

# A Variational Approach to Online Road and Path Segmentation with Monocular Vision

Lina Maria Paz

Pedro Piniés

Paul Newman

*Abstract*— In this paper we present an online approach to segmenting roads on large scale trajectories using only a monocular camera mounted on a car. We differ from popular 2D segmentation solutions which use single colour images and machine learning algorithms that require supervised training on huge image databases. Instead, we propose a novel approach that fuses 3D geometric data with appearance-based segmentation of 2D information in an automatic system. Our contribution is twofold: first, we propagate labels from frame to frame using depth priors of the segmented road avoiding user interaction most of the time; second, we transfer the segmented road labels to 3D laser point clouds. This reduces the complexity of state-of-the-art segmentation algorithms running on 3D Lidar data. Segmentation fails in only 3% of the cases over a sequence of 13,600 monocular images spanning an urban trajectory of more than 10km.

## I. INTRODUCTION

Road detection is an essential and challenging task that plays an important role for supporting advanced driver assistance systems, such as road following or vehicle and pedestrian detection. Moreover, the estimation of the road geometry (slopes and borders) together with the localisation of the vehicle are essential tasks in this context since they aid the lateral and longitudinal control of the vehicle. On-board vision systems [1, 2, 3] have been widely used as they offer many advantages over other active sensors such as Radar or Lidar (higher resolution, low power consumption, low cost, easy aesthetic integration, and nonintrusive nature), that allow to reduce time and computational effort.

In this paper, we propose a novel framework for segmenting roads from outdoor images. We combine the use of geometric priors with statistical colour segmentation. Unlike popular approaches, we do not make any assumptions about the type of road, but instead assume that the road images can be captured from an arbitrary camera orientation.

Our contribution is twofold: first, we automatically propagate labels from frame to frame using depth priors of the segmented road; second, we transfer the segmented road labels to 3D laser point clouds. Figure 1 illustrates our method. Less than 20% of the image pixels are required to initialise multilabelling optimisation. Moreover, segmentation is only reset in 3% of the cases over a sequence of 13,600 monocular images spanning a trajectory of more than 10km.

## II. RELATED WORK

At first glance the problem of detecting the road geometry from visual information seems simple. However, complexity

Mobile Robotics Group, University of Oxford, Oxford, England; {linapaz, ppiniés, pnewman}@robots.ox.ac.uk

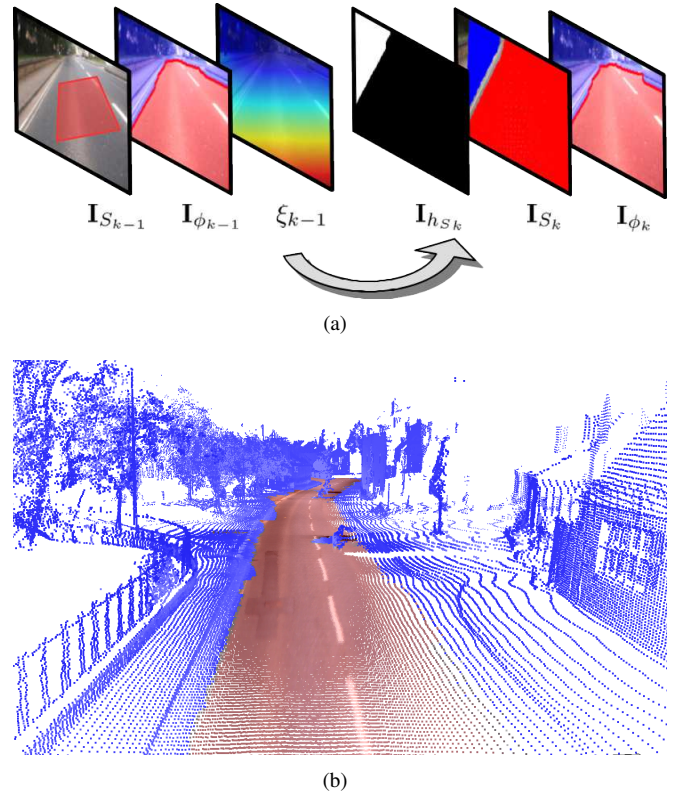


Fig. 1. Transfer of road labels from 2D to 3D. a) We run 2D segmentation and propagate the labels from frame to frame on a sequence of collected images. Time propagation of labels is performed by exploiting geometric information: dense depth estimation is run using only a pair of sequential monocular images. b) A dense 3D point cloud is rendered using data from an inexpensive 2D pushbroom laser. Given the camera-laser calibration and proper sensor synchronisation, we project the 3D point cloud onto the corresponding frame. As a result, a label is assigned to each projected point using the label at the corresponding pixel coordinate of the segmented image.

is introduced as the road is imaged from a mobile vehicle/camera with a constantly changing background, under the presence of different objects like vehicles and pedestrians, whilst being exposed to varying ambient illumination and weather conditions. In [4] a comprehensive review of vision-based road detection systems is presented.

A particularly difficult scenario manifests when the road surface has both shadowed and non-shadowed areas. [2] proposes an approach to vision-based road detection that is robust to shadows. The approach relies on using a shadow-invariant feature space combined with a model-based classifier. The model is built online to improve the adaptability

of the algorithm to the current lighting conditions and the presence of other vehicles in the scene. Similarly, the work of [1] presents a low cost approach to ego-lane detection on illumination-invariant colour images. Interestingly, the authors show that, employing the segmented road region as a prior for extracting lane markings, significantly improves the execution time and the success rate of their detection algorithm. [3] addresses road detection within a general dataset using a single image. The process is split into two steps: the estimation of the vanishing point associated with the central (straight) part of the road, followed by the segmentation of the corresponding road area based upon that vanishing point using a technique for detecting and refining road boundaries.

In this paper we propose the use of per pixel inverse depth estimates as a geometric prior for road segmentation from a sequence of images. Our approach is similar in spirit to [5], where self-supervised learning is achieved by combining the camera image with a laser range finder to identify a drivable surface area in the near vicinity of the vehicle. Once identified, this area is used as training data for vision-based road classification. In contrast, our approach uses only a monocular camera to estimate dense depth maps and propagate road labels from a previously segmented reference image to the next observed images.

We perform dense inverse depth map estimation from a variational method perspective. An energy function is optimised based on a data fidelity term that measures the photo-consistency over a set of small-baseline, monocular frames. In [6] the total variation (TV) is used as a regularisation term to preserve sharp depth discontinuities whilst simultaneously enforcing smoothness of homogeneous surfaces. The solution is based on a primal-dual formulation successfully applied in solving variational convex functions that arise in many image processing problems [7]. Unlike [6], where usually sideways or longitudinal movements are applied to the camera in bounded office-like scenarios, we have a forwards-facing camera mounted on a car travelling forwards and sensing distant objects with a low parallax. This leads us to rely on an improved regularisation method to reinforce depth on critical parts of the scene. In our case, a suitable assumption is to expect to find many affine surfaces in the environment, like roads, pathways, building facades or vehicle surfaces. In this paper we employ a Total Generalised Variation (TGV) regularisation [8] which has shown to favour piecewise linear regions of a structure.

With regards to image segmentation, we find a multitude of algorithms widely used for specific tasks such as image editing, detection of interest regions in medical images or object tracking in video sequences. The most popular approaches try to efficiently compute minimum energy solutions for cost functions, using graph cuts [9], level set methods [10], random walks [11], and convex relaxation techniques [12]. These methods combine two important concepts in their energy models: first, a data fidelity term that measures how well a pixel fits to each label; second, a regularisation term that measures the consistency

of the segmentation with respect to some prior knowledge. Examples of priors are the object boundary length, the number of labels, specific intra-label cost functions and label co-occurrence. In practice, all these algorithms require the external input of a user to initialise pixels with a given label.

In this paper we use an probabilistic approach to reduce the intra-label variability caused by different road textures and lighting conditions.

As a second contribution, we extend current vision-based approaches for online road registration by spatiotemporally relating camera and laser data. We address the problem of transferring the road labels learnt on the 2D image sequence to 3D laser point clouds under the assumption of perfect synchronisation and accurate camera-laser extrinsic calibration.

The paper is organised as follows: in section III we describe the general pipeline of our vision-based system for road segmentation. We also explain how the depth prior can be used to propagate labels between consecutive frames. Section III-A illustrates our dense depth map estimation approach with TGV regularisation from monocular images. Image segmentation with spatially varying colour distributions is introduced in III-B. In section IV we show the mechanism to transfer labels from 2D images to large scale 3D laser point clouds. Evaluation of our approach over a 10km dataset is shown in V. Finally, we draw our conclusions in VI.

### III. SYSTEM DESCRIPTION

Our algorithm for road detection from multiple input images is visualised schematically in Fig. 2. From a front-facing camera mounted on a car we assume that we obtain a sequence of  $n$  RGB images  $I_1, \dots, I_n : \Omega \subset \mathbb{R}^2 \mapsto \mathbb{R}$  collected during a driving trial along a semi-urban environment. The corresponding camera poses  $T_1, \dots, T_n \in SE(3)$  are estimated in practice from our on-board scaled Visual Odometry system [13].

Our algorithm initially creates a dense inverse depth map  $\xi(\mathbf{u})$  from consecutive monocular images. Unlike other approaches where longer sequences of images are integrated for accurate depth map estimation, we use only a pair of images  $I_{k-1}, I_k$ . This choice enables us to estimate the depth of dynamic objects (particularly important in urban environments), which could be potentially disregarded by popular long sequence integration approaches. Then, we run semi-automatic road labelling on the image reference  $I_{k-1}$  where the depth map is created.

This process can be carried out by performing a simple projection of a viewing frustum onto the ground plane using the depth estimates of the central image pixels.

The output of this process is a labelled image  $I_{S_{k-1}}$  that serves as an initial seed for a two-region segmentation. In this step we explicitly consider the spatial variation of colour distribution in a general Bayesian MAP estimation approach, allowing us to deal with different textures and lighting conditions on the road. At this point, a new image  $I_{\phi_{k-1}}$  with the desired road segment is available. When a new

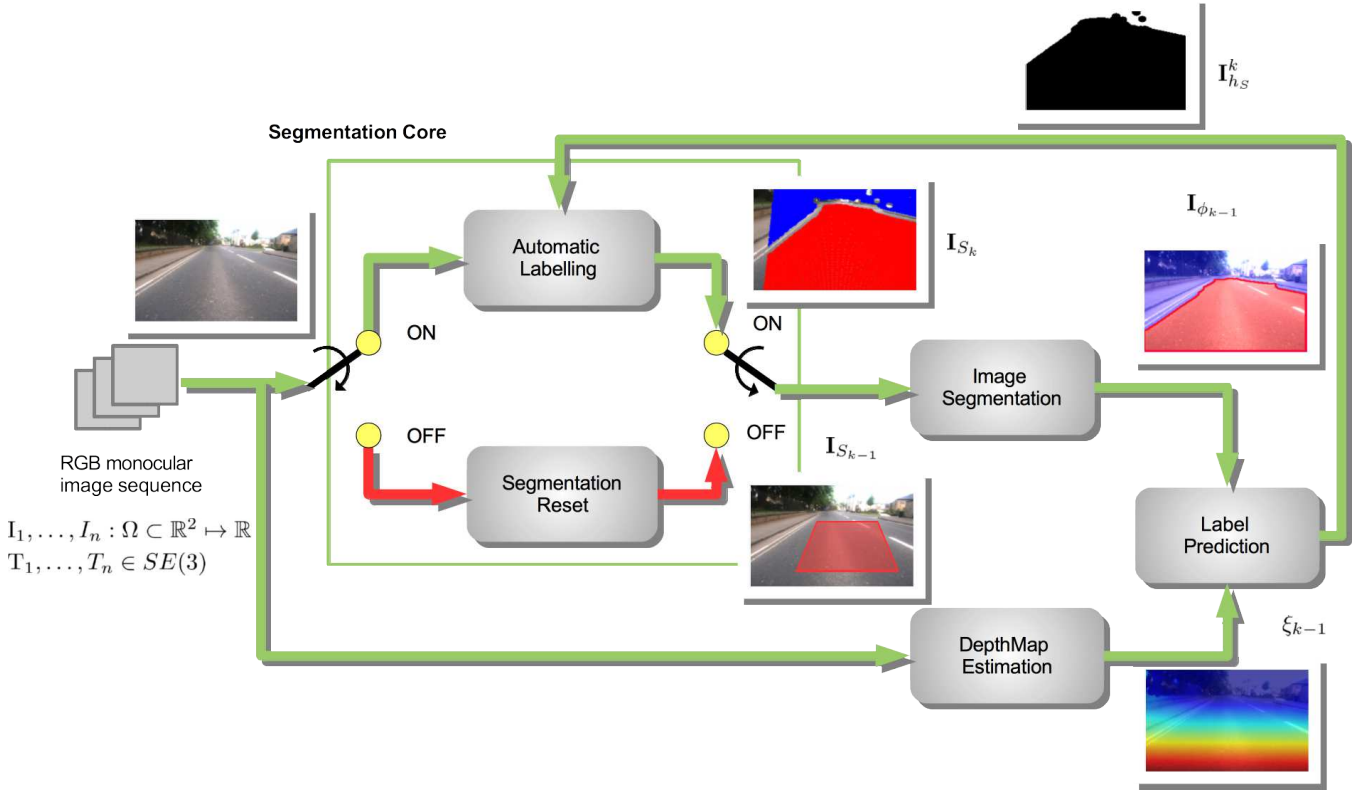


Fig. 2. Road segmentation with monocular images. Our algorithm initiates by creating a dense inverse depthmap  $\xi_k(\mathbf{u})$  from consecutive monocular images. We then run automatic road labelling on the image reference where the depth map is created. We perform a simple projection of a viewing frustum onto the ground plane, using the depth estimates of central image pixels. The output of this process is a labelled image  $I_{S_{k-1}}$  that serves as initial seed for a multilabelling optimisation process. At this point, a new image  $I_{u_{k-1}}$  with the desired road segment is available. When a new image arrives, both the background and road labels are propagated into the next frame by predicting the pixel locations. To prevent incorrect propagation of labels we carry out morphology regularisation on pixels at the label boundaries. Finally, the output image  $I_{S_k}$  is fed into the multilabelling optimisation process. Although the process is implemented to run automatically, static camera motions (usually due to the fact that the car stops) can drastically affect the accuracy of the depth map and therefore, the label prediction. To overcome this problem we provide the possibility to reset the process.

image arrives, both background and road labels are warped into the next frame as follows:

- First, we backproject the classified pixels to 3D space

$$\mathbf{p}_{\mathbf{u}_r, \mathbf{u}_b}^k = \pi^{-1}(\xi_{k-1}(\mathbf{u}_r, \mathbf{u}_b), K^{-1}) \quad (1)$$

where  $\pi^{-1}$  refers to the back-projection of a pixel  $\mathbf{u}$  with inverse depth  $\xi_i$  and intrinsic calibration matrix  $K$ .

- Then, we calculate the perspective projection of the labelled points

$$h_k(\mathbf{u}_r, \mathbf{u}_b) = \pi(\mathbf{p}_{\mathbf{u}_r, \mathbf{u}_b}^k, K, T_{k,k-1}) \quad (2)$$

- Eq.3 provides the new set of labelled pixels that will be used as initialisation for road segmentation on image frame  $I_k$ . To prevent the incorrect propagation of labels owing to depth inaccuracies, we reduce the intra-label variability by using morphology regularisation on pixels at the label boundaries. This operation is synthesised as

$$I_{S_k} = I_{h_k} \ominus B = \{w \in \Omega | B_w \in I_{h_{k-1}}\} \quad (3)$$

where  $I_{h_{k-1}}$  is a binary image rendered from the predicted label pixels,  $B_w$  is the translation of  $B$  by the vector  $w$ . In practice, we assume that the structure

of  $B$  is isotropic with the centre located at the origin of  $\Omega$ .

- The resulting image  $I_{S_k}$  is fed into the multilabelling optimisation process producing the desired segmented image  $I_{\phi_k}$ .

Although the process is designed to run automatically, static camera motions (usually due to the fact that the car stops) can drastically affect the accuracy of the depth map and therefore, the label prediction. To overcome this problem, we provide the possibility to reset the process when segmentation fails. In the following sections, we review the techniques used in the pipeline to create dense depth maps as well as to perform multilabelling optimisation.

#### A. TGV Depthmap calculation

To create a depthmap  $\xi(\mathbf{u})$  from a pair of consecutive images  $[I_{k-1}(\mathbf{u}), I_k(\mathbf{u})]$  we follow a variational approach similar to the one presented in [6]. However, in our application we have a major constraint: our camera is mounted on the front of a moving car and facing a distant horizon. This means that for most of the pixels (particularly those that correspond to distant points) we will not have sufficient parallax. Therefore, for estimating the depth, we will have

to rely on a regulariser that should contain some prior knowledge of the environment. In our case, a very reasonable assumption is to find affine surfaces in the environment, such as roads, pathways, building facades or vehicle surfaces. For this reason, we have chosen to implement our regulariser as a Total Generalised Variation (TGV) norm that favours piecewise linear solutions. The energy function to be minimised is given by

$$\min_{\xi} E_R(\xi) + E_D(\xi) \quad (4)$$

where  $E_D(\xi)$  is a nonconvex data term that calculates the photometric error  $\rho$  between the new image  $I_k(\mathbf{u})$  and a warping of the previous one  $I_{k-1}(\mathbf{u})$

$$E_D(\xi) = \lambda \int_{\Omega} \rho(I_k(\mathbf{u}), I_{k-1}(\mathbf{u}), \xi(\mathbf{u})) d\mathbf{u} \quad (5)$$

and  $E_R(\xi)$  is the TGV regularisation term given by

$$\min_{\xi} \int_{\Omega} \alpha_1 \|\nabla_{\xi} \xi - \mathbf{v}\|_1 + \int_{\Omega} \alpha_2 \|\nabla_v \mathbf{v}\|_1 \quad (6)$$

By introducing an additional variable  $\mathbf{v}$ , TGV can intrinsically yield a balance between the zero and first order derivatives of the solution signal. This property allows us to generalise the piecewise constant behaviour of the classical TV norm and favour instead the reconstruction of piecewise affine surfaces. The TGV regularisation term depends on two constants  $\alpha_1$  and  $\alpha_2$  that control the piecewise smoothing.

Equation 6 gives us some intuition about why TGV regularisation favours piecewise affine functions. Think of  $\mathbf{v}$  as the slope of  $\xi$ . If  $\xi$  is piecewise linear then  $\mathbf{v}$  should be a piecewise constant signal which explains the TV  $\|\nabla_v \mathbf{v}\|_1$  penalty term for  $\mathbf{v}$ . Regarding the first term  $\|\nabla_{\xi} \xi - \mathbf{v}\|_1$ , if  $\mathbf{v}$  properly estimates the slope of an affine region of  $\xi$  then the contribution to the energy cost will be zero because  $\nabla_{\xi} \xi = \mathbf{v}$  and the only cost due to the penalty term will be the TV of  $\mathbf{v}$ .

The optimisation problem in Eq.(4) is solved using an iterative alternating optimisation method, explained in [6], that is based on an exhaustive search step which involves the data term  $E_D(\xi)$ , and a Primal-Dual algorithm [7] involving the regularisation term  $E_R(\xi)$ .

### B. Multilabel optimisation for road detection

We start by considering the reference image  $I_k$  and the corresponding depth map  $\xi_k$ . At this point our task is to split up the image into  $n_l$  pair-wise disjoint regions (e.g. the road and background):

$$\Omega = \bigcup_{i=1}^{n_l} \Omega_i, \quad \Omega_i \cap \Omega_j = \emptyset, \quad \forall i \neq j \quad (7)$$

The problem is solved by assigning a label  $i \rightarrow \{1, \dots, n_l\}$  to each pixel  $u$ , such that  $\Omega_i = \{u | \phi_i(u) = i\}$ , where  $\phi_i$  is an indicator function defined as

$$\phi_i(u) = \begin{cases} 1 & \text{if } u \in \Omega_i \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

The major challenge is to find an optimal label configuration for all image pixels among all possible configurations. To efficiently solve the problem we adopt a general Bayesian estimation approach to compute the segmentation in a MAP sense,

$$\arg \max_u P(I|\phi)P(\phi) \quad (9)$$

where  $P(I|\phi)$  and  $P(\phi)$  are the likelihood and prior probabilities over the colour and region functions.

To deal with different textures and lighting conditions on the path, we model the likelihood as

$$P(I|\phi) = \prod_{i=1}^{n_l} \prod_{u \in \Omega_i} (P(I(u), u | \phi(u) = i)) \quad (10)$$

The term inside the product denotes the joint probability for observing a colour value  $I$  at spatial location  $u \in \Omega_i$  which can be approximated using Gaussian kernels with  $\sigma$  and  $\rho$  variances on the colour and location variables respectively. Here we use our initial seed  $I_{S_k}$  calculated automatically from the prediction  $I_{h_k}$ . In the following, we drop the index  $k$  to keep a simple notation. Let us consider the set of points in label  $i$  as  $S_i := \{I_{S_i}, \mathbf{u}_{S_i}\}$ . The joint probability is computed as

$$\hat{P}(I(u), u | \phi(u) = i) \propto \sum_{j \in S_i} k_{\rho_i}(u - u_{S_i,j}) k_{\sigma}(I(u) - I_{S_i,j}) \quad (11)$$

where  $\sigma$  and  $\rho$  are set up using spatially adaptive kernel functions that depends on the distance to the nearest closest point in each label. The segmentation problem in Eq. (9) also requires the specification of a prior  $P(\phi)$  over all regions. A common choice is to use priors that favour regions of shorter length such as:

$$P(\phi) \propto \frac{1}{2} \sum_{i=1}^{n_l} \int_{\Omega} g \|\nabla \phi_i\|_1 \quad (12)$$

Eq. (12) considers the perimeter of each region measured by  $\int_{\Omega} g \|\nabla \phi_i\|_1$ , also known as the Total Variation of the region represented by  $\phi$ . In practice,  $g$  is a function of the form  $\exp(-\gamma \|\nabla I(u)\|_2)$  typically used to promote edges. A more general formulation of the Total Variation is

$$\int_{\Omega} g \|\nabla \phi_i\|_1 = \max_{\Psi(u) \leq 1} \int_{\Omega} g \nabla \phi \cdot \Psi du \quad (13)$$

where  $\Psi$  is the dual variable of region  $\phi$ .

By substituting Eqs. (10-13) in Eq. (9), we arrive at the equivalent optimisation problem with linear constraints

$$\min_{\phi} \max_{\Psi} \frac{1}{2} \sum_{i=1}^{n_l} \int_{\Omega} g \nabla \phi \cdot \Psi du + \lambda_l \sum_{i=1}^{n_l} \int_{\Omega_i} f_i(u) du \quad (14)$$

$$s.t. \quad \sum_{i=1}^{n_l} \phi_i(u) = 1 \quad (15)$$

$$f_i = -\log \hat{P}(I(u), u | \phi(u) = i) \quad (16)$$

Recently, [14] has shown that a suboptimal stable solution for the multilabelling problem (and optimal for a two-region

case) can be found by introducing the Lagrange multipliers  $\Gamma(u)$ , such that the energy function can be rewritten as follows:

$$\begin{aligned} \min_{\phi} \max_{\Psi(u) \leq 1} & \frac{1}{2} \sum_{i=1}^{n_i} \left( \int_{\Omega} g \nabla \phi \cdot \Psi du + \lambda_l \int_{\Omega_i} f_i(u) du + \right. \\ & \left. + \int_{\Omega} \Gamma(u) \left( \sum_{i=1}^{n_i} \phi_i(u) - 1 \right) du \right) \end{aligned} \quad (17)$$

An iterative primal dual algorithm, applied to the saddle point formulation in Eq. (17), is summarised in the following set of equations:

$$\begin{aligned} \bar{\Psi}^{t+1} &= \Psi^t + \omega \nabla \bar{\phi}^t \\ \bar{\Psi}^{t+1} &= \bar{\Psi}^{t+1} / \max(1, \|\bar{\Psi}^{t+1}\|) \\ \phi^{t+1} &= \phi^t - \tau (\nabla^T \Psi^{t+1} + \lambda_l f + \Gamma^t) \\ \Gamma^{t+1} &= \Gamma^t + \mu \left( \sum_{i=1}^{n_i} \phi_i(u) - 1 \right) \\ \bar{\phi}^{t+1} &= \phi^{t+1} + \theta_r (\phi^{t+1} - \phi^t) \end{aligned} \quad (18)$$

where  $\tau, \omega$  and  $\mu$  are the lengths of the gradient steps. In practice, each variable is updated by performing pixel wise calculations while the gradient operator  $\nabla$  is approximated by finite differences. The parameters are set up to  $1/2, 1/4$  and  $1/5$  respectively through the use of preconditioning [14].

Analogous to the depth map estimation problem, the primal dual approach followed in this section allows us to take advantage of general purpose GPUs hardware for parallel computing. For a detailed derivation of these update equations, we refer the interested reader to [14].

#### IV. TRANSFER OF 2D ROAD LABELS TO 3D LASER POINT CLOUDS

We envisage autonomous transport systems in which vehicles equipped with monocular cameras and an inexpensive 2D pushbroom laser are provided with a labelled 3D laser point cloud as a prior to support advanced driver assistance systems. To this end, we generate metrically consistent local 3D swathes from a pushbroom laser using a subset of camera pose estimates  $T_w \in SE3$  in an active time window  $w$  as follows:

$$\mathcal{M}_i = f(T_w, T_{CL}, \mathbf{x}_i)$$

where  $f$  is a function of the total set of collected laser points  $\mathbf{x}_i$  in the same time interval, and  $T_{CL}$  is the extrinsic calibration between camera and laser. The transfer of labels is carried out by re-projection of the 3D points onto the image plane at the camera reference  $T_r$ ,  $r \in w$  where the point cloud is represented. A label is assigned to each laser point from the corresponding image  $I_{\phi_r}$  using the pixel coordinates of the projected points. Figure 3 depicts the labels propagated over three 3D generated swathes.

## V. EXPERIMENTS

In order to demonstrate the robustness and scalability of our road segmentation system, we ran experiments on an outdoor image sequence gathered from a  $65 \times 50$  FOV monocular camera at 25Hz. The camera was mounted on roof of our Wildcat car looking forwards in the moving vehicle direction. The sequence is composed of 13600 images of resolution  $512 \times 384$  collected in a village nearby our home town. The full trajectory is approximately 10km long from the initial position.

The dense mapping approach as well as the multilabelling optimisation process are implemented in CUDA C++. The whole pipeline runs on a laptop equipped with an i7 Intel processor at 2,3GHz and a GeForce GT 750M NVIDIA Graphics card with 2048 MB of device memory. Figure 4 shows a time assessment of the online road segmentation. The graph depicts the contribution of the depth map creation (red circles) and the label propagation stages to the total running time. The former requires 10 TGV primal dual iterations (less than 500ms in average) to estimate a depth map. The second considers the label warping induced by equations (1-3) and the primal dual iterations, run to achieve the final segmented image. In practice, we execute less than 100 primal dual iterations during multilabelling optimisation (3.5s average time).

Figure 6 shows the places along the whole trajectory where we had to reset the system. Around 172 places we restart the roads segmentation over the total number of places (i.e less than 3% of the time). In such cases, the running time it is only due to the segmentation from our simple label initialisation (projection of the viewing frustum onto ground plane).

We expand our analysis to study the performance of the propagation of road labels on consecutive images. We are interested in knowing if a label can be propagated continuously over long periods of time. To this end, we introduce  $P(X \leq x_d)$ , a probability that denotes how unlikely the propagation has to be reset before a minimum distance  $x_d$ .  $X$  is defined as the travelled distance with respect to the vehicle localisation. Figure 6 shows that, for instance, after label propagation for more than 80m there is 95% of probability of requiring a reset. In our case, it is very unlikely that the system will require reset below a minimum distance of 20m.

In the supplemental material, we provide a video that shows the extensive results running our road segmentation on the complete sequence.

## VI. CONCLUSIONS

In this paper we propose the use of per pixel inverse depth estimates as geometric prior for road segmentation. The presented approach uses only a monocular camera to estimate dense inverse depth map and propagate road labels from a previously segmented reference image to the next observed images. A second contribution is the extension of current vision based approaches for online road registration by relating image-laser data spatiotemporally. We address the problem of transferring the road labels learnt on the 2D



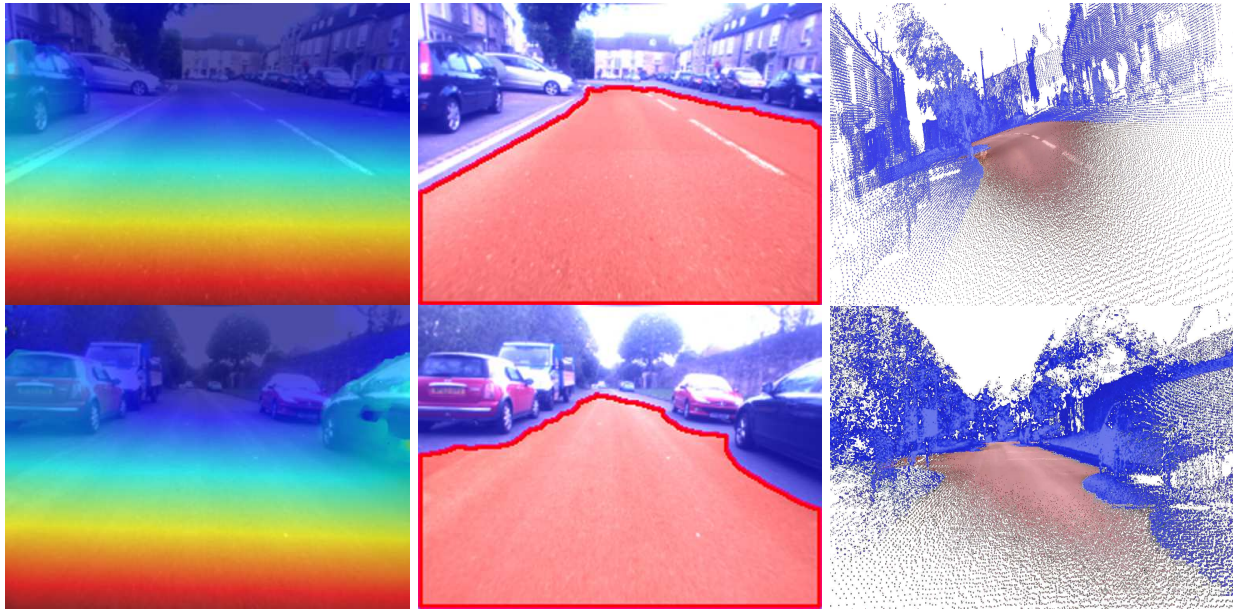


Fig. 3. Transfer of 2D labels to 3D point clouds rendered from an inexpensive 2D pushbroom laser. Given the camera-laser calibration and proper sensor synchronisation, we project the 3D point cloud onto the image plane of the reference frame. As a result, a label is assigned to each projected point with the label of the pixel at the corresponding coordinates in the segmented image. First column, depth map estimate. Second column, 2D road segmentation on the reference image. Third column, result of the transfer to 3D point clouds.

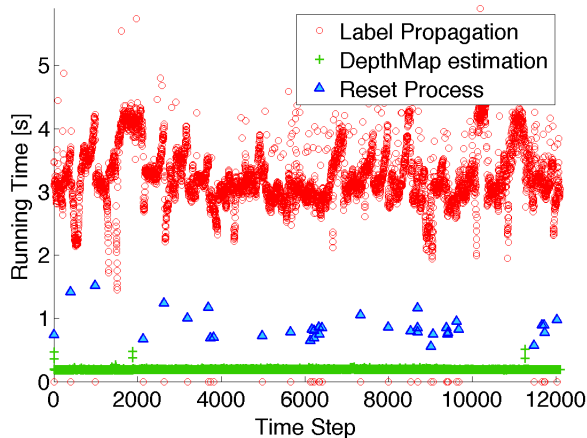


Fig. 4. Running time per step of all associated processes. The graph depicts the exclusive times for the depth map estimation approach (green crosses) and the propagation of labels (red circles). Blue triangles denote the running time in those steps in which the system was reset. In those cases, segmentation is achieved by the simple label initialisation (projection of the viewing frustum onto ground plane).

image sequence to large 3D laser point clouds under the assumption of perfect synchronisation and accurate camera-laser extrinsic calibration. Qualitative experiments on real-world road sequences with normal traffic show that the method is robust to shadows and lighting variations, thus helping to preserve the road pattern over tens of kilometres. In further research we will exploit the recurring features that emerge from datasets where a vehicle follows the same route more than once. We will study the case of transferring the road labels learnt in a first journey into data

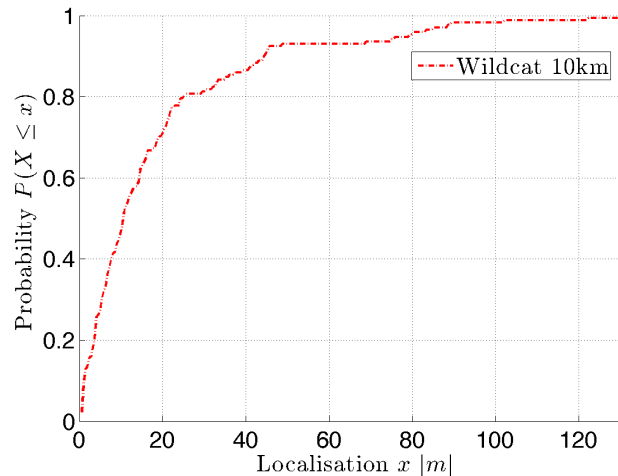


Fig. 5. Probability that denotes how likely propagation has to be reset for a given traversed distance.

from other journeys driven at different times and varying appearance conditions. We will study the fusion of depth map and segmentation estimation processes into a single energy optimisation.

## REFERENCES

- [1] M. Beyeler, F. Mirus, and A. Verl, "Vision-based robust road lane detection in urban environments," in *IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, 2014, pp. 243–252.
- [2] J. M. Alvarez and A. M. Lopez, "Road detection based on illuminant invariance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 184–193, 2010.

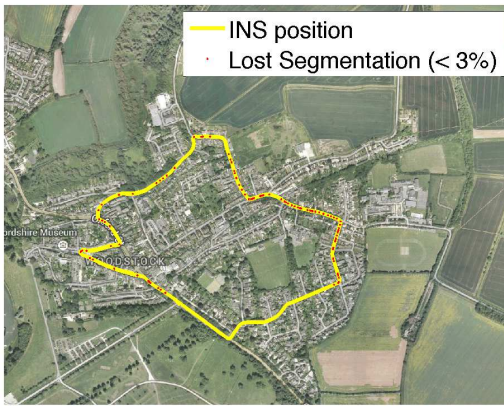


Fig. 6. An aerial view of the workspace and continuous paths driven (multiple loops are shown). The car trajectory is shown according to the INS recordings. Only 3% of the places along the trajectory required the system reset.

- [3] H. Kong, J.-Y. Audibert, and J. Ponce, "General road detection from a single image," *Trans. Img. Proc.*, vol. 19, no. 8, pp. 2211–2220, Aug. 2010.
- [4] S. Yenikaya, G. Yenikaya, and E. Düven, "Keeping the vehicle on the road: A survey on on-road lane detection systems," *ACM Comput. Surv.*, vol. 46, no. 1, pp. 2:1–2:43, Jul. 2013.
- [5] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G. R. Bradski, "Self-supervised monocular road detection in desert terrain," in *Robotics: Science and Systems II, August 16-19, 2006. University of Pennsylvania, Philadelphia, Pennsylvania, USA*, 2006.
- [6] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "DTAM: Dense tracking and mapping in real-time," in *Proceedings of the 2011 International Conference on Computer Vision ICCV*, ser. ICCV '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 2320–2327.
- [7] A. Chambolle and T. Pock, "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, May 2011.
- [8] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Img. Sci.*, vol. 3, no. 3, pp. 492–526, Sep. 2010.
- [9] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, 2001.
- [10] T. Brox, M. Rousson, R. Deriche, and J. Weickert, "Colour, texture, and motion in level set based segmentation and tracking," *Image Vision Comput.*, vol. 28, no. 3, pp. 376–390, 2010.
- [11] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, 2006.
- [12] T. Pock, A. Chambolle, D. Cremers, and H. Bischof, "A convex relaxation approach for computing minimal partitions," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*. IEEE, 2009, pp. 810–817.
- [13] M. Smith, I. Baldwin, W. Churchill, R. Paul, and P. Newman, "The new college vision and laser data set," *The International Journal of Robotics Research*, vol. 28, no. 5, pp. 595 – 599, May 2009. [Online]. Available: <http://www.robots.ox.ac.uk/NewCollegeData/>
- [14] T. Pock and A. Chambolle, "Diagonal preconditioning for first order primal-dual algorithms in convex optimization," in *IEEE Int. Conf. on Computer Vision (ICCV)*, 2011, pp. 1762–1769.