A GENERATIVE TRAVERSABILITY MODEL FOR MONOCULAR ROBOT SELF-GUIDANCE

Michael Sapienza[1], Kenneth P. Camilleri[1]
[1] *Department of Systems & Control Engineering, University of Malta, Msida MSD 2080, Malta*
{*mikesapi@gmail.com, kenneth.camilleri@um.edu.mt*}

Keywords: Traversability Detection; Autonomous Robotics; Self-Guidance.

Abstract: In order for robots to be integrated into human active spaces and perform useful tasks, they must be capable of discriminating between traversable surfaces and obstacle regions in their surrounding environment. In this work, a principled semi-supervised (EM) framework is presented for the detection of traversable image regions for use on a low-cost monocular mobile robot. We propose a novel generative model for the occurrence of traversability cues, which are a measure of dissimilarity between *safe-window* and image superpixel features. Our classification results on both indoor and outdoor images sequences demonstrate its generality and adaptability to multiple environments through the online learning of an exponential mixture model. We show that this appearance-based vision framework is robust and can quickly and accurately estimate the probabilistic traversability of an image using no temporal information. Moreover, the reduction in *safe-window* size as compared to the state-of-the-art enables a self-guided monocular robot to roam in closer proximity of obstacles.

1 INTRODUCTION

Giving autonomous robots the ability to explore and navigate through their environment using CCD/CMOS cameras has become a major area in mobile robotics research (DeSouza and Kak, 2002). This paper addresses the fundamental problem of determining terrain traversability for a mobile robot equipped with a single camera (Lorigo et al., 1997; Ulrich and Nourbakhsh, 2000; Santosh et al., 2008; Katramados et al., 2009).

Even though autonomous guidance has achieved relative success using active sensors, this task still remains challenging for robots equipped with vision sensors (Santosh et al., 2008). However, in addition to providing local depth information in the vicinity of the robot, camera information has the potential to provide long-range traversability information and environmental semantics (Hadsell et al., 2009; Hoiem et al., 2007), making it ideal for mobile robot exploration and navigation.

Multiple camera vision has been used for depth estimation and image analysis (Roning et al., 1990), however in practice, 3D reconstruction works well for close objects, with the accuracy diminishing significantly with distance from the camera (Michels et al., 2005; Hadsell et al., 2009). Recent self-guided vehicles used in the DARPA LAGR programme have led to significant advances in robotic perception systems (Hadsell et al., 2009; Sofman et al., 2006; Kim et al., 2007), however the multiple sensors and complexity of these systems do not address the needs of low-cost autonomous robots (Katramados et al., 2009; Murali and Birchfield, 2008). A commonly available web camera presents a desirable alternative that will be used in this work, and is motivated by the human ability to interpret 2D low resolution images (Murali and Birchfield, 2008).

A self-guided monocular robot needs to extract information on surrounding objects in order to identify areas through which it can move. This problem has been approached by providing the robot with a detailed description of its environment, usually an explicit geometric or topological map built manually or extracted from the stereo/monocular vision sensors on the robot (Kosaka and Kak, 1992; Meng and Kak, 1993; Ohno et al., 1996). Building such representations of the environment is time-consuming, and constrains the limits of operation of the robot to the particular environment in which the hard-data was collected. For mobile robots that can be used in dynamically changing environments, a basic form of scene traversability understanding must be available to the robot (Kim et al., 2007). This computer vision task

forms a basic building block that intelligent systems will need to gain autonomy, and upon which more complex behaviours can be built.

Motivated by autonomous Martian landscape exploration, Lorigo (Lorigo et al., 1997) proposed an appearance-based approach to traversability detection in which a rectangular *safe-window* of pixels towards the bottom of the captured image is assumed to be traversable. This appearance-based technique was shown to work in real-time and was implemented in various indoor (Ulrich and Nourbakhsh, 2000; Santosh et al., 2008), and outdoor (Lorigo et al., 1997; Katramados et al., 2009) environments. The size of the *safe-window* determines the closest safe distance between the robot and obstacles, for a given camera pose and optical properties. Thus, a smaller *safe-window* will allow greater robot agility and manoeuvrability between obstacles in a cluttered environment. Moreover, reducing the size of the *safe-window* allows dynamic obstacles to move closer to the robot without the risk of being captured in this window.

Due to the real-time requirement of mobile robotic systems, image region classification must be computationally efficient. Ulrich (Ulrich and Nourbakhsh, 2000) used a static threshold on the feature histograms of the *safe-window* pixels to determine image traversability, making this approach unsuitable for other novel environments that may require different thresholds. Santosh (Santosh et al., 2008) based his method on that of Ulrich stating that the histogram threshold is determined from the histogram entropy, but does not provide any details how this is achieved. On the other hand, Katramados (Katramados et al., 2009) determines a classification threshold from the *safe-window* histogram peak and mean level, thus clearly showing that the classification method allows the robot to be used in other novel environments. However, a histogram with multiple peaks may result from a *safe-window* over composite surfaces, thereby employing multiple thresholds. An interesting simplified alternative is proposed by Lorigo (Lorigo et al., 1997), where the area of overlap between the feature histograms of the safe window and that of rectangular patches is computed. However a *static* threshold on this area is used to assign a traversable or non-traversable label to each image patch, making it unsuitable to novel environments without threshold tuning. In this work, we develop a framework that allows the robot to be used in novel environments, as in (Katramados et al., 2009), and that uses a dissimilarity measure between image regions as a cue for traversability, as in (Lorigo et al., 1997). However, instead of using the feature distributions directly (Lorigo et al., 1997; Ulrich and Nourbakhsh, 2000; Santosh et al., 2008; Katramados et al., 2009), we propose to model the feature *dissimilarity* distribution. This allows a probabilistic framework to be used in which the dissimilarity model parameters are self-learned in a semi-supervised manner, allowing the robot to be used in new environments.

The main contribution of this work is a novel generative model for the classification of traversable image regions with the *safe-window* approach (cf. Section 2.5). In addition to the inherent environment adaptability provided by the *safe-window* (Lorigo et al., 1997; Ulrich and Nourbakhsh, 2000; Santosh et al., 2008; Katramados et al., 2009), our traversability classification method (cf. Section 2) is based upon a principled framework in which a mobile robot can also self-learn its model parameters in *any* novel environment where the traversable region differs by some degree to the appearance of obstacles. Similarly to previous *safe-window* approaches, once the robot is initialized in a particular environment, it cannot make a transition to another traversable surface with different appearance properties, unless the *safe-window* is reinitialized manually or automatically by means of active sensors. Once initialized however, our method will allow the model parameters to adapt to the present ground/obstacle dissimilarity and varying lighting conditions of the ground in a semi-supervised manner. In the experimental section (cf. Section 3) we demonstrate that our approach allows robust traversability classification on single image frames without requiring temporal information (cf. Section 3.1). This means that the algorithm may be used intermittently alongside other computationally intensive algorithms such as human gesture recognition. Furthermore, the method is robust to the reduction in *safe-window* size, without loss in classification performance (cf. Section 3.2). This allows a mobile robot to guide-itself safely and maneuver in a tight corridor space cluttered with obstacles (cf. Section 3.3).

2 Methods

In this work, traversability detection is accomplished by adopting the well-known principled probabilistic framework based on Bayes' rule to infer the class label of image regions from *traversability cues*. The traversability cues $\mathbf{X} = \langle X_1, X_2, \ldots X_j, \ldots X_n \rangle$ are found by comparing descriptive feature distributions (cf. Section 2.1) from oversegmented regions called superpixels (cf. Section 2.2) to those in the *safe-window*, by using a dissimilarity metric (cf. Section 2.3). Initially, n descriptive feature distributions

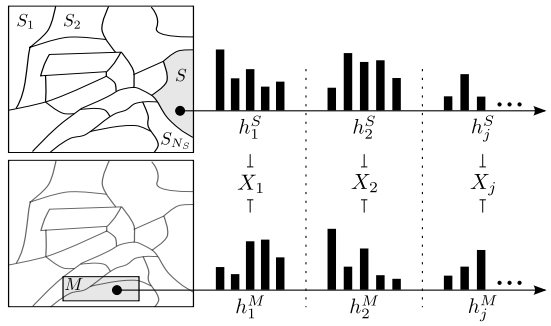


Figure 1: The discrete distributions of the descriptive features in each superpixel S are compared to those of the model in M. For each traversability cue j, a random variable X_j is found by comparing the query and model distributions with a dissimilarity metric.

h_j^S and h_j^M, where $j = 1,...,n$, are extracted from image superpixel S and from the *safe-window* M respectively, as illustrated in Fig. 1. The traversability cue values $X_j = x_j = d(h_j^S \| h_j^M)$ are a measure of dissimilarity between the descriptive feature distributions in S and M. The occurrence of traversability cue values are modelled by a generative traversability model (cf. Section 2.5), whose parameters are updated from the traversability cue statistics by means of the Expectation-Maximization algorithm (cf. Section 2.6). Finally, Bayes' rule is used to calculate the posterior probability $P(\Theta|\mathbf{X})$ of a superpixel being traversable, *given* a vector of traversability cues $\mathbf{X}$:

$$P(\Theta = \theta_l|\mathbf{X}) = \frac{P(\mathbf{X}|\Theta = \theta_l)P(\Theta = \theta_l)}{\sum_m P(\mathbf{X}|\Theta = \theta_m)P(\Theta = \theta_m)}, \quad (1)$$

where $\Theta = \theta_l$ is a Boolean random variable representing the l^{th} class label of superpixel S, and which can take on values defined by $\Theta \in \{\theta_1, \theta_2\}$. In this case, θ_1 and θ_2 represent the traversable and non-traversable classes respectively.

Assuming conditional independence between traversability cues X_j, the most likely class label is chosen using the *maximum a posteriori* (MAP) decision rule by

$$\Theta^* \leftarrow \arg\max_{\theta_l} P(\Theta = \theta_l) \prod_{j=1}^{n} P(X_j|\Theta = \theta_l). \quad (2)$$

Note that this probabilistic framework allows a high degree of flexibility in the choice of descriptive features, image primitives, comparison metrics, prior, and modelling of the traversability cue values. Though it is not the purpose of this work to compare different techniques in various stages of the algorithm, this framework affords a straightforward way to do so. The following subsections will describe the design choices for each element of the framework.

Table 1: Descriptive Features.

No. j	Features	Dim.	Type
1	Hue colour channel	32	Colour
2	Saturation colour channel	32	
3	Illumination Invariant Channel	32	
4	Edge gradient magnitudes	32	Texture
5	Edge gradient orientations	9	
6	Local Binary Patterns (LBP)	32	

2.1 Descriptive features

The set of descriptive features that are considered in this work were inspired from (Lorigo et al., 1997; Ulrich and Nourbakhsh, 2000; Katramados et al., 2009; Davidson and Hutchinson, 2003), and are listed in Table 1. The colour features have been chosen to minimize the effects of shadows and reflections that can confuse the classifier (Katramados et al., 2009; Ulrich and Nourbakhsh, 2000; Lorigo et al., 1997). The three illumination invariant colour cues considered are the hue (H) and Saturation (S) channels of the HSV colour space, and a combination of intensity invariant channels from the YCbCr and LAB colour space. This Illumination Invariant Colour channel is found by a weighted combination of the *Cb*, *Cr*, and *A* colour channels, as suggested by (Katramados et al., 2009).

Texture features are also important where colour information is not sufficient or even present at all. We consider the edge gradient magnitudes and orientations as is common in object recognition (Dalal and Triggs, 2005), and the local texture distributions provided by the Local Binary Pattern Operator (LBP), which can be computed very efficiently (Davidson and Hutchinson, 2003; Mäenpää et al., 2003).

2.2 Oversegmentation

The basic image regions used for traversability classification have classically been pixels (Ulrich and Nourbakhsh, 2000; Katramados et al., 2009) or rectangular patches (Lorigo et al., 1997). Using pixels may result in a noisy/spotty classification since pixel neighbourhoods are not considered. Patch based classification allows local feature distributions to be extracted, but may contain multiple object boundaries within the same region. An oversegmented representation of the image into superpixels overcomes these shortcomings since superpixels delineate homogeneous pixel regions whilst preserving the image structure. Thus rich pixel statistics can be extracted from more perceptually meaningful regions (Kim et al., 2007; Santosh et al., 2008; Hoiem et al., 2007).

In this implementation, the initial pixel grouping is done using the fast oversegmentation technique from (Felzenszwalb and Huttenlocher, 2004). We have used the code publicly released by the authors with the parameters $\sigma = 0.5, k = 100, min = 100$, where σ is a smoothing constant, k is a threshold which determines how readily image regions are joined together, and *min* is minimum superpixel size (Hoiem et al., 2007).

2.3 Dissimilarity metric

In the current implementation, the dissimilarity measure used to compare the superpixel and *safe-window* feature distributions is the G-statistic. This dissimilarity measure is based on the Kullback–Leibler cross entropy measure, and was inspired from (Mäenpää et al., 2003) where it was used to compare the distributions of local binary patterns.

2.4 Simple prior

Using the exponential function, a heuristic prior is constructed which favours superpixels towards the lower parts of the image to be traversable. Let the prior likelihood functions for the expectation of traversable superpixels before seeing the data be exponentially distributed with rate λ:

$$P(C|\theta_1) = \frac{1}{Y_1} e^{-\lambda C}, \quad P(C|\theta_2) = \frac{1}{Y_2}\left(1 - e^{-\lambda C}\right). \tag{3}$$

Thus, the probability of a superpixel being traversable given its centre pixel row position C (the mean superpixel pixel height in the image) may be expressed by the following equation from Bayes' rule:

$$P(\Theta = \theta_1|C) = \frac{1}{1 + \frac{Y_1}{Y_2}(e^{\lambda C} - 1)}, \tag{4}$$

where Y_1 and Y_2 are normalization values updated on each iteration to ensure the probability over all possible height values sums to one. The rate parameter λ fully describes the exponential distribution, and its Maximum Likelihood Estimate (MLE) $\hat{\lambda}$ is the reciprocal of the sample mean $\bar{C}$ of traversable superpixels (Garthwaite et al., 2002).

2.5 Generative traversability model

The superpixels S in image I originate from either the traversable (θ_1) or non-traversable (θ_2) class. Since the *safe-window* M is *a priori* traversable, when it is compared to a traversable superpixel, a low dissimilarity is expected (approaching zero). On the other hand, if the safe window is compared to a superpixel from the non-traversable class, a large dissimilarity score is expected (approaching a maximum g_{max}). However, it is also possible to have obstacle regions similar in appearance to the ground, although with a lower probability. This reasoning can be captured in a generative model in which the probability of a random traversability cue X_j is a mixture of two truncated exponential distributions, as shown in Fig. 2.

The likelihood functions for the random variable X_j, given it was generated by matching a superpixel from class label θ_l to the *safe-window* M is a (one-sided) truncated exponential which can be expressed as

$$P(X_j|\theta_1) = \alpha_{j1} e^{-\alpha_{j1} X_j} \left(1 - e^{-\alpha_{j1} g_{max}}\right)^{-1}, \tag{5}$$

$$P(X_j|\theta_2) = \alpha_{j2} e^{\alpha_{j2} X_j} \left(e^{\alpha_{j2} g_{max}} - 1\right)^{-1}, \tag{6}$$

where $0 \leq X_j \leq g_{max}$, and $\alpha_{jl} > 0$ is the rate parameter of the distribution. The rate parameters need to be learned from the data and this is achieved using the EM algorithm which is discussed next.

2.6 Expectation Maximization (EM)

The learning task required here is to output a hypothesis $h_j = \langle \alpha_{j1}, \alpha_{j2} \rangle$ for each traversability cue j, that describes the rate parameters of the exponential mixture. Since it is not known which distribution gave rise to the current observation, the EM algorithm will be used to iteratively re-estimate the parameters given some current hypothesis:

***E*-Step:** Calculate the expected value $E[\theta_l|\mathbf{x}^k, h_j]$ that the l^{th} truncated exponential distribution was responsible for the j^{th} traversability cue originating from superpixel k, assuming h_j holds. We will denote this responsibility by r_l^k.

***M*-Step:** Calculate the new maximum likelihood hypothesis $h_j^{[t+1]} = \langle \alpha_{j1}^{[t+1]}, \alpha_{j2}^{[t+1]} \rangle$ assuming that the values for the responsibility r_l^k were those calculated from Step 1.

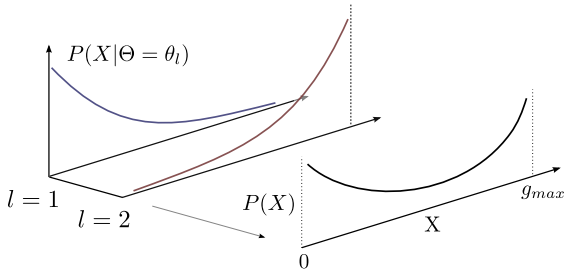


Figure 2: Mixture of two truncated exponential distributions. The mixture density created by marginalizing over the hidden variable l, which acts as an identifier for each truncated exponential distribution.

Table 2: Safe window dimensions used by various authors.

Author	Shape	Dimensions	% Area
Our Model	Rectangular	$\frac{W}{3}, \frac{H}{8}$	4.2
Katramados	Rectangular	$\frac{W}{2}, \frac{H}{4}$	12.5
Lorigo	Rectangular	$\frac{W}{3.2}, \frac{H}{6.4}$	15.6
Ulrich	Trapezoidal	$a=\frac{W}{2}, b=\frac{W}{1}, h=\frac{H}{3.3}$	22.3

It can be shown that the maximum likelihood estimate for a single exponential distribution parametrized by α_{jl} given the observed data instances $\langle \mathbf{x}^1, \mathbf{x}^2, ...\mathbf{x}^k, ...\mathbf{x}^{N_S} \rangle$ is the reciprocal of the sample mean (Garthwaite et al., 2002):

$$\hat{\alpha}_{jl}^{[t+1]} = \frac{1}{\bar{x}_{jl}^{[t+1]}}, \quad \text{where } \bar{x}_{jl}^{[t+1]} = \frac{\sum_k r_l^k x_j^k |S_k|}{\sum_k r_l^k |S_k|}, \quad (7)$$

and $|S_k|$ is the size of superpixel S_k in pixels ($k = 1, ..., N_S$). This expression is a weighted sample mean of x_j^k, where each instance is weighted by the expected value that it was generated by one of the two exponential distributions (Mitchell, 1997; Prince, 2011). Note that since the exponential distribution is truncated, the mean of the distribution becomes

$$\bar{x}'_{jl} = \frac{1}{\hat{\alpha}_{jl}} - g_{max}\left(e^{\hat{\alpha}_{jl} g_{max}} - 1\right)^{-1}, \text{ where } \hat{\alpha}'_{jl} = \frac{1}{\bar{x}'_{jl}}, \quad (8)$$

and $\hat{\alpha}'_{jl}$ will be a MLE only if $0 < \bar{x}'_{jl} < \frac{g_{max}}{2}$ (Al-Athari, 2008).

3 Experimental Results & Discussion

In order to test the validity of our approach, the vision framework was tested on three datasets: i) a *Static Traversability dataset* containing still images, ii) the *Cranfield University dataset* containing video sequences acquired from a teleoperated robot (Katramados et al., 2009), and iii) a *Self-Guided dataset* captured from a low-quality webcam during robot autonomous guidance. All images were reduced to a resolution of 160×120 and the size of the *safe-window* was set to: $\frac{W}{3}, \frac{H}{8}$, whose top left corner is located at position $\frac{W}{3}, \frac{H}{7}$, where W, H are the width and height of the image respectively. The various *safe-window* sizes used in the literature are compared in Table 2.

3.1 Static Traversability dataset

This dataset is made up of 100 challenging images of indoor and outdoor scenes picked from the Internet. These images contain scenes with highly reflective surfaces, specular reflections and traversable regions with varying amount of colour and texture, as shown in Fig. 3. Ground truth information was obtained through manual labelling of the traversable image regions by a human observer, and is made available online[1].

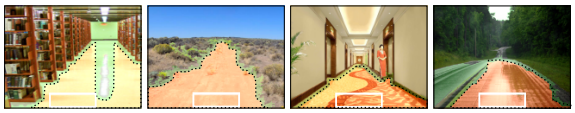


Figure 3: A sample of surfaces classified in the Static Traversability dataset including indoor tiling, dirt roads, textured carpets and wet roads with specular reflections. The red highlighted area shows the accessible traversable region. The white box towards the bottom of the image denotes the position of the *safe-window*.

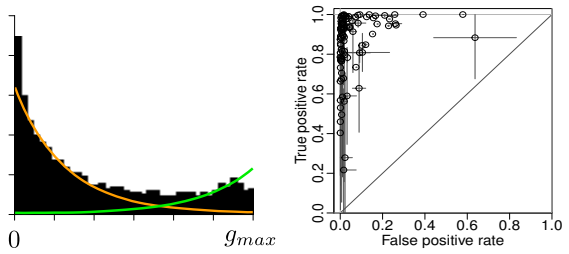
3.1.1 Generative model validation

In this experiment, the traversability cues obtained from the 100 images in the Static Traversability dataset are accumulated in a histogram, shown in Fig. 4(a). The shape of the histogram shows that the distribution of the traversability cues $\mathbf{X}$ can in fact be modelled as the joint mixture of two truncated exponential distributions and demonstrates the suitability of our proposed generative traversability model.

3.1.2 Image classification & EM initialization sensitivity

The objective of this experiment was to test the accuracy of traversability classification on images from multiple scenes with no temporal information. Each test image was subjected to 100 random initializations where the rate parameters α_{jl} and truncation point g_{max} were randomly initialized to values within the range [0.01 - 10.01]. The algorithm was allowed to iterate until the parameters converged, or up to a maximum of 10 iterations. The mean TPR and FPR obtained for each image, together with one standard deviation is shown in Fig. 4(b). Images which always converged to the same result have zero standard deviation. Those points with a large cross-size resulted from images which converged to different TPR-FPR results. Overall, 89% of the images converged to 1 - 2 identical FPR-TPR values, and the mean classification accuracy was 91.62% with a standard deviation of 8.78%. The model parameters α_{jl} typically converged within 3-5 iterations, with respective processing times varying from 80-150ms. This result demonstrates the model robustness to random initialization,

[1] Visit: https://sites.google.com/site/mikesapi



(a) (b)

Figure 4: (a) Histogram of normalized traversability cues **X** accumulated from the Static Traversability dataset, which demonstrates the suitability of our generative traversability model (cf. Section 2.5). (b) ROC plot showing the mean and one standard deviation resulting from the static image classification.

and the adaptability of its classification model parameters to multiple types of indoor and outdoor environments with no temporal model, in contrast to (Katramados et al., 2009) which uses a temporal memory model. Note that the output of our algorithm can be extended by a tracking algorithm which incorporates temporal and kinematic information. Errors occurred where more than one ground type was present in the image, of which one was not represented in the *safe-window*, and when obstacle regions had a similar appearance to the ground.

3.2 Cranfield University dataset

This dataset provided by Katramados *et al.* (Katramados et al., 2009), consists of eight outdoor video sequences captured from a teleoperated mobile robot. Ground truth labelling was available from the author at a rate of one frame per second. The video sequences were captured over a wide range of outdoor conditions (cloudy, wet, sunny, shadows), and a range of terrain types (concrete, grass, soil, tarmac, snow), as shown in Fig. 5.

This experiment shows that our method can easily be extended to video sequences and operate in real-time. Instead of allowing the EM algorithm to converge on each video frame (cf. Section 3.1.2), the correlation between frames allows the algorithm to converge *across* frames. This reduces the computational cost to 30-50ms on each image (20-30fps). The accuracy results obtained using a 12.5% area *safe-window*, a 4.2% area *safe-window*, and the result *Safe-4.2|25* of discarding 24 out of 25 frames in each sequence are listed in Table 3. The similar results from *Safe-12.5%* and *Safe-4.2%* demonstrate that the reduction in *safe-window* size from 12.5% to just 4.2% of the image area did not reduce the accuracy of the algorithm. This makes it more suitable for self-guided robots to

Table 3: Cranfield University dataset results

	Safe-12.5%		**Safe-4.2%**		**Safe-4.2|25**	
Conditions	*%Acc*	*%Std*	*%Acc*	*%Std*	*%Acc*	*%Std*
Cloudy Dry	95.11	2.11	**95.35**	2.57	94.91	4.40
Cloudy Wet	93.54	3.94	**93.74**	4.53	92.15	11.12
Cloudy Muddy	**91.63**	4.00	89.30	6.62	81.49	15.28
Sunny Wet	82.13	9.49	**82.32**	8.88	71.24	12.71
Complex Shadow	83.98	9.10	**85.77**	6.35	85.20	8.46
Sunny Dry	89.64	4.61	**89.71**	4.53	84.73	13.99
Strong Shadows	85.23	15.62	**88.33**	12.42	87.03	14.39
Snow	88.50	10.95	**89.19**	10.29	89.18	9.93
Mean	88.72	7.48	**89.21**	**7.02**	85.74	11.29

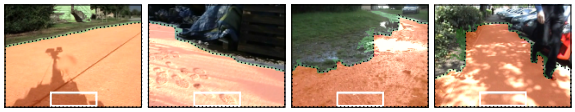


Figure 5: Sample images of the classification results taken from the Cranfield University dataset. The samples were taken from videos (starting from left): Strong Shadows, Snow, Cloudy Muddy, Cloudy Wet, Complex Shadows.

move about in the proximity of obstacles, as detailed in the *Self-guided* experiment. The ROC plots for the 4.2% area *safe-window* results are shown in Fig. 6, where it is seen that our model achieves better performance in environments where the ground and obstacles have a contrasting appearance (e.g. Cloudy Dry, Cloudy Wet, Sunny Dry, Snow), than environments where the ground obstacle boundary is not easily discriminable (e.g. Cloudy Muddy, Sunny Wet). Dropping 24 out of 25 frames in each video sequence, the accuracy results in *Safe-4.2%|25* were negatively affected, more so in Sunny Wet where the appearance of the ground is changing very quickly. In the other sequences however, this robustness makes it possible to use the traversability algorithm intermittently alongside other computationally expensive algorithms. Although the authors of the dataset (Katramados et al., 2009) reported higher classification rates (97.6% Acc, 2.7% Std) on these video sequences, Katramados *et al.* included a temporal model in their framework, thus a direct comparison of the results is not possible. The results in (Katramados et al., 2009) are given for offline classification; in the next sub-section we provide performance results for online autonomous robot self-guidance, which is the ultimate purpose of this system.

3.3 Self-Guided dataset

An experiment was set up in which our ERA-MOBI mobile robotic platform named `VISAR01` was placed in a previously unknown indoor corridor environment cluttered with a variety of objects typically found in-

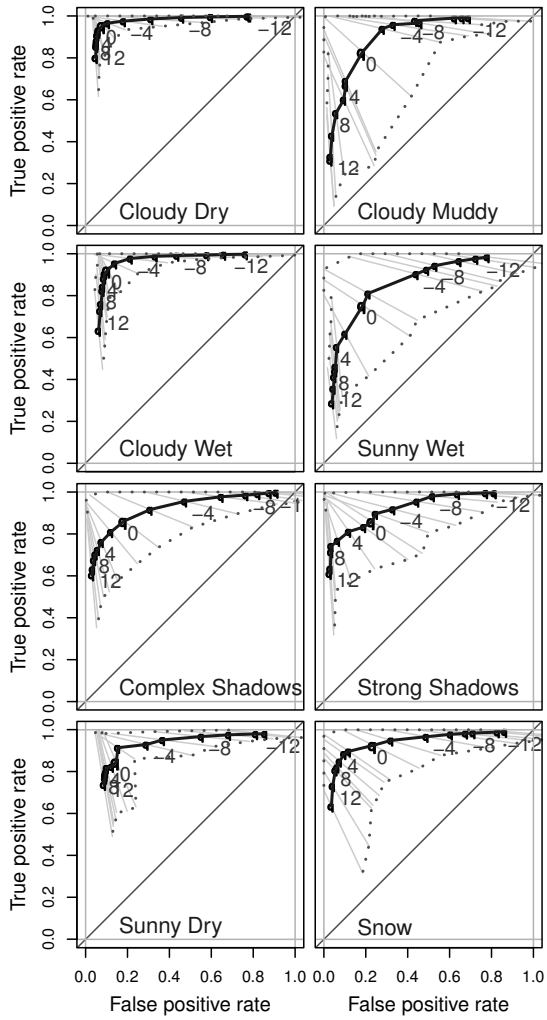


Figure 6: Mean ROC (black dots) and one standard deviation (grey lines) for the classification of video sequences in the Cranfield University dataset. The curves were plotted by varying the log posterior ratio at which classification is decided.

doors: a chair, plant, box, waist paper basket, tool box, and a person standing in the way. The floor tiles are a pale grey, making them difficult to distinguish from the white, untextured wall. In this experiment, the horizon boundary that results from the classification was used to steer the robot towards the largest open space (Santosh et al., 2008). The robot moved at 0.15m/sec and the algorithm was run on every frame at 25 fps. A video frame from the camera together with its corresponding on-line classification result was saved every 25^{th} frame, for a total of 99 frames in the 99 second sequence. Ground truth data was collected by asking a human observer to manually label the image regions as traversable or non-traversable.

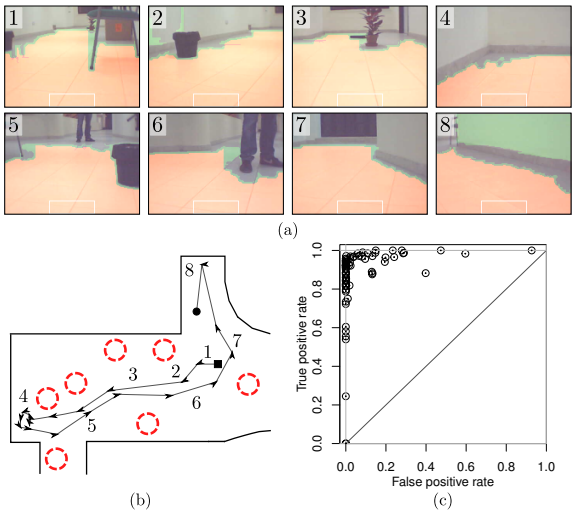


Figure 7: (a) Samples image results from the Self-Guided dataset. (b) Illustration of path taken by the robot moving autonomously through a faculty corridor dotted with static obstacles (dashed red circles). (c) ROC scatter plot for the traversable ground classification during self-guidance.

A sample of the resulting traversability classification and the path taken by the robot during the autonomous movement can be seen in Fig. 7a&b. The resulting ROC scatter plot is shown in Fig. 7c. Overall, the mean and standard deviation of the classification accuracy for this successful run were 89.71% and 9.68% respectively. The low levels of colour and texture in this sequence made the floor and wall indistinguishable at times, however when the robot approached an obstacle their difference to the ground became more apparent and therefore this did not confuse the motion of the robot. Despite a high but imperfect classification accuracy, the robot successfully managed its way around the corridor and obstacles using only a monocular web-camera and turned back on itself when it encountered a dead-end. As the ground appearance changes *gradually* due to changing lighting and reflections, the classification model was able to adapt to these changes. The robot may transit to a surface with different appearance characteristics by re-initializing the traversability model and allowing new model parameters to be self-learned.

4 Conclusion and Future work

A real-time vision algorithm has been designed for a mobile robotic platform to detect traversable areas and guide itself safely in proximity of obstacles using the smallest reported *safe-window*. We have modelled the feature *dissimilarity* distribution with a

truncated exponential mixture model and showed the model's competence without the need for a temporal model, prior training, or manual adjustments to the system parametrization. The robustness of the generative model to initialization, and its ability to learn the model parameters for textured/untextured, indoor/outdoor environments have been demonstrated through experimental analysis and from the many hours VISAR01 has been allowed to roam out into the faculty corridors, avoiding both static and dynamic objects.

Future work will see the inclusion of structure-from-motion depth estimation to allow the robot to transition from one type of surface to another automatically, and new exploration behaviours based on the probability of traversability, rather than simple binary classification. This means that instead of merely moving towards an obstacle-free path determined by a hard decision (Santosh et al., 2008), the robot may decide to take the path that has the highest probability of being traversable.

ACKNOWLEDGEMENTS

The research work disclosed in this publication is partially funded by the Strategic Educational Pathways Scholarship (Malta). The scholarship is part-financed by the European Union - European Social Fund (ESF) under the Operational Programme II - Cohesion Policy 2007-2013, Empowering People for More Jobs and a Better Quality of Life.

REFERENCES

Al-Athari, F. (2008). Estimation of the mean of truncated exponential distribution. *Journal of Mathematics and Statistics*, 4(4):284–288.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 886–893.

Davidson, J. and Hutchinson, S. (2003). Recognition of traversable areas for mobile robotic navigation in outdoor environments. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 297–304.

DeSouza, G. and Kak, A. (2002). Vision for mobile robot navigation: A survey. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(2):237–267.

Felzenszwalb, P. and Huttenlocher, D. (2004). Efficient graph-based image segmentation. *Int. Journal of Computer Vision*, 59(2):167–181.

Garthwaite, P., Jolliffe, I., and Jones, B. (2002). *Statistical Inference*. Oxford University Press, Inc., second edition.

Hadsell, R., Sermanet, P., Ben, J., Erkan, A., Scoffier, M., Kavukcuoglu, K., Muller, U., and LeCun, Y. (2009). Learning long-range vision for autonomous off-road driving. *Journal of Field Robotics*, 26(2):120–144.

Hoiem, D., Efros, A., and Hebert, M. (2007). Recovering surface layout from an image. *Int. Journal of Computer Vision*, 75(1):151–172.

Katramados, I., Crumpler, S., and Breckon, T. (2009). Real-time traversable surface detection by colour space fusion and temporal analysis. In *Int. Conf. Computer Vision Systems*, volume 5815, pages 265–274.

Kim, D., Oh, S., and Rehg, J. (2007). Traversability classification for UGV navigation: a comparison of patch and superpixel representations. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 3166–3173.

Kosaka, A. and Kak, A. (1992). Fast vision-guided mobile robot navigation using model-based reasoning and prediction of uncertainties. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 2177–2186.

Lorigo, L., Brooks, R., and Grimson, W. (1997). Visually-guided obstacle avoidance in unstructured environments. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 373–379.

Mäenpää, T., Turtinen, M., and Pietikäinen, M. (2003). Real-time surface inspection by texture. *Real-Time Imaging*, 9(5):289–296.

Meng, M. and Kak, A. (1993). NEURO–NAV: A neural network based architecture for vision-guided mobile robot navigation. In *IEEE Int. Conf. on Robotics and Automation*, pages 750–757.

Michels, J., Saxena, A., and Ng, A. (2005). High speed obstacle avoidance using monocular vision and reinforcement learning. In *Proceedings 22nd Int. Conf. on Machine Learning*, pages 593–600.

Mitchell, T. (1997). *Machine Learning*. The McGraw-Hill Companies, Inc., first edition.

Murali, V. and Birchfield, S. (2008). Autonomous navigation and mapping using monocular low-resolution grayscale vision. In *IEEE Workshop on Computer Vision and Pattern Recognition*, pages 1–8.

Ohno, T., Ohya, A., and Yuta, S. (1996). Autonomous navigation for mobile robots referring pre-recorded image sequence. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 672–679.

Prince, S. (2011). Computer vision models.

Roning, J., Taipale, T., and Pietikainen, M. (1990). A 3-d scene interpreter for indoor navigation. In *IEEE Int. Workshop on Intelligent Robots and Systems*, pages 695–701.

Santosh, D., Achar, S., and Jawahar, C. (2008). Autonomous image-based exploration for mobile robot navigation. In *IEEE Int. Conf. on Robotics and Automation*, pages 2717–2722.

Sofman, B., Lin, E., Bagnell, J., Cole, J., Vandapel, N., and Stentz, A. (2006). Improving robot navigation through self-supervised online learning. *Journal of Field Robotics*, 23(11-12):1059–1075.

Ulrich, I. and Nourbakhsh, I. (2000). Appearance-based obstacle detection with monocular color vision. In *AIII Conf. on Artificial Intelligence*, pages 866–871.