Outdoor Ride: Data Fusion of a 3D Kinect Camera installed in a Bicycle.

Alicia Robledo, Stephen Cossell and Jose Guivant
School of Mechanical and Manufacturing Engineering
The University of New South Wales, Sydney, Australia

Abstract

This paper presents the results of an experimental test that applies a low cost 3D sensor, i.e. a Kinect camera, for the perception of the terrain quality for the synthesis of a safe path. The 3D images are processed for generating a local 3D map, in front of the bicycle. The real time processing is based on a Piece Wise Multi-Linear (PWML) approximation of the 3D surfaces. The PWML patches are classified for inferring the traversability of the terrain. The acquired datasets are also made public for the educational and research community.

1 Introduction

The synthesis of 3D images for the perception of the terrain traversability is a relevant resource for the autonomous operation of a platform that needs to operate in a cluttered and not uniform terrain. The resource is also useful as an aiding process where a human operator can also control a platform based on information describing the slope of the terrain, discontinuities and other features of the terrain surface, [Elfes, 1989] [Thrun and Bucken, 1996].

In this experiment a set of sensors were installed in a standard push-bike. The sensor set is composed by a low cost 3D camera (Kinect), an IMU and a GPS.

The camera is installed as shown in Figure 1, with the aim of scanning the terrain ahead the bike. The IMU is used for providing raw angular velocities and accelerations.

Each 3D image is processed in several stages. A Piece Wise Multi-Linear (PWML) approximation is performed for approximating the surface by patches that are then used for estimating the normal vectors at different parts of the terrain. The PWML representation provides a number of advantages. Firstly it allows a straightforward numerical evaluation of the normal vectors of the terrain surface, it reduces the effect of noise in the 3D samples that appear as high frequency components of low amplitude. The PWML approximation also allows an extremely efficient compression of information for being stored and/or transmitted to other processing nodes.

The approximated PWML representation also allows an efficient scan matching process, where the “data” is composed by 3D points and the “model” by a set of patches (a surface). The scan matching process is a version of the well known ICP approach, as presented in [Melkumyan, 2009]. In our application the main goal of the scan matching process is for providing information for estimating the attitude of the platform, in combination with the angular velocities provided by the IMU sensor.

Figure 1. The 3D Kinect sensor installed at the front of the bike. The inclination (in elevation) in order to capture the terrain in front of the bike.

Figure 2 The IMU and GPS. The IMU sensor provides 3D accelerations and angular velocities in the platform coordinate frame. The on-board computer is inside one of the bags. A 12volts battery for powering the Kinect camera is located on the second bag.

Figure 3. Full view of the platform. Different camera's elevation angles were tested in order to deal with the limited range and field of view of the sensor.

2 PWML Approximation of 3D frames

Each 3D frame can be understood as a discrete 2D function, as an image where the pixels are represented by a 1D property.

The PWML approximation for a 2D function is defined as follows,

$$\left|\hat{f}(u,v)-f(u,v)\right|<\tau$$
$$\forall(u,v)\in\Omega$$
$$\hat{f}(u,v)=a_k\cdot u\cdot v+b_k\cdot u+c_k\cdot v+d_k$$
$$\forall(u,v)\in\Omega_k$$
$$\{\Omega_k\}_{k=1}^{N} \quad / \quad \bigcup_{k=1}^{N}\Omega_k=\Omega$$

Given a function $f(u,v)$, evaluated in a domain Ω, it can be approximated by a Piece Wise Multi Linear function $\hat{f}(u,v)$. The domain Ω can be partitionated in rectangular regions $\{\Omega_k\}_{k=1}^{N} \quad / \quad \bigcup_{k=1}^{N}\Omega_k=\Omega$. In each region the function $\hat{f}(u,v)$ is implemented by a MultiLinear expression, $\hat{f}(u,v)=a_k\cdot u\cdot v+b_k\cdot u+c_k\cdot v+d_k$, that is defined by the set of parameters $[a_k,b_k,c_k,d_k]$.
By defining a distance function, e.g. $E=\max_{(u,v)\in\Omega}\left(f(u,v)-\hat{f}(u,v)\right)$, an optimal approximating function $\hat{f}(.,.)$ can be found. The optimality criterion focuses in minimizing the "size" of the approximating function, i.e. the amount of data needed for storing the representation. The "size" of the approximating representation is proportional to the number of regions, ***N***. The free parameter in this optimization process is the set $\{\Omega_k\}_{k=1}^{N}$
The optimization problem can be simplified if the selection of regions is constrained to be defined by a Quad-Tree approach and the regions being square. This new constraint is introduced in order to make the optimization process real-time feasible although the solution would be suboptimal though still highly efficient.

Other distance functions can be considered, such as any distance of the family $E=\int_{(u,v)\in\Omega}\left(f(u,v)-\hat{f}(u,v)\right)^{\alpha}\cdot du\cdot dv$ for a positive α.
In this application the case of $\alpha=\infty$ was considered, i.e. $E=\max_{(u,v)\in\Omega}\left(f(u,v)-\hat{f}(u,v)\right)$.
This optimization problem can be expressed as $\min_{\{\Omega_k\}_{k=1}^{N}}(N)$ under the constraint of Equation 1.
An example of a 2D function is presented in the next figures where a 2D image representing depths is approximated by a PWML representation.
Figure 4 shows a Depth image shown as a 2D function, being the domain the horizontal and vertical pixels. The same information is shown in Figure 5 as an image, in grayscale. A PWML approximating representation of the data is presented in **Figure 6** (figures 1 and 3 are shown in the same scale and orientation). The domain divided according to the evaluated quad-tree is presented in **Figure 7**.

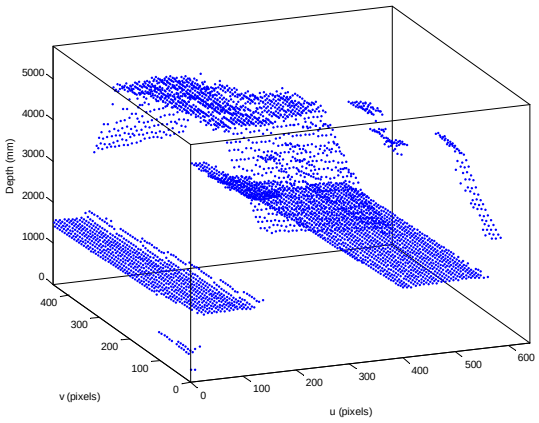

Figure 4. Depth values shown as points. The image only shows 2% of the original points for simplicity of the presentation in the picture.

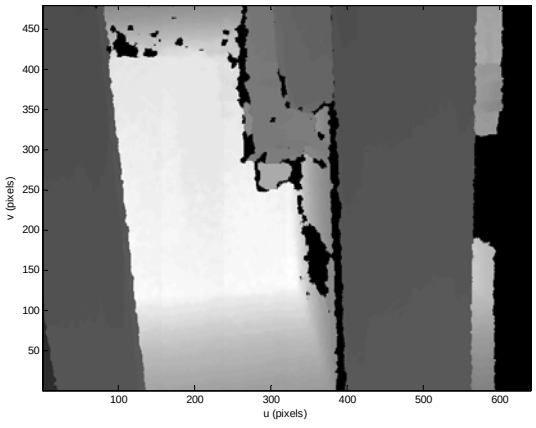

Figure 5. Depth information shown as an grayscale image.

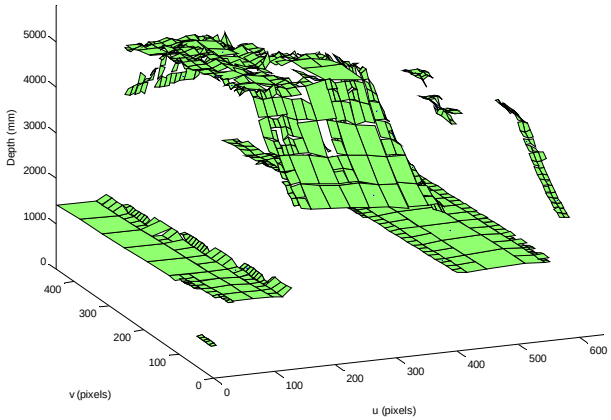

Figure 6. Relevant PWML patches for the Range image in Figure 4.. The PWLM approximated the range values treating those as a 2D function.

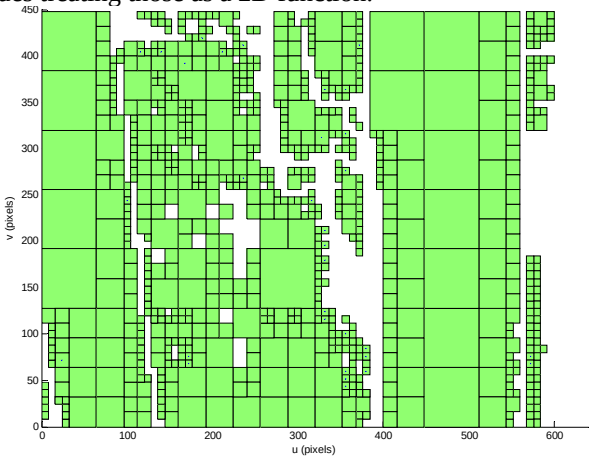

Figure 7. Relevant PWML Regions in the original 2D domain (image pixels) for the Range image in Figure 4. The small cells are not shown for simplicity of the picture. Some cells do not exist because the original Range image reported unknown values (e.g. the sensor did no generated valid values for those pixels). The density of cells increases in zones of transition (e.g. discontinuities in the range values).

Synthesis of the PWML patches in 3D.

Given a Multi-linear expression that represents the ranges associated to a set of pixels contained in a rectangular region of the image,

$$R(u,v) = a \cdot u \cdot v + b \cdot u + c \cdot v + d$$

$$(u,v) \in I \quad / \quad I = [u_i, u_f] \times [v_i, v_f]$$

where $R(u,v)$ represents the range associated to the pixel (u,v).

That set of pixels contained in the set I can be projected to its 3D equivalent through the following approximation

$$p(u,v) = (x,y,z) = (x(u,v), y(u,v), z(u,v))$$

$$p(u,v) = \mu_1 + \frac{(\mu_2 - \mu_1)}{u_f - u_i} \cdot (u - u_i)$$

$$\mu_1 = p(u_i,v_i) + \frac{p(u_i,v_f) - p(u_i,v_i)}{v_f - v_i} \cdot (v - v_i)$$

$$\mu_2 = p(u_f,v_i) + \frac{p(u_f,v_f) - p(u_f,v_i)}{v_f - v_i} \cdot (v - v_i)$$

Where the vertex points of the 3D multi linear patch, $\{p(u_i,v_i), p(u_f,v_i), p(u_i,v_f), p(u_f,v_f)\}$ are the projections of the triplets

$$\{(u_i,v_i,R(u_i,v_i)),(u_f,v_i,R(u_f,v_i)),(u_i,v_f,R(u_i,v_f)),(u_f,v_f,R(u_f,v_f))\}$$

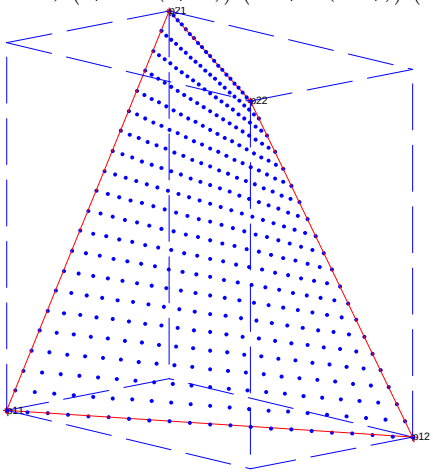

Figure 8. A Multi Linear patch defined by its four vertex points. This is an unusual case of a twisted patch.

A by-product of the PWML: Data Compression

In this application the PWML is mainly used for classifying the terrain characteristic. Additionally the PWML representation provides excellent compression properties. In the following figure the compressiopn performance of the PWML optimizer is shown.

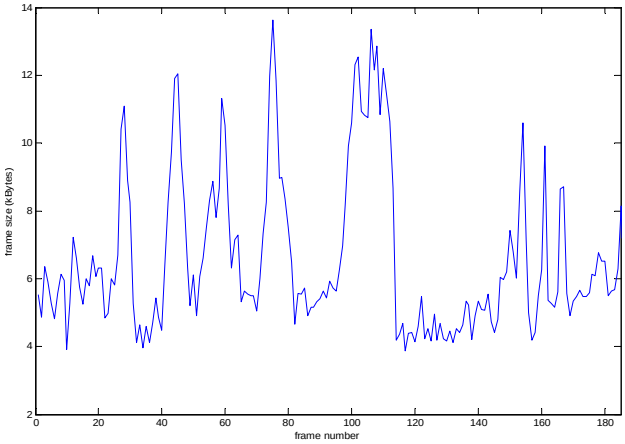

Figure 9. PWML compression ratio for an error tolerance of 5cm. This figure shows the size needed for storing/transmitting full 3D frames. A raw depth image is stored in 412KB, what means the size (in memory requirements) of the compressed version is less than 5% of the original size.

3 Typical Cases of Terrain Perception

Some characteristic cases that usually occur when riding the bike are shown in this section. The processing and perception process is explained through the figures. The first figure shows the raw data as acquired by the 3D sensor. It includes the Depth image and the associated RGB image (not used in the perception process). The second figure presents the PWML approximation. In some cases the normal vectors are also shown in the figure. Those vectors are always used in the perception process. And in the last figure the labelled PWML patches are shown in colours for indicating the location of risks and traversable regions.

Case 1: Detecting a Kerb

This case describes how the perception process performs in inferring the presence of kerbs

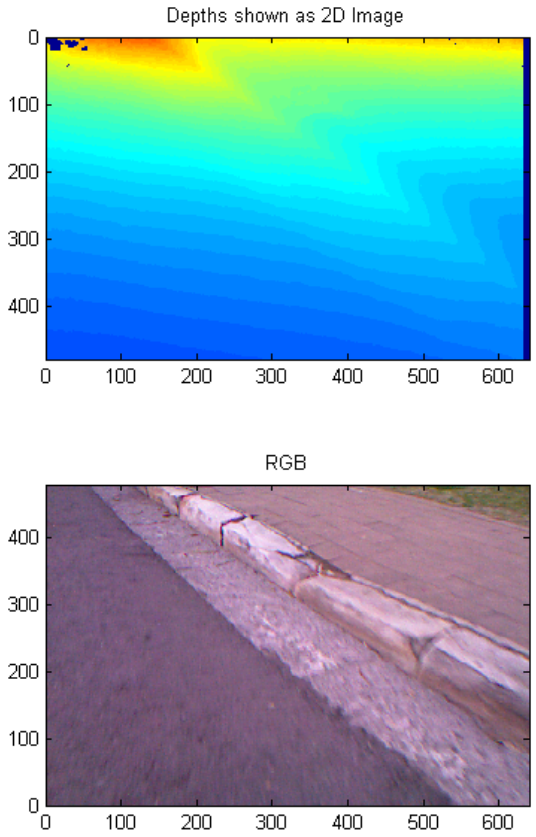

Figure 10. Typical image. Depth and RGB images capturing the road and part of the kerb at one moment of the trip.

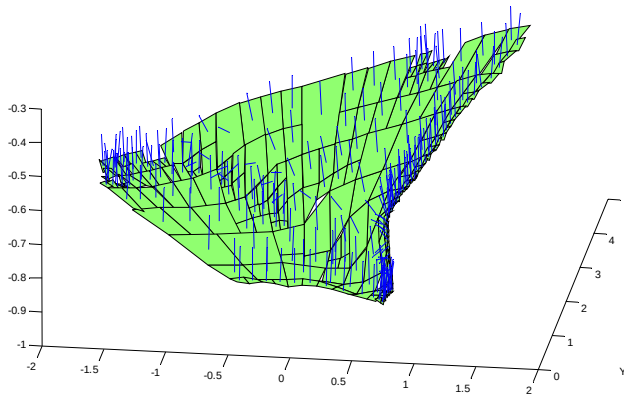

Figure 11. The PWML representation and the associated normal vectors. Each patch that approximates a region of the original surface has an associated normal vector. The associated vector corresponds to the normal vector of the multi-linear functional at its middle point.

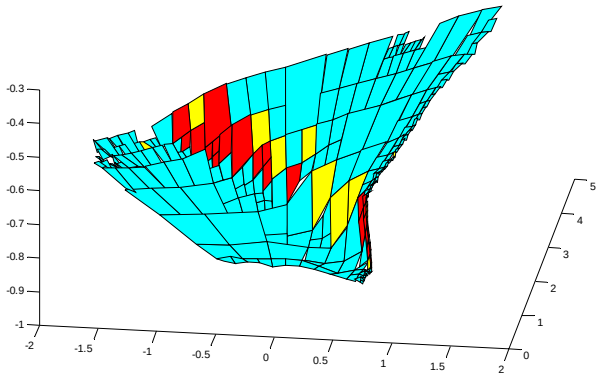

Figure 12. The PWML patches labelled according to their difficulty level, i.e. almost horizontal patches in cyan, representing traversable regions. Red patches are not traversable. Yellow are less aggressive features but still not traversable regions.

Case 2: Detecting Poles

This section shows the case where objects such as poles are present in the area of operation of the platform. The PWML representation captures the regions that are a risk for the platform. The classification is the risk is highly reliable as shown in the related figures.

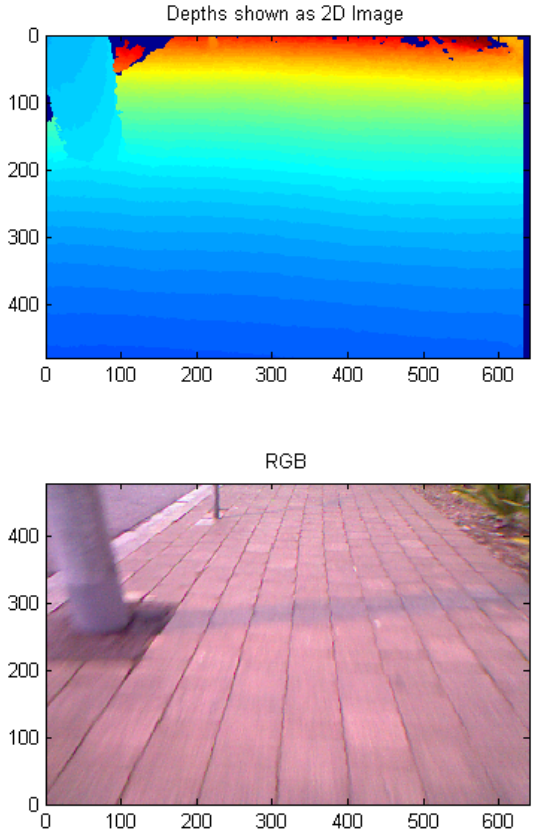

Figure 13. Context: a flat sidewalk and some pole. The top subfigure shows the Depth image as provided by the sensor. The lower picture shows the RGB image. It can be seen that a pole appears on the left of the image.

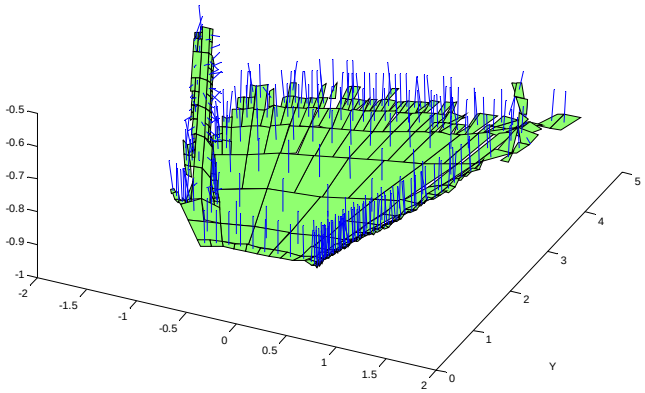

Figure 14. The PWML representation and the associated normal vectors.

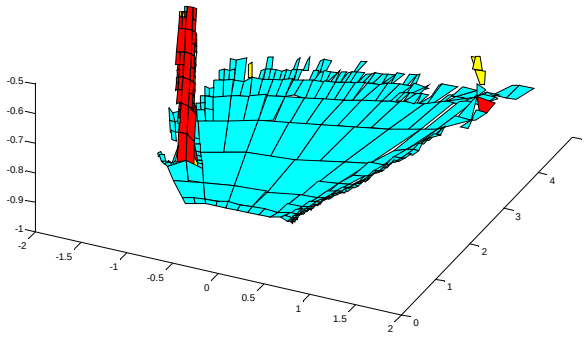

Figure 15. The labelled PWML representation indicating the patches inferred to be risks in red and safe regions in cyan. A clear indication of the pole can be seen on the left.

Case 3: Modelling a Narrow Passage

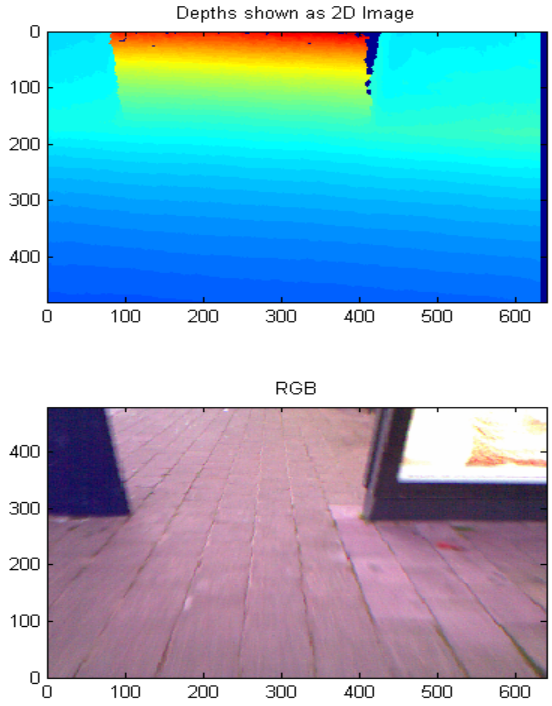

Figure 16. Image taken when the bike was approaching a narrow passage.

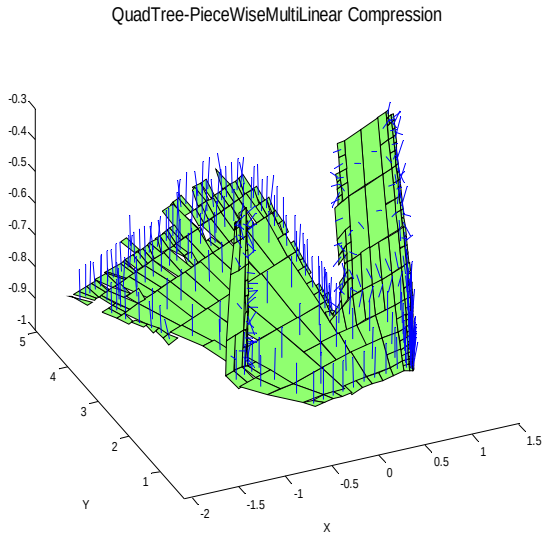

Figure 17. The PWML approximation showing the patches and the associated normal vectors.

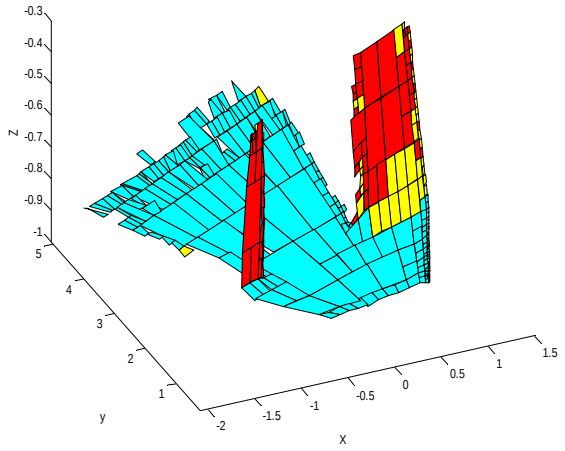

Figure 18. A figure showing the PWML patches and the associated labels as colours. Clear indication of the passage constraints can be seen in read and yellow.

3 Path Planning Process

The planning process is implemented by a PPQ-Dijstra as presented in [Robledo , 2010]. The optimization approach is a modified version of the well known Dijkstra's algorithm [Dijkstra, 1959].
The implemented planning process needs an Occupancy Grid (OG) for performing the optimization. Occupancy Grid is an efficient and widely applied method for representing the belief of the environment [Elfes, 1989] [Thrun and Bucken, 1996].
The OG is obtained by projecting the PWML patches into a 2D plane. OG's cells that are invaded by the projections of risky patches are considered to be obstacles. Risky patches are those having vertical inclinations, terrain discontinuities and other non traversable characteristics.

Traversable cells are those that fully intersect the projections of PWML patches considered traversable. Unknown cells are those that have not been invaded by any projected PWML patch.

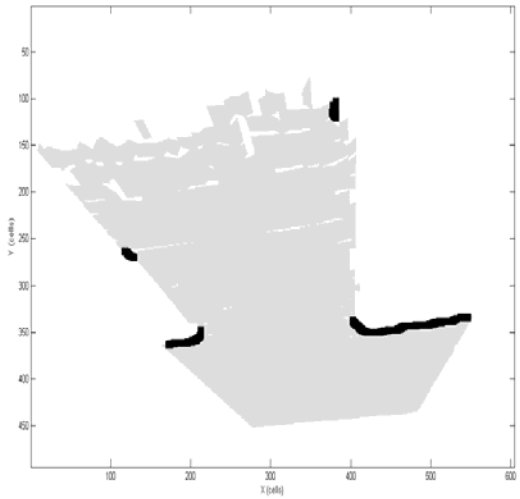

Figure 19. Occupancy grid for the Narrow Passage example. It was generated through the projection of the labelled PWML surface. In the OG definition three labels have been considered. Obstacle (black cells), traversable (gray) and unknown (white).

The PPQ-Dijkstra approach was used for solving the short term planning task. The destination position was usually set approximately 6 meters ahead the bike with the aim of generating a path in order of avoiding short term obstacles

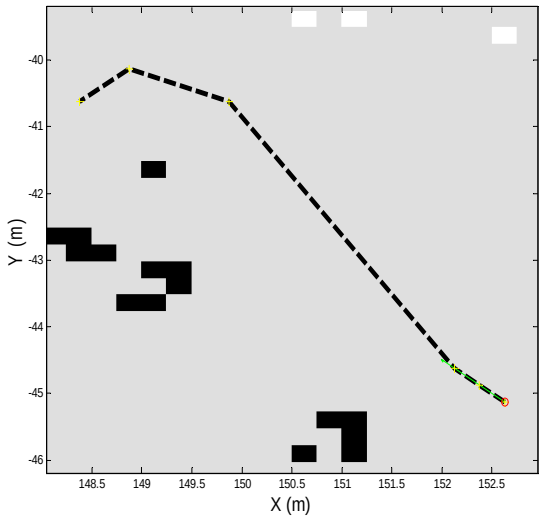

Figure 20. An Occupancy Grid generated based on the classified terrain. The planer produces a "holonomic" path based on the OG. The black cells are regions of the surface that have been labelled as risks (potholes or salient objects). Risks are expanded in size in order to generate a safety band surrounding the obstacles as shown in the next figure.

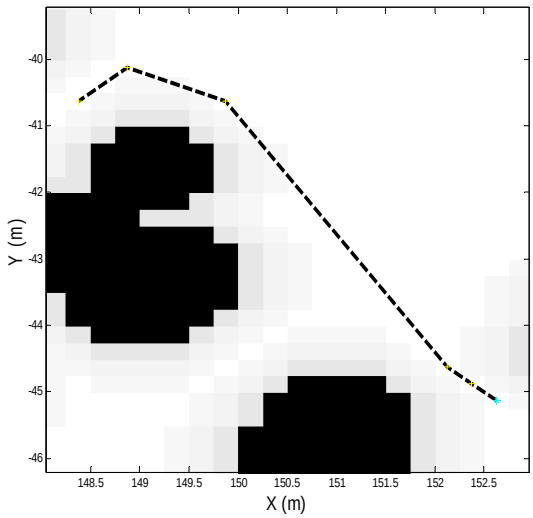

Figure 21. Expanded Occupancy Grid. The input to the planner is this modified version of the OG. The illegal regions are surrounded by a band of high cost in order to punish the path if approaching the risks. In the image, the black cells are expanded obstacles, white is traversable and light gray is traversable but more expensive.

4 Public Dataset

The dataset acquired for this test is available for students and researchers.
The dataset is composed by 4 sources of measurements:

1) Kinect camera operating at 5Hz.
2) IMU (MicroStrain GX3) providing 3D accelerometers, 3D gyros and 3D magnetometers. Sampling frequency f=200hz.
3) Low quality GPS (at f=1Hz)
4) Extra camera (Webcam) operating at 5hz

The outdoor experiments were performed during the day but in illumination conditions that did not affect the performance of the 3D camera that is intended for indoor contexts.
All the data sources are have high quality timestamps referenced to the same clock what make them feasible of being used in a data fusion process.
The dataset includes a playback module and a Matlab API for off-line and real-time accessing of the data.
In addition to the raw measurements other sources of data are included such as the 3D pose (position and attitude) estimates; the PWML version of the 3D data and Google Earth paths.
The full datasets used in this paper are available at [Dataset and Video]

5 Conclusion

This paper has presented a perception approach, and its implementation, for the modelling of the terrain based on Piece Wise Multilinear representation and the data generated by a commercial low cost 3D camera, mounted on a bike.
By application of the Piece Wise Multi-linear (PWML) it is possible to efficiently extract relevant features of the terrain. These features are used for the synthesis of a safe path.
The results show that the approach is adequate for solving the problem and it is also applied in real-time at the expense of low processing cost.
The produced datasets are a valuable public resource for the educational and research community.

References

[Robledo et al., 2010] Robledo, A., Cossell, S., and Guivant, J. E. (2010). Pseudo priority queues for Real-Time performance on dynamic programming processes applied to path planning. In Proceedings of the 2010 Australasian Conference on Robotics & Automation, Brisbane.

[Katz, et al., 2006] Katz, R., Melkumyan, N., Guivant, J., Bailey, T., Nieto, J., Nebot, E. Integrated sensing framework for 3D mapping in outdoor navigation (2006) IEEE International Conference on Intelligent Robots and Systems, art. no. 4058721, pp. 2264-2269.

[Thrun and Bucken, 1996] Sebastian Thrun and Arno Bucken. Integrating grid-based and topological maps for mobile robot navigation. In Proceedings of the Thirteenth National Conference on Artificial Intellegence, Portland, Oregon, USA, August 1996

[Dijkstra, 1959] E. W. Dijkstra. A note on two problems in connexion with graphs. Numerische Mathematik, 1(1):269{271, 1959.

[Tomatis et al., 2003] Nicola Tomatis, Illah Nourbakhsh, and Roland Siegwart. Hybrid simultaneous localization and map building: a natural integration of topological and metric. Robotics and Autonomous Systems, 44(1):3{14, February 2003.

[Elfes, 1989] Alberto Elfes. Using occupancy grids for mobile robot perception and navigation. Computer, 22(6):46{57, June 1989

[Melkumyan, 2009] Melkumyan, N, PhD Thesis, "Surface-based Mapping for Unstructured Environments" The Univerisity of Sydney, 2009

[Whitty et al., 2010] Mark Whitty, Stephen Cossell, Kim Son Dang, Jose Guivant, and Jayantha Katupitiya. Autonomous navigation using a real-time 3d point cloud. In Australasian Conference of Robotics and Automation, Brisbane, Australia, December 2010.

[Dataset and Video] http://www.robotics.unsw.edu.au/index.php?option=com_jresearch&task=show&view=researcharea&id=14&Itemid=54