Localization is a key task for autonomous cars; systems such as the Google driverless car rely on precise and detailed maps for safe operation. Light detection and ranging (LIDAR) sensors are capable of providing rich information—including metric range and point appearance. Robust methods can use this data for vehicle localization by extracting the ground-plane for alignment to a prior map, as in [2].

Vision sensors as part of the localization pipeline can be a great enabler for autonomous platforms. Contrary to LIDAR methods, identifying the ground-plane from a camera image is a more challenging task. In our previous work [3], we considered localizing with a monocular camera by aligning the image to a prior map. As we demonstrated, this can be difficult as the ground-plane can be obscured by obstacles within view of the camera. In this work, we are interested in partitioning an image stream into obstacles and prior map, as shown in Fig. 1, so we can mask out obstacles during registration.

Similar to previous work [1, 4], we use a 1D-Markov random field (MRF) to model a horizontal image partition between obstacles and ground-plane, as in Fig. 1. However, rather than formulating our MRF potentials using image appearance alone (using learned [1] or hand-tuned features [4]), we instead consider the temporal stream of images and inferred parallax. We probabilistically evaluate optical flow against expected optical flow derived from known scene structure and camera egomotion, as in Fig. 2.

Our approach is evaluated on a challenging urban dataset with grayscale imagery, where lighting is non-uniform. We demonstrate our proposed algorithms by looking at errors with respect to hand-labeled groundtruth and present results showing improved image registration when obstacle masks are used.

Acknowledgements: This work was supported by a grant from Ford Motor Company via the Ford-UM Alliance under award N015392.