - semantic geo-registration
  - MCWL, DARPA ASPN, DARPA Squad-X
    - collaborative navigation through ranging radios

Current Navigation Hardware

© 2017 SRI International. All Rights Reserved. Confidential