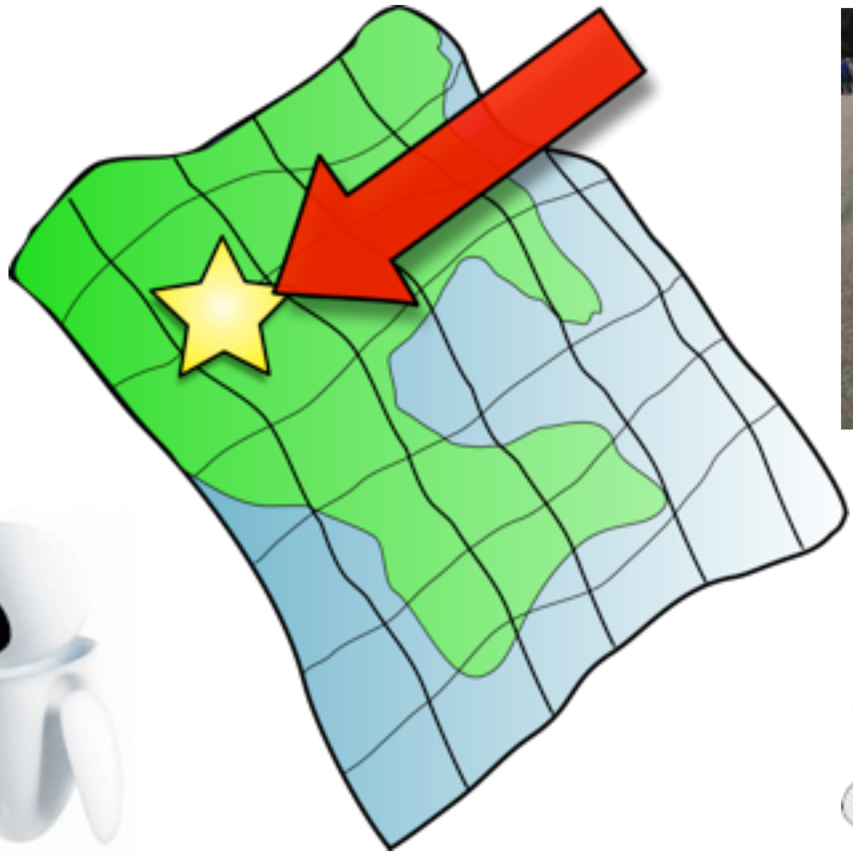
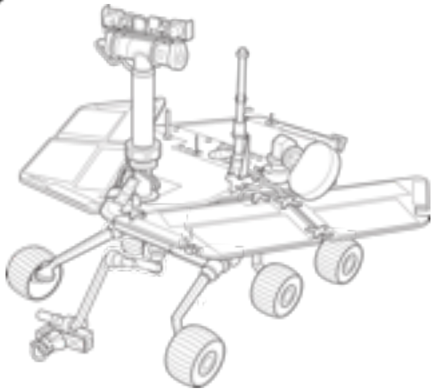
Lost! Leveraging the Crowd for Probabilistic Visual Self-Localization

Marcus A. Brubaker (Toyota Technological Institute at Chicago)
Andreas Geiger (Karlsruhe Institute of Technology & MPI Tübingen)
Raquel Urtasun (Toyota Technological Institute at Chicago)

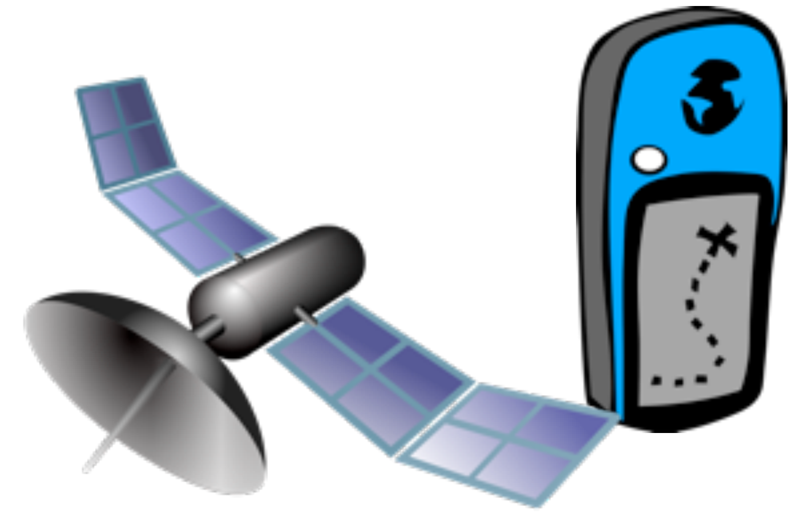
Introduction





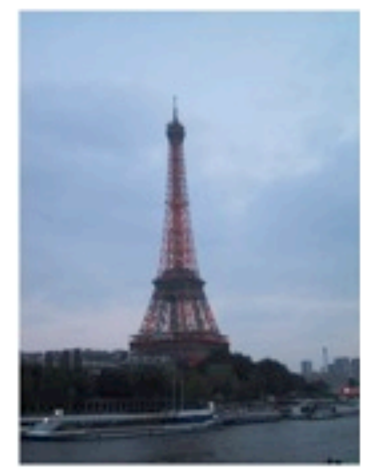
- Localization is crucial for autonomous systems

Introduction



- GPS has limitations in terms of reliability and availability

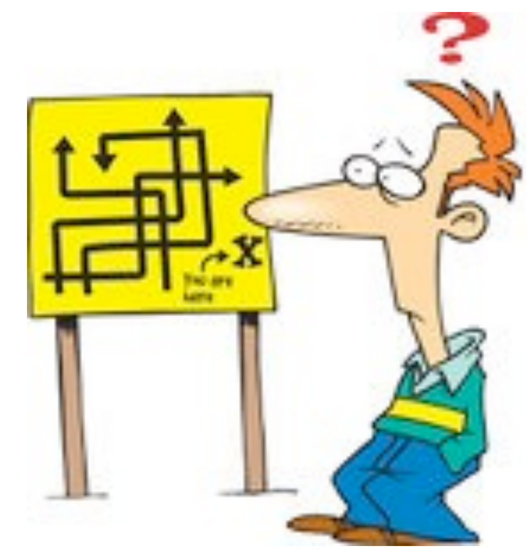
- Place recognition techniques use image features and a database of previously collected images
  - [Dellaert et al, ICRA 1999; Thrun et al, AI 2001; Hays and Efros, CVPR 2008; Schindler et al, CVPR 2008; Crandall et al, WWW 2009; Kalogerakis et al, ICCV 2009]



- We develop an inexpensive technique for localizing to ~**3m** in **unseen regions**

Introduction

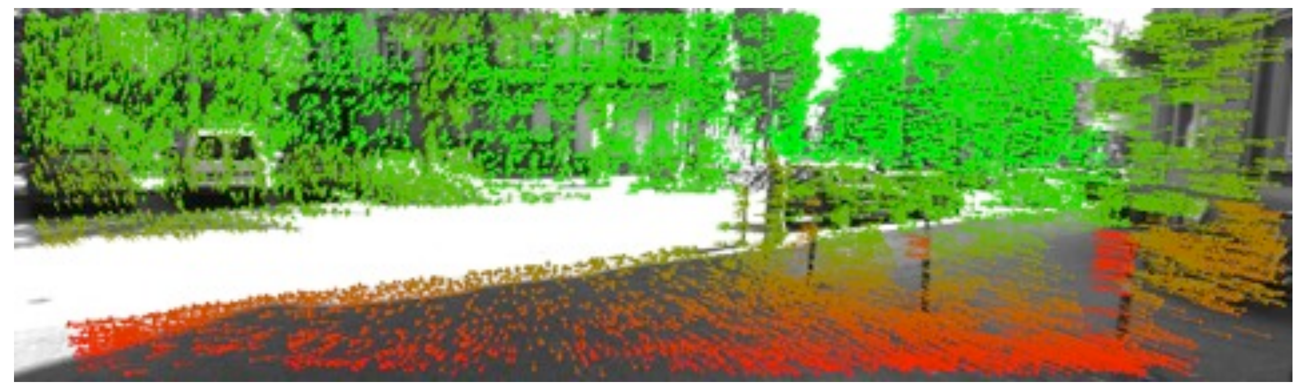
- Humans are able to use a map, combined with visual input and exploration, to localize effectively
- Detailed, community developed maps are freely available (OpenStreetMap)



How can we exploit maps, combined with visual cues, to localize a vehicle?

Probabilistic Localization using Visual Odometry

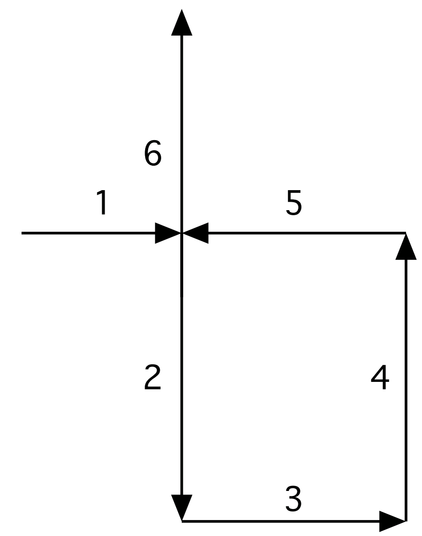
- Visual odometry provides a strong source of information for localization
- Visual odometry has some issues
  - Over short time periods it can be noisy and highly ambiguous
  - Over long time periods it drifts when integrated
- We adopt a probabilistic approach to represent and maintain this uncertainty

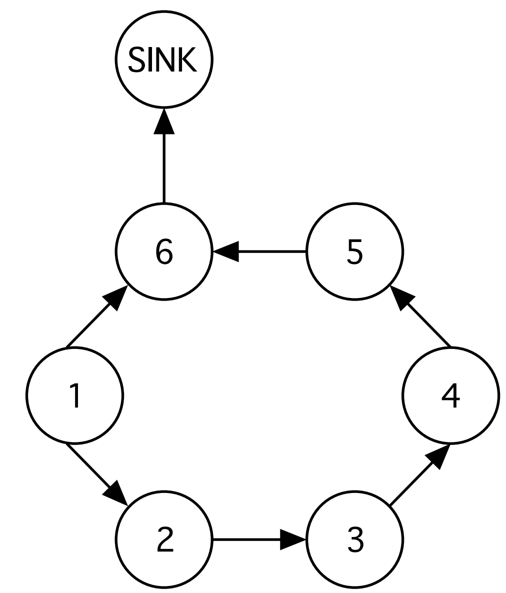


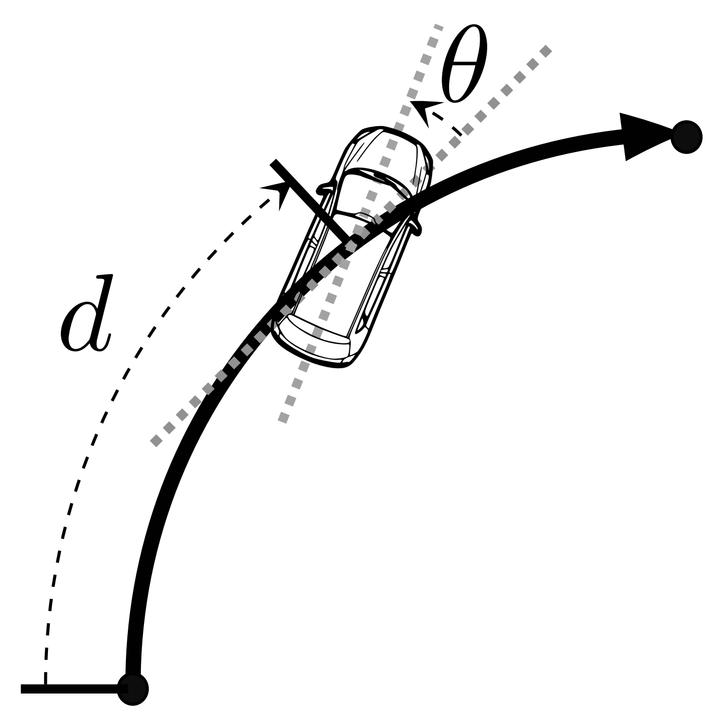
[Geiger et al, IV 2011]

Probabilistic Localization using Visual Odometry

- Maps can be considered as a graph
  - Nodes of the graph represent street segments
  - Edges represent intersections and allowed transitions between these segments
- Position is defined by the current street and the distance travelled, d, and orientation, θ, on that street

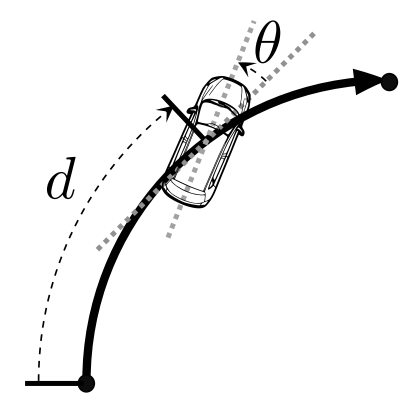






Probabilistic Localization using Visual Odometry

- The complete state includes
  - the current street segment u_t, and
  - the current and previous position and orientation on the street segment, $\mathbf{s}_t = (d_t, \theta_t, d_{t-1}, \theta_{t-1})$
- Odometry observations $\mathbf{y}_{1:t} = (\mathbf{y}_1, \ldots, \mathbf{y}_t)$
- Localization is formulated as posterior inference



$$p(u_t, \mathbf{s}_t | \mathbf{y}_{1:t})$$

$$\propto \underbrace{p(\mathbf{y}_t | u_t, \mathbf{s}_t)}_{\text{Likelihood}} \sum_{u_{t-1}} \int \overbrace{\underbrace{p(u_t | u_{t-1}, \mathbf{s}_{t-1})}_{\text{Street Segment Transition}} \underbrace{p(\mathbf{s}_t | u_t, u_{t-1}, \mathbf{s}_{t-1})}_{\text{Pose Transition}}}^{\text{State Transition}} \overbrace{p(u_{t-1}, \mathbf{s}_{t-1} | \mathbf{y}_{1:t-1})}^{\text{Previous Posterior}} d\mathbf{s}_{t-1}$$

Probabilistic Localization using Visual Odometry

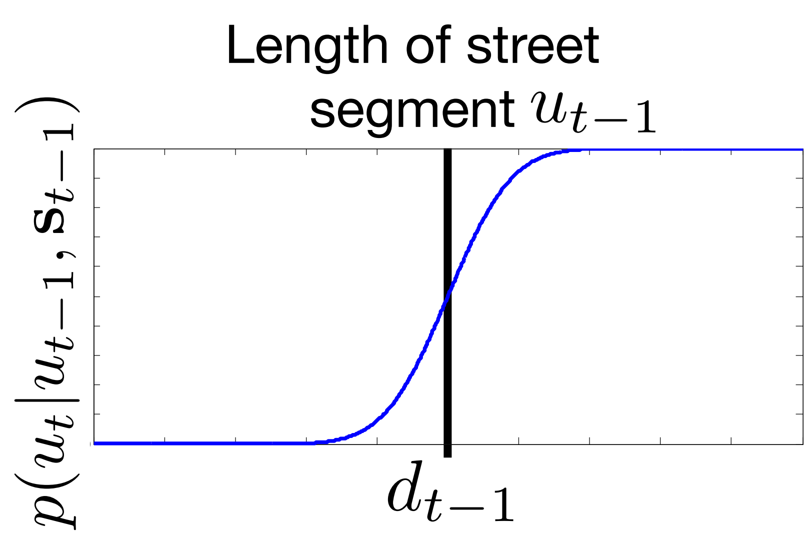
- Likelihood: $p(\mathbf{y}_t | u_t, \mathbf{s}_t)$

$$\mathbf{y}_t = \mathbf{M}\mathbf{s}_t + \eta \qquad \eta \sim \mathcal{N}(0, \Sigma_{\mathbf{y}})$$

- Pose transition: $p(\mathbf{s}_t | u_t, u_{t-1}, \mathbf{s}_{t-1})$

$$\mathbf{s}_t = \mathbf{A}\mathbf{s}_{t-1} + \mathbf{b} + \zeta \qquad \zeta \sim \mathcal{N}(0, \Sigma_{\mathbf{s}})$$

- Street segment transition:



- Parameters (e.g., variances) estimated from data
- Model is nearly Gauss-Linear which we exploit to derive a custom inference algorithm

Probabilistic Localization using Visual Odometry

- To represent the posterior we factorize it

Discrete distribution over street segments

$$p(u_t, \mathbf{s}_t | \mathbf{y}_{1:t}) = p(\mathbf{s}_t | u_t, \mathbf{y}_{1:t}) p(u_t | \mathbf{y}_{1:t})$$

Posterior over pose, given the street segment

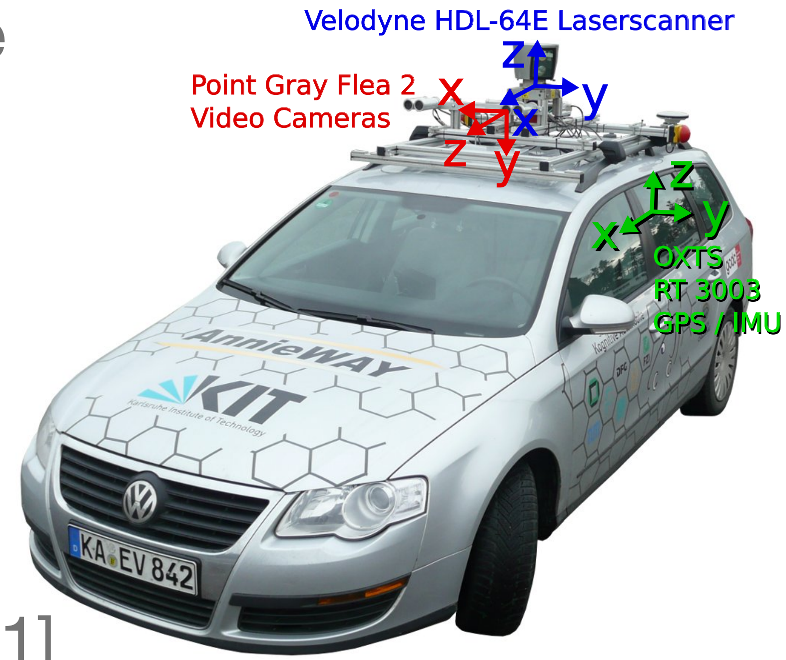
- The posterior over pose is represented with a Mixture of Gaussians for each street segment

$$p(\mathbf{s}_t | u_t, \mathbf{y}_{1:t}) = \sum_{i=1}^{N_{u_t}} \pi_{u_t}^{(i)} \mathcal{N}\left(\mathbf{s}_t | \mu_{u_t}^{(i)}, \Sigma_{u_t}^{(i)}\right)$$

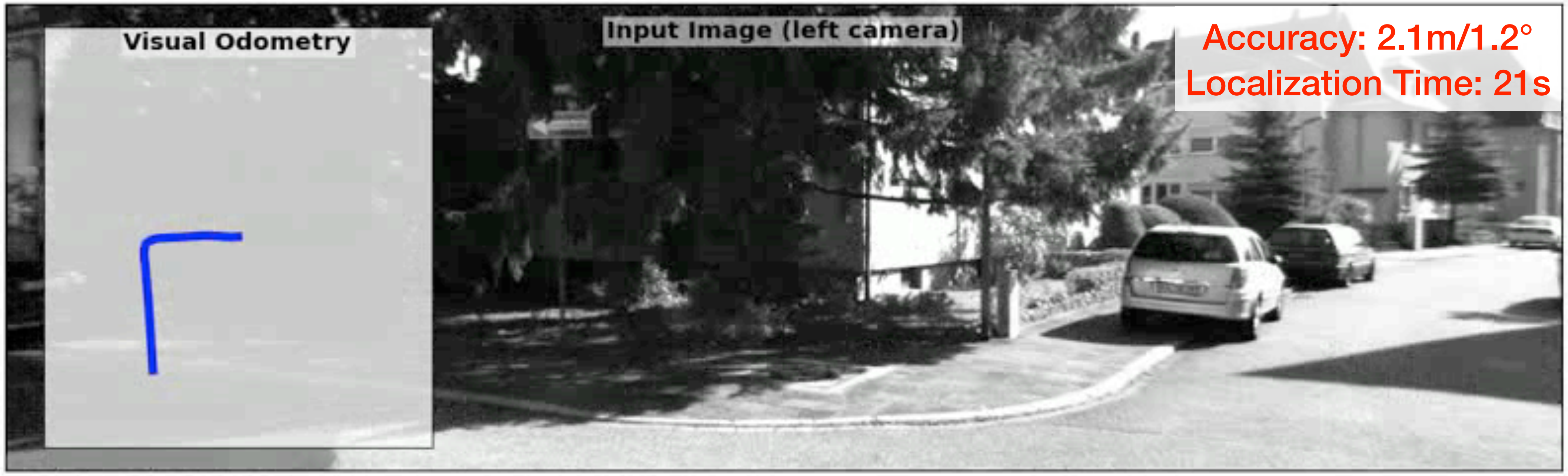
- We've derived a general algorithm for simplifying mixture models and do this periodically to reduce computation

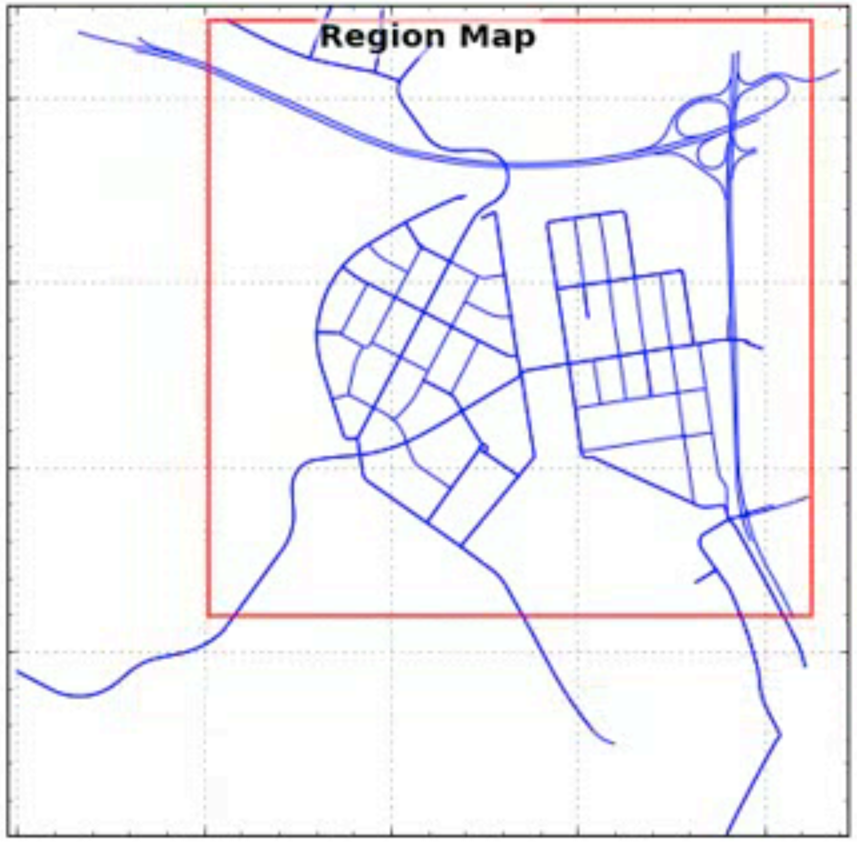
Experiments

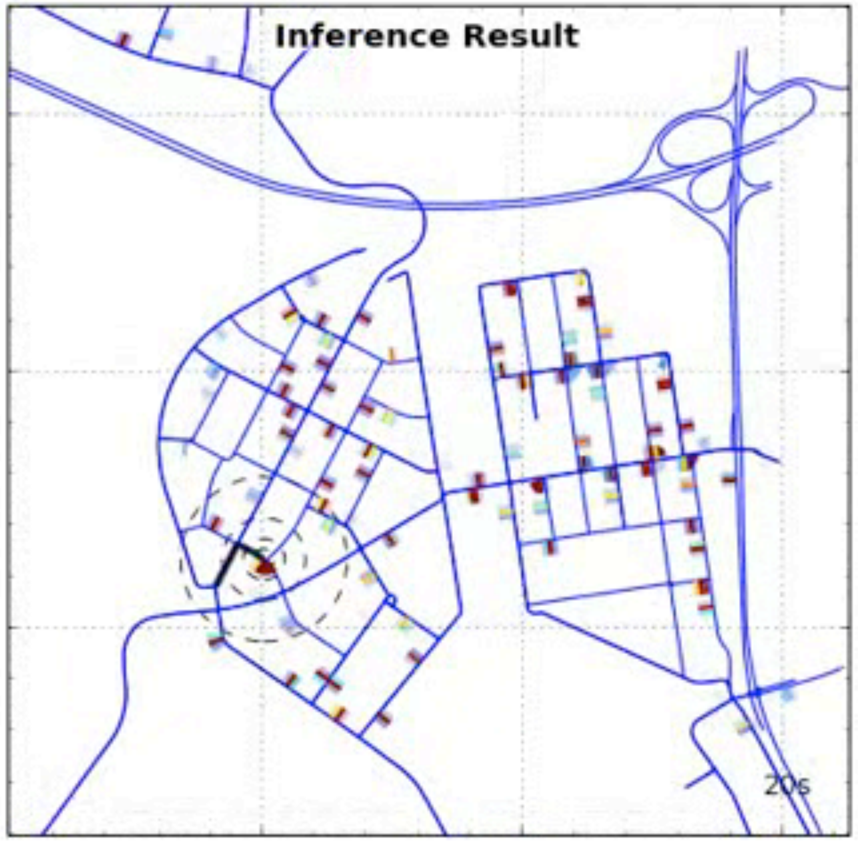
- We used the Visual Odometry sequences from the KITTI dataset [Geiger et al, CVPR 2012]
  - Video captured from car-mounted cameras
  - 11 sequences captured in a variety of settings (e.g., urban, highway, rural, etc)
  - Monocular and Stereo visual odometry computed using LIBVISO2 [Geiger et al, IV 2011]
  - Errors computed in position and heading
  - Parallelized implementation runs at frame rate on average on 16 cores



Experiments







11

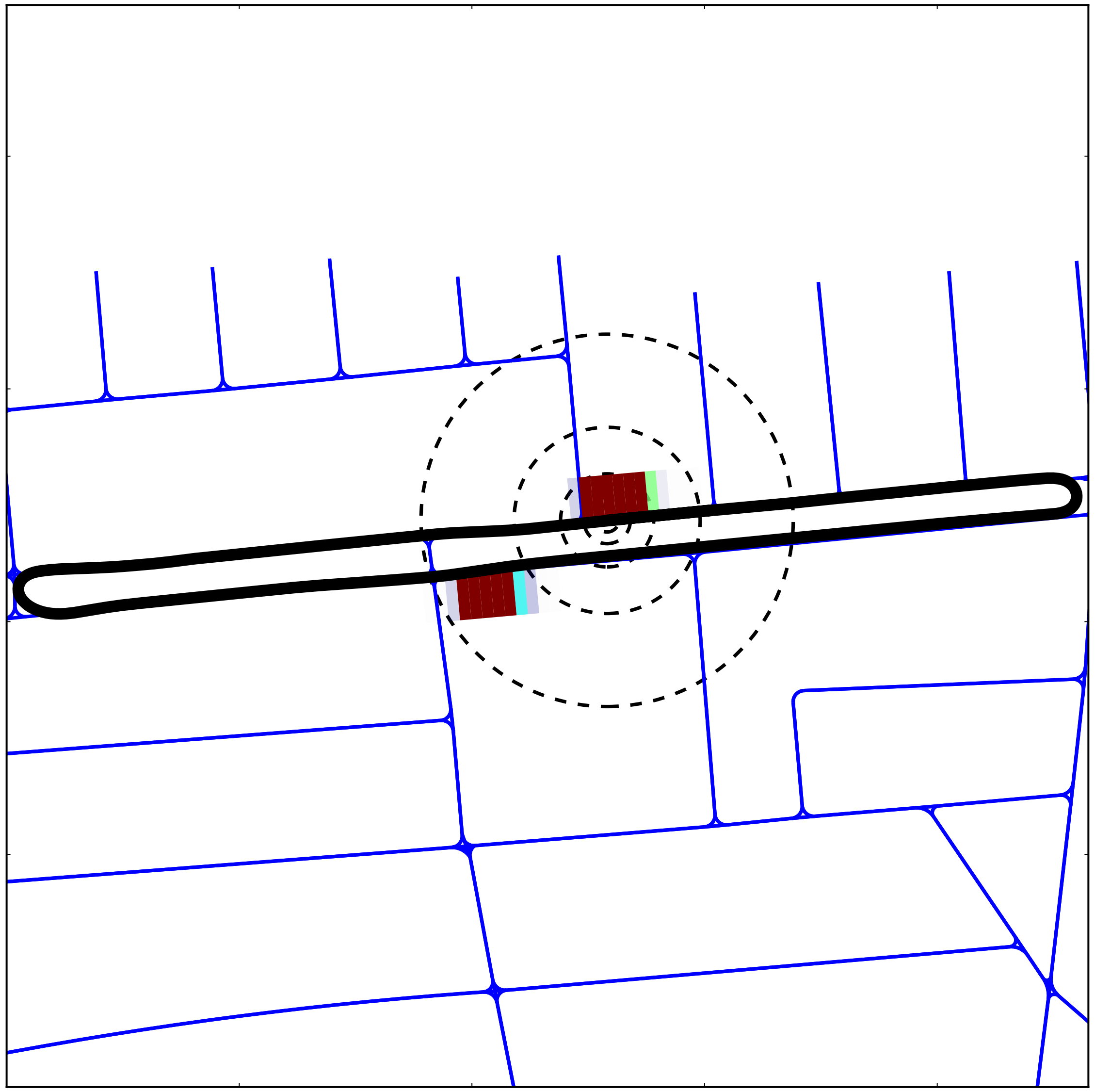
Experiments

12

Experiments: Quantitative Results

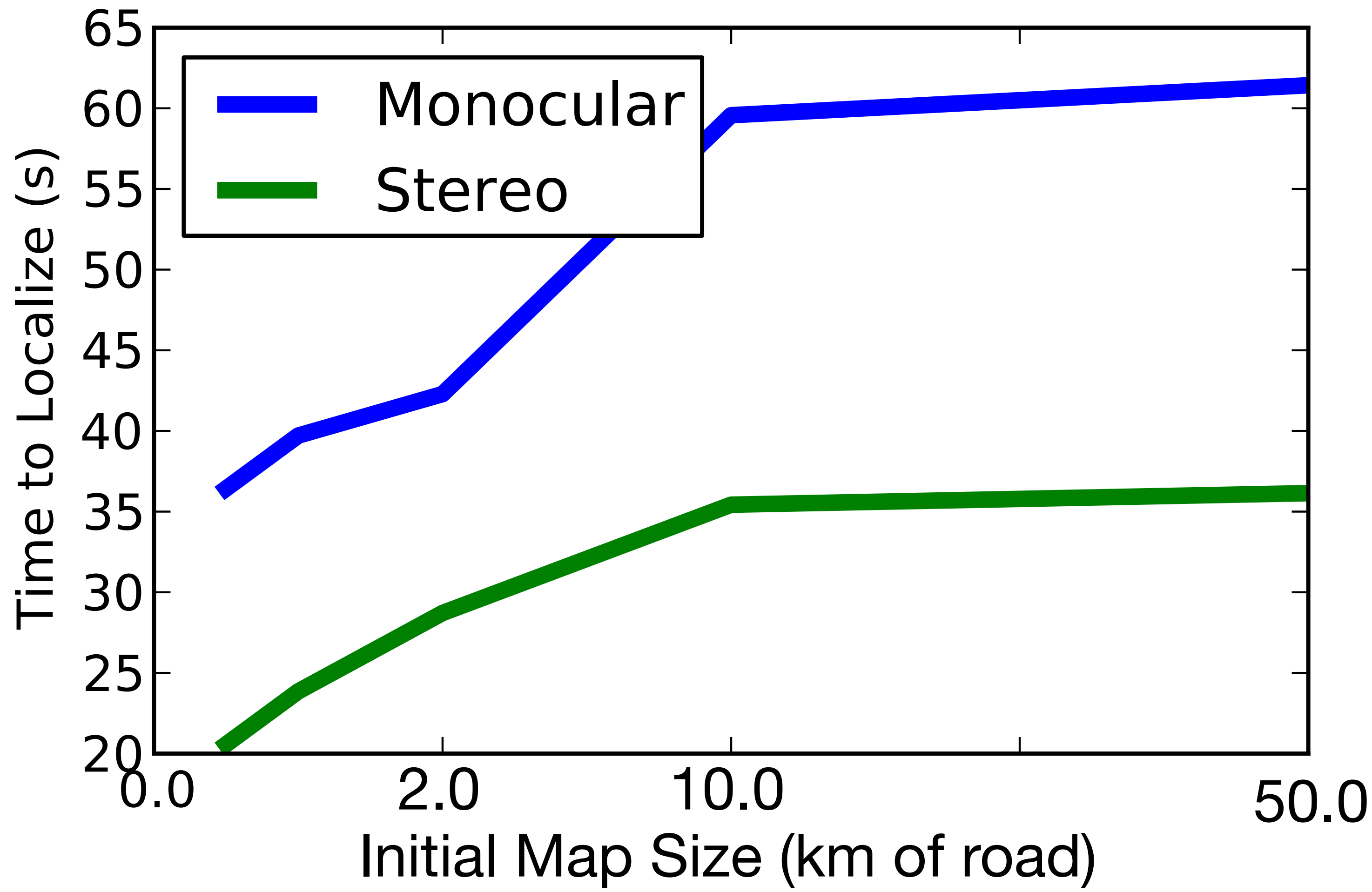
Average	Stereo Odometry	Monocular Odometry	Map Projection
Position Error	3.1m	18.4m	1.4m
Heading Error	1.3°	3.6°	-
Localization Time	36s	62s	-

Experiments: Ambiguous Sequences

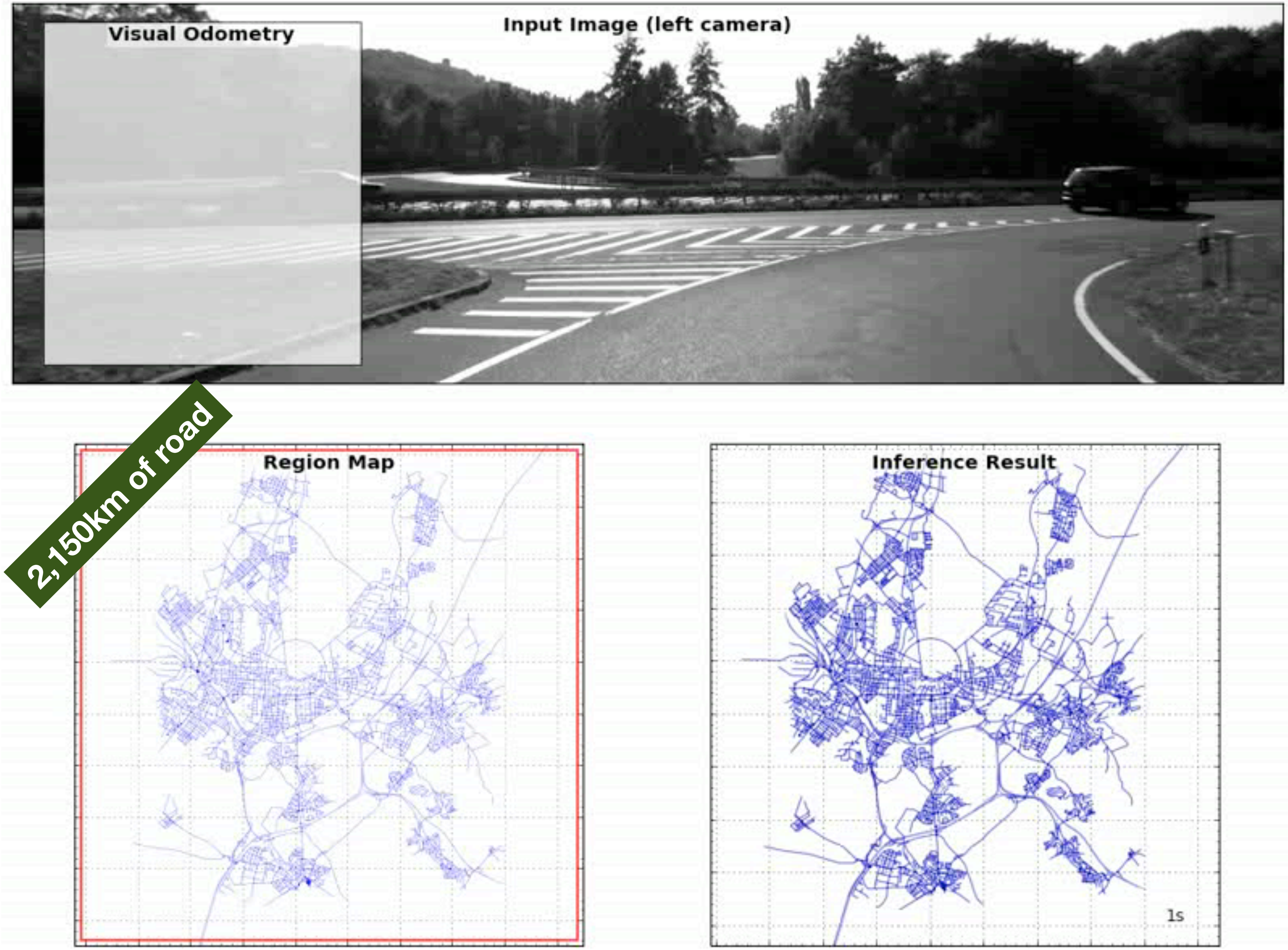


14

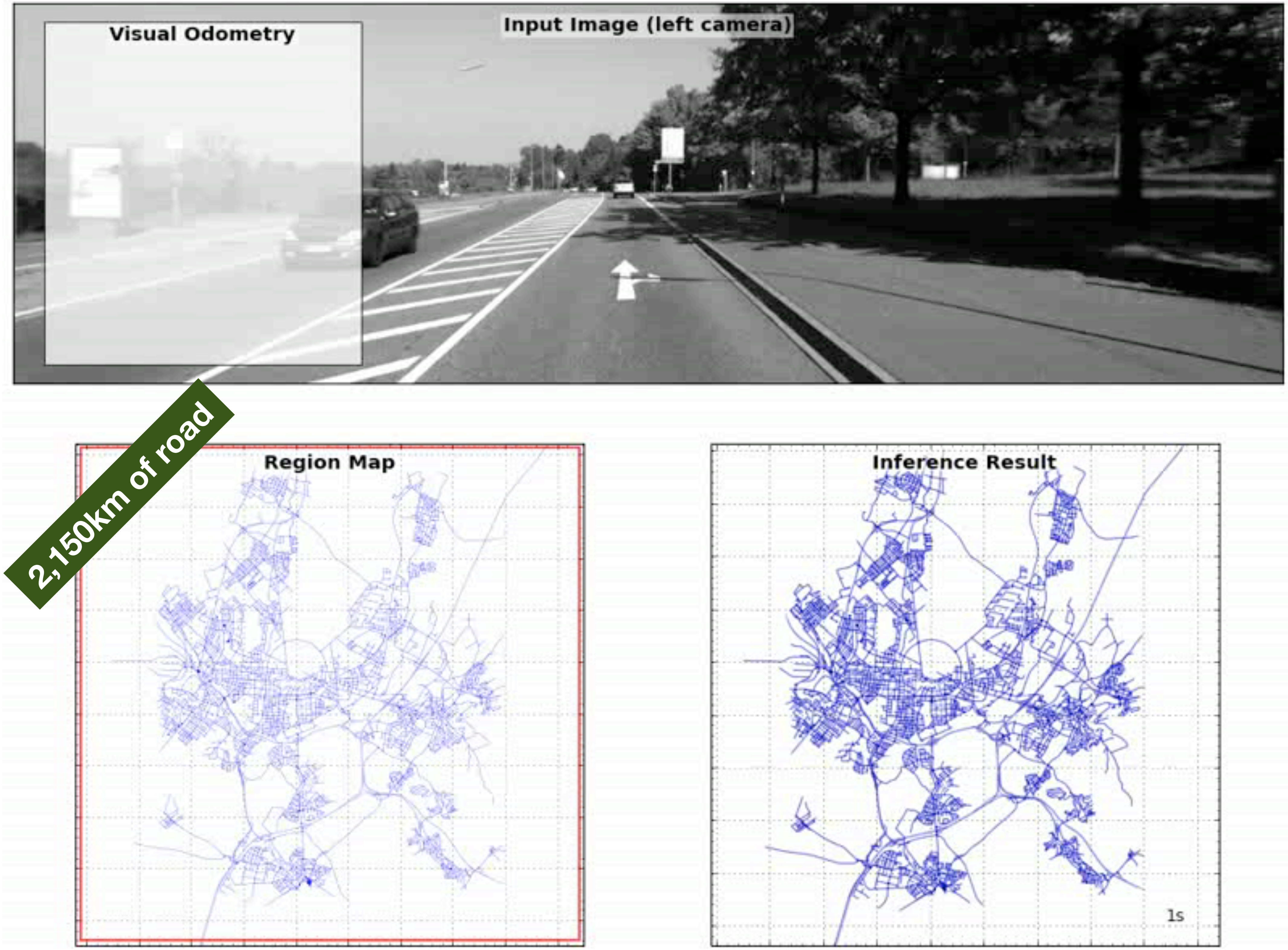
Experiments: Initial Map Size



Experiments: Full City Maps



Experiments: Full City Maps



17

Conclusions

- Fast, accurate map-based localization using only visual odometry
  - Accuracy of 3.1m/1.3° in 36 seconds of driving time on average
  - Highly parallelizable, runs at real-time on average w/ 16 cores
  - **Code will be available:** http://www.cs.toronto.edu/~mbrubake
- Future work
  - Exploiting other map information, e.g., landmarks, speed limits
  - Integration of other sensors, e.g., accelerometers