Probabilistic Obstacle Partitioning of Monocular Video for Autonomous Vehicles

Ryan W. Wolcott
rwolcott@umich.edu
Ryan M. Eustice
eustice@umich.edu

Ford Motor Company
Dearborn, MI
Perceptual Robotics Laboratory
University of Michigan
Ann Arbor, MI, USA

Abstract

This paper reports on visual obstacle detection from a monocular camera for autonomous vehicles. By leveraging a textured prior map, we propose a probabilistic formulation for finding the optimal image partition that separates obstacles from ground-plane. Our key insight is the use of a prior map that enables ground appearance models conditioned on prior map texture and a probabilistic optical flow vector formulation derived from known scene structure and camera egomotion. We evaluate our methods on a challenging urban setting using data collected on our autonomous platform and we demonstrate that a notion of obstacles in the camera frame can improve visual localization quality.

1 Introduction

Localization is a key task for autonomous cars; systems such as the Google driverless car rely on precise and detailed maps for safe operation [18]. 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 done by Levinson et al. [11, 12].

Due to decreased cost and the ability to have robust, redundant sensing, vision sensors as part of the localization pipeline can be a great enabler for autonomous platforms. Contrary to LIDAR approaches, identifying the ground-plane from a camera image is a much more challenging task. In our previous work [21], we considered localizing with just a monocular camera by aligning the whole image to a prior map. This can be problematic as the ground-plane can frequently be obscured by obstacles within view of the camera. We showed that our visual localization system can be distracted when the image is dominated by obstacles, leading to a degradation in localization.

In this work, we are interested in partitioning an image stream into obstacles and prior map as shown in Fig. 1, with the goal of only using the portions of the image containing the prior map for localization. This addition will lend itself to a more robust end-to-end visual localization system.

We propose to leverage our textured prior map, consisting of a ground-plane mesh, to formulate a Markov random field (MRF) that models the image partition between the obstacles and the ground-plane. We present several probabilistically motivated energy functions

Pages 128.1-128.12
DOI: https://dx.doi.org/10.5244/C.30.128

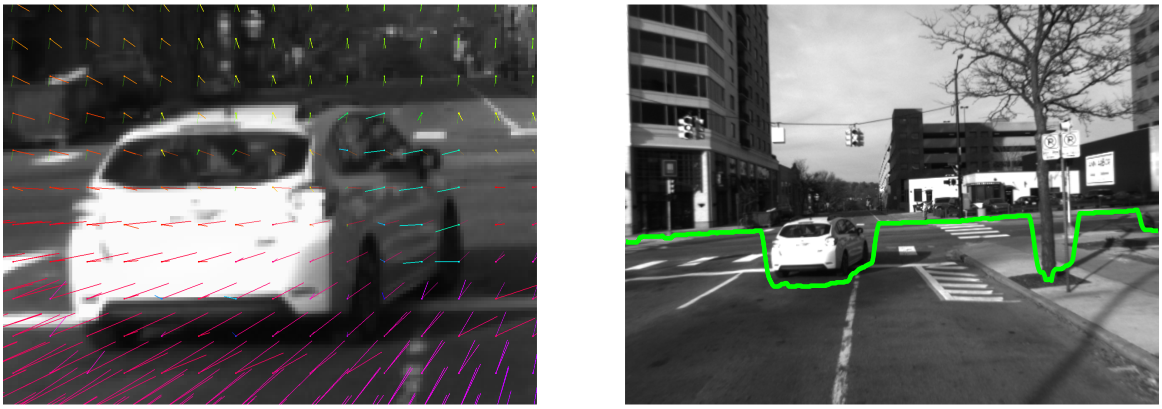


Figure 1: In this work, we extract optical flow vectors and probabilistically evaluate them against expected flow vectors (left). We present an MRF framework that fuses various energy potentials, including an optical flow derived potential, such that minimizing the energy results in a partition of the image into ground-map and obstacles, as shown on the right.

that can be fused in this MRF framework. Specifically, the prior map allows us to evaluate ground likelihood by conditioning our belief on the expected appearance from the prior map. Moreover, we present a probabilistic method to evaluate optical flow likelihood against our three-dimensional (3D) prior map, taking into account expected motion parallax.

Our proposed approach is evaluated on a challenging urban dataset where lighting is non-uniform and our camera is an 8-bit monochrome sensor; note, explicitly no color imagery is used to demonstrate effectiveness of our approach that relies more on observed *motion* than visual appearance. 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.

1.1 Related Work

Modeling the ground plane appearance distribution directly from image data has been successful in many domains. Ulrich and Nourbakhsh [17] build a histogram appearance model for the ground plane, learning this distribution with the assumption that the bottom of the image is mostly ground plane. Dahlkamp et al. [4] improves on this by restricting appearance learning to within co-registered laser range finder returns. In addition, they and Álvarez and Ĺopez [1] use an RGB colorspace transform to minimize the effect of shadows by actively removing them from their appearance model. These works heavily rely on color images that are clearly more discriminative than grayscale images.

Others have looked to exploit camera motion to infer scene structure and motion. Considering a temporal stream of images, Zhang et al. [23] looked at the residual error from focus of expansion estimation. Similar to our proposed work, others have assumed a locally planar ground in which motion can be inferred [13] or provided via odometry [3]. Moreover, Wedel et al. [20] proposed classifying between foreground and background by warping sequential images onto multiple plane hypotheses.

In this work, we are interested in computing dense optical flow fields and evaluating them against a likelihood measure. The use of optical flow for obstacle detection from a moving vehicle was first looked at in [5] and [9], where optical flow vectors are *sparsely* extracted and compared against estimated model flow vectors. Roberts and Dellaert [15] performed a similar classification employing dense flow fields, though found problems when faced with textureless image regions.

Similar to our work, McManus et al. [14] applied optical flow for background detection on an autonomous vehicle assuming an already known localization within a 3D prior map. They evaluate the likelihood of this optical flow by computing optical flow twice—first on the raw images then on the image warped via the 3D prior map—and comparing the flow vectors.

Badino et al. [2] proposed a novel idea called the “Stixel World” in which image processing demands can be significantly reduced under the context of on-vehicle cameras. The representation is such that the world can be decomposed into a set of vertical *stixels* that directly correspond to a column in image space. A key insight here is that the pixels between the bottom of the image and the first obstacle in each column is strictly identified as free-space—thus imposing a 1D image space partitioning that can be efficiently solved for using dynamic programming.

Our work is quite similar in underlying machinery to more recent work by Yao et al. [22] and Levi et al. [10], which closely resemble the stixel-world formulation with a monocular camera. In [22], they propose inference in a 1D MRF that incorporates various cues including pixel appearance, image edges, temporal consistency, and spatial smoothness. However, many of these cues are severely biased towards the bottom of the images, leading to a brittle system when faced with difficult imagery (e.g., shadows). In [10], they use a convolutional neural network (CNN) to offline learn the appearance of the image partition. Both of these methods rely on learning the appearance of the image partition and do not leverage the temporal stream of images; thus, we propose a new set of cues that are probabilistically motivated to jointly reason over appearance and perceived motion from optical flow.

2 Preliminaries

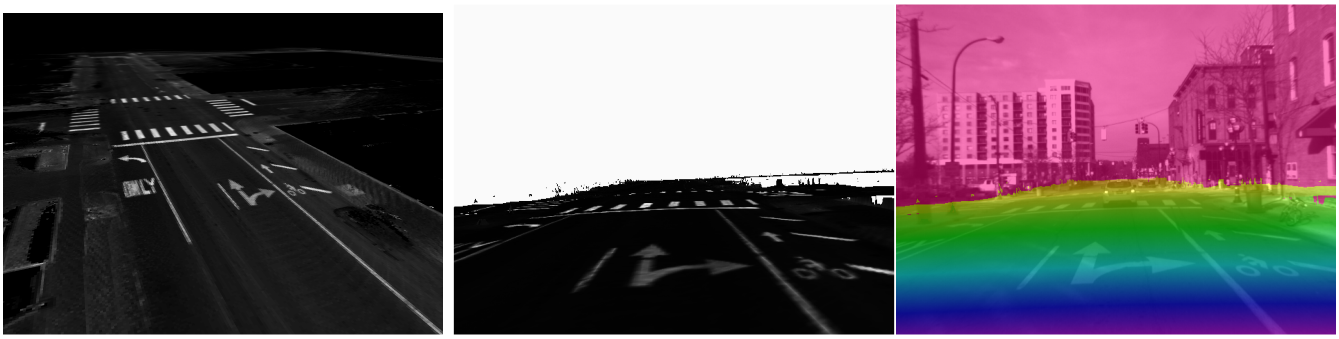
In our work, we use a survey vehicle equipped with 3D LIDAR scanners to construct a detailed prior map for localization. As presented in [21], we build a 3D mesh of the ground-plane that we texturize using reflectivity measurements from the LIDAR, as shown in Fig. 2.

We then localize an image, I_t, taken at time t from a monocular camera within this prior map, $\mathcal{M}$, by exploiting the statistical dependency between camera intensity values and LIDAR reflectivities. Using a coarse prior (such as that from GPS), we generate several synthetic views of the prior map, maximizing normalized mutual information (NMI):

$$\hat{\mathbf{x}}_t = \operatorname*{argmax}_{\mathbf{x}} \mathrm{NMI}\left(I_t, L_t\right), \tag{1}$$

where $L_t = \texttt{proj}(\mathcal{M}, \mathbf{x})$ is the synthetic LIDAR image generated by projecting $\mathcal{M}$ into the camera frame at $\mathbf{x} = [x, y, z, r, p, h]^\top$, using the standard pinhole camera model. NMI is a normalized variant of mutual information that is maximized by minimizing the dispersion between two random variables (a metric that is evaluated with the entropy of the joint and marginal histograms of the two signals).

The projections for localization can be done efficiently within OpenGL using custom shaders. Further, the OpenGL rendering process populates a depth buffer to determine screen ordering of drawn triangles. This depth buffer can be scaled by the *near* and far clipping planes to generate an expected depth image $\hat{Z}_t$. Thus, the localization process provides expected depths for a given camera location, which we leverage for obstacle partitioning.



(a) Prior Map (b) Synthetic Image (c) Image with Expected Depth

Figure 2: Using a survey vehicle equipped with 3D LIDAR scanners, we can offline generate a rich mesh of the ground-plane colored by LIDAR reflectivity, as shown in (a). OpenGL is used to generate synthetic viewpoints and expected depth of this prior map, (b) and (c). The synthetic image and depth are relied upon for obstacle partitioning.

In the following sections, we detail how we can estimate prior map likelihood and then how we can incorporate these likelihoods into a Markov random field (MRF) smoothing framework.

3 Probabilistic Obstacle Partitioning

Our proposed formulation is heavily motivated by the Stixel World presented by Badino et al. [2] and the similar monocular approach for free space estimation by Yao et al. [22]. Realizing the structure of the roadway as viewed in a camera image, they assume that there is a distinct separation between free space and obstacles. This defined partition regularizes the task of identifying obstacles in a camera image. We propose to use a sequence of camera images to derive probabilistic appearance and motion likelihoods to find this partition.

Given an image I_t taken at time t, probabilistic obstacle partitioning seeks an optimal seam that traverses the image left-to-right, $S = \{s_i\}_{i=1}^{w}$, where s_i can take the value of $h+1$ labels, $s_i \in \{0,\cdots,h\}$ (w and h denote the width and height of I_t). Considering the i^{th} column of I_t, $\mathbf{c_i} = \{I_t(i,j)\}_{j=1}^{h}$, the cut s_i implies a partitioning of this column into two disjoint sets such that $\{I_t(i,j)\}_{j=1}^{s_i}$ is sampled from the obstacle set, $\mathcal{O}$, and $\{I_t(i,j)\}_{j=s_i+1}^{h}$ is sampled from the prior map, $\mathcal{M}$; here, the prior map refers to the ground-only prior map (as static 3D structure is not in our maps). In our framework, $i=1$ indicates the leftmost column and $j=1$ indicates the topmost row of the image. An illustration of this is provided in Fig. 3.

We formulate obstacle partitioning as the maximum *a posteriori* (MAP) estimation of the set of column seams conditioned on the previous n camera images,

$$S^* = \operatorname*{argmax}_S p(S|I_t,\cdots,I_{t-n+1}). \tag{2}$$

Assuming a Markov factorization, we can factor the posterior as

$$\begin{aligned} p(S|I_t,\cdots,I_{t-n+1}) &\propto p(I_t,\cdots,I_{t-n+1}|S)p(S) \\ &= \prod_i p\left(I_t,\cdots,I_{t-n+1}|s_i\right)\prod_j p\left(s_j|s_{j-1}\right), \end{aligned} \tag{3}$$

where we assume independence between columns $\mathbf{c_i}$, given the column partition s_i. Applying the negative log-likelihood, the MAP inference results in the following energy function to be

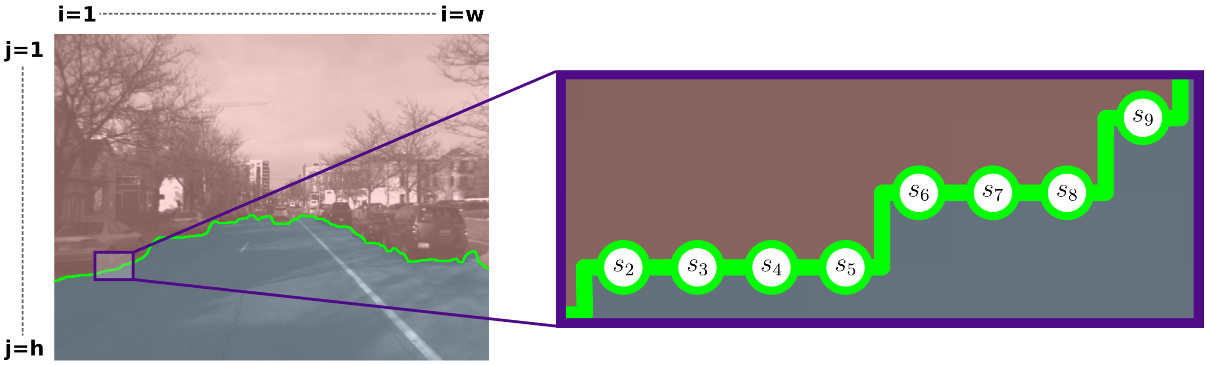


Figure 3: We use a 1D MRF to partition images into two sets: ground-plane (*blue*) and not ground-plane (*red*). Each variable in our MRF (*green* nodes) models a partition point for each column in the image, and has unary potentials computed from a set of partition likelihoods and pairwise potentials to enforce smoothness.

minimized:

$$E = \sum_i \sum_{k \in K} \underbrace{w_k \phi_k (s_i)}_{\text{unary}} + \sum_j \underbrace{w_p \phi_p (s_j, s_{j-1})}_{\text{pairwise}}, \tag{4}$$

where K represents the set of unary potentials, $\{a, f, e, l, r\}$, and w_n represents the weighting for each potential, which can be learned using training data as in [22]. The MRF forms a chain connecting neighboring columns and can be efficiently solved using dynamic programming as a Viterbi problem [19].

The pairwise potential is modeled as a truncated quadratic to enforce smoothness across the seam, $\phi_p \left(s_j, s_{j-1}\right) = \min(|s_j - s_{j-1}|, T_p)^2$, where T_p is a threshold that allows the potential to enforce local smoothness without penalizing large jumps, as should be allowed with objects near the camera. The remainder of this section details the unary potentials that are used in this energy function, which exploit appearance and motion in the images.

3.1 Appearance Potential

We derive an appearance based potential that can be learned online using a monochrome camera. The theory could easily be applied to color imagery, though we opted against to demonstrate the effectiveness of our motion potential presented next (color can be an extremely discriminative feature in this context).

The motivation for this potential is to maximize the likelihood of the class assignments (obstacle and prior map) using image intensities. The potential is defined as

$$\phi_a (s_i) = -\log p \left(\mathbf{c_i} | s_i\right), \tag{5}$$

where $\mathbf{c_i}$ is the set of pixels in the i^{th} column and the likelihood term is derived assuming independence and recalling the strict partitioning of the data at s_i:

$$p\left(\mathbf{c_i}|s_i\right) = \prod_{j=1}^{h} p(I_t(i,j)|s_i) \tag{6}$$

$$= \prod_{j=1}^{s_i} \underbrace{p(I_t(i,j)|\mathcal{O})}_{\substack{\text{obstacle} \\ \text{likelihood}}} \prod_{j=s_i+1}^{h} \underbrace{p(I_t(i,j)|\mathcal{M})}_{\substack{\text{prior map} \\ \text{likelihood}}}. \tag{7}$$

The obstacle appearance model is maintained using a joint histogram for $p(I_t(i,j)|\mathcal{O}) = p(I_t(i,j)|\mathcal{O},i)$. This can be thought of as a 2D histogram with image intensity on one axis and image column on the other. We convolve this with a Gaussian kernel so as to avoid over-fitting and smooth the likelihoods.

The prior map appearance model is more intricate in that intensity is conditioned on the LIDAR reflectivity in the prior map; therefore, $p(I_t(i,j)|\mathcal{M}) = p(I_t(i,j)|L_t(i,j))$. This conditional distribution is managed via the joint histogram over image intensity and LIDAR reflectivity—this is the same distribution used to compute NMI for localization. Note, this requires the camera to be localized as outlined in §2 to condition on the expected appearance from the LIDAR prior, which is a reasonable assumption on an autonomous car.

Both of these conditional histograms are learned online using the previous n images and extracted seams. Combined with the other potentials and the smoothing pairwise potential, the appearance prior continuously learns the obstacle and prior map distributions. In this work, we used a sliding time window over the last several seconds of data—this should be kept short so distributions can quickly adapt to lighting changes.

3.2 Optical Flow Potential

Appearance potentials alone can perform quite poorly in complex environments where partial illumination can distract the measure. Moreover, 8-bit grayscale imagery makes it difficult to differentiate between cars and roadways, resulting in weak appearance models. In this section, we present a motion potential derived from evaluating the likelihood of optical flow vectors—with the expectation that this can invalidate distracted areas due to parallax and physically moving objects. This illumination robust measure can further aid the appearance potential by maintaining the partition through complex lighting transitions so that the appearance models can adapt to new lighting distributions.

Optical Flow Likelihood: We first extract optical flow vectors $\mathcal{U}_t = \{\mathbf{u_1},\dots\mathbf{u_w}\}$, where $\mathbf{u_i}$ denotes a column of optical flow vectors, $\mathbf{u_i} = \{\mathbf{f}_{i,1},\dots,\mathbf{f}_{i,h}\}$ and $\mathbf{f}_{i,j} = [u_{i,j}, v_{i,j}]^\top$ is the optical flow at pixel (i,j). We use known egomotion $\mathbf{x_e} = [x,y,z,r,p,h]^\top$ derived from vehicle odometry, an estimate on the motion uncertainty Σ_e, and the expected scene depth, $\hat{Z}_t$, as outlined in §2, to calculate the expected optical flow measurement using the homogeneous point transfer [8]:

$$\mathbf{v}_{t-1} = \mathrm{KRK}^{-1}\mathbf{v}_t + \mathrm{K}\mathbf{t}/\hat{Z}_t(i,j), \tag{8}$$

where $\mathbf{v}_{t-1} = [x,y,1]^\top$ represents the expected homogeneous pixel location in I_{t-1} of $\mathbf{v}_t = [i,j,1]^\top$ (a homogeneous pixel in the current image I_t). Further, K represents the pinhole camera calibration matrix and $[\mathrm{R}|\mathbf{t}]$ is the camera motion derived from $\mathbf{x_e}$. Therefore, the expected optical flow measurement is $\hat{\mathbf{f}}_{i,j} = \mathbf{v}_{t-1} - \mathbf{v}_t$.

Additionally, we can use the unscented transform (UT) to propagate motion uncertainty, Σ_e, and scene depth uncertainty at each pixel, σ_z^2, through the nonlinear point transfer of (8), yielding Σ_{UT}. This uncertainty estimate only accounts for optical flow uncertainty induced by errors in odometry or expected scene depth. We extend this by estimating the uncertainty of measuring optical flow at each pixel considering the spatial image gradients and uncertainties in the spatio-temporal gradients, yielding Σ_g—we adopted the method proposed by Simoncelli et al. [16]. This allows us to make use out of poorly constrained flow vectors (such as those on image edges), yet still fully account for its inaccuracies. We can finally

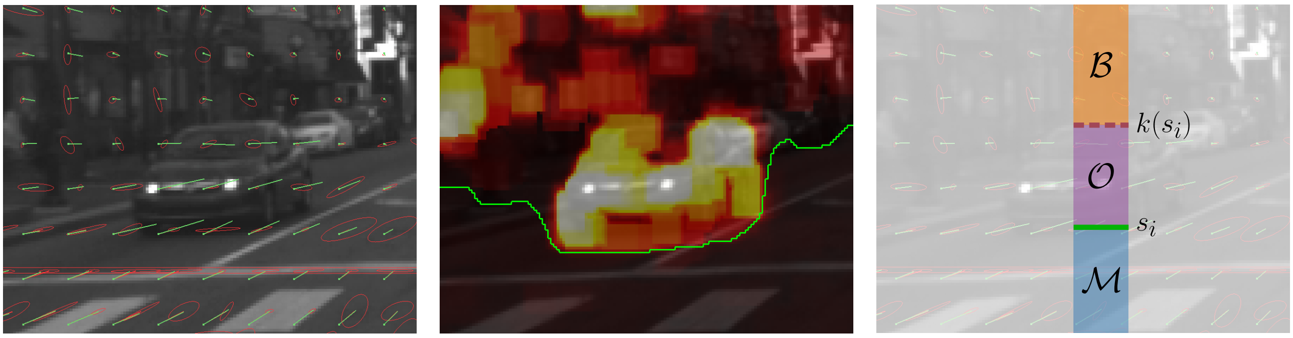


Figure 4: This figure details our proposed optical flow likelihood and potential. (Left) the optical flow vectors (*green*) and the expected optical flow vectors and uncertainties (*red*), note the shape of uncertainty ellipses along gradients. (Middle) Flow likelihood computed against expected, brighter indicates a lower likelihood. (Right) Partition potential considers segmenting image column into background, obstacle, and ground-map.

characterize the expected optical flow as a normally distributed measurement of the form:

$$\mathbf{f}_{i,j} \sim \mathcal{N}\left(\mathbf{v}_{t-1} - \mathbf{v}_t, \Sigma_{\mathrm{UT}} + \Sigma_g\right). \tag{9}$$

See Fig. 4 for visual depictions of this distribution.

Optical Flow Partition: Similar to appearance, the optical flow partition potential is formulated as a function of the likelihood of class assignments (obstacle and prior map), such that minimization of the potential maximizes the associated likelihood:

$$\phi_f\left(s_i\right) = -\log p_f\left(\mathbf{u}_i | s_i\right). \tag{10}$$

Following a similar derivation as (7), we arrive at the likelihood decomposition,

$$p\left(\mathbf{u_i} | s_i\right) = \prod_{j=1}^{s_i} p(\mathbf{f}_{i,j} | \neg\mathcal{M}) \prod_{j=s_i+1}^{h} p(\mathbf{f}_{i,j} | \mathcal{M}). \tag{11}$$

The prior map likelihood, $p(\mathbf{f}_{i,j}|\mathcal{M})$, can be computed by evaluating against the Gaussian in (9). However, the term on the left we decompose even further into,

$$\prod_{j=1}^{s_i} p(\mathbf{f}_{i,j} | \neg\mathcal{M}) = \prod_{j=1}^{k(s_i)} p(\mathbf{f}_{i,j} | \mathcal{B}) \prod_{j=k(s_i)}^{s_i} p(\mathbf{f}_{i,j} | \mathcal{O}), \tag{12}$$

to partition the non-map elements into a background set, $\mathcal{B}$, and an obstacle set, $\mathcal{O}$, at $k(s_i)$. This split at $k(s_i)$ is necessary to divide the very dissimilar flow sets generated by $\mathcal{B}$ and $\mathcal{O}$; these 3 disjoint sets are visualized in Fig. 4.

Given a world-frame height in meters of target obstacles, H_{obs}, we use known camera geometry and scene depth to calculate the height in pixels of the obstacle, $h(s_i) = f \cdot H_{\mathrm{obs}} / \hat{Z}_t(i, s_i)$, where f is the camera focal length. This can then be used to determine the pixel location for splitting $\mathcal{B}$ and $\mathcal{O}$, $k(s_i) = s_i - h(s_i)$. This world-frame obstacle height is a tuning parameter, though we have found the algorithm to be insensitive to selection of H_{obs} and is chosen based on minimum acceptable obstacle height ($H_{\mathrm{obs}} = 1.5$m in our experiments).

Within an image column, elements of an obstacle are at a constant depth and, thus, flow vectors are quite similar over the column. Therefore, we estimate $p(\mathbf{f}_{i,j}|\mathcal{B})$ and $p(\mathbf{f}_{i,j}|\mathcal{O})$ by fitting a uniform distribution over the flow vectors within their respective column segment.

We compute this potential over multiple image sequences so that we can capture fast moving objects, yet still maintain observability for slow moving objects (such as those within the focus of expansion). We use Farneback's optical flow algorithm [6] and perform forward-backward flow to discard inconsistent measurements (these discarded measurements provide no influence in the likelihood computations). It is important to note that while stationary, the optical flow potential only provides input to the MRF if something else is moving (a roughly uniform prior over all partitionings otherwise)—this is a byproduct of the formulation as stationary flow vectors observed from a stationary platform yields near constant likelihoods derived in (11).

3.3 Additional Potentials

Edge Potential: There is typically a strong gradient between obstacles and the road, thus we introduce an edge potential to bias cutting along image gradients: $\phi_e(s_i) = -\nabla I_t(i, s_i)^2$.
LIDAR Potential: While our primary motivation is an image-only solution, the MRF provides a convenient method to fuse online LIDAR measurements. Given a LIDAR point in the camera frame, $p = [x, y, z]^\top$, we project into the camera frame, $[i, j]^\top$. Using the estimated ground-plane depth image, $\hat{Z}_t$, we find the expected ground point $\hat{s}_i$ by minimizing $\left\|\hat{Z}_t(i, \hat{s}_i) - z\right\|$. The resulting potential is a truncated quadratic: $\phi_l(s_i) = \min(|s_i - \hat{s}_i|, T_l)^2$, where T_l is a threshold controlling the region of influence of the LIDAR potential.
Recursive Potential: The recursive potential propagates the full energy functional from the previous time step into the current frame, as in [22]. With a known ground-model and egomotion, we use the homogeneous point transfer (8) to propagate the sum over unary potentials of the previous frame into the current frame, generating $\phi_r(s_i)$.

4 Results

We evaluated our proposed method on our autonomous platform, a TORC ByWire XGV, that is equipped with Velodyne LIDAR scanners and a Point Grey Flea3 monochrome camera. The LIDAR scanners, unless otherwise specified, were used only offline for generating prior maps. Majority of the algorithms presented were implemented in CUDA and all experiments were run on a laptop equipped with a Core i7-4910MQ and a laptop GPU (NVIDIA Quadro K4100), resulting in an implementation that runs at 5-8 Hz. Throughout our experiments, we assume that our platform is localized within our prior maps as detailed in §2.

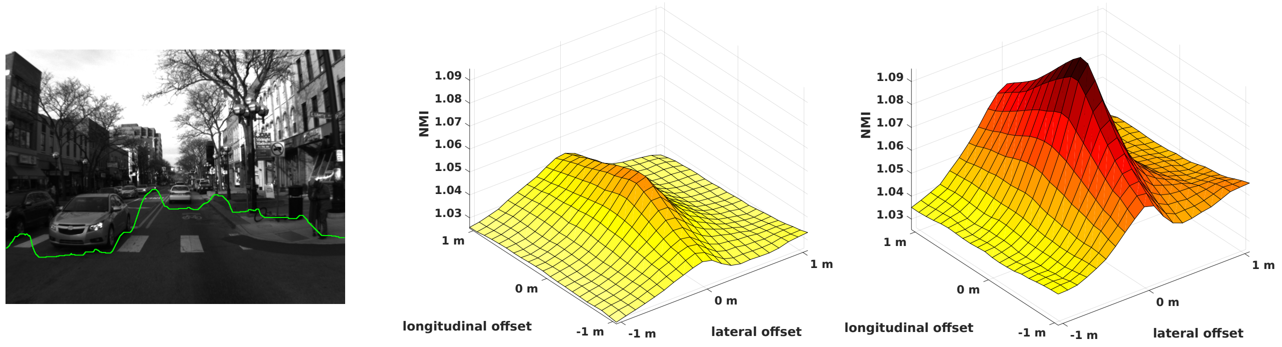
4.1 Quantitative Analysis

We first present experiments on a hand-labeled dataset in which we have 240 ground-truth image partitions. In addition to our vision only solution, we also demonstrate the effectiveness of including a simple 2D LIDAR scanner to our system. Note that while our platform is not equipped with such a planar scanner, we simulate this with the Velodyne scanners by only using point returns within a 40 cm window, 1 m off the *body frame* ground.

Following the metrics presented by Fritsch et al. [7], we project our image partition into the world to create a Bird's Eye View (BEV) before calculating the F1 measure, precision, recall, and false positive rate and results are tabulated in Table 1. Overall, we see that our method performs quite well on a fairly difficult dataset and the addition of the LIDAR scanner dramatically improves obstacle detection.

Method	F1	Precision	Recall	FPR
Proposed	87.85 %	90.07 %	85.74 %	9.93 %
Proposed+2D LIDAR	93.18 %	94.65 %	91.75 %	5.35 %

Table 1: The F1-score, precision, recall, and false positive rate for our proposed method and our proposed method with the addition of 2D LIDAR measurements.



(a) Partitioned Image (b) Cost surface without masks (c) Cost surface with masks

Figure 5: Sample of improvement of the NMI cost surface when obstacle masks are used.

4.2 Qualitative Analysis

To demonstrate the contributions of each unary potential that is a part of the MRF model, we present several candidate image partitions along with an overlay of each potential, see Fig. 6—in these images, lighter (white) colors indicate a lower energy state. All potentials presented in this paper were enabled *except* the LIDAR potential.

In the first row, we see our platform exiting a bright region into an area in shadow. Through the illuminated region, the appearance models fit to this bright distribution and hallucinate obstacles at the bright to shaded transition. Despite this, the image partitioning is still successful because of the optical flow potential. Several frames later, depicted in row 2, we see the appearance models have quickly adjusted to the new lighting.

The second and third row demonstrates the flexibility of our model to be able to perfectly follow the sharp contours of a pedestrian and a lightpost, respectively. One significant drawback of the optical flow potential is the effect of cast shadows from moving platforms, as shown in the third row. There is a gap of falsely detected obstacles triggered by the moving shadow to the right of the vehicle.

Finally, in the fourth row, we see a slight error on the right half of the road, where the appearance potential beats out the flow potential as this is during a turn where there is limited motion parallax. The remaining rows are samples of typical performance of the system.

4.3 Localization with Obstacle Partitions

Next, we incorporated our image partitions into our localization pipeline. To do so, we only used pixels *below* the obstacle partition when performing registrations against our prior map—all other pixels are discarded. As in [21], we performed a set of registration attempts from randomly initialized offsets within a 3 m window around known ground-truth.

Overall, we see a modest improvement in median absolute deviation from 12.4 cm to 11.6 cm longitudinally, and 14.3 cm to 9.1 cm laterally. Further, we see that the cost function is much more peaked when obstacle masks are used (Fig. 5). In future work, we hope this distinct improvement can help improve registration efficiency.

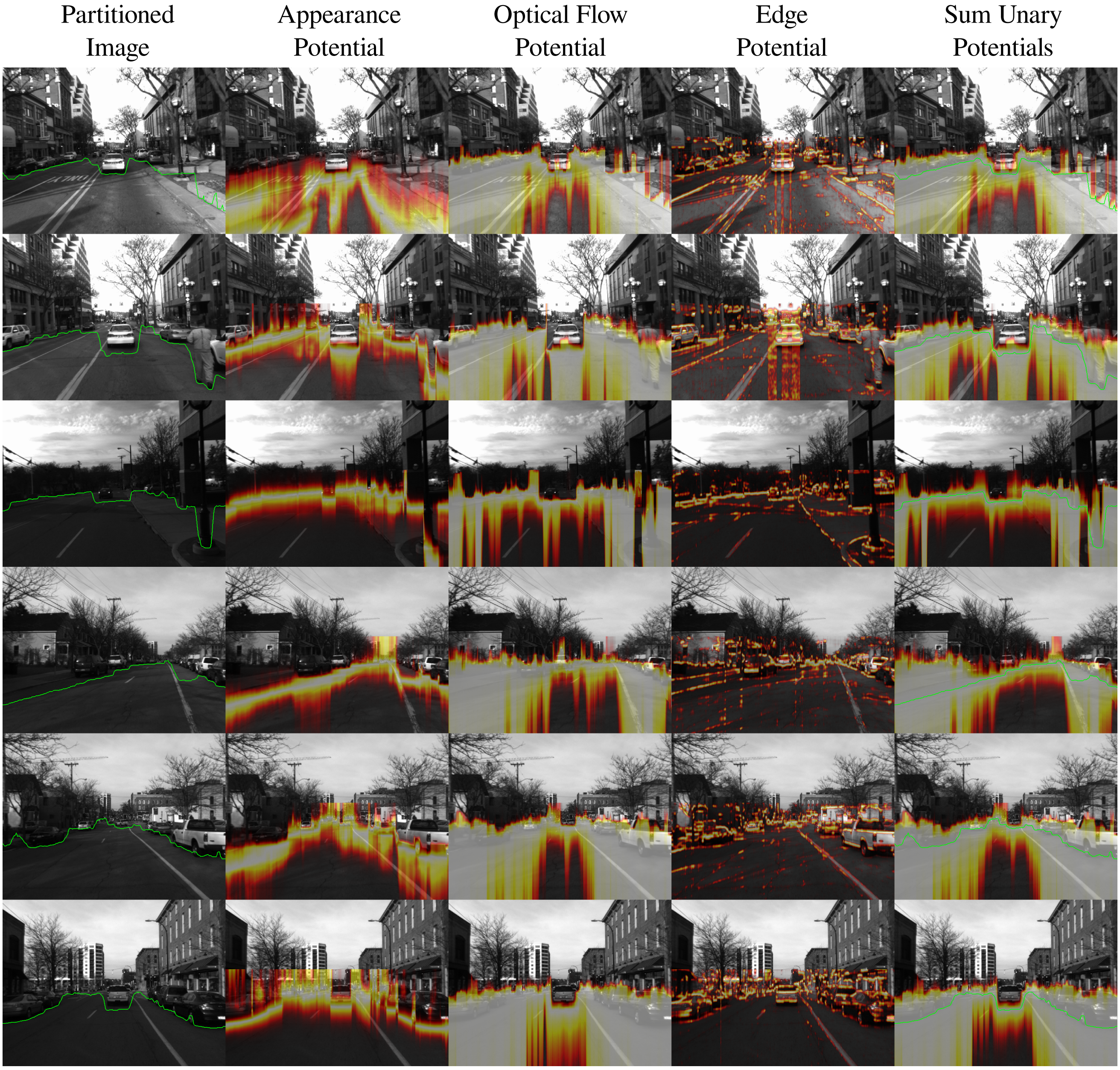


Figure 6: This figure shows sample results of our probabilistic obstacle partitioning. In each of these, lighter (white) colors indicate a lower energy state. See the text for more discussion; best viewed in color.

5 Conclusion

In this paper, we showed that a grayscale, monocular camera can be used to partition an image into disjoint sets of obstacles and the ground plane. We utilized a textured prior map to derive appearance models and optical flow likelihoods that could be integrated into an MRF. The resulting formulation can be solved at a framerate of 5–8 Hz. Furthermore, we integrated this into our visual localization pipeline and demonstrated improved robustness when obstacle partitions are considered during registration.

In the future, we hope to use the extracted optical flow vectors to segment objects lying above the image partition, which can further be used to improve the recursive potential with a motion model. Additionally, we plan to expand this to simpler prior maps where a full 3D ground prior may not be available.

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

References

[1] José M Álvarez and Antonio M Ĺopez. Road detection based on illuminant invariance. *IEEE Transactions on Intelligent Transportation Systems*, 12(1):184–193, October 2011.

[2] H. Badino, U. Franke, and D. Pfeiffer. The stixel world – a compact medium level representation of the 3d-world. In *Proceedings of the DAGM Symposium on Pattern Recognition*, volume 5748, pages 51–60, Jena, Germany, September 2009.

[3] Christophe Braillon, Cédric Pradalier, James L Crowley, and Christian Laugier. Real-time moving obstacle detection using optical flow models. pages 466–471, Tokyo, Japan, June 2006.

[4] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G. Bradski. Self-supervised monocular road detection in desert terrain. In *Proceedings of Robotics: Science and Systems*, Philadelphia, PA, USA, August 2006.

[5] W Enkelmann, V Gengenbach, W Kruger, S Rossle, and W Tolle. Obstacle detection by real-time optical flow evaluation. pages 97–102, Paris, France, October 1994.

[6] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th Scandinavian Conference on Image Analysis*, pages 363–370, Berlin, Heidelberg, June 2003. Springer-Verlag.

[7] Jannik Fritsch, Andreas Geiger, and Tobias Kühnl. A new performance measure and evaluation benchmark for road detection algorithms. In *IEEE Conference on Intelligent Transportation Systems*, pages 1693–1700, The Hague, Netherlands, October 2013.

[8] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[9] W Krüger, W Enkelmann, and S Rössle. Real-time estimation and tracking of optical flow vectors for obstacle detection. pages 304–309, Detroit, MI, USA, September 1995.

[10] Dan Levi, Noa Garnett, and Ethan Fetaya. Stixelnet: A deep convolutional network for obstacle detection and road segmentation. In *Proc. British Mach. Vis. Conf.*, pages 109.1–109.12, Swansea, United Kingdom, September 2015.

[11] Jesse Levinson and Sebastian Thrun. Robust vehicle localization in urban environments using probabilistic maps. In *Proc. IEEE Int. Conf. Robot. and Automation*, pages 4372–4378, Anchorage, AK, May 2010.

[12] Jesse Levinson, Michael Montemerlo, and Sebastian Thrun. Map-based precision vehicle localization in urban environments. In *Proc. Robot.: Sci. & Syst. Conf.*, Atlanta, GA, June 2007.

[13] Manolis IA Lourakis and Stelios C Orphanoudakis. Visual detection of obstacles assuming a locally planar ground. In *Proceedings of the Asian Conference on Computer Vision*, pages 527–534. Hong Kong, China, January 1998.

[14] Colin McManus, Winston Churchill, Ashley Napier, Ben Davis, and Paul Newman. Distraction suppression for vision-based pose estimation at city scales. In *Proc. IEEE Int. Conf. Robot. and Automation*, pages 3762–3769, Karlsruhe, Germany, May 2013.

[15] Richard Roberts and Frank Dellaert. Optical flow templates for superpixel labeling in autonomous robot navigation. In *IROS Workshop on Planning, Perception, and Navigation for Intelligent Vehicles*, Tokyo, Japan, November 2013.

[16] E P Simoncelli, E H Adelson, and D J Heeger. Probability distributions of optical flow. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 310–315, Maui, HI, USA, June 1991.

[17] Iwan Ulrich and Illah Nourbakhsh. Appearance-based obstacle detection with monocular color vision. In *Proc. AAAI Nat. Conf. Artif. Intell.*, pages 866–871, Austin, TX, USA, July 2000.

[18] Chris Urmson. How a driverless car sees the road. TED Talks, March 2015. URL http://www.ted.com/talks/chris_urmson_how_a_driverless_car_sees_the_road.

[19] Andrew J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, IT-13(2):260–269, April 1967.

[20] Andreas Wedel, Thomas Schoenemann, Thomas Brox, and Daniel Cremers. Warpcut–fast obstacle segmentation in monocular video. In *Pattern Recognition*, pages 264–273. Springer Berlin Heidelberg, 2007.

[21] Ryan W. Wolcott and Ryan M. Eustice. Visual localization within LIDAR maps for automated urban driving. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, pages 176–183, Chicago, IL, USA, September 2014.

[22] Jian Yao, Srikumar Ramalingam, Yuichi Taguchi, Yohei Miki, and Raquel Urtasun. Estimating drivable collision-free space from monocular video. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, pages 420–427, Waikoloa Beach, HI, USA, January 2015.

[23] Yan Zhang, Stephen J Kiselewich, William A Bauson, and Riad Hammoud. Robust moving object detection at distance in the visible spectrum and beyond using a moving camera. In *Computer Vision and Pattern Recognition Workshop*, page 131, New York, NY, USA, June 2006.