A generative model for 3D urban scene understanding from movable platforms Supplementary Material

Andreas Geiger and Martin Lauer
Department of Measurement and Control
Karlsruhe Institute of Technology
{geiger,martin.lauer}@kit.edu

Raquel Urtasun
Toyota Technological Institute at Chicago
rurtasun@ttic.edu

In this supplementary material we first present a review on sampling methods and details on the bijection we employ for learning a non-parametric prior over the orientations. We then show inference results for all the sequences in our dataset.

1. A Review on Sampling Methods

Bayesian inference often requires the calculation of posterior distributions. Unfortunately, only in special cases (e. g., conjugate priors) it is possible to compute analytically the posterior. Typical approximations to this computation are Laplace approximations and sampling. In sampling-based approaches the integral is approximated numerically. Markov chain Monte Carlo methods (MCMC) can be used to generate samples from a complex graphical model efficiently. The basic idea is to create an artificial ergodic Markov chain so that the posterior distribution of interest becomes a stationary distribution of the Markov chain. Hence, after a burn-in period sampling from the Markov chain yields samples from the posterior distribution [1].

The main challenge in MCMC methods is to create an adequate transition kernel for the Markov chain. Two major strategies have been developed, Metropolis-Hastings and Gibbs sampling. Metropolis-Hastings [4] uses an arbitrary proposal distribution $q(x'|x)$ to sample a new candidate state x' from the current state x of the Markov chain. x' is accepted as the new state of the Markov chain with a certain probability $\mathcal{A}(x, x')$, otherwise the current state x is replicated. The acceptance probability $\mathcal{A}(x, x')$ is designed in such a way that the stochastic transition kernel $T(\cdot|x)$ defined by this procedure meets the *detailed balance condition* for Markov chains, i. e.,

$$T(x'|x) \cdot p(x) = T(x|x') \cdot p(x')$$

where $p(\cdot)$ denotes the posterior distribution of interest. The detailed balance condition ensures that p is a stationary distribution of the transition kernel T.

Gibbs sampling [2] is an alternative way to implement transition kernels, where the proposal distributions are the conditional distributions $p(x_i|x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_N)$ for all i[1]. Several transition kernels with the same stationary distribution can be combined by applying them in sequential order or by randomly selecting one of them in each step. Note that for these type of transition kernel, the acceptance probability is always one.

More than just allowing stochastic inference on models of fixed size, MCMC methods are able to sample over models of varying size and varying topology. This is important to us since in our model the number of parameters varies with the number of arms in the intersection. Green [3] proposed an extension of Metropolis-Hastings, named *reversible jumps*, that allows transdimensional jumps between models of different size and topology. Again, meeting the *detailed balance condition* ensures that the resulting reversible jump sampler converges in distribution to the posterior distribution of interest.

For simplicity, lets consider an example where we are interested in two different model topologies with states X_1 and X_2 respectively. Let p_1 and p_2 be the posterior distribution of interest, and let π_1, $\pi_2 = 1 - \pi_1$ be priors to control the amount of samples from each topology. Unfortunately a direct comparison of the densities is misleading since the probability measures on X_1 and X_2 might be different. To overcome this problem, reversible jumps introduce additional states U_1 and U_2 to have a one to one mapping between the augmented state spaces $\tau : X_1 \times U_1 \to X_2 \times U_2$. Using a proposal distribution of the

[1]here we assume x to be a vector $x = (x_1, \ldots, x_N)$

form $q_1(u_1|x_1)$, we can create a vector (x_1, u_1), transform it into (x_2, u_2) applying the mapping τ and neglect u_2 to obtain a candidate for x_2. The acceptance probability of a jump from x_1 to x_2 becomes

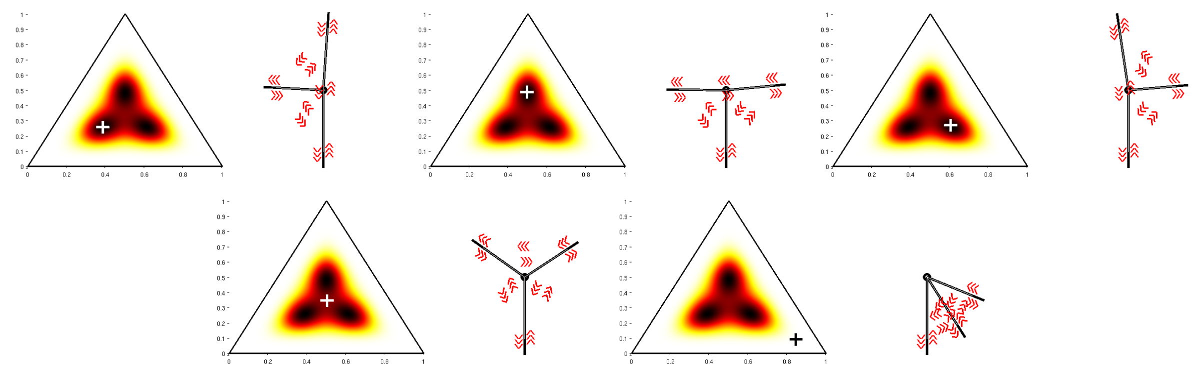
$$\mathcal{A}(x_1, x_2) = \min\{1, \frac{\pi_2 \cdot p_2(x_2) \cdot q_2(u_2|x_2) \cdot |det(\mathcal{J}_\tau(x_1, u_1))|}{\pi_1 \cdot p_1(x_1) \cdot q_1(u_1|x_1)}\}$$

where $det(\mathcal{J}_\tau(x_1, u_1))$ denotes the determinant of the Jacobian of τ. Analogously, we can switch from X_2 to X_1 using a proposal distribution $q_2(u_2|x_2)$ and compute the corresponding acceptance probability. Note that it is important to implement forward and backward steps together to meet the *detailed balance condition*.

When using reversible jumps, it is important to combine transition kernels that perform transitions within the present subspace and those that jump between subspaces. Typically, one of the kernels is selected randomly in every step. Note that if the selection probabilities are different for a transdimensional jump and its inverse jump, the ratio of the selection probabilities has to be considered in the acceptance probability $\mathcal{A}(x_1, x_2)$.

2. Non-parametric Models of Distributions on a Simplex

This section elaborates in more details on the bijection we employ to obtain a non-parametric distribution estimate over the orientation vector $\mathbf{o}$. First, lets illustrate the behavior of $\mathbf{o}$ for $k = 3$. The following figure shows different orientation configurations. Note that for the sake of this illustration, a Mixture-of-Dirichlet distribution has been specified manually, while in our approach we *learn* a multimodal prior distribution from labeled data. We will come back to the learning later in the supplementary material (section 2.3).



2.1. A bijection between the simplex Δ^{k-1} and $\mathbb{R}^{k-1}$

We now show how we express a multimodal distribution over the simplex Δ^{k-1} via a multimodal distribution over $\mathbb{R}^{k-1}$. Let $\mathbf{o} = (o_1, ..., o_k)^T$ be an orientation vector as described in the paper submission, with

$$\mathbf{o} \in \Delta^{k-1} = \left\{ \mathbb{R}^k | \ \forall_i : o_i \geq 0 \wedge \sum_{i=1}^{k} o_i = 1 \right\} \tag{1}$$

This vector can be interpreted as a multinomial distribution and samples can be drawn from the Dirichlet distribution

$$Dir(\mathbf{o}|\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{k} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{k} \alpha_i)} \prod_{i=1}^{k} o_i^{\alpha_i - 1},$$

hence the Dirichlet distribution qualifies as a prior for $\mathbf{o}$. However, since the Dirichlet distribution can not handle multiple distinct modes for $\alpha_i \geq 1$, learning a mixture of Dirichlet distributions will be neccessary to truthfully capture the data statistics. Unfortunately, the conjugate prior of the Dirichlet distribtion has no standard form and thus it is hard to employ in practice. Instead, we use a commonly used trick in the statistics community and assume a softmax model for $\mathbf{o}$, which implies that the preimage $\tilde{\mathbf{o}}$ (of $\mathbf{o}$) lives in $\mathbb{R}^{k-1}$, while $\mathbf{o}$ lives in Δ^{k-1}. This allows for the use of mixture distributions on $\mathbb{R}^{k-1}$, e.g., Gaussian mixture, which has analytic solutions to all relevant quantities. In the following we give a bijection

$\tau(\tilde{\mathbf{o}}) : \mathbb{R}^{k-1} \rightarrow \Delta^{k-1}$ which provides a unique mapping between the two spaces. Using the softmax, we have for all i that the

$$\tau(\tilde{\mathbf{o}})_i = o_i = \frac{\exp(\tilde{o}_i)}{\sum_{j=1}^{k} \exp(\tilde{o}_j)} = \frac{e_i}{\sum_{j=1}^{k} e_j} \tag{2}$$

with $e_i = \exp(\tilde{o}_i)$. The inverse mapping is more complicated. Rewriting the above equation, gives

$$\underbrace{\begin{pmatrix} o_1 - 1 & o_1 & \cdots & o_1 \\ o_2 & o_2 - 1 & \cdots & o_2 \\ \vdots & \vdots & \ddots & \vdots \\ o_k & o_k & \cdots & o_k - 1 \end{pmatrix}}_{=\mathbf{O}} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{3}$$

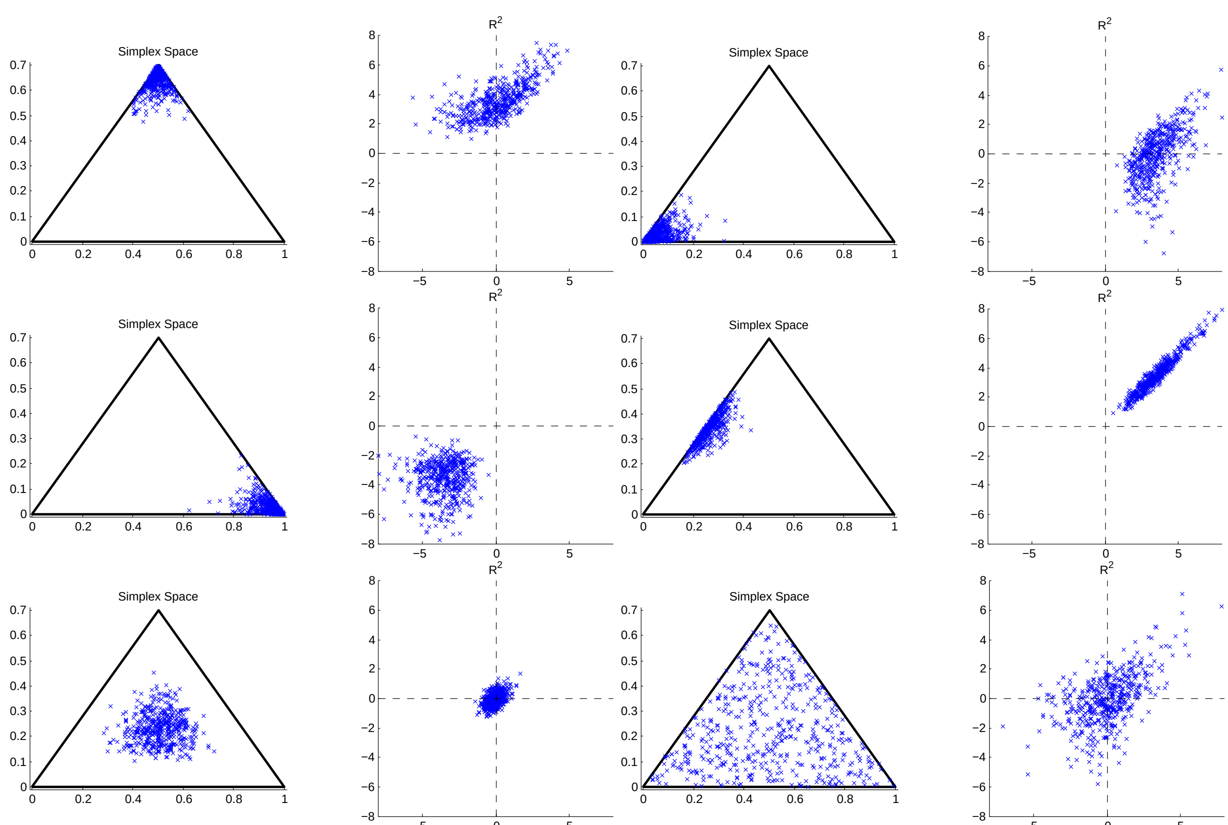
Note that $\mathbf{O}$ is of rank $k-1$, giving rise to infinitly many solutions $\mathbf{e}$. In order to make the inverse mapping τ^{-1} unique we therefor introduce an additional constraint, $e_k = 1$, resulting in

$$\begin{pmatrix} o_1 - 1 & o_1 & \cdots & o_1 \\ o_2 & o_2 - 1 & \cdots & o_2 \\ \vdots & \vdots & \ddots & \vdots \\ o_{k-1} & o_{k-1} & \cdots & o_{k-1} - 1 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_{k-1} \end{pmatrix} = \begin{pmatrix} -o_1 \\ -o_2 \\ \vdots \\ -o_{k-1} \end{pmatrix} \tag{4}$$

Thus, with $\mathbf{e}$ the solution to Eq. 4, the inverse mapping τ^{-1} is given by

$$\tau^{-1}(\mathbf{o})_i = \tilde{o}_i = \log e_i \tag{5}$$

The following plots show samples from a Dirichlet distribution in Δ^2 (left) and the transformed samples in $\mathbb{R}^2$:



2.2. Change of Variables

Mapping probability density functions into each other is also possible by change of variables. Again, we have

$$\tau(\tilde{\mathbf{o}})_i = o_i = \frac{\exp(\tilde{o}_i)}{\sum_{j=1}^{k} \exp(\tilde{o}_j)} = \frac{\exp(\tilde{o}_i)}{\sum_{j=1}^{k-1} \exp(\tilde{o}_j) + 1}$$

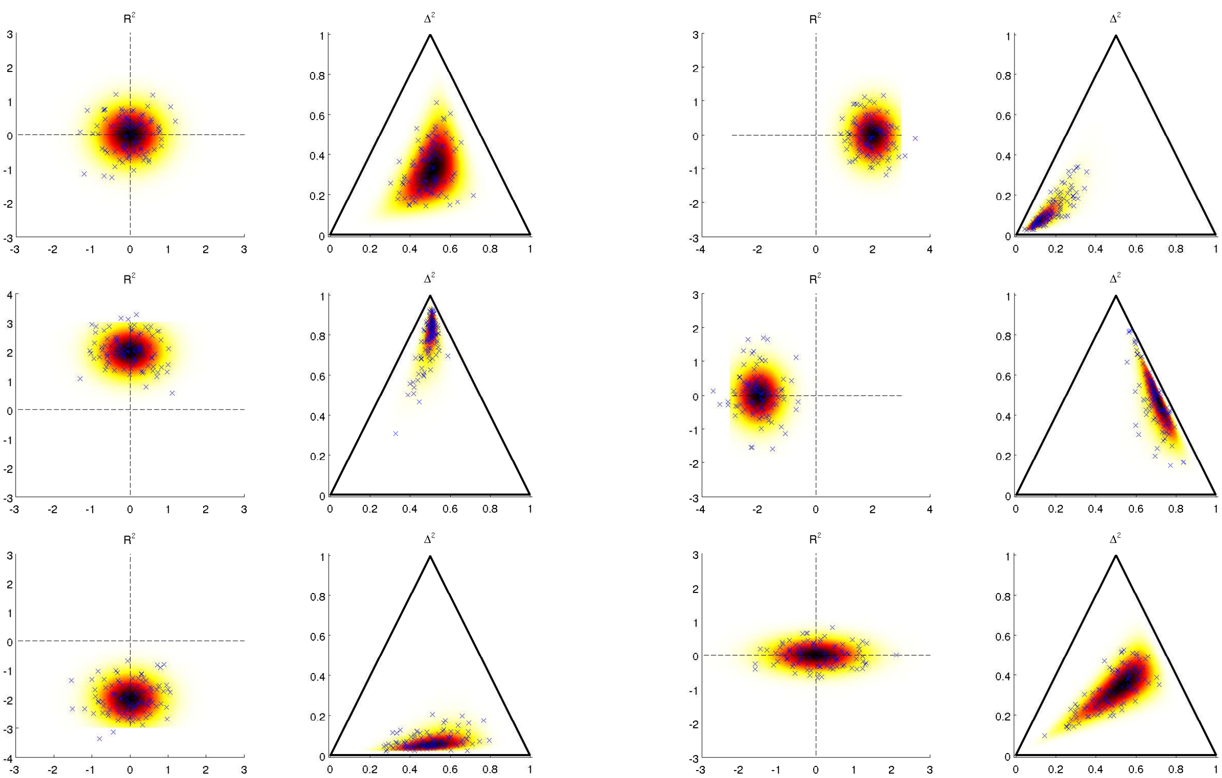
and the Jacobian of $\tau(\tilde{\mathbf{o}})$ is given by

$$J_\tau = \frac{d\tau}{d\tilde{\mathbf{o}}} = \bar{o} \cdot \begin{pmatrix} \exp(\tilde{o}_1) - \exp(2\tilde{o}_1)\bar{o} & \cdots & -\exp(\tilde{o}_1 + \tilde{o}_{k-1})\bar{o} \\ -\exp(\tilde{o}_2 + \tilde{o}_1)\bar{o} & \cdots & -\exp(\tilde{o}_2 + \tilde{o}_{k-1})\bar{o} \\ \vdots & \ddots & \vdots \\ -\exp(\tilde{o}_{k-1} + \tilde{o}_1)\bar{o} & \cdots & \exp(\tilde{o}_{k-1}) - \exp(2\tilde{o}_{k-1})\bar{o} \end{pmatrix}$$

with $\bar{o} = \left[\sum_{j=1}^{k-1} \exp(\tilde{o}_j) + 1\right]^{-1}$. Lets assume now that there exists a probability density f on $\tilde{\mathbf{o}}$, for instance $f(\tilde{\mathbf{o}}) = \mathcal{N}(\tilde{\mathbf{o}}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and also a probability density g on $\mathbf{o}$. Then, by laws of change of variables these two densities are related via

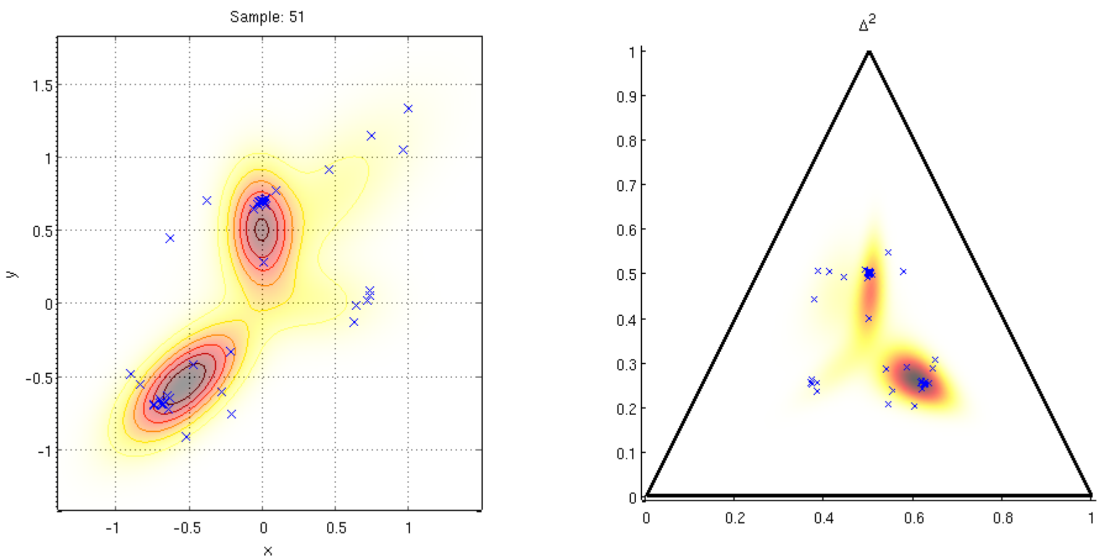
$$g(\mathbf{o}) = \frac{f(\tilde{\mathbf{o}})}{|\det(J_\tau)|}$$

The following example shows how a 2-dimensional Gaussian density translates into a non-standard density in Δ^2:



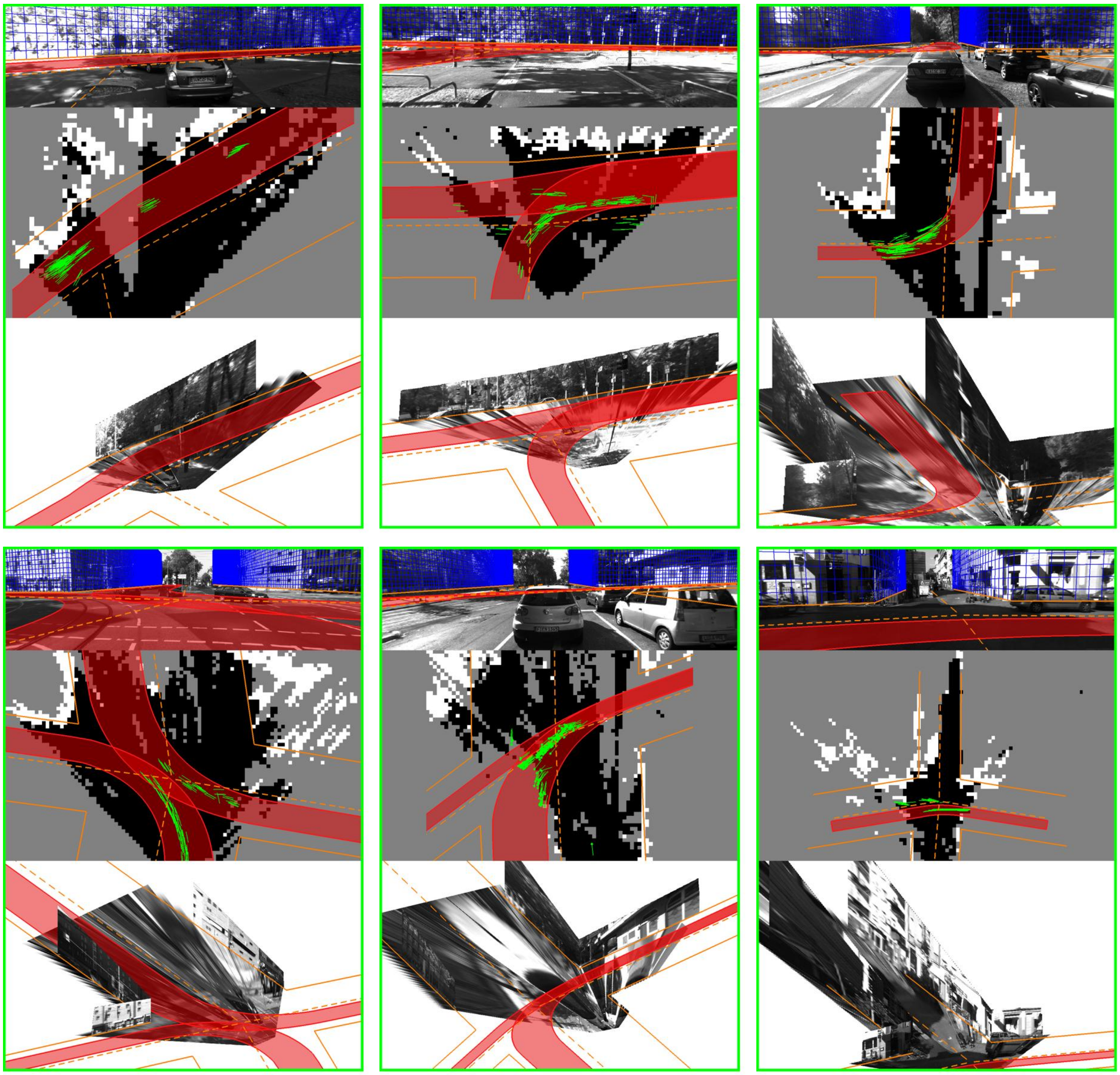
2.3. Sampling a Gaussian Mixture in $\mathbb{R}^{k-1}$ for data from Δ^{k-1}

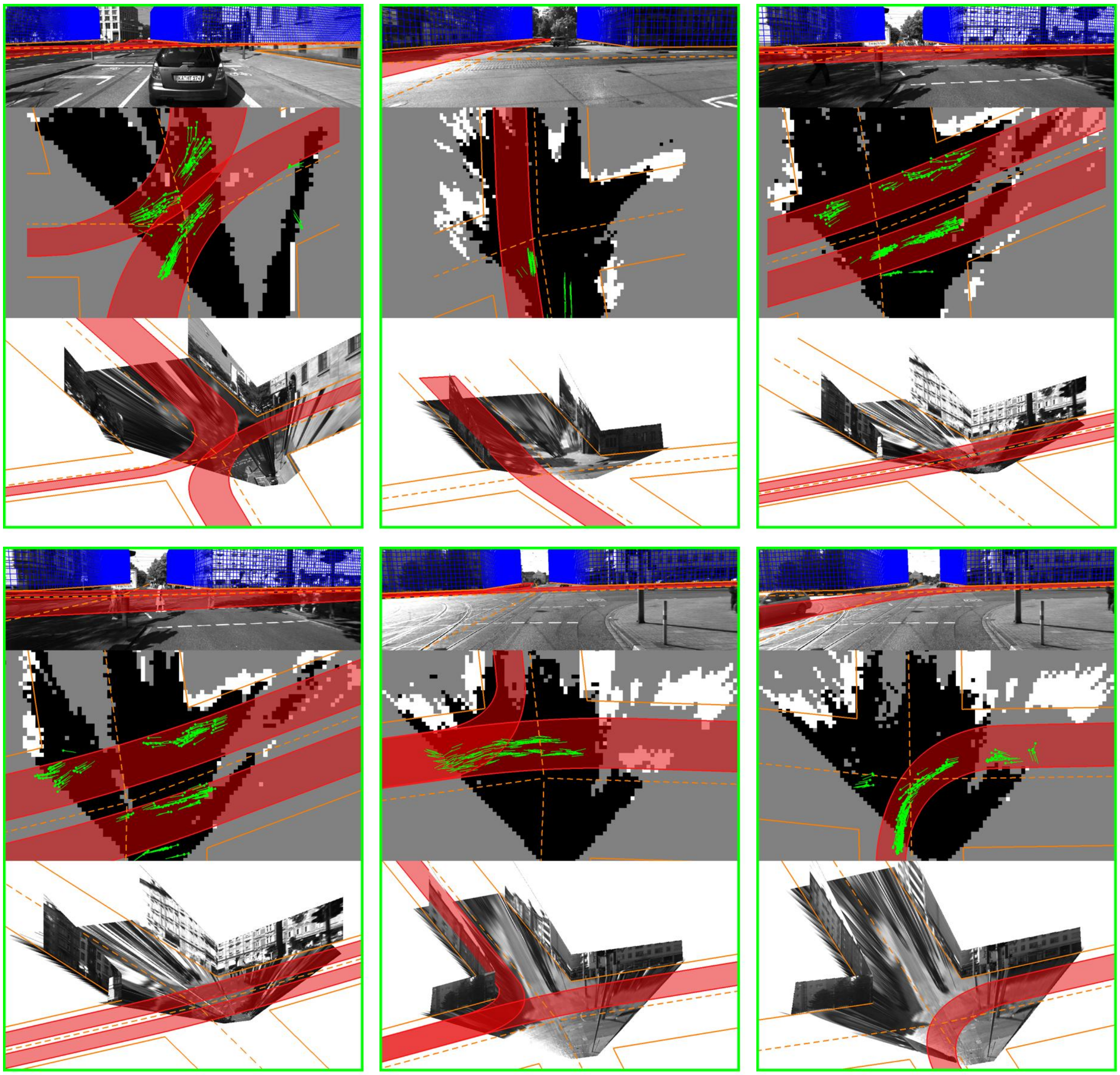
For each k, we use the derived bijection in order to map $\mathbf{o}$ to $\tilde{\mathbf{o}}$ for each data point and learn a non-parametric multimodal distribution on $\tilde{\mathbf{o}}$ using a Dirichlet Process Mixture model via a transformed Gaussian Mixture Model. The following figure illustrates, for $k = 3$, a sample from the posterior over $\tilde{\mathbf{o}}$ in both $\mathbb{R}^{k-1}$ and Δ^{k-1}:

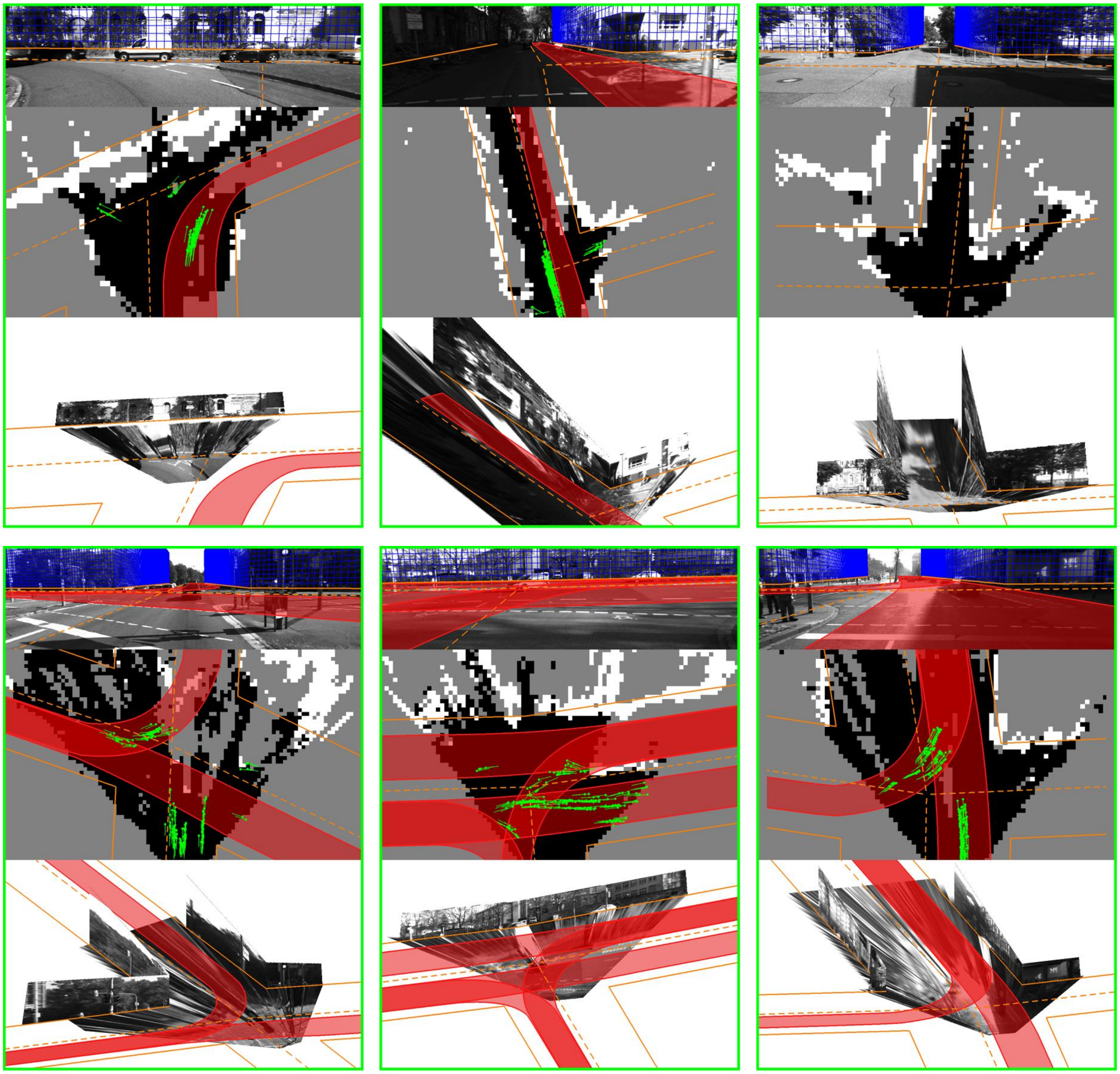


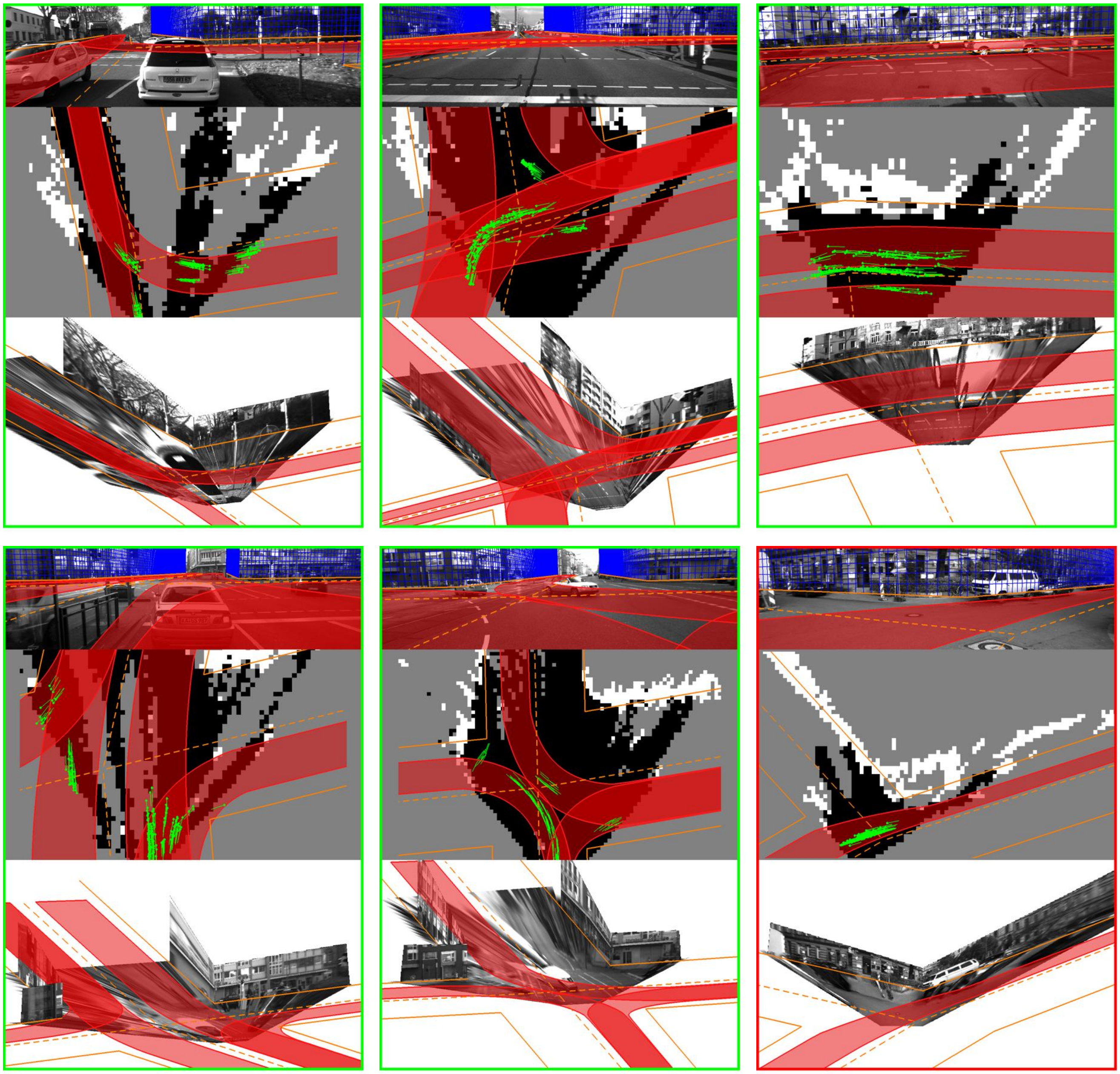
3. Inference Results on all 113 Video Sequences

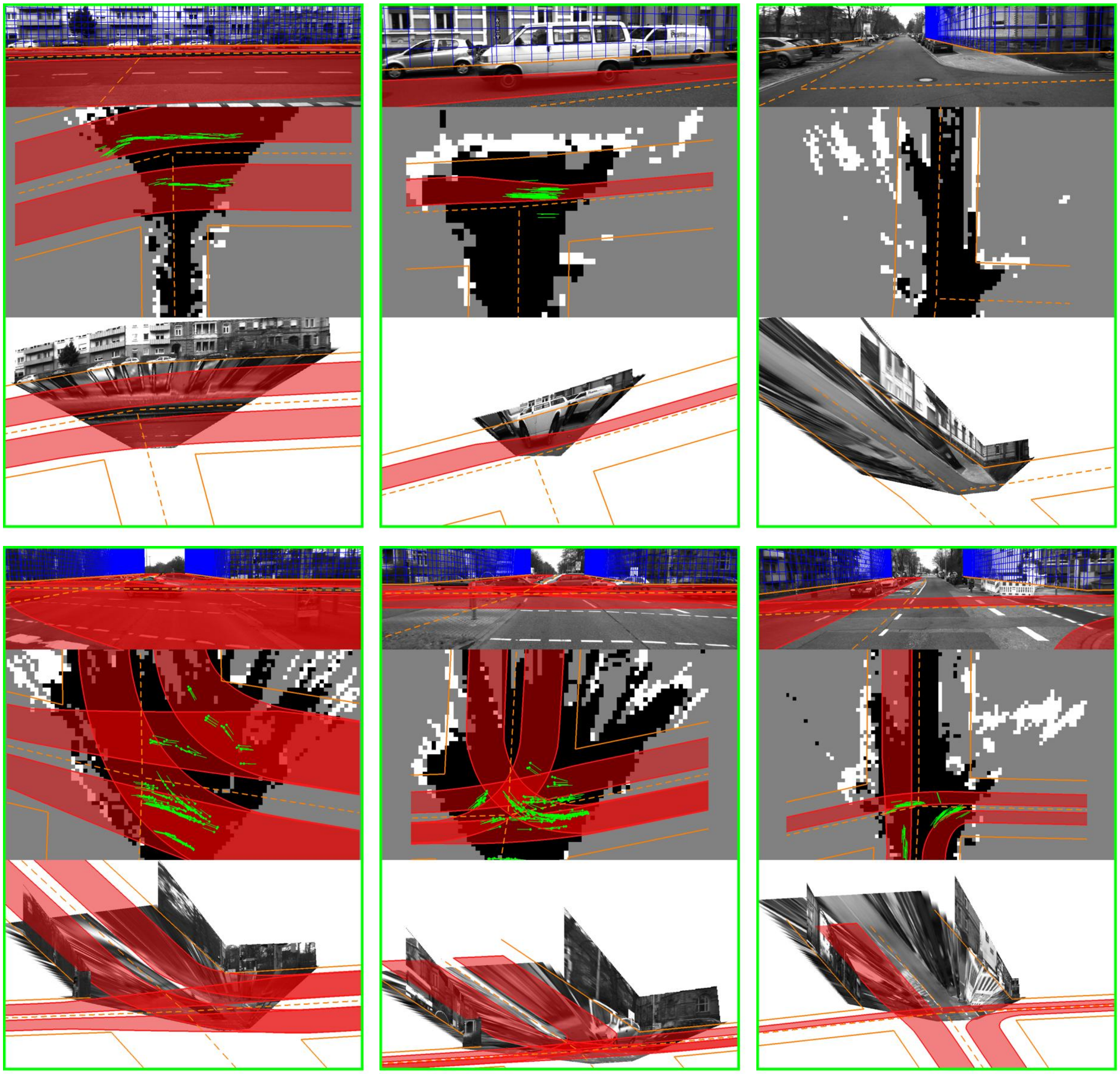
In this section, we depict the inference results of our approach on all 113 video sequences from the experimental evaluation of the paper. All sequences, for which k is estimated correctly are marked with a *green* frame, while the other ones are depicted in *red*. Note that in most of the latter cases, the estimated geometry is still plausible, given the observed features and the limited angular aperture.

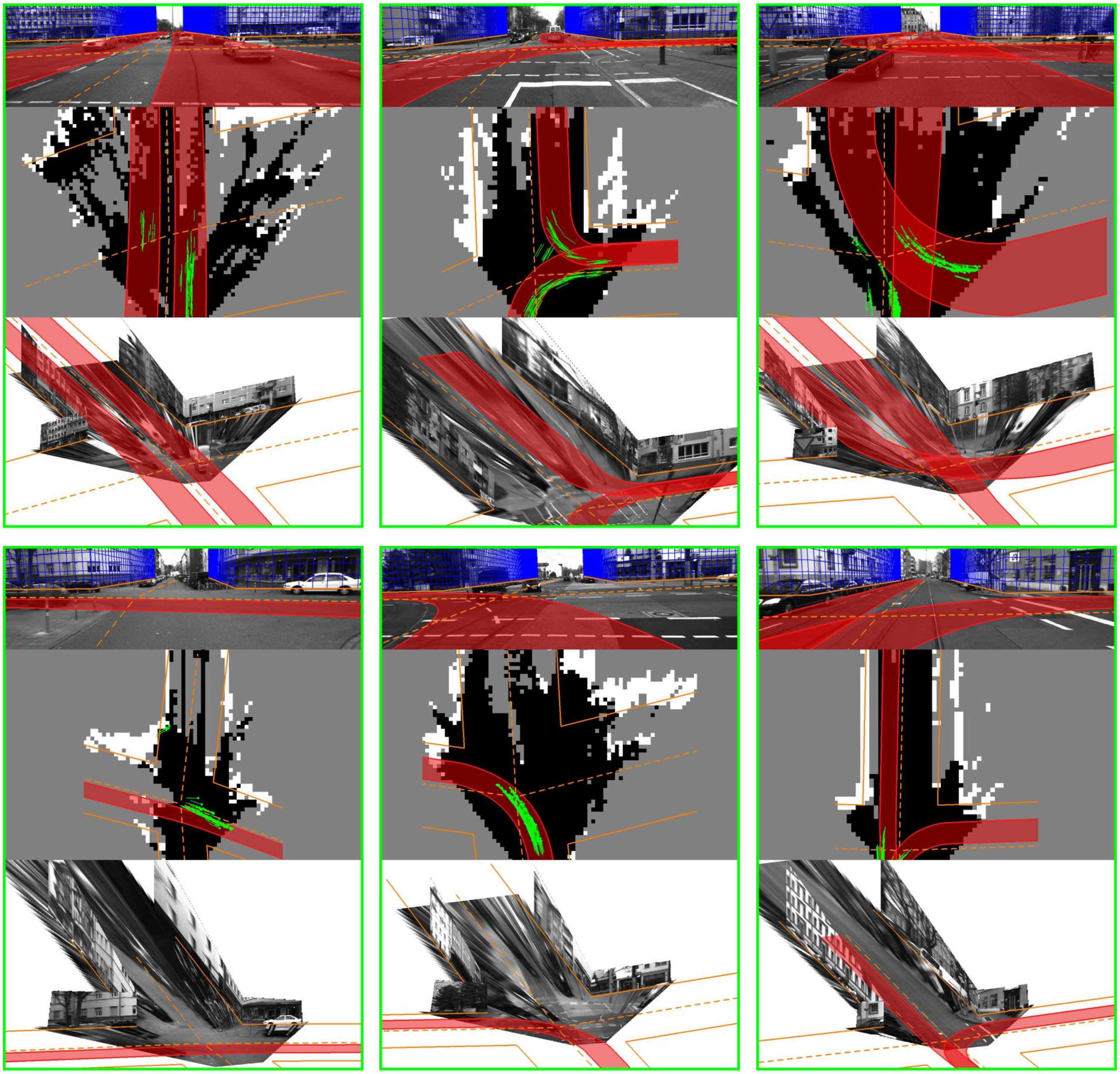


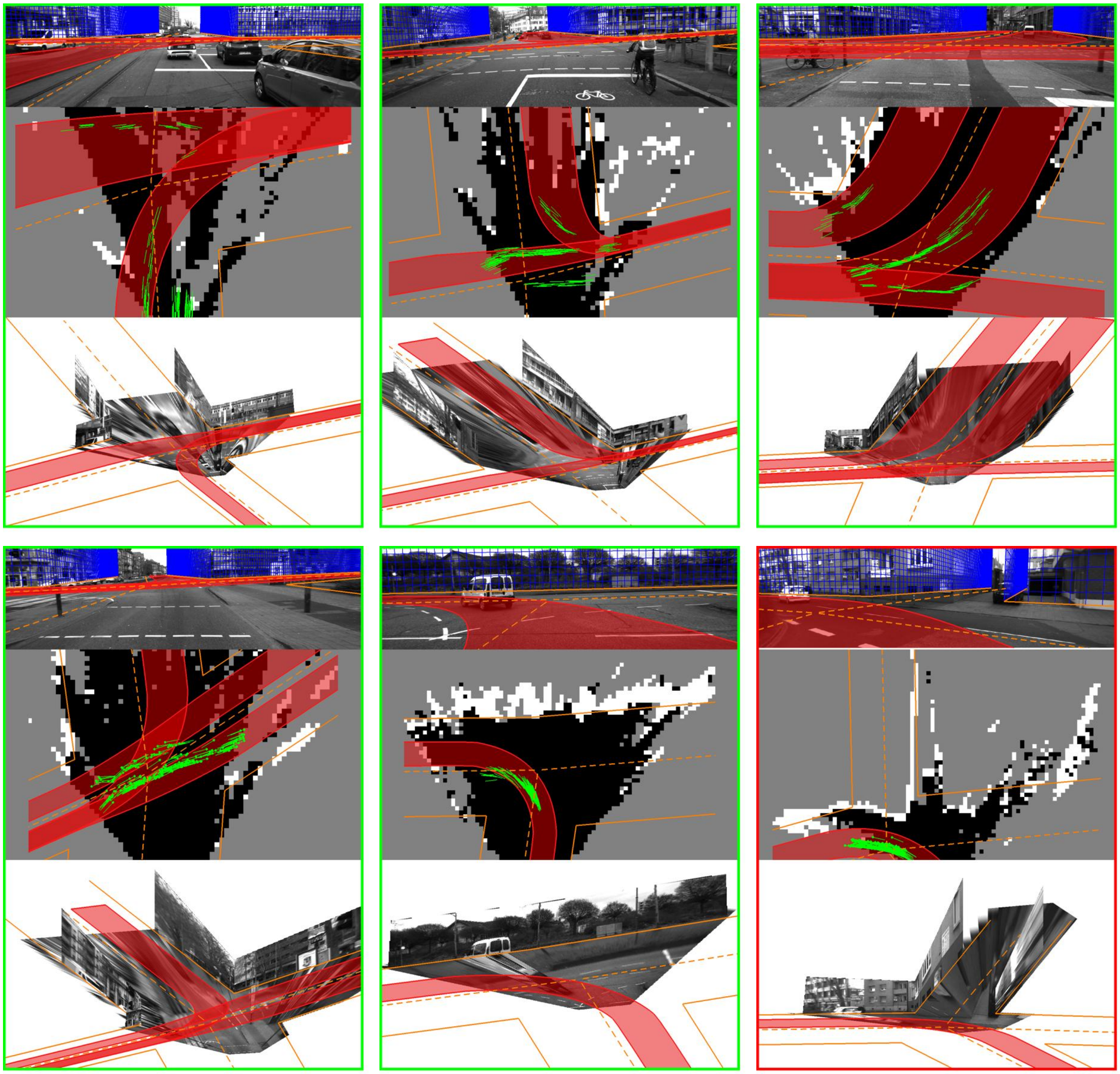


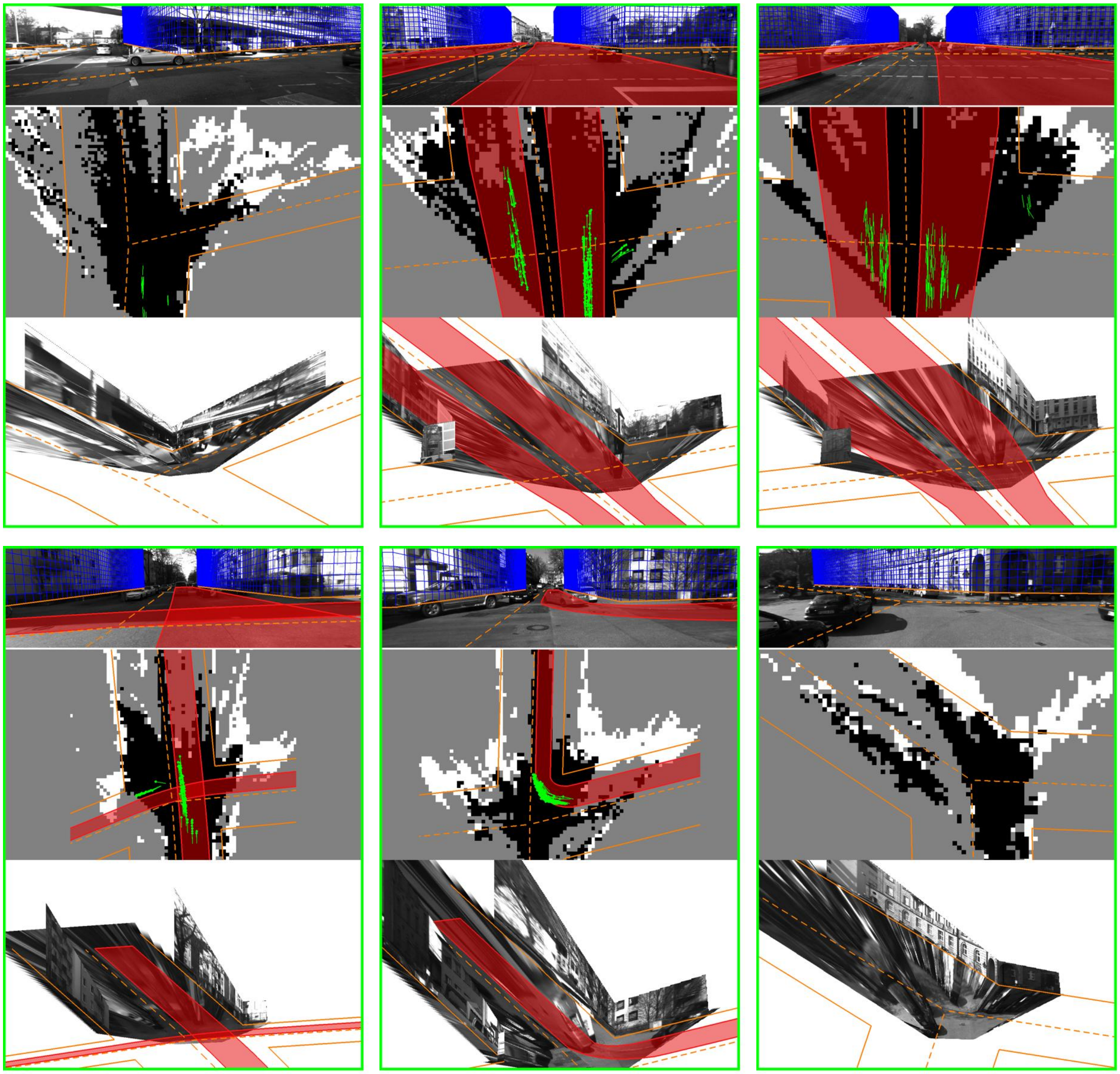


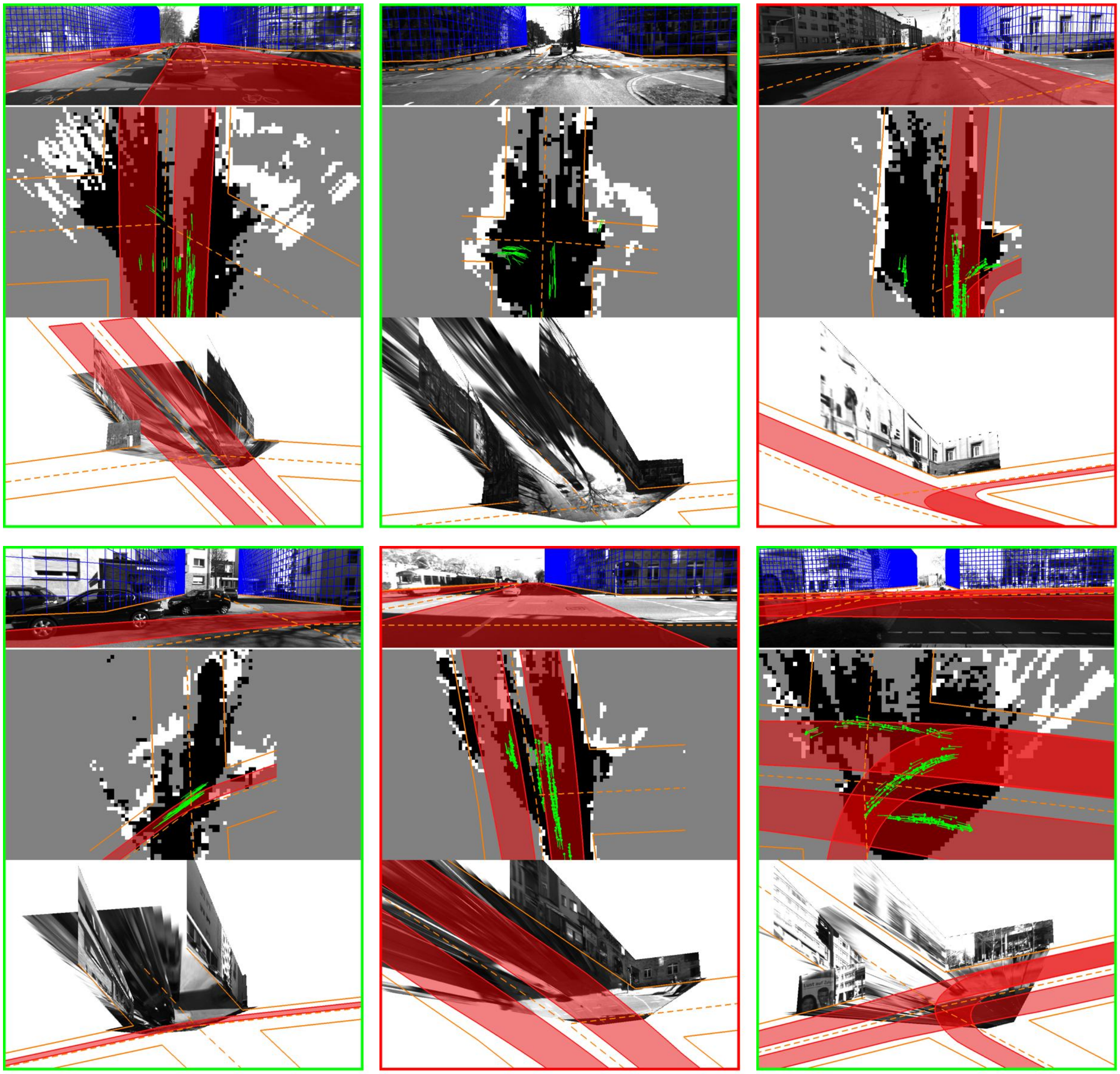


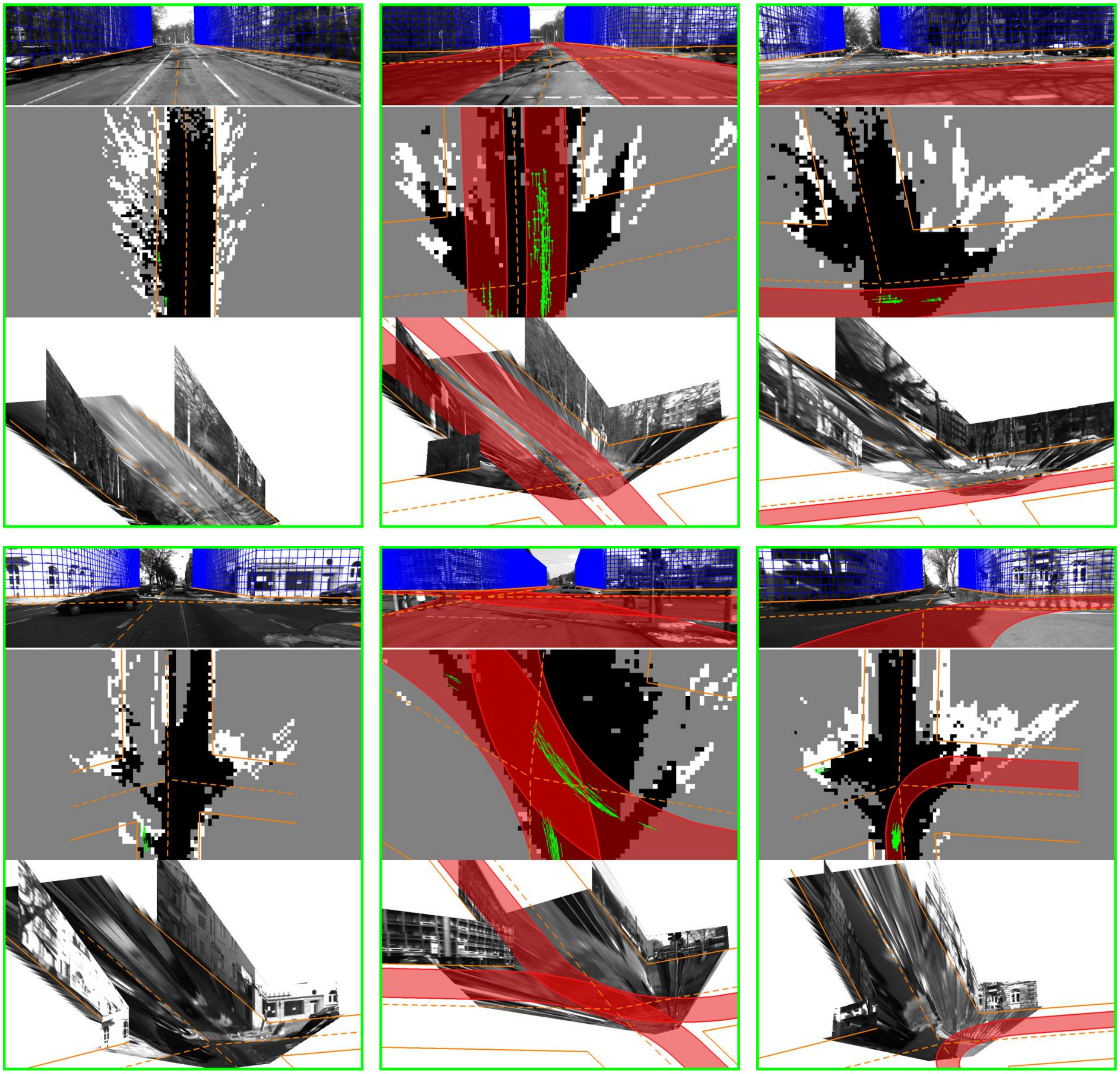


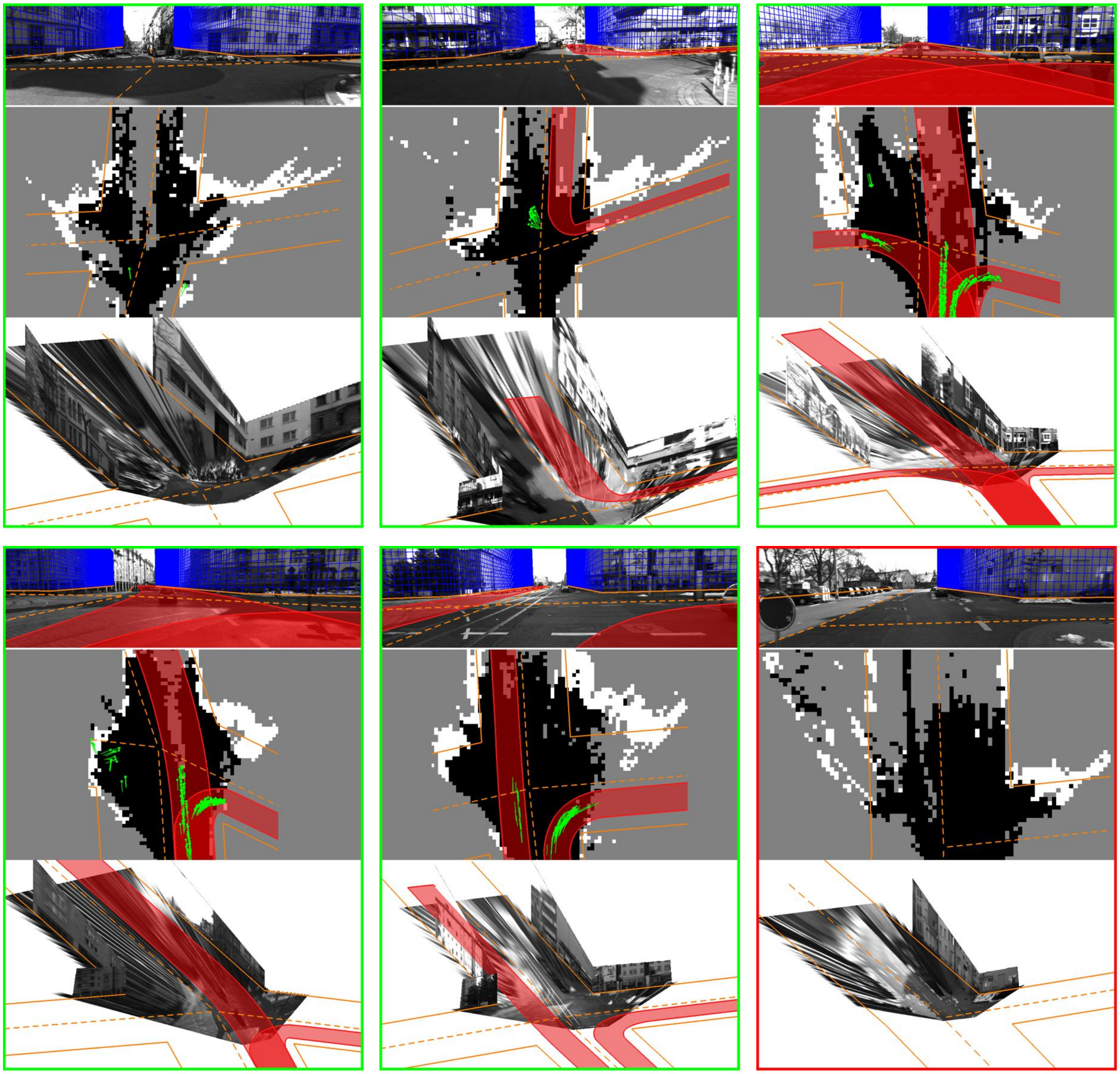


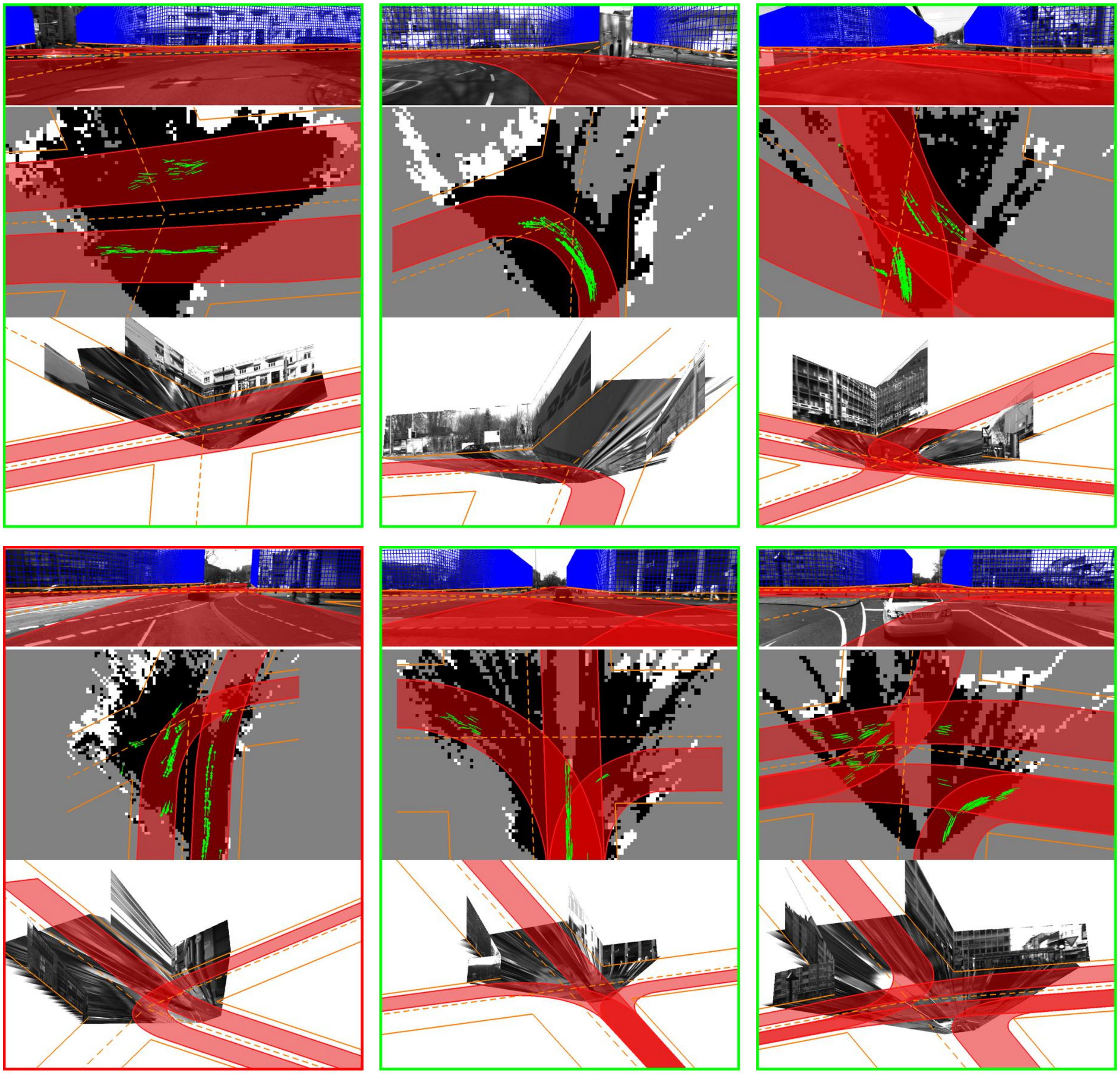


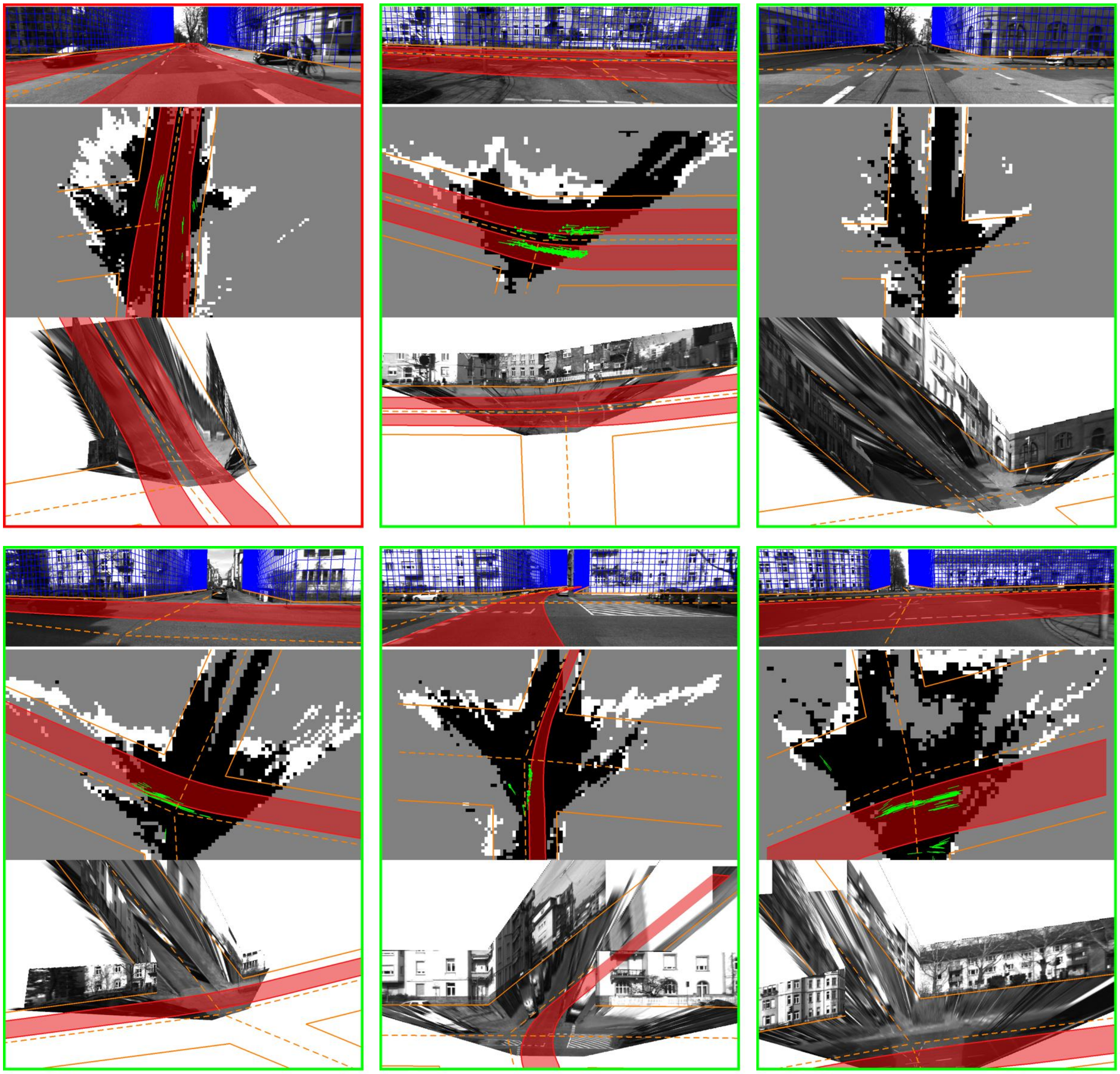


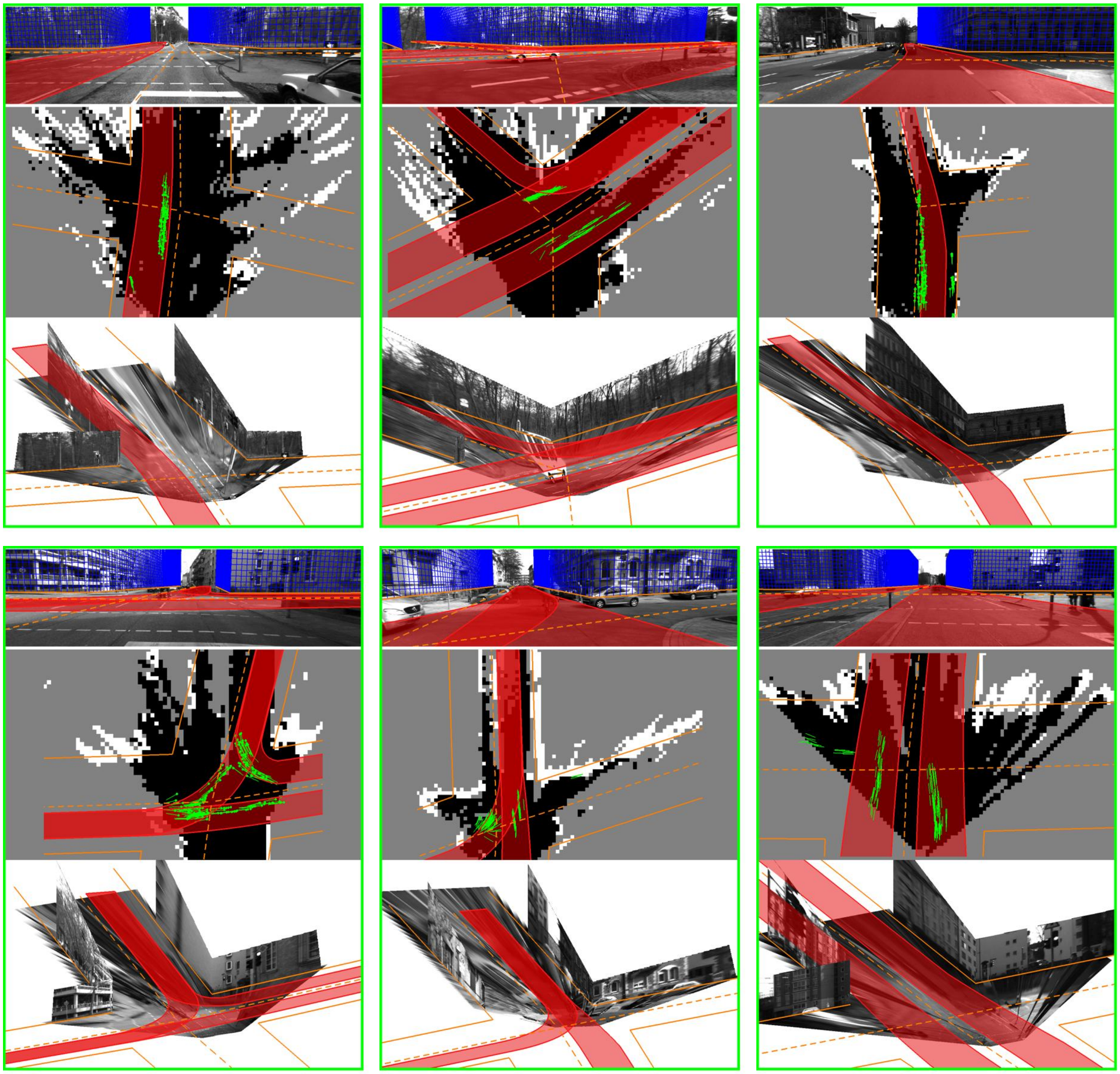


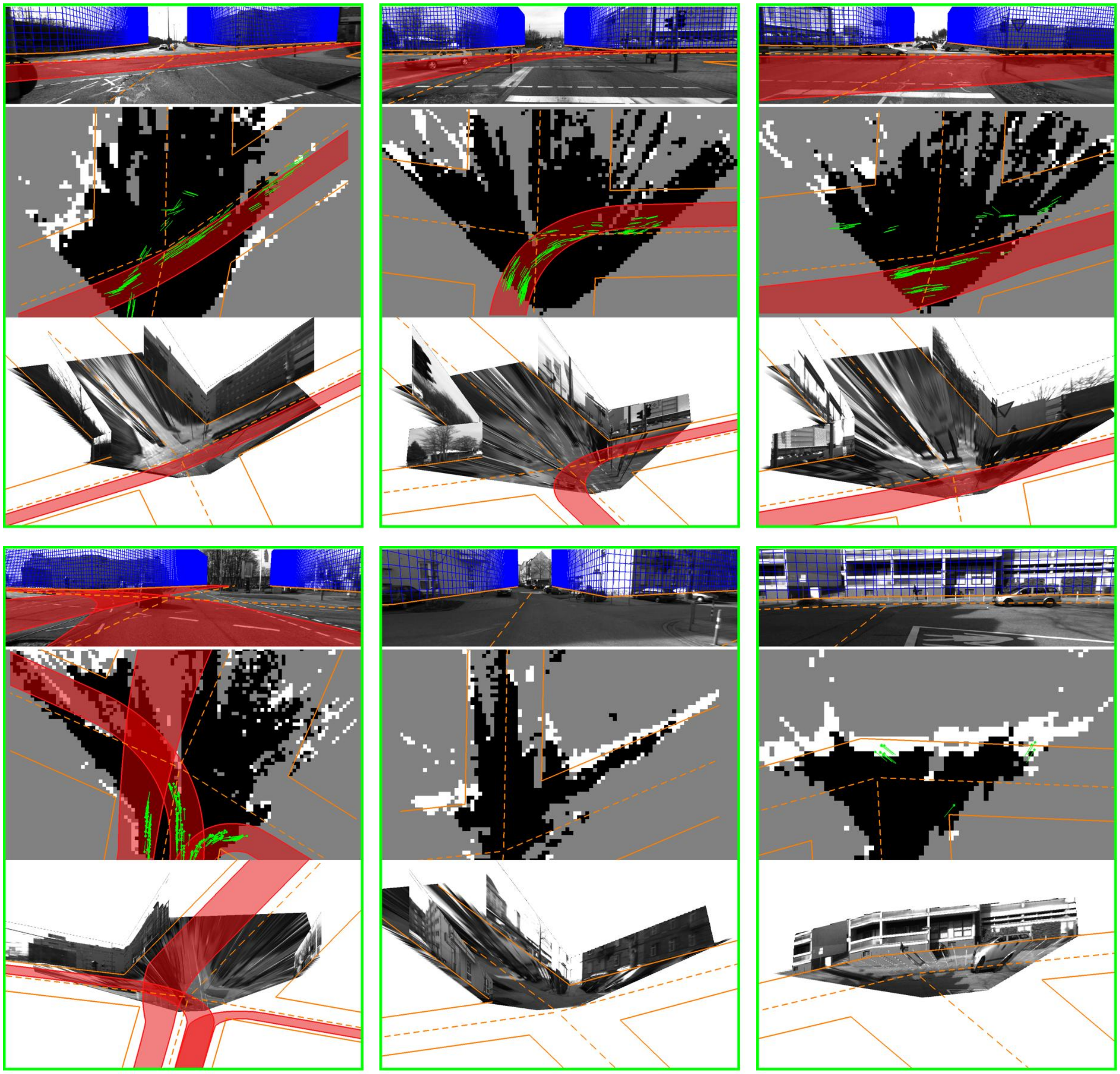


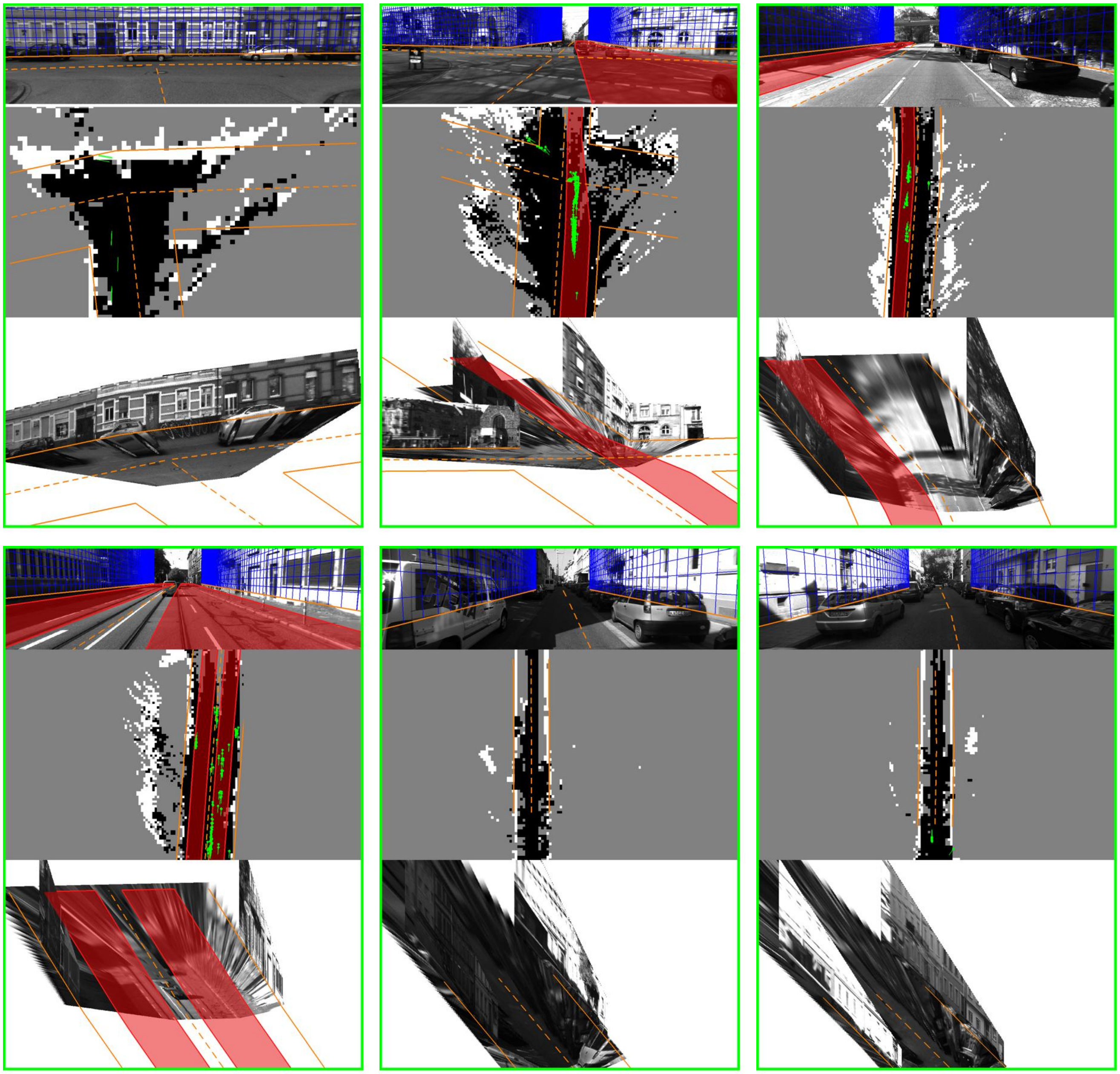


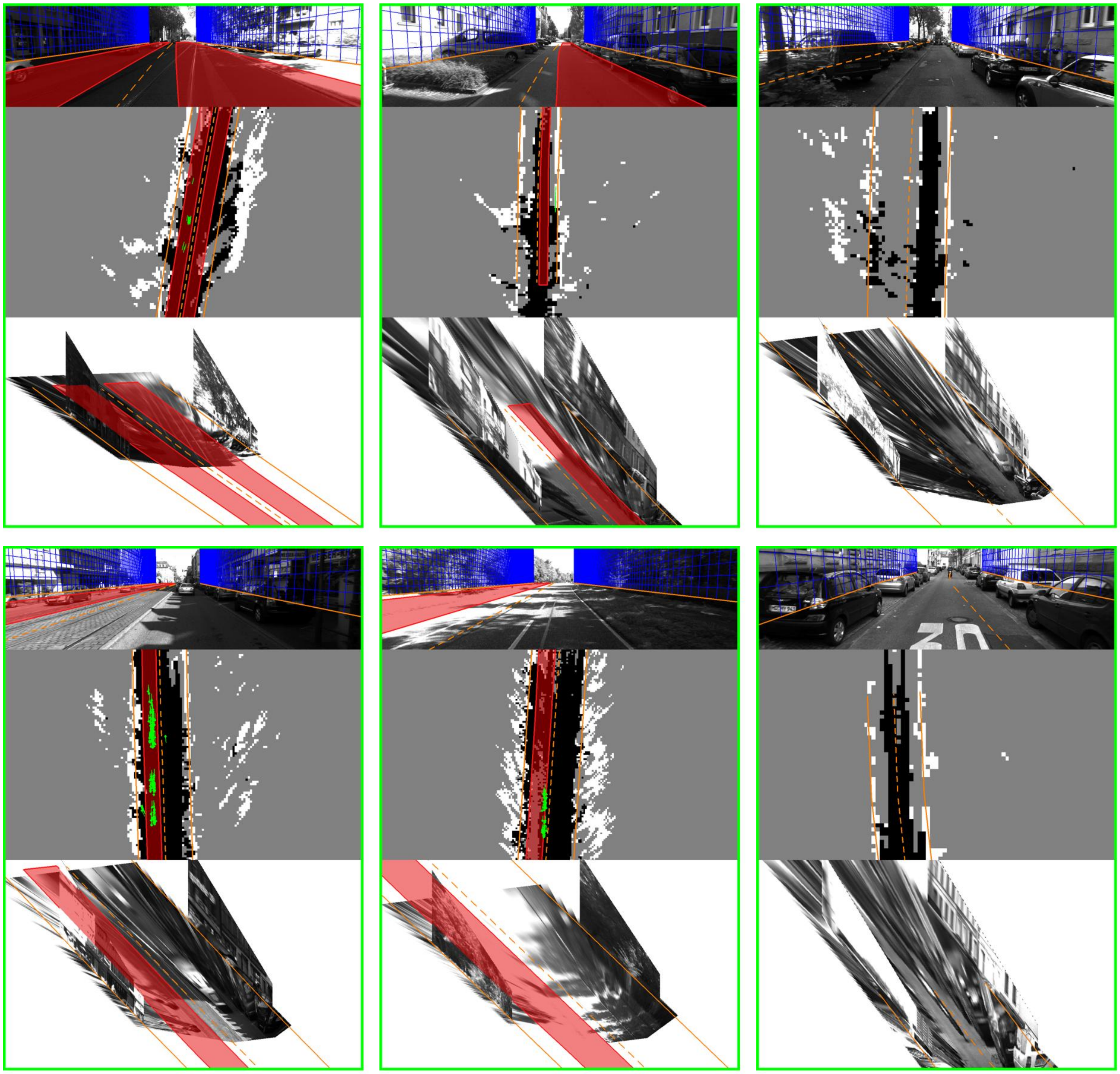


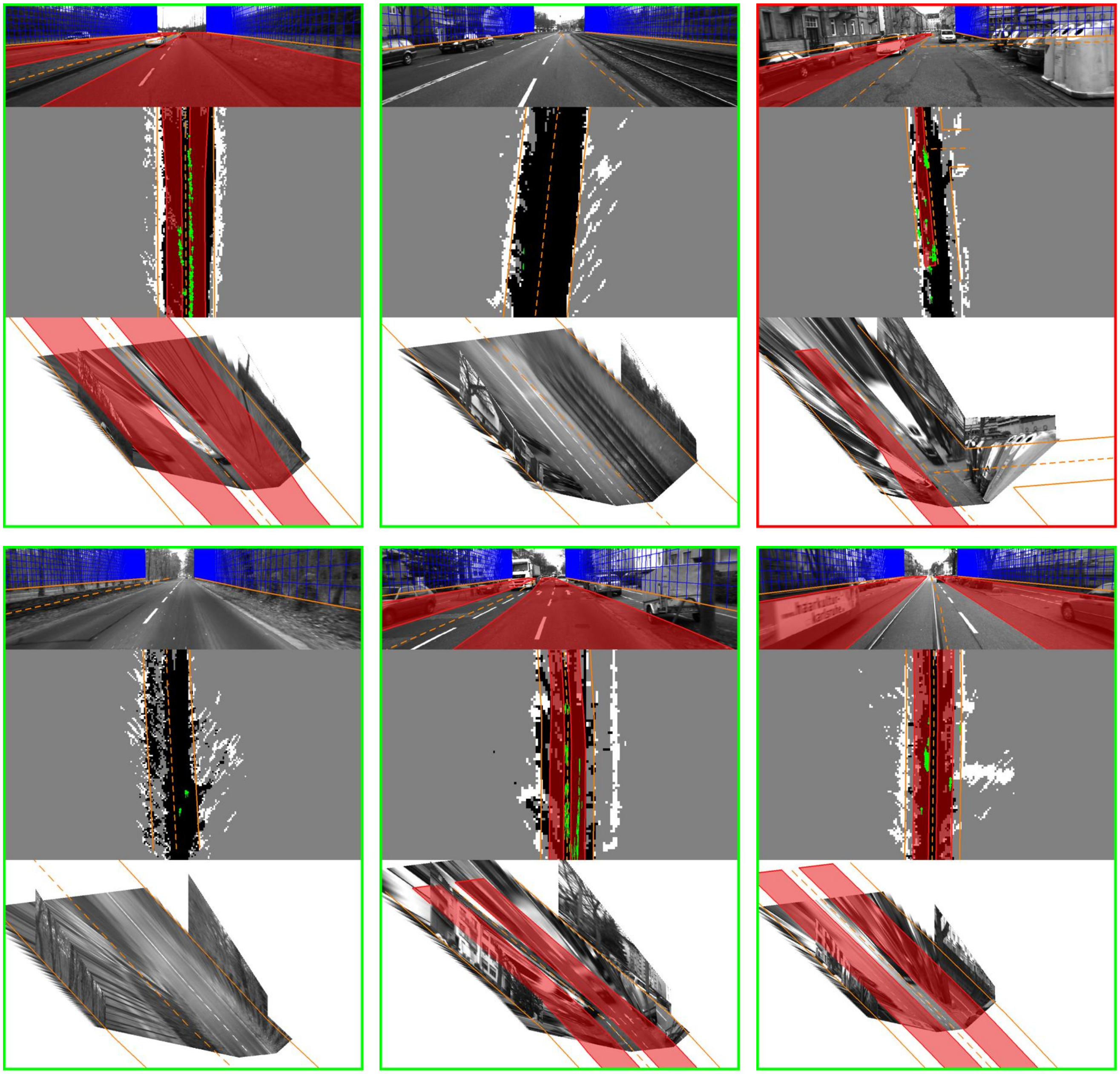


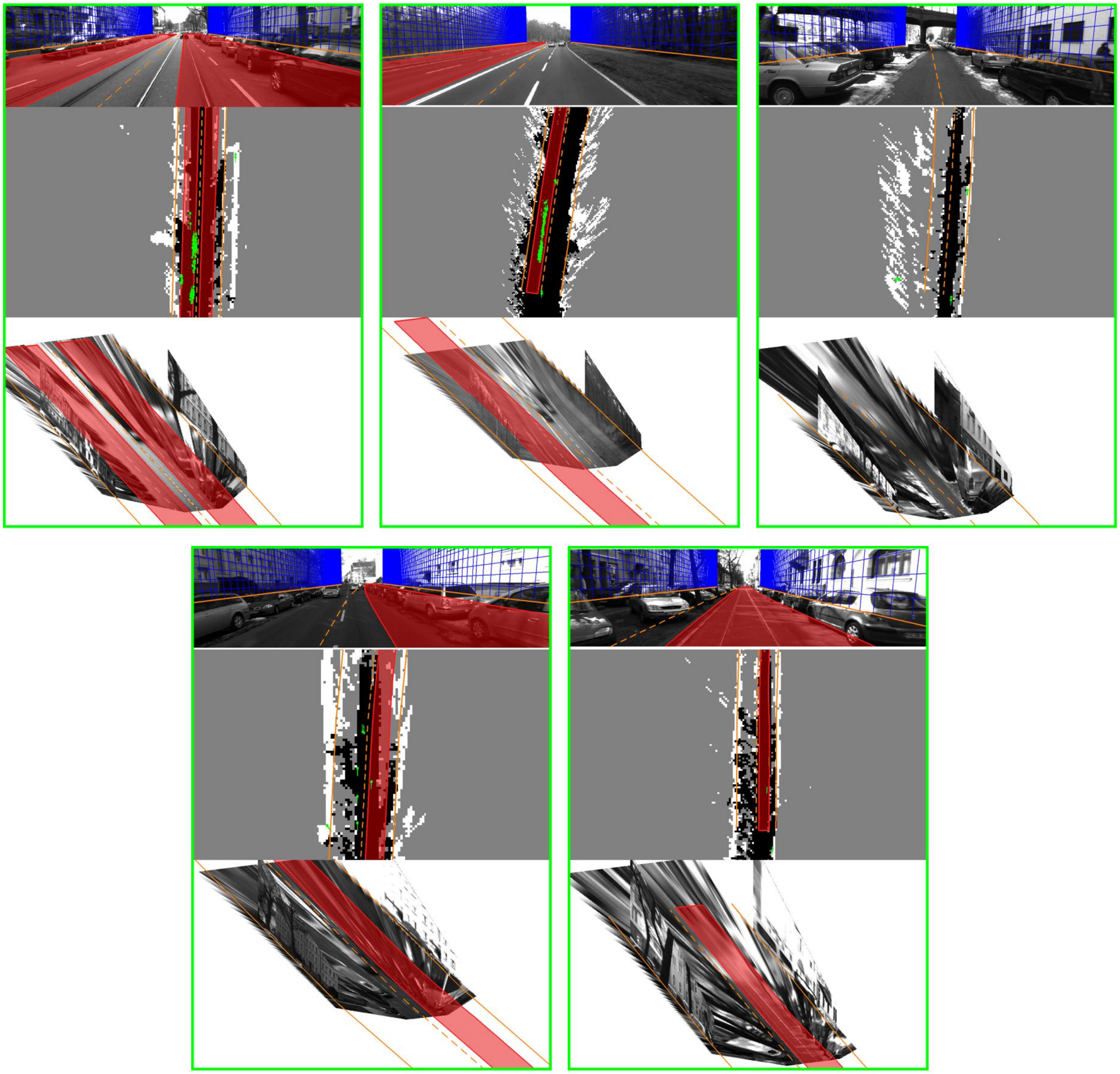












References

[1] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1-2):5–43, 2003. 1

[2] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restauration of images. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 6:721–741, 1984. 1

[3] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995. 1

[4] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970. 1