Supplemental Material for
Lost! Leveraging the Crowd for Probabilistic Visual Self-Localization

Marcus A. Brubaker
TTI Chicago
mbrubake@cs.toronto.edu

Andreas Geiger
KIT & MPI Tübingen
geiger@kit.edu

Raquel Urtasun
TTI Chicago
rurtasun@ttic.edu

1. Overview

This document includes supplementary material to *Lost! Leveraging the Crowd for Probabilistic Visual Self-Localization*. Included are detailed versions of the algorithms and other results.

2. Additional Results

Included in this document are additional results which could not fit in the main paper. See Figure 1 for frames from the remaining training sequences and Figures 2 and 3 for frames from the test sequences where ground truth data is not availble.

3. GMM Simplification

As described in the appendix of the main paper, given a Gaussian mixture model $f(x) = \sum_a \pi_a \mathcal{N}(x|\mu_a, \Sigma_a)$ we seek $g(x) = \sum_b \omega_b \mathcal{N}(x|\mu_b, \Sigma_b)$ with the least number of components such that $D(f\|g) < \epsilon$ where $D(f\|g)$ is the KL divergence. We begin with $g(x) = f(x)$ and successively remove the lowest weight component of $g(x)$ and update the remaining components to better fit $f(x)$ so long as the KL divergence (or its upper bound) is below the threshold ϵ. As explained in the paper, experimentation found an optimal value of this parameter to be $\epsilon = 10^{-2}$nats for our problem.

The upper bound [1] used is

$$\hat{D}(\phi, \psi, f, g) = \sum_{a,b} \phi_{a,b} \left(\log \frac{\phi_{a,b}}{\psi_{a,b}} + D(f_a \| g_b) \right) \tag{1}$$

where $\phi_{a,b}$ and $\psi_{a,b}$ are variational parameters and

$$D(f_a \| g_b) = \frac{1}{2} \left[\log |\Sigma_b| - \log |\Sigma_a| + Tr(\Sigma_b^{-1} \Sigma_a) - d + (\mu_a - \mu_b)^T \Sigma_b^{-1} (\mu_a - \mu_b) \right] \tag{2}$$

is the KL divergence between the Gaussian mixture components f_a and g_b. To minimize this objective (and improve the fit of a given approximation) we note that the global optima can be found for each set of parameters individually.

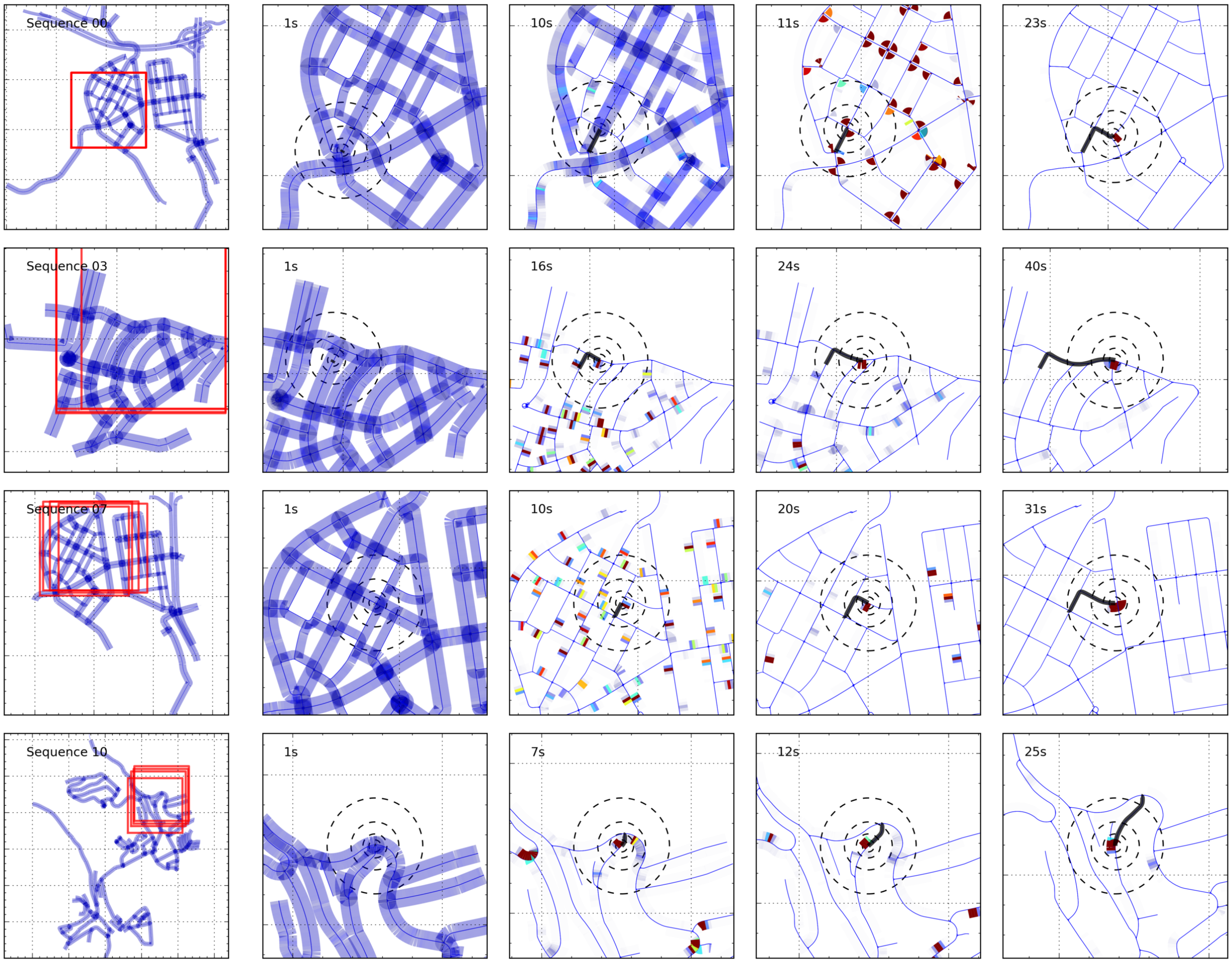

Figure 1. **Selected Frames:** Inference results for the remaining training sequences. The left most column shows the full map region for each sequence, followed by zoomed in sections of the map showing the posterior distribution over time. The black line is the GPS trajectory and the concentric circles indicate the current GPS position. Grid lines are every 500m.

That is, by setting the derivatives of (1) equal to zero and solving, one can find

$$
\begin{aligned}
\mu_b &= \left(\sum_a \phi_{a,b}\right)^{-1} \sum_a \phi_{a,b}\mu_a \\
\Sigma_b &= \left(\sum_a \phi_{a,b}\right)^{-1} \sum_a \phi_{a,b}\left(\Sigma_a + (\mu_a - \mu_b)(\mu_a - \mu_b)^T\right) \\
\omega_b &= \sum_a \phi_{a,b} \\
\psi_{a,b} &= \omega_b \frac{\phi_{a,b}}{\sum_{a'} \phi_{a',b}} \\
\phi_{a,b} &= \pi_a \frac{\psi_{a,b} \exp\left(-D(f_a \| g_b)\right)}{\sum_{b'} \psi_{a,b'} \exp\left(-D(f_a \| g_{b'})\right)}
\end{aligned}
$$

Note that it is necessary to enforce the constraints $\sum_b \omega_b = 1$, $\sum_b \phi_{a,b} = \pi_a$ and $\sum_a \psi_{a,b} = \omega_b$ through the use of Lagrange multipliers. This procedure is summarized in Algorithm 1.

Algorithm 1 GMM Simplification

Input: $f(x) = \sum_a \pi_a \mathcal{N}(x|\mu_a, \Sigma_a)$
Initialize $g = f$, $\phi \leftarrow diag(\pi)$, $\psi \leftarrow diag(\pi)$
loop
  Select a component to remove $\hat{b} = \arg\min_{b'} \omega_{b'}$
  for $b \neq \hat{b}$ **do**
    Remove $\hat{b}$ *from the variational parameters*
    $\omega'_b \leftarrow \omega_b + \frac{\omega_{\hat{b}}}{|g|-1}$
    $\mu'_b \leftarrow \mu_b$
    $\Sigma'_b \leftarrow \Sigma_b$
    $\forall a$: $\phi'_{a,b} \leftarrow \phi_{a,b} + \frac{\phi_{a,b'}}{|g|-1}$
    $\forall a$: $\psi'_{a,b} \leftarrow \omega'_b \frac{\phi'_{a,b}}{\sum_{a'} \phi'_{a'|b}}$
  end for
  while $\hat{D}(\phi', \psi', f, g') \geq \epsilon$ and not converged **do**
    Minimize $\hat{D}(\phi', \psi', f, g')$ *w.r.t.* ϕ', ψ', ω', μ', Σ'.
    $\forall b \neq \hat{b}$: $\omega'_b \leftarrow \sum_a \phi'_{a,b}$
    $\forall a, b \neq \hat{b}$: $\psi'_{a,b} \leftarrow \omega'_b \frac{\phi'_{a,b}}{\sum_{a'} \phi_{a',b}}$
    $\forall a, b \neq \hat{b}$: $\phi'_{a,b} \leftarrow \pi_a \frac{\psi'_{a,b} \exp\left(-D(f_a \| g'_b)\right)}{\sum_{b'} \psi_{a,b'} \exp\left(-D(f_a \| g'_{b'})\right)}$
    $\forall b \neq \hat{b}$: $\mu'_b \leftarrow \left(\sum_a \phi'_{a,b}\right)^{-1} \sum_a \phi'_{a,b} \mu_a$
    $\forall b \neq \hat{b}$: $\Sigma'_b \leftarrow \left(\sum_a \phi'_{a,b}\right)^{-1} \sum_a \phi'_{a,b} \left(\Sigma_a + (\mu_a - \mu'_b)(\mu_a - \mu'_b)^T\right)$
  end while
  if $\hat{D}(\phi', \psi', f, g') \geq \epsilon$ **then**
    Return: g
  else
    $g \leftarrow g'$
  end if
end loop

4. Filtering Algorithm

Due to space constraints, the equations necessary to implement the filtering algorithm could not be included. Instead, they are presented here in Algorithm 2.

The function used to determine whether to use a Monte Carlo approximation or an analytic approximation is:

$$g(\mu, \Sigma) = \int p(u_t|u_{t-1}, \mathbf{s}_{t-1}) \mathcal{N}(\mathbf{s}_{t-1}|\mu, \Sigma) d\mathbf{s}_{t-1} \tag{3}$$

In the model we have that

$$p(u_t|u_{t-1}, \mathbf{s}_{t-1}) = \xi_{u_t, u_{t-1}} \int_{\ell_{u_{t-1}}}^{\ell_{u_{t-1}} + \ell_u} \mathcal{N}(x|\mathbf{a}_d^T \mathbf{s}_{t-1}, \mathbf{a}_d^T \Sigma^{\mathbf{x}}_{u_{t-1}} \mathbf{a}_d) dx \tag{4}$$

where $\xi_{u_t,u_{t-1}}$ is a constant. Substituting this in we have

$$\begin{aligned} g(\mu,\Sigma) &= \xi_{u_t,u_{t-1}} \int\int_{\ell_{u_{t-1}}}^{\ell_{u_{t-1}}+\ell_u} \mathcal{N}(x|\mathbf{a}_d^T\mathbf{s}_{t-1}, \mathbf{a}_d^T\Sigma_{u_{t-1}}^{\mathbf{x}}\mathbf{a}_d)\mathcal{N}(\mathbf{s}_{t-1}|\mu,\Sigma)dxd\mathbf{s}_{t-1} \\ &= \xi_{u_t,u_{t-1}} \int_{\ell_{u_{t-1}}}^{\ell_{u_{t-1}}+\ell_u} \int \mathcal{N}(x|\mathbf{a}_d^T\mathbf{s}_{t-1}, \mathbf{a}_d^T\Sigma_{u_{t-1}}^{\mathbf{x}}\mathbf{a}_d)\mathcal{N}(\mathbf{s}_{t-1}|\mu,\Sigma)d\mathbf{s}_{t-1}dx \\ &= \xi_{u_t,u_{t-1}} \int_{\ell_{u_{t-1}}}^{\ell_{u_{t-1}}+\ell_u} \mathcal{N}(x|\mathbf{a}_d^T\mu, \mathbf{a}_d^T(\Sigma_{u_{t-1}}^{\mathbf{x}}+\Sigma)\mathbf{a}_d)dx \\ &= \xi_{u_t,u_{t-1}} \left(\Phi(\ell_{u_{t-1}}+\ell_u|m,s^2) - \Phi(\ell_{u_{t-1}}|m,s^2)\right) \end{aligned}$$

where $m = \mathbf{a}_d^T\mu$, $s^2 = \mathbf{a}_d^T(\Sigma_{u_{t-1}}^{\mathbf{x}}+\Sigma)\mathbf{a}_d$ and $\Phi(x|m,s^2)$ is the CDF of a univariate normal distribution with mean m and variance s^2 evaluated at x.

References

[1] J. Hershey and P. Olsen. Approximating the Kullback-Leibler Divergence Between Gaussian Mixture Models. In *ICASSP*, volume 4, pages 317–320, 2007. 1

Algorithm 2 Filter

Input: Posterior at $t-1$, $\{P_u^{t-1}, \mathcal{M}_u^{t-1}\}$, and observation at t, $\mathbf{y}_t$
Initialize mixtures, $\mathcal{M}_u^t \leftarrow \varnothing$, for all u
for all streets u_{t-1} **do**
  for all streets u_t reachable from u_{t-1} **do**
    $\mathcal{M}' \leftarrow \varnothing$
    for all $(\omega, \mu, \Sigma) \in \mathcal{M}_{u_{t-1}}^{t-1}$ **do**
      if $\|\frac{d}{d\mu} g(\mu, \Sigma)\| < \eta$ **then**
        Analytically approximate $c_{pred}\mathcal{N}(\mu_{pred}, \Sigma_{pred})$
        $c_{pred} \leftarrow p(u_t | u_{t-1}, \mathbf{s}_{t-1} = \mu)$
        $\mu_{pred} \leftarrow \mathbf{A}_{u_t, u_{t-1}} \mu + \mathbf{b}_{u_t, u_{t-1}}$
        $\Sigma_{pred} \leftarrow \Sigma_{u_t}^{\mathbf{x}} + \mathbf{A}_{u_t, u_{t-1}} \Sigma \mathbf{A}_{u_t, u_{t-1}}^T$
      else
        Sample to compute $c_{pred}\mathcal{N}(\mu_{pred}, \Sigma_{pred})$
        for $j = 1, \ldots, M$ **do**
          $\mathbf{s}_{t-1}^{(j)} \sim \mathcal{N}(\mu, \Sigma)$
          $\mathbf{s}_t^{(j)} \leftarrow \mathbf{A}_{u_t, u_{t-1}} \mathbf{s}_{t-1}^{(j)} + \mathbf{b}_{u_t, u_{t-1}}$
        end for
        $c_{pred} \leftarrow M^{-1} \sum_{j=1}^{M} p(u_t | u_{t-1}, \mathbf{s}_{t-1}^{(j)})$
        $\mu_{pred} \leftarrow (M c_{pred})^{-1} \sum_{j=1}^{M} p(u_t | u_{t-1}, \mathbf{s}_{t-1}^{(j)}) \mathbf{s}_t^{(j)}$
        $\Sigma_{pred} \leftarrow \Sigma_{u_t}^{\mathbf{x}} + \sum_{j=1}^{M} \frac{p(u_t | u_{t-1}, \mathbf{s}_{t-1}^{(j)})}{M c_{pred}} (\mathbf{s}_t^{(j)} - \mu_{pred})(\mathbf{s}_t^{(j)} - \mu_{pred})^T$
      end if
      Incorporate $\mathbf{y}_t$ *to compute* $c_{upd}\mathcal{N}(\mu_{upd}, \Sigma_{upd})$
      $\Sigma_{upd} \leftarrow \left(\mathbf{M}_{u_t}^T {\Sigma_{u_t}^{\mathbf{y}}}^{-1} \mathbf{M}_{u_t} + \Sigma_{pred}^{-1} \right)^{-1}$
      $\mu_{upd} \leftarrow \Sigma_{upd} \left(\mathbf{M}_{u_t}^T {\Sigma_{u_t}^{\mathbf{y}}}^{-1} \mathbf{y}_t + \Sigma_{pred}^{-1} \mu_{pred} \right)$
      $c_{upd} \leftarrow \frac{\omega c_{pred} |\Sigma_{upd}|^{0.5}}{|\Sigma_{pred}|^{0.5} |\Sigma_{u_t}^{\mathbf{y}}|^{0.5}} \exp \left(-\frac{1}{2} \|\mathbf{y}_t - \mathbf{M}_{u_t} \mu_{pred}\|^2_{\Sigma_{u_t}^{\mathbf{y}} + \mathbf{M}_{u_t} \Sigma_{pred} \mathbf{M}_{u_t}^T} \right)$
      if $u_t \neq u_{t-1}$ **then**
        Add $(c_{upd}, \mu_{upd}, \Sigma_{upd})$ to $\mathcal{M}'$
      else
        Add $(c_{upd}, \mu_{upd}, \Sigma_{upd})$ to $\mathcal{M}_{u_t}^t$
      end if
    end for
    Compute (c, μ, Σ) to approximate components of $\mathcal{M}'$ and add to $\mathcal{M}_{u_t}^t$
  end for
end for
for all streets u **do**
  Set P_u^t to the sum of the weights of mixture $\mathcal{M}_u^t$
  Normalize the weights of mixture $\mathcal{M}_u^t$
  if $\frac{\ell_u}{|\mathcal{M}_u^t|} < 10$ meters **then**
    Simplify $\mathcal{M}_u^t$ with Algorithm 1
  end if
end for
Normalize P_u^t so that $\sum_u P_u^t = 1$.
For all u, if $P_u^t < 10^{-50}$ set $P_u^t \leftarrow 0$ and $\mathcal{M}_u^t \leftarrow \varnothing$
Return: Posterior at t, $\{P_u^t, \mathcal{M}_u^t\}$

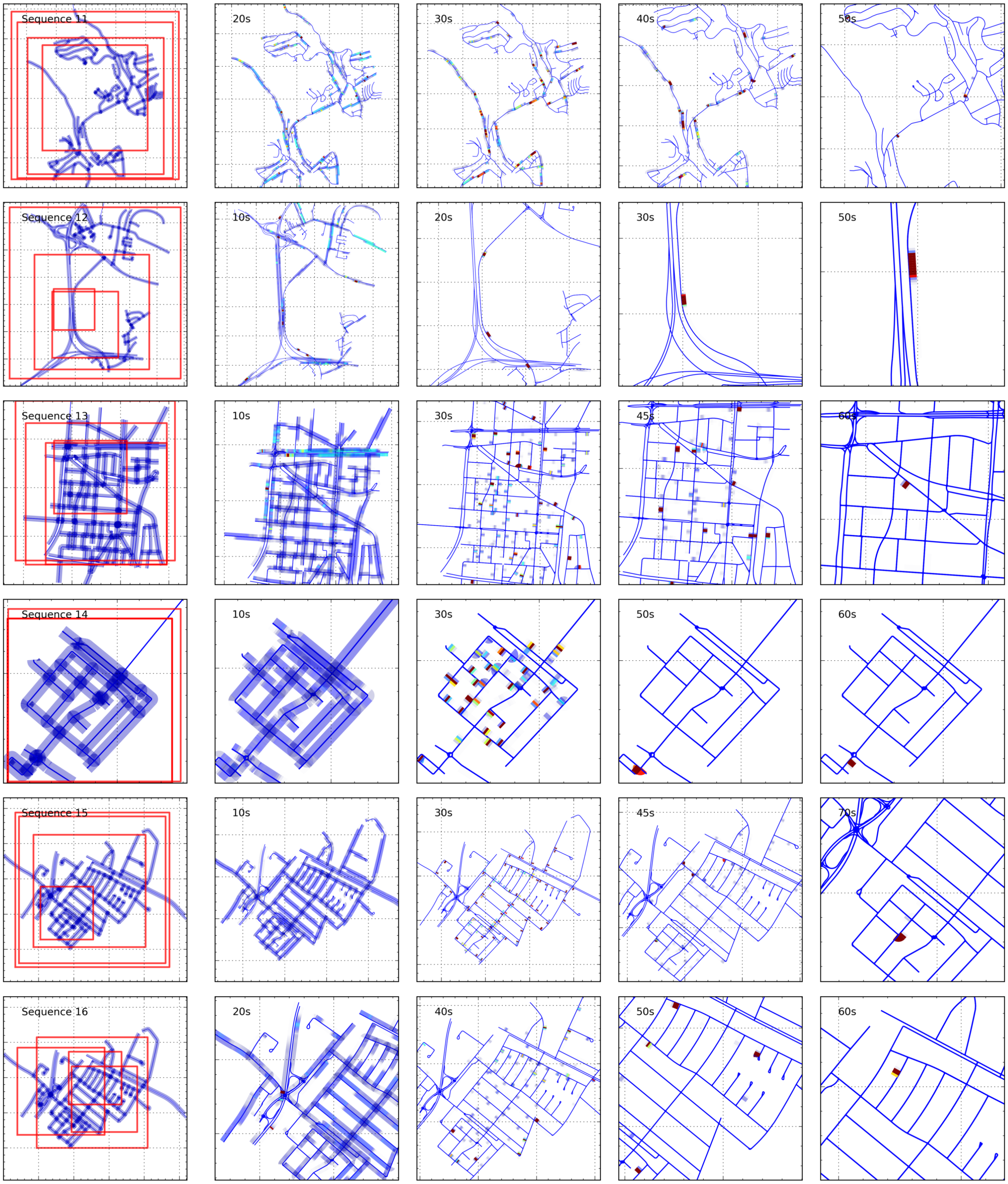

Figure 2. **Selected Frames of Test Sequences:** Inference results for the testing sequences. The left most column shows the full map region for each sequence, followed by zoomed in sections of the map showing the posterior distribution over time. GPS data is unavailable for these sequences. Zooming is done based on a smoothed bounding box of the dominant modes of the posterior. Grid lines are every 500m.

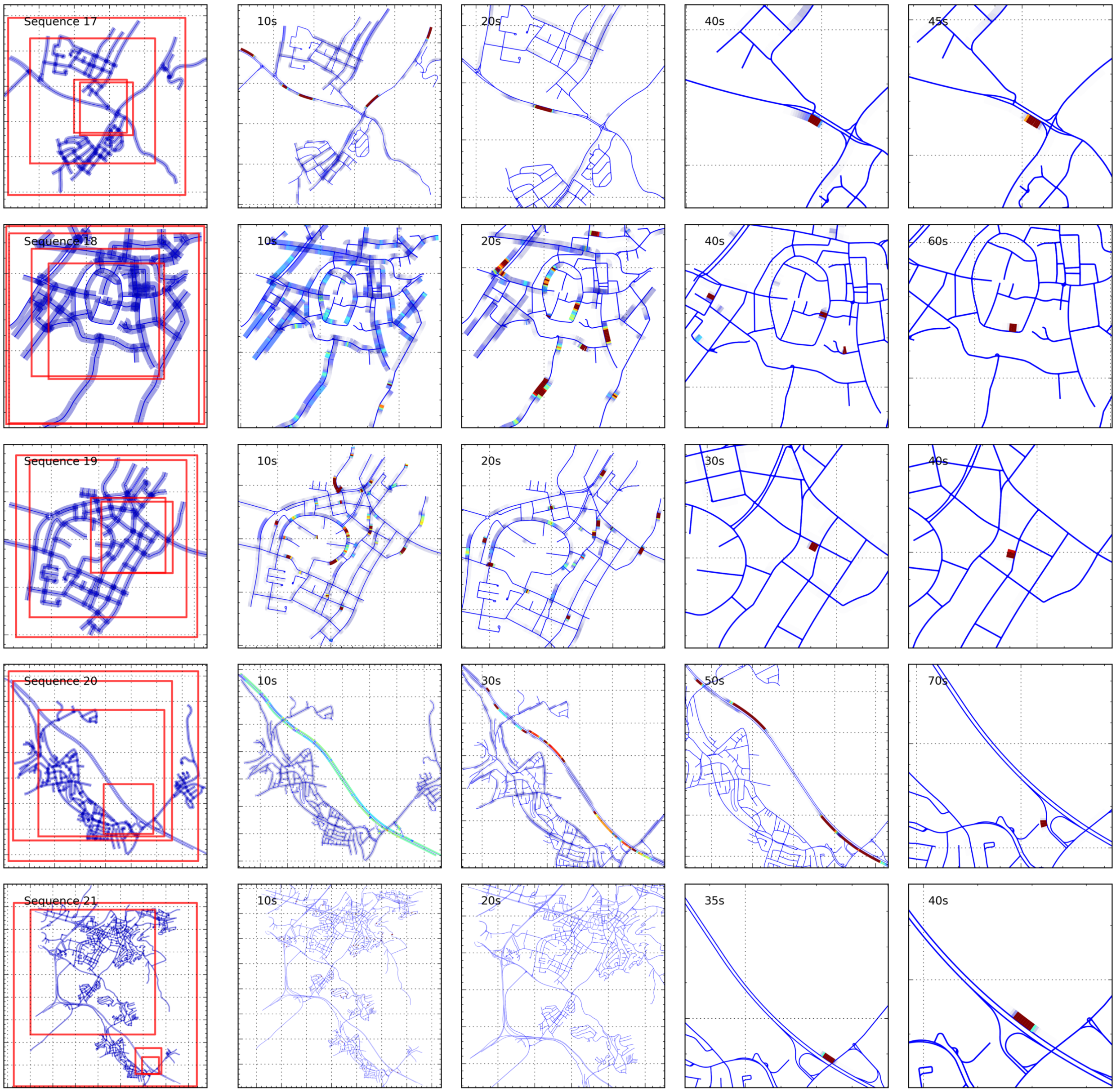

Figure 3. **Selected Frames of Test Sequences:** Inference results for the testing sequences. The left most column shows the full map region for each sequence, followed by zoomed in sections of the map showing the posterior distribution over time. GPS data is unavailable for these sequences. Zooming is done based on a smoothed bounding box of the dominant modes of the posterior. Grid lines are every 500m.